

NVIDIA GPU TECHNOLOGY THEATER AT SC13



** Information subject to change. All times listed in Mountain Time Zone (GMT-7)*

Monday, November 18 – Thursday, November 21 during exhibition hours | NVIDIA Booth #613

The NVIDIA GPU Technology Theater is hosting an amazing lineup of industry luminaries, scientists, and developers at SC13, covering a wide range of topics on HPC and accelerated computing.

Come back often for updates!

Monday, November 18 | Booth #613

Accelerated Computing: What's Coming Next.

7:30 PM – 8:00 PM

Ian Buck
General Manager, Accelerated
Computing
NVIDIA

Tuesday, November 19 | Booth #613

Towards Performance-Portable Applications Through Kokkos: A Case Study with LAMMPS

10:30 AM – 11:00 AM

Christian Trott
Postdoctoral Appointee
Sandia National Laboratories

In this talk we demonstrate how LAMMPS uses the many-core device performance portability library Kokkos to implement a single code base for CPUs, NVIDIA GPUs and Intel Xeon Phi co-processors. This portable code base has equal or better performance compared to LAMMPS' current generation of hardware specific add-on packages.

Massively Parallel Computing and the Search for New Physics at the LHC

11:00 AM – 11:30 AM

Valerie Halyo
Assistant Professor of Physics
Princeton University

The quest for rare new physics phenomena at the LHC leads us to evaluate a Graphics Processing Unit (GPU) enhancement of the existing High-Level Trigger (HLT), made possible by the current flexibility of the trigger system, which not only provides faster and more efficient event selection, but also includes the possibility of new complex triggers that were not previously feasible. A new tracking algorithm is evaluated on a NVIDIA Tesla K20c GPU, allowing for the first time the reconstruction of long-lived particles or displaced black holes in the silicon tracker system at real time in the trigger.

** Information subject to change. All times listed in Mountain Time Zone (GMT-7)*

<p>Being Very Green with Tsubame 2.5 Towards 3.0 and Beyond to Exascale</p> <p>11:30 AM – 12 Noon</p> <p>Satoshi Matsuoka Professor Global Scientific Information and Computing Center (GSIC), Tokyo Institute of Technology</p>	<p>TSUBAME 2.5 succeeded TSUBAME 2.0 by upgrading all 4224 Tesla M2050 GPUs to Kepler K20x GPUs, achieving 5.76 / 17.1 Petaflops peak in double / single point precision respectively, latter the fastest in Japan. By overcoming several technical challenges, TSUBAME 2.5 exhibits x2-3 speedup and multi-petaflops performance for many applications, leading to TSUBAME 3.0 in 2015-16.</p>
<p>Accelerated Computing with OpenACC</p> <p>12:00 Noon – 12:30 PM</p> <p>Michael Wolfe Compiler Engineer The Portland Group, NVIDIA</p>	<p>The OpenACC API provides a high-level, performance portable programming mechanism for parallel programming accelerated nodes. Learn about the latest additions to the OpenACC specification, and see the PGI Accelerator compilers in action targeting the fastest NVIDIA GPUs.</p>
<p>Efficiency and Programmability: Enablers for Exascale</p> <p>12:30 PM – 1:00 PM</p> <p>Bill Dally Chief Scientist & SVP Research NVIDIA</p>	<p>HPC and data analytics share challenges of power, programmability, and scalability to realize their potential. The end of Dennard scaling has made all computing power limited, so that performance is determined by energy efficiency. With improvements in process technology offering little increase in efficiency, innovations in architecture and circuits are required to maintain the expected performance scaling. The large scale parallelism and deep storage hierarchy of future machines poses programming challenges. This talk will discuss these challenges in more detail and introduce some of the technologies being developed to address them.</p>
<p>New Features in CUDA 6 Make GPU Acceleration Easier</p> <p>1:00 PM – 1:30 PM</p> <p>Mark Harris Chief Technologist, GPU Computing NVIDIA</p>	<p>The performance and efficiency of CUDA, combined with a thriving ecosystem of programming languages, libraries, tools, training, and services, have helped make GPU computing a leading HPC technology. Learn how powerful new features in CUDA 6 make GPU computing easier than ever, helping you accelerate more of your application with much less code.</p>

