GPU Technology Theater @ SC12 (Booth #2217)

SCHEDULE SUBJECT TO CHANGE - **All times are Mountain Time Zone (GMT-7).**

Time	Talk Title	Presenter Name	Presenter Title	Presenter Organization	Talk Abstract
Monday, November 12, 2012					
19:30	**Inside the Kepler Architecture**	**Stephen Jones**	CUDA Software Engineer	NVIDIA	This presentation looks into the features of NVIDIA's latest Kepler GPU architecture. Join us as one of CUDA's language architects explains what's new, why it's exciting, and demonstrates the power of Kepler GPU accelerators with a real-time cosmology simulation in full 3D.
20:00	**Evolution of GPU Computing**	**Steve Scott**	CTO, Tesla GPU Computing Business Unit	NVIDIA	The GPU has evolved rapidly from its invention in 1999 as a "VGA Accelerator", becoming a massively parallel general purpose accelerator for heterogeneous computing systems. This talk will focus on significant milestones in GPU architecture and programming models, covering several key concepts that demonstrate why advances in GPU-accelerated computing performance and power efficiency will continue to outpace CPUs.
20:30	**Titan: ORNL's New Computer System for Science**	**Arthur (Buddy) Bland**	Projector Director - Oak Ridge Leadership Computing Facility	Oak Ridge National Laboratory	The Oak Ridge Leadership Computing Facility is deploying the Titan supercomputer in support of the U.S. Department of Energy's Office of Science programs. This talk will describe the Titan system and its use of NVIDIA's latest GK110 processor.
Tuesday, November 13, 2012					
10:30	**Inside the Kepler Architecture**	**Stephen Jones**	CUDA Software Engineer	NVIDIA	This presentation looks into the features of NVIDIA's latest Kepler GPU architecture. Join us as one of CUDA's language architects explains what's new, why it's exciting, and demonstrates the power of Kepler GPU accelerators with a real-time cosmology simulation in full 3D.
11:00	**"Big Data" Astronomical Data Analysis and Visualization**	**Amr Hassan**	Senior Software Engineer and Development Leader	Swinburne University of Technology	I will present a high-performance; graphics processing unit (GPU)-based framework for the efficient analysis and visualization of ``big data'' astronomical data cubes. Using a cluster of 96 GPUs, we demonstrate for a 0.5 TB image: volume rendering at 10 fps; computation of basic statistics in 1.7 s; and evaluation of the median in 45s. The framework is one of the first solutions to the image analysis and visualization requirements of next-generation telescopes, including the forthcoming SKA pathfinder telescopes.
11:30	**Real-time Triggering Using GPUs in High Energy Physics**	**Felice Pantaleo**	Physicist	CERN	In the field of high energy physics, several groups are pursuing the use of GPUs for data analysis and for Monte Carlo simulations of particle interactions. The use of GPUs presented in this seminar is different: GPUs are employed for taking decisions in a trigger system, both as coprocessors in high level software trigger or "embedded" in real-time, fixed-latency hardware trigger.
12:00	**Many-GPU Calculations in Lattice Quantum Chromodynamics**	**Justin Foley**	Postdoctoral Research Associate	University of Utah	Lattice Quantum Chromodynamics (QCD) is a computational approach to the theory of the strong nuclear force. State-of-the-art calculations involve integrals over a billion variables or more, which are evaluated using Monte Carlo methods. Such calculations are typically performed on large-scale distributed systems. In this talk, we outline the main steps in a lattice calculation, and describe multi-GPU implementations of the core routines. We focus in particular on the sparse-matrix linear solves which dominate lattice QCD calculations, and involve significant inter-GPU communication. Preconditioning methods that substantially reduce inter-GPU communication, and hence improve processor utilization, are discussed.
12:30	**Running the FIM and NIM Weather Models on GPUs**	**Mark Govett**	Computer Scientist	NOAA Earth System Research Laboratory	Two U.S. global-scale weather models, developed at NOAA, are running on GPUs. The FIM runs at 15 KM resolution and is expected to be run by the U.S. National Weather Service in the next year. The NIM is a next-generation forecast model designed to run at 4KM resolution. This presentation will give an update on our efforts to parallelize and run these models on GPUs.
13:00	**Bringing GPU Computing to the Masses with OpenACC**	**John Urbanic**	Parallel Computing Specialist	Pittsburgh Supercomputing Center	The OpenACC standard has emerged as a solution for GPU computing in projects where CUDA programming resources are not available, or where code maintenance issues prevent its use. We have been teaching OpenACC to scientists and others with great success and will discuss this approach.
13:30	**Turbulence Visualization at the Terascale on Desktop PCs**	**Marc Treib and Kai Bürger**	PhD Student and Research Assistant	Technische Universität München	We present a high-performance, throughput-oriented compute system using CUDA to enable the visually guided interactive exploration of large-scale turbulent flows. Our system works on a compressed data representation, employing a wavelet-based compression scheme including run-length and entropy encoding, and efficiently intertwines on-the-fly data decoding and volume ray-casting.
14:00	**MAGMA - a New Generation of Linear Algebra Libraries for GPU and Multicore Architectures**	**Jack Dongarra**	Professor of Electrical Engineering and Computer Science	University of Tennessee and Oak Ridge National Lab	This talk will highlight the latest 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. This methodology is applied in MAGMA to develop high-performance fundamental linear algebra routines, such as the one-sided dense matrix factorizations (LU, QR, and Cholesky) and linear solvers, two-sided dense matrix factorizations (bidiagonal, tridiagonal, and Hessenberg reductions) for singular and eigenvalue problems, in addition to iterative linear and eigenvalue solvers. MAGMA is designed to be similar to LAPACK in functionality, data storage, and interface, in order to allow scientists to effortlessly port any of their LAPACK-relying software components to take advantage of the new architectures.

