March 05, 2012 | By Jan Zverina

SDSC's 'Gordon' Supercomputer: Ready for Researchers

Initial Projects Range from Storm Predictions to Stock Market Data

Accurately predicting severe storms, or what Wall Street's markets will do next, may become just a bit easier in coming months as *Gordon*, a unique supercomputer at the San Diego Supercomputer Center (SDSC) at the University of California, San Diego, begins helping researchers delve into these and other data-intensive projects.

Following acceptance testing in January, *Gordon* has now begun serving University of California and national academic researchers as well as industry and government agencies. Named for its massive amounts of flash-based memory, *Gordon* is part of the National Science Foundation's (NSF) Extreme Science and Engineering Discovery Environment, or XSEDE program, a nationwide partnership comprising 16 supercomputers and high-end visualization and data analysis resources.

The first round of allocations recently approved by XSEDE includes about a dozen separate research projects that will use a combined 10 million processing hours on *Gordon*, whose data-intensive computational capabilities will help scientists find, among other things, ways to create more robust molecular "force fields" to find new drugs for diseases, and explore how particles in atmospheric aerosols directly affect air quality.

Along with its innovative supernodes that create multi-terabyte shared-memory systems, *Gordon*, the result of a five-year, $20 million National Science Foundation (NSF) grant, can process data-intensive problems about 10 times faster than other supercomputers because it employs massive amounts of flash-based memory – common in smaller devices such as cellphones or laptop computers – instead of slower spinning disks. Full specifications of SDSC's *Gordon* can be found here.

"To enable the basic and breakthrough science and scholarship of this new century, UC San Diego has made a significant investment in research cyberinfrastructure, an investment that will benefit our university, the University of California system, our state and our nation," said UC San Diego Vice

Chancellor for Research Sandra A. Brown. "*Gordon* is a powerful and promising part of that infrastructure, and should quickly demonstrate its value as a research resource."

"*Gordon* is now assisting the research community in a number of very diverse and data-intensive projects, many of them which could not be addressed previously because scientists simply didn't have the computer capability to do so – and some of which are rather unconventional applications that could have a broader, societal impact," said SDSC Director Michael Norman, the principal investigator for the *Gordon* project.

Currently the most powerful supercomputer in all of Southern California based on speed and capability, *Gordon* was recently ranked among the top 50 supercomputers in the world in terms of speed of doing pure math (see top500.org). However, it is the most powerful high-performance computing (HPC) system anywhere when it comes to accessing data, according to Allan Snavely, SDSC's associate director and project leader along with Norman. In recent validation tests, *Gordon* achieved an unprecedented 36 million input/output operations per second, or IOPS, a critical measure for doing data-intensive computing such as sifting through huge datasets to find a mere sliver of meaningful information.

"*Gordon's* unique ability to provide extremely fast random access to very large data sets is opening new doors for researchers across a wide range of disciplines, some outside the traditional realm of science," said Snavely. "Such areas could include analyzing social-media data and seeing a political event such as the recent Arab Spring emerge in cyberspace before we saw people emerging in the street."

Here are a few of the projects included in the first round of allocations for *Gordon* that were recently approved by XSEDE's Resource Allocation Committee (XRAC). The XRAC consists of researchers and computational scientists who review the submissions and make allocation recommendations for all XSEDE resources:

On August 30, 2011, about 3 million orders were submitted to the NASDAQ market to trade the stock SPDR S&P 500 Trust (ticker symbol SPY). This image shows that 18.3% of the orders were canceled within one millisecond. Combined with the light blue section, 42.5% of orders had a lifespan of less than 50 milliseconds, less time than it takes to transfer a signal between New York and California. Therefore, more than 40% of orders disappeared before a trader in California could react. Image courtesy of Mao Ye, University of Illinois

Gordon Takes on Wall Street

So-called high-frequency trading has become commonplace in the U.S. stock market, and has been considered by the press to have contributed to the "flash crash" in May 6, 2010 that sent the Dow Jones Industrial average down 998.5 points in just 20 minutes. The Securities and Exchange Commission recently said it is looking to curb high-frequency traders' huge influence on stock trading, and is considering charging fees for buy and sell orders that are later canceled, among other options.