<p>Titan: Accelerating Computational Science and Engineering with Leadership Computing</p> <p>1:30 PM – 2:00 PM</p> <p>Jack Wells Director of Science National Center for Computational Science Oak Ridge National Laboratory</p>	<p>Modeling and simulation with petascale computing has supercharged the process of innovation, dramatically accelerating time-to-discovery. This presentation will focus on early science from the Titan supercomputer at the Oak Ridge Leadership Computing Facility, with results from scientific codes discuss, e.g., LAMMPS and WL-LSMS. I will also summarize the lessons we have learned in preparing applications to move from conventional CPU architectures to a hybrid, accelerated architecture, and the implications for the research community as we prepare for exascale computational science.</p>
<p>A Productive Framework for Generating High-Performance, Portable, Scalable Applications for Heterogeneous Computing</p> <p>2:00 PM – 2:30 PM</p> <p>Wen-mei Hwu Professor and Sanders-AMD Chair, ECE University of Illinois at Urbana-Champaign</p>	<p>I will present two synergistic systems that enable productive development of scalable, Efficient data parallel code. Triolet is a Python-syntax based functional programming system where library implementers direct the compiler to perform parallelization and deep optimization. Tangram is an algorithm framework that supports effective parallelization of linear recurrence computation.</p>
<p>Fighting HIV with GPU-Accelerated Petascale Computing</p> <p>2:30 PM – 3:00 PM</p> <p>John Stone Senior Research Programmer, Associate Director, CUDA Center of Excellence Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign</p>	<p>Molecular dynamics simulations provide a powerful tool for probing the dynamics of cellular processes at atomic and nanosecond resolution not achievable by experimental methods alone. We describe how GPU-accelerated petascale supercomputers are enabling studies of large biomolecular systems such as the HIV virus in all-atom detail for the first time.</p>

<p>Flying Snakes on GPUs</p> <p>3:00 PM – 3:30 PM</p> <p>Lorena Barba Associate Professor of Engineering and Applied Science Department of Mechanical and Aerospace Engineering The George Washington University</p>	<p>It would be hard to put a flying snake in a wind tunnel. So we are trying to put them in GPUs instead—via computational fluid dynamics. Our initial success is to see that a flying snake's cross-section can in fact create quite some lift: it even has a favorite angle of attack for which it gives extra lift. We don't know if this is the secret of flying snakes, but we do know that looking at nature can teach engineers some new tricks.</p>
<p>Navigating The Transition to Heterogeneous Architectures</p> <p>3:30 PM – 4:00 PM</p> <p>James Hack Director National Center for Computational Sciences, Oak Ridge National Laboratory</p>	<p>The National Center for Computational Sciences (NCCS) at Oak Ridge National Laboratory (ORNL) was selected as the Leadership Computing Facility (LCF) by the U.S. Department of Energy in 2004. In 2011 ORNL began a physical upgrade to Jaguar to convert it from a Cray XT5 into a Cray XK6 system to be named Titan, which was placed in the number one position on the November 2012 TOP500 list demonstrating 17.59 PF on the HPL benchmark, a 10-fold increase over the 2009 Jaguar performance. The migration to a heterogeneous architecture has presented new challenges to the application community, most notably the development of a programming strategy that allows applications to run efficiently on the hybrid architecture, while maintaining portability with other more conventional computer architectures. This talk will review the deployment of petascale capabilities at ORNL that has led to the current architectural direction and will discuss the preparations aimed at ensuring a successful transition to heterogeneous architectures for some key simulation problems, including global atmospheric modeling.</p>
<p>Exploring Emerging Technologies in the HPC Co-Design Space</p> <p>4:00 PM – 4:30 PM</p> <p>Jeff Vetter Future Technologies Group Leader and Professor Oak Ridge National Laboratory and Georgia Tech</p>	<p>New architectures, such as novel heterogeneous cores and NV-RAM memory systems, are often radically different from today's systems. Our team has recently developed a number of techniques for modeling, simulating, and measuring these future systems. Aspen, our performance modeling language, allows users to compose and answer modeling questions quickly.</p>

