

Application Frameworks for High Performance and Grid Computing

Gabrielle Allen

Assistant Director for Computing Applications, Center for Computation & Technology

Associate Professor, Department of Computer Science

Louisiana State University

Edward Seidel

Director, Center for Computation & Technology

Floating Systems Professor, Departments of Physics and Computer Science

Louisiana State University

Advanced Computation

From engineering gas turbine blade cooling methods for more efficient engines to determining the optimal location for oil wells in the Gulf of Mexico, advanced computational techniques are becoming increasingly important as we study and interpret the world. From predicting complex processes in nature such as hurricane-induced storm surges to the gravitational waves emitted by the collision of two black holes, solving these Grand Challenge problems cannot be effectively tackled by a single researcher, or even a single group. Today's complex problems require that interdisciplinary research groups combine their expertise in computer science, applied mathematics, engineering, science and the humanities.

The scope and accuracy of modern computational approaches are limited, however, compared to what is required by such complex applications. Going beyond high-end computers, which are still too small to adequately compute many processes at the level of detail we need, a much more comprehensive approach involving large-scale computational power, networks, data, and storage, as well as application and system software, is needed to tackle these problems.

The Center for Computation & Technology at Louisiana State University

Louisiana has funded various initiatives to build up large-scale computer resources such as those housed at LSU's new Center for Computation & Technology (CCT), and the LONI network, which is an advanced 40Gbps optical network and distributed set of computer servers connecting the state's research institutes to the National Lambda Rail. Louisiana is also building a critical mass of faculty and staff with the spectrum of expertise required to advance complex application areas.

CCT was established in 2003 with a mission to tackle the issues involved in solving complex scientific problems. Today, CCT is home to more than two dozen faculty from a dozen different LSU departments and plans to recruit another fifteen faculty to form complementary research groups in different areas of computation and technology. For example, Thomas Sterling is researching the next generation of computer architectures, Gabrielle Allen is designing computational frameworks to support application development on large scale and distributed supercomputers and grids, Tefik Kosar is investigating how new generations of data-intensive applications can locate and store petabytes of data. Additionally, Seung-Jong "Jay" Park is developing new protocols for leveraging high speed optical networks and sensor networks, Paul Saylor and Burak Aksoylu are researching mathematical algorithms and techniques for solving increasingly complex systems of governing equations, and Bygg Ullmer is investigating human-computer interactions and scientific visualization. Such technical expertise in scientific computing has to be driven by the applications themselves.

CCT includes research teams in numerical relativity and computational fluid dynamics that team with the many other application areas at LSU such as petroleum engineering, coastal modeling and computational chemistry.

Application Frameworks for High End Computing

Many large-scale applications interface with supercomputers and high-end computing environments through advanced frameworks and toolkits, which can provide both generic tools for parallel computing and custom libraries to support different domain communities. One such programming environment co-developed at the CCT and the Albert Einstein Institute in Potsdam, Germany, is the open source Cactus Framework [1, 2]. Cactus allows scientists and engineers to develop and reuse numerical software targeted for high-end parallel computing in a collaborative, portable environment. Using Cactus, scientists can develop their codes on laptop computers, using C, C++, or Fortran 77/90 but then run them on virtually any supercomputer. Cactus also provides access to myriad computational tools, such as advanced numerical techniques, adaptive mesh refinement, parallel I/O, live and remote visualization, and remote steering. The various computational layers and functionalities, as well as different scientific modules, can be selected at run time via parameter choices.

Cactus provides application groups with access to cutting edge technologies, such as Grid computing, remote visualization, and high speed parallel I/O; just as important, however, working with a single modular framework such as Cactus helps to develop communities and enable code sharing and toolkit building. Such code sharing also happens between disciplines; at CCT the computational fluid dynamics (CFD) and numerical relativity groups collaborate and share expertise in high- accuracy numerical methods and boundary conditions. Cactus is used by many different application domains, including numerical relativity and astrophysics, CFD, climate modeling, chemical engineering, and financial modeling. In particular, toolkits for building applications in relativity and CFD are openly available and used by groups around the world.