Distribution of SPY Quote Life for Orders Cancelled on 08/30/11

18.3% cancelled in 0.001s

26.1% cancelled > 1s

24.2% cancelled in 0.001 – 0.05s

31.4% cancelled in 0.05 – 1s
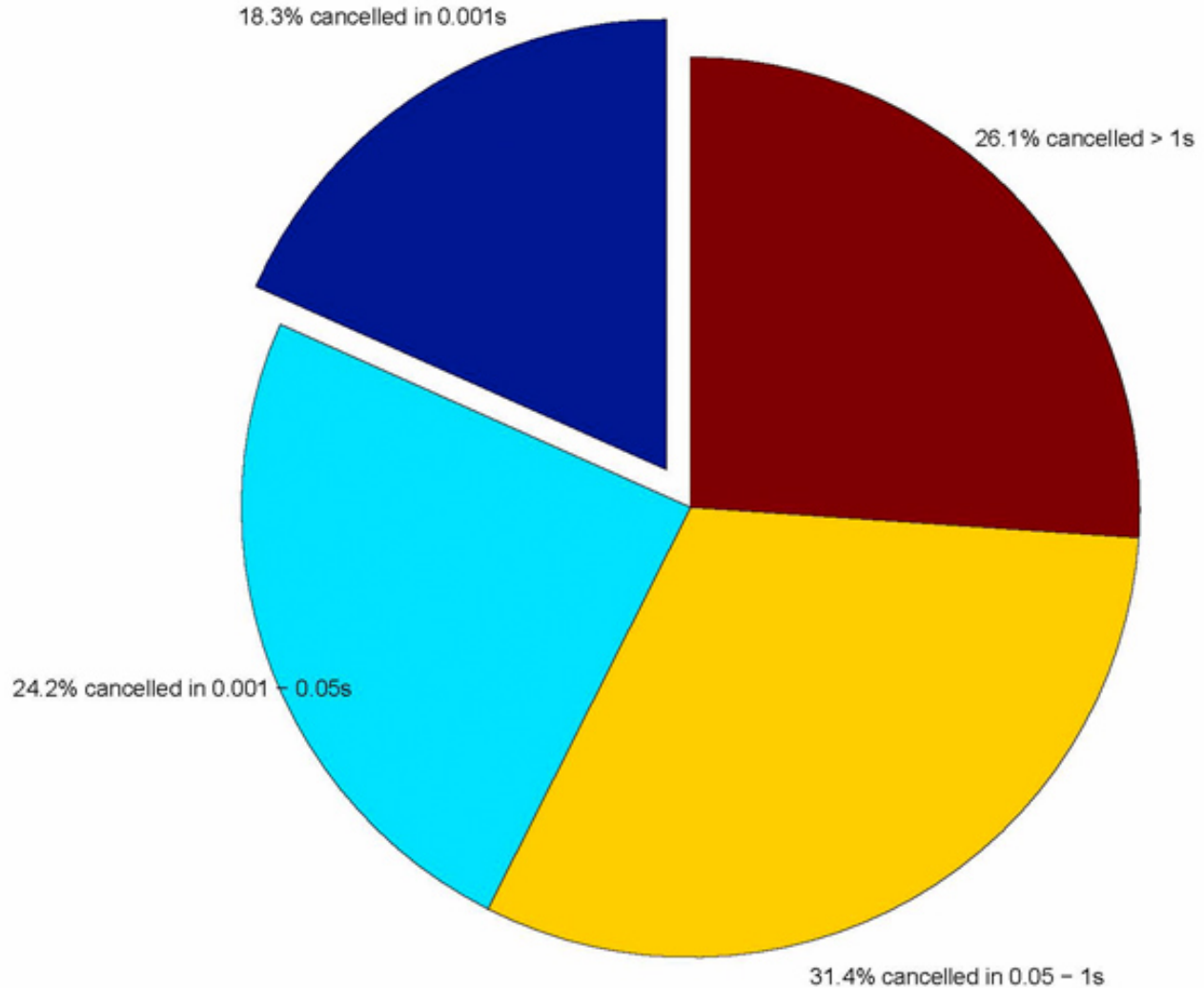
Mao Ye, an assistant professor of finance at the University of Illinois at Urbana-Champaign, is using *Gordon* to sift through massive amounts of NASDAQ historical ITCH market data to better understand the practices and patterns in high-frequency trading.