<p>Map-D: GPU-Powered Databases and Interactive Social Science Research in Real Time</p> <p>4:30 PM – 5:00 PM</p> <p>Tom Graham Researcher Harvard Law School and Todd Mostak Research Programmer MIT's Computer Science and Artificial Intelligence Laboratory</p>	<p>Map-D (Massively Parallel Database) uses multiple NVIDIA GPUs to interactively query and visualize big data in real-time. Map-D is an SQL-enabled column store that generates 70-400X speedups over other in-memory databases. This talk discusses the basic architecture of the system, the advantages and challenges of running queries on the GPU, and the implications of interactive and real-time big data analysis in the social sciences and beyond.</p>
<p>Using a Hybrid Cray Supercomputer to Model Non-Icing Surfaces for Cold-Climate Wind Turbines</p> <p>5:00 PM – 5:30 PM</p> <p>Masako Yamada Physicist Advanced Computing Lab GE Global Research</p>	<p>We have been awarded 80 million CPU hours on Titan, a hybrid Cray supercomputer, to model the freezing behavior of water droplets. By optimizing the three-body mW water potential in LAMMPS, we have achieved 5x acceleration in the hybrid CPU/GPU environment, relative to previous Jaguar performance.</p>
<p>20 Petaflops Simulation of Protein Suspensions in Crowding Conditions</p> <p>5:30 PM – 6:00 PM</p> <p>Simone Melchionna Researcher IPCF - National Research Council of Italy</p>	<p>This talk describes the recent simulation of ~18,000 proteins in suspension, reproducing the crowding conditions of the cell interior. The simulations were obtained with MUPHY, a computational platform for multi-scale simulations of real-life biofluidic problems. The same software has been used in the past to simulate blood flows through the human coronary arteries and DNA translocation across nanopores. The simulations were performed on the Titan system at the Oak Ridge National Laboratory, and exhibits excellent scalability up to 18, 000 K20X NVIDIA GPUs, reaching 20 Petaflops of aggregate sustained performance with a peak performance of 27.5 Petaflops for the most intensive computing component. In this talk I will describe how the combination of novel mathematical models, computational algorithms, hardware technology and parallelization techniques allowed reproducing for the first time such a massive amount of proteins.</p> <p><i>ACM Gordon Bell Finalist</i></p>

* Information subject to change. All times listed in Mountain Time Zone (GMT-7)

<p>Skin-Barrier Investigation Using GPU-Enhanced Molecular Dynamics</p> <p>10:30 AM – 11:00 AM</p> <p>Russell Devane Scientist Proctor & Gamble</p>	<p>GPU enabled molecular dynamics are being used to investigate permanent crossing of the primary skin barrier, the stratum corneum (SC). This work is helping to identify the molecular characteristics that dictate the ability of a compound to cross the SC barrier in order to build more accurate skin penetration models.</p>
<p>Emerging Technologies for High-Performance Computing</p> <p>11:00 AM – 11:30 AM</p> <p>Jack Dongarra Professor University of Tennessee</p>	<p>This talk will highlight the emerging technologies in high performance computing. We will look at the development of accelerators and some of the accomplishments in the Matrix Algebra on GPU and Multicore Architectures (MAGMA) project. We use a hybridization methodology that is built on representing linear algebra algorithms as collections of tasks and data dependencies, as well as properly scheduling the tasks' execution over the available multicore and GPU hardware components.</p>
<p>New Features in CUDA 6 Make GPU Acceleration Easier</p> <p>11:30 AM – 12 Noon</p> <p>Mark Harris Chief Technologist, GPU Computing NVIDIA</p>	<p>The performance and efficiency of CUDA, combined with a thriving ecosystem of programming languages, libraries, tools, training, and services, have helped make GPU computing a leading HPC technology. Learn how powerful new features in CUDA 6 make GPU computing easier than ever, helping you accelerate more of your application with much less code.</p>
<p>Applications of Programming the GPU Directly from Python Using NumbaPro</p> <p>12 Noon – 12:30 PM</p> <p>Travis Oliphant Co-Founder and CEO Continuum Analytics</p>	<p>NumbaPro is a powerful compiler that takes high-level Python code directly to the GPU producing fast-code that is the equivalent of programming in a lower-level language. It contains an implementation of CUDA Python as well as higher-level constructs that make it easy to map array-oriented code to the parallel architecture of the GPU.</p>