Grid and Distributed Computing

Beyond large-scale parallel clusters and supercomputers lies the emerging field of Grid computing. Conventionally defined as coordinated resource sharing for complex problem solving, Grids will provide scientists with distributed access not only to computer resources but also to high speed networks, information repositories and data archives, and experimental and observational devices; Grids also provide a host of tools for service discovery, notification, etc. The term Cyberinfrastructure has been coined to describe the software and hardware layers that will enable application domains to make good use of distributed resources and Grids, leading to new, powerful scenarios for problem solving. Cyberinfrastructure includes middleware tools for application development; services for data, information, and knowledge management; and secure interfaces for interaction, collaboration, and visualization. At CCT, such cyberinfrastructure is being developed in several fields, including CFD, petroleum engineering, coastal ocean observation and prediction, numerical relativity, and computational chemistry. We are also developing generic cyberinfrastructure tools for applications using the Cactus Framework.

The field of Grid computing is rapidly maturing, with interfaces moving from proprietary protocols to new web service standards. New services are being developed by research groups and industry, with varying degrees of functionality, reliability, security, and usability. To allow for application development in such a dynamic field, the CCT uses the Grid Application Toolkit (GAT) [3], the generic API that provides a standard programming interface for the Grid. As MPI provides a stable

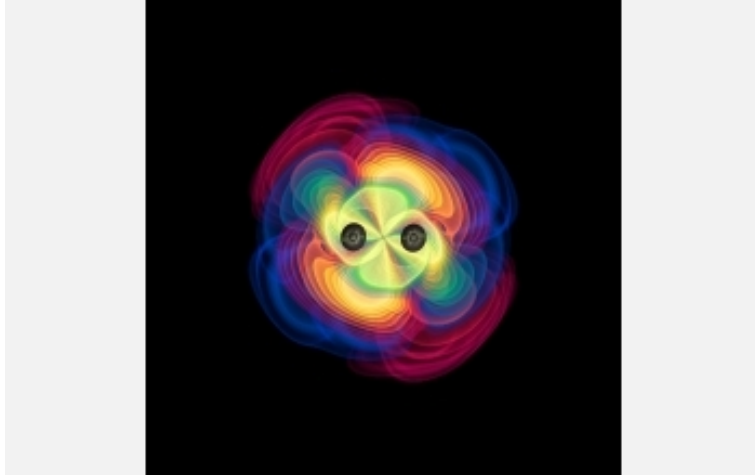


Figure 1: We show an image from a visualization of a binary black hole simulation, displaying the coalescing black holes and the resulting gravitational waves. Such visualizations can be generated as the simulations are underway, and displayed through portal technologies to the collaborating scientists, to better understand and coordinate large scale simulation efforts. [*Visualization: Werner Bengert (CCT/AEI/ZIB)*].

interface to different MPI implementations, the GAT provides a simple, standardized interface to many different Grid services. The CCT researches and develops the GAT along with more than 10 European partners from the GridLab project [4], for which Allen and Seidel were investigators. The programming interface innovated by the GridLab team has led to a working group in the Grid standards and best practices body, the Global Grid Forum. This group is developing a more widely adopted interface, termed the Simple API for Grid Applications, or SAGA [5]. An implementation of SAGA is now being developed at CCT that will be compatible with the GAT.

User Interfaces

Scientists today typically have access to a large diverse and geographically distributed array of computers. For example, the numerical relativists at CCT use machines locally at LSU, across the state in the new LONI network, across the country at national centers via a large NSF allocation, and machines in Europe and Korea. Web-based portals are being developed to track, monitor and manage a community's resources, software codes, simulations and data. Visualizations of the ongoing simulations can be computed while they are carried out and viewed by all collaborators in real-time. The GridLab project also developed a popular, modular framework for building portals, called GridSphere. This new framework contains a range of portlets that CCT is extending and customizing to provide not only access to Grid capabilities, but also community tools for different application areas.