iversity of Illinois at Urbana-Champaign

According to Ye's research, many of these traders place an order, only to cancel it within 0.001 second or less. "We want to find out if these cancelled orders are being done to manipulate the market in some way," said Ye, whose research also has focused on odd lots – or trades of less than 100 shares – which do not have to be reported to the consolidated tape.
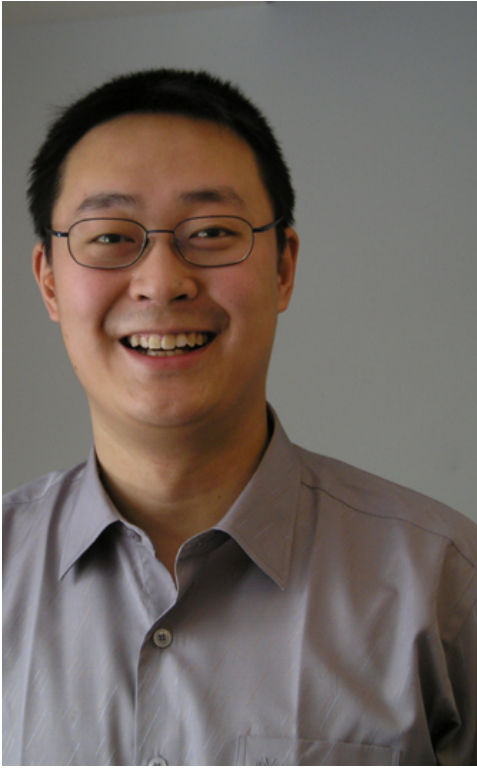
"Because the minimum report requirement is 100 shares, we believe that people trade 99 shares for a strategic reason: they may be trying to hide their information," said Ye. "Since so many orders are cancelled within such a short time, a natural question to ask is whether orders are cancelled due to legitimate reasons, or if these orders are entered for deceptive and manipulative reasons. Also, are these orders contributing to the liquidity and efficiency of the financial market, or are they the causes of the recent flash crash?"

One recent hypothesis is called "quote stuffing," which states that high-frequency traders send huge amount of messages to purposely slow down the exchange's trading system. "We'd like to know if these kinds of orders contribute to liquidity of the market, or whether they lead to abnormal volatility."

Ye and his research team were allocated 500,000 processing hours of time on *Gordon* to process two years of data containing about 400 million order messages for each day. Ye's research could be important for recent policy debates on whether a minimum quote life should be set for the orders.

Weather Warnings and Water Management

Two of the first research projects using *Gordon* are in the area of climate simulations, and severe weather predictions in particular. Ming Xue, a professor at the School of Meteorology, and Director of the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma, was granted 2,000,000 processor hours on *Gordon* to better understand and improve the accuracy of predictions of high-impact weather such as severe thunderstorms, tornadoes, and tropical cyclones.

"High-speed computing and data storage will allow us to create very high-resolution simulations of tornadoes for analysis at numerous output times, and allow us to use radar data at precise time intervals to create accurate numerical simulations and predictions of severe weather," said Xue, whose research is funded by multiple NSF grants. "These data files are extremely large, so using traditional storage devices during the simulations can be extremely time-consuming. *Gordon* will help us significantly reduce the input/output time, thus speeding-up the overall research process."