<p>Twinkle, Twinkle Little Star - Using 18.000 GPUs to Simulate Jets in the Cosmos</p> <p>12:30 PM – 1:00 PM</p> <p>Michael Bussmann Junior Group Leader Computational Radiation Physics Helmholtz-Zentrum Dresden-Rossendorf</p>	<p>In order to understand what happens when jets of hot, streaming gas are ejected at high speed into the cosmos, we are bound to rely on measuring the radiation emitted by the particles in the jet. Astrophysical jets can originate from a variety of sources such as stars, black holes and even galaxies. In such jets, the plasma flow can become unstable, generating characteristic patterns of particle flows. Using our particle-in-cell code PIconGPU utilizing the complete TITAN supercomputer system at Oak Ridge National Laboratory, we were able, for the first time, to not only simulate the particle dynamics but also the radiation emitted during the formation of such an instability, the Kelvin-Helmholtz instability.</p> <p><i>ACM Gordon Bell Finalist</i></p>
<p>AMR Based on Space-Filling Curve for Stencil Applications</p> <p>1:00 PM – 1:30 PM</p> <p>Takayuki Aoki Professor/Deputy Director Global Scientific Information and Computing Center (GSIC)/ Tokyo Institute of Technology</p>	<p>AMR is an efficient method to assign a mesh with a proper resolution to any local areas. By using bigger leaves than those of CPU, we can assign a CUDA block to a leaf with enough thread numbers. We show a GPU implementation in which the leaves are connected by a space-filling curve.</p>
<p>Can You Really Learn To Use Accelerators in One Morning?</p> <p>1:30 PM – 2:00 PM</p> <p>John Urbanic Information Technology Manager Pittsburgh Supercomputing Center</p>	<p>OpenACC provides a friendly learning curve for using GPUs and other accelerators. We describe how we have been able to create hundreds of capable new users with half-day and two-day hands-on workshops. Come knowing nothing about accelerators, leave knowing how you and your colleagues can get in on this important new paradigm.</p>

<p>Tomorrow's Exascale Systems: Not Just Bigger Versions of Today's Peta-Computers</p> <p>2:00 PM – 2:30 PM</p> <p>Thomas Sterling Executive Associate Director & Chief Scientist Indiana University</p>	<p>To get to Exascale is going to require more innovation than simply extending Petaflops machine structures with Moore's Law. Power, reliability, user productivity, generality, and cost all demand dramatic advances in all aspects of supercomputer design, operation, and programming. A new synergy of dynamic adaptive techniques, architecture, and programming interfaces are driving research towards a new generation of HPC. This talk will introduce these emerging ideas and illustrate how they will impact the future of our field.</p>
<p>Piz Daint: A Productive, Energy Efficient Supercomputer with Hybrid CPU-GPU Nodes</p> <p>2:30 PM – 3:00 PM</p> <p>Thomas Schulthess Professor of Computational Physics & Directors CSCS ETH Zurich, Swiss National Supercomputing Center (CSCS)</p>	<p>Piz Daint: a productive, energy efficient supercomputer with hybrid CPU-GPU nodes We will discuss the makings of Piz Daint, a Cray XC30 supercomputer with hybrid CPU-GPU nodes. The presentation will focus on quantitative improvements in time and energy to solution due to the use of GPU technology in full climate, materials science and chemistry simulations.</p>
<p>Interactively Visualizing Next Generation Science</p> <p>3:00 PM – 3:30 PM</p> <p>Kelly Gaither Director of Visualization Texas Advanced Computing Center, The University of Texas at Austin</p>	<p>This presentation will cover the design and deployment of a remote interactive visualization and data analysis resource for the national open science community, and the motivating science drivers behind it.</p>
<p>Deploying Clusters with NVIDIA® Tesla® GPUs</p> <p>3:30 PM – 4:00 PM</p> <p>Dale Southard Senior Solution Architect HPC/Cloud NVIDIA</p>	<p>Tips and techniques for deploying high-performance computing clusters using NVIDIA® Tesla® GPUs.</p>

* Information subject to change. All times listed in Mountain Time Zone (GMT-7)

<p>Pinning Down the Superconducting Transition Temperature in the Hubbard Model</p> <p>4:00 PM – 4:30 PM</p> <p>Peter Staar PhD Student ETH Zurich</p>	<p>With massive improvement in algorithms and how they map onto modern hardware platforms, as well as availability of efficient multi-petaflops supercomputers like Titan, simulation-based solution of one the most sought after problems in condensed matter theory has become possible. Implications on studies of high-temperature superconductivity will be discussed.</p> <p><i>ACM Gordon Bell Finalist</i></p>
<p>The NVIDIA Co-Design Lab for Hybrid Multicore Computing at ETH</p> <p>4:30 PM – 5:00 PM</p> <p>Peter Messmer Professor of Computational Physics & Directors CSCS NVIDIA</p>	<p>Developing successful scientific software becomes increasingly a collaborative endeavor, joining the talents of from a multitude of disciplines. NVIDIA and ETH Zurich are forming a Co-Design Lab for Hybrid Multicore Computing as a joint effort to develop and optimize scientific applications for hybrid computing architectures. In this talk, I will introduce the lab and present some early successes of this new collaboration.</p>
<p>PARALUTION - Library for Iterative Sparse Methods on Multi-core CPU and GPU Devices</p> <p>5:00 PM – 5:30 PM</p> <p>Dimitar Lukarski Post-Doctoral Researcher Dept. of Information Technology, Div. of Scientific Computing Uppsala University</p>	<p>PARALUTION is a library which enables you to perform various sparse iterative solvers and preconditioners on multi/many-core CPU and GPU devices. Based on C++, it provides generic and flexible design which allows seamless integration with other scientific software packages, and gives you full portability of your code.</p>
<p>COBALT: A New Correlator for LOFAR</p> <p>5:30 PM – 6:00 PM</p> <p>Chris Broekema Researcher ASTRON (Netherlands Institute for Radio Astronomy)</p>	<p>We've designed and built a new GPU based correlator for the LOFAR radio telescope in the Netherlands. In this talk I will discuss the design process of the GPU based streaming and real-time correlator, the problems we faced in finding a suitable solution, and how we solved these.</p>