Coastal and Environmental Modeling

An application area of particular interest in Louisiana, that ties all these elements together, is coastal and environmental modeling. The recent catastrophes in the southeast United States following the triad of Hurricanes Katrina, Rita, and Wilma have highlighted the need, not only for timely and accurate forecasts, but also for improved coordination and information transfer between domain experts, policy makers, and emergency responders. Coastal modeling involves many issues

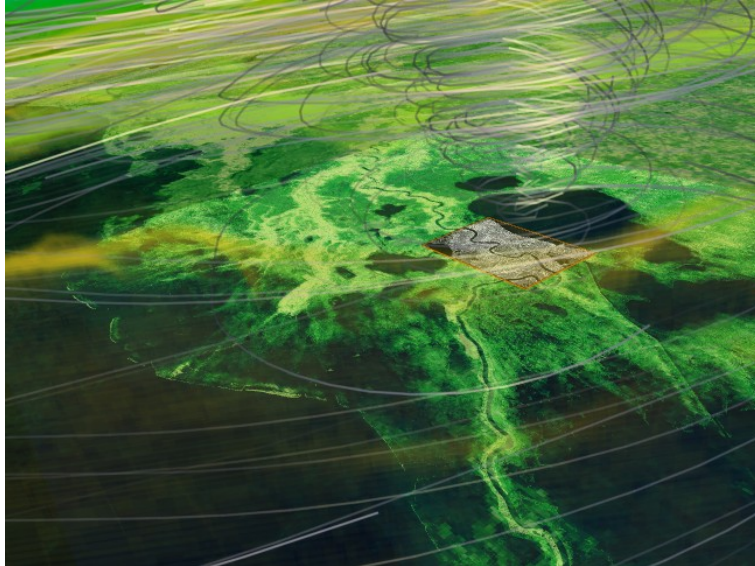


Figure 2: An image from a visualization of Katrina developed at CCT which integrates different fields such as wind vectors, surge height and air temperature, and different scales. Such visualizations can help researchers interpret simulation data, and the general public better comprehend the impact on the region. [*Visualization: Werner Bengler, Shalini Venkataraman (CCT)*].

that are priority areas targeted by federal funding agencies, namely: multiscale models capable of covering large-scale global climate domains including turbulence and other fine scale features within regional models; and coupled models incorporating the interface between wind, ocean circulation and surge models, and various models of land and sea ecology. Coastal modeling motivates another emerging area of computer science: Dynamic Data Driven Application Systems, or DDDAS. This system replaces conventional self contained simulations with living applications that interact with and respond to realtime data from sensors, satellites, and other simulations, not just reacting to their environment, but effecting needed change.

To progress in this complex field requires the use of modular, community frameworks, such as the Cactus Framework with well defined interfaces for coupling models and exploiting cutting edge computer science technologies. It also requires tools such as the Grid Application Toolkit for building the Cyberinfrastructure needed to interact with distributed data sources and deploy dynamic data driven models across brokered resources. End users need secure community portals, such as are being built with the GridSphere framework to deploy and track their simulations, discover data and services, and build custom views for policy makers and emergency responders.

In summary, modern computational approaches to complex problems facing our communities require a comprehensive mix of high-end computing, network, sensor, data, and visualization technologies, novel multiscale algorithms, as well as application scientists and engineers from multiple disciplines. Advanced system and application software tools, such as modeling, grid, and portal development frameworks, are advancing the fields by integrating and simplifying these different technologies for the application developers and users.

References

- [1] Cactus Framework. <http://www.cactuscode.org>.
- [2] Tom Goodale, Gabrielle Allen, Gerd Lanfermann, Joan Massó, Thomas Radke, Edward Seidel, and John Shalf. The Cactus framework and toolkit: Design and applications. In *High Performance Computing for Computational Science - VECPAR 2002, 5th International Conference, Porto, Portugal, June 26-28, 2002*, pages 197–227, Berlin, 2003. Springer.
- [3] G. Allen, K. Davis, T. Goodale, A. Hutanu, H. Kaiser, T. Kielmann, A. Merzky, R. Van Nieuwpoort, A. Reinefeld, F. Schintke, T. Schuett, E. Seidel, and B. Ullmer. The grid application toolkit: Toward generic and easy application programming interfaces for the grid. *Proceedings of the IEEE*, 93:534–550, 2005.
- [4] GridLab: A Grid application toolkit and testbed project home page: <http://www.gridlab.org>.
- [5] SAGA Research Group. See web site at: <https://forge.gridforum.org/projects/saga-rg/>.