The image at left is a computerized simulation of how the Florida Everglades area looked circa the 1850s, before the sheet-water flow was interrupted by man-made canals in order to keep the land dry for human development. Since then, the entire ecosystem has changed for the worse (right), as runoff containing pesticides/fertilizers/toxins was flushed into Florida Bay. Image courtesy of South Florida Water Management District (SFWMD)

Craig Mattocks, an atmospheric research scientist at the Rosenstiel School of Marine and Atmospheric

Pre-Drainage System (1850's) | Managed System (~2003)
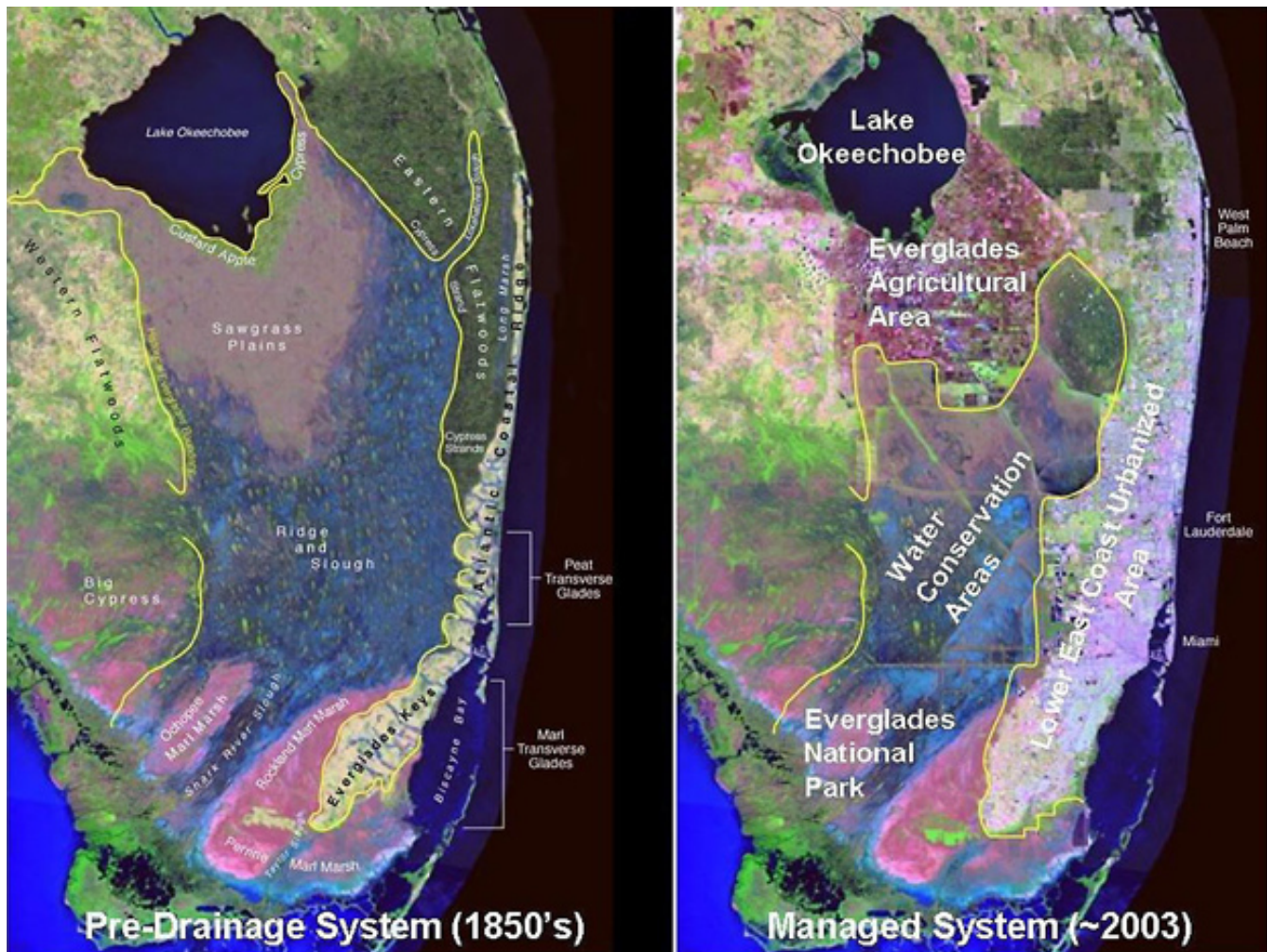
Science at the University of Miami, was awarded 3,500,000 processor hours on *Gordon* to deploy the Ocean Land Atmosphere Model (OLAM) Earth System on XSEDE computational resources as a Science Gateways project for teaching graduate-level meteorology, climate, and predictability courses. Mattocks' research is also focused on generating the most detailed regional climate-change projections and simulations to date, to help guide water resource management decisions throughout South Florida.