CUDA Implementation of the Weather Research and Forecasting (WRF) Model

10:30 AM – 11:00 AM

Bormin Huang
Research Scientist
Space Science and Engineering
Center, University of Wisconsin-
Madison

The Weather Research and Forecasting (WRF) Model is a next-generation mesoscale numerical weather prediction system designed to serve both atmospheric research and operational forecasting needs. The inherently parallel problem of weather forecasting can be effectively solved using GPUs, each with hundreds or thousands of compute cores. In this talk, we present the latest progress in our massively parallel implementation of the WRF model on NVIDIA GPUs using CUDA.

From Brain Research to High-Energy Physics: GPU-Accelerated Applications in Jülich

11:00 AM – 11:30 AM

Dirk Pleiter
Leading Scientist
Jülich Supercomputing Centre

In 2012 the NVIDIA Application Lab at Jülich has been established to work with application developers on GPU enablement. In this talk we will tour through a variety of applications and evaluate opportunities of new GPU architectures and GPU-accelerated HPC systems, in particular for data-intensive applications.

Accessing New NVIDIA® CUDA® Features from CUDA Fortran

11:30 AM – 12 Noon

Brent Leback
Compiler Engineering Manager
The Portland Group, NVIDIA

This talk will present examples of how to take advantage of new CUDA features, specifically those introduced in CUDA 5.0 and CUDA 5.5, from CUDA Fortran. Details of overloading one specific Fortran intrinsic function which obtains near peak performance will be given. A preview of CUDA 6.0 features and how CUDA Fortran will evolve to enable them for Fortran programmers will also be shown.

Earthquake Simulations with AWP-ODC on Titan, Blue Waters and Keeneland

12 Noon – 12:30 PM

Yifeng Cui
Lab Director, HPGeOC
San Diego Supercomputer
Center/UC San Diego

We simulate realistic 0-10 Hz earthquake ground motions relevant to building engineering design, and accelerate SCEC CyberShake key strain tensor calculations on Titan and Blue Waters. Performance improvements of AWP-ODC, coupled with co-scheduling CPUs and GPUs, make a California statewide seismic hazard model a goal reachable with existing supercomputers.

<p>Common Use Cases and Performance for NVAMG: A Fast Linear Solver Toolkit on the GPU</p> <p>12:30 PM – 1:00 PM</p> <p>Joe Eaton Manager, NVAMG CUDA Library NVIDIA</p>	<p>We discuss some of common use cases for NVAMG, our toolkit for fast linear solvers on the GPU. NVAMG includes Algebraic Multi-Grid methods, Krylov methods, nesting preconditioners, and allows complex composition of the solvers and preconditioners. We also present some recent performance results on NVIDIA® Tesla® K20 and K40 GPUs for large-scale CFD problems of industrial relevance.</p>
<p>HACC: Extreme Scaling and Performance Across Architectures</p> <p>1:00 PM – 1:30 PM</p> <p>Salman Habib Senior Scientist Argonne National Laboratory</p>	<p>Cosmological simulations of the evolution of structure in the universe are among the most challenging of supercomputing tasks. Running on Titan, we use HACC (Hardware/Hybrid Accelerated Cosmology Code) to demonstrate how the use of GPUs to accelerate the force evaluations can lead to significantly improved performance.</p> <p>ACM Gordon Bell Finalist</p>
<p>New Features in NVIDIA® CUDA® 6 Make GPU Acceleration Easier</p> <p>1:30 PM – 2:00 PM</p> <p>Mark Harris Chief Technologist GPU Computing NVIDIA</p>	<p>The performance and efficiency of CUDA, combined with a thriving ecosystem of programming languages, libraries, tools, training, and services, have helped make GPU computing a leading HPC technology. Learn how powerful new features in CUDA 6 make GPU computing easier than ever, helping you accelerate more of your application with much less code.</p>