Mattocks' group has teamed with hydrology modelers from the South Florida Water Management District (SFWMD), which is responsible for managing the flow of water throughout natural and artificial structures such as rivers, canals, floodgates, and water conservation areas. SFWMD must meet the diverse needs of 7.7 million people along with demands from tourism, industrial, and agricultural concerns — all while simultaneously trying to restore the Florida Everglades to their natural state.

The SFWMD has reconstructed the land-cover distribution of the natural system in order to visualize the dramatic changes that occurred in the Florida Everglades between the 1850s and present day. "When the sheet-water flow through the Everglades — or the flow of a thin layer of water that's not concentrated into channels — was interrupted by drainage from canals in order to keep the land dry

for human development, the entire ecosystem changed for the worse," he said. "Much of the marshland was converted to agriculture, and runoff containing pesticides, fertilizers, and toxins was flushed into Florida Bay, killing the pristine coral reef systems there."

According to Mattocks, it has been a challenge to restore the flow to something even remotely resembling the area's earlier, natural system because of commercial agriculture and other man-made impacts.

"There's been a war going on between environmentalists and commercial enterprises, so progress is slow and funding for restoration efforts is difficult to maintain," he said. "SFWMD needs predictions of precipitation and evapotranspiration under different climate scenarios to drive their hydrology models in order to plan future water-management strategies in times of rapid climate change. With *Gordon*, we hope to be able to zoom these simulations down to a one-kilometer resolution, and run longer climate 'timeslice' experiments to provide more accurate information to make south Florida's water supply more resilient."

More Robust Computer Simulations for Drug Design

Other researchers are relying on *Gordon* to help scientists expose weaknesses in molecular dynamics simulations, which are instrumental in developing new drugs to combat diseases such as AIDS. Junmei Wang, an assistant professor in biochemistry at the UT Southwestern Medical Center in Dallas, Texas, was awarded 300,000 hours on *Gordon* to study large-scale biological systems or dynamic events in a long time scale using the state-of-the-art molecular dynamics simulation (MD) techniques.

"For us, *Gordon* is ideal to evaluate our molecular mechanical force fields, which are the foundation of MD technique, and help reveal any problems of a force field for studying biological systems/events, which can only be identified through long time scale simulations," said Wang, who is using *Gordon* to conduct *ab initio* (or, from first principles) calculations to get high-quality structural and dynamic data for model compounds used to develop molecular-mechanical force fields.

"Without a high quality molecular mechanical force field, it is impossible to successfully identify promising inhibitors of a protein or nucleic acid receptor from millions of compounds in libraries," said Wang. "These promising inhibitors are potential drug leads, and in principle such structure-based drug design can be applied to battle any disease, such as AIDS."

About SDSC

As an Organized Research Unit of UC San Diego, SDSC is considered a leader in data-intensive computing and cyberinfrastructure, providing resources, services, and expertise to the national research community, including industry and academia. Cyberinfrastructure refers to an accessible, integrated network of computer-based resources and expertise, focused on accelerating scientific

inquiry and discovery. SDSC supports hundreds of multidisciplinary programs spanning a wide variety of domains, from earth sciences and biology to astrophysics, bioinformatics, and health IT. With its two newest supercomputer systems, *Trestles* and *Gordon*, SDSC is a partner in XSEDE (Extreme Science and Engineering Discovery Environment), the most advanced collection of integrated digital resources and services in the world.

MEDIA CONTACT

Warren R. Froelich, 858 822-3622, froelich@sdsc.edu

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.