GPU Technology Theater @ SC12 (Booth #2217)

SCHEDULE SUBJECT TO CHANGE - **All times are Mountain Time Zone (GMT-7).**

Time	Talk Title	Presenter Name	Presenter Title	Presenter Organization	Talk Abstract
14:30	**Beyond Tsubame 2.0**	**Satoshi Matsuoka**	Professor	Tokyo Institute of Technology	Tsubame2.0 has been in successful production for the last 2 years, producing numerous research results and accolades. With possible upgrade of the GPUs to Kepler 2s, it will have the capability to surpass the 10 petaflops-class supercomputers in single-precision applications, without any increase in the power consumption of 1MW average.
15:00	**CUDA Development Using NVIDIA Nsight, Eclipse Edition**	**David Goodwin**	Manager – CUDA Visual Tools	NVIDIA	NVIDIA Nsight, Eclipse Edition for Linux and Mac is an all-in-one development environment that lets you develop, debug and optimize CUDA code in an integrated UI environment. This talk provides a detail usage walk-through of the fully CUDA aware source editor, build integration of the CUDA tool chain, graphical debugger for both CPU and GPU, and graphical profiler to enable performance optimization. If you've been waiting for a CUDA IDE on Linux and Mac then this talk is for you.
15:30	**CUDA-Accelerated Libraries**	**Levi Barnes**	DevTech Software Engineer	NVIDIA	This presentation will be an overview of several libraries in the CUDA SDK and other third-party libraries including cuBLAS, cuRAND, NPP and Thrust. Using these libraries can often significantly shorten the development time of a GPU project while leading to high-performance, high-quality software. We will discuss common use cases and the strengths of individual libraries and also provide guidance for selecting the best library for your project.
16:00	**High-Productivity Development with the Thrust Parallel Algorithms Library**	**Thomas Bradley**	DevTech Software Engineer	NVIDIA	Thrust is a parallel algorithms library which resembles the C++ Standard Template Library (STL). Thrust's high-level interface greatly enhances developer productivity while enabling performance portability between GPUs and multicore CPUs. Interoperability with established technologies (such as CUDA, TBB and OpenMP) facilitates integration with existing software. In this talk we'll walk through the library's main features and explain how developers can build high-performance applications rapidly with Thrust.
16:30	**Compiling Parallel Languages with the NVIDIA Compiler SDK**	**Mark Harris**	Chief Technologist, GPU Computing Software	NVIDIA	NVIDIA's CUDA C/C++ Compiler (NVCC) is based on the widely used LLVM open source compiler infrastructure. This open foundation enables developers to create or extend programming languages with support for GPU acceleration using the CUDA Compiler SDK. In this talk you will learn how to use the NVIDIA Compiler SDK to generate high-performance parallel code for NVIDIA GPUs.
17:00	**Compiling Python to the GPU with Numba**	**Travis E. Oliphant**	CEO	Continuum Analytics, Inc.	GPUs can offer orders of magnitude speed-ups for certain calculations, but programming the GPU remains difficult. Using NVIDIA's new support of LLVM, Continuum Analytics has built an array-oriented compiler for Python called Numba that can target the GPU. In this talk, I will demonstrate how Numba makes programming the GPU as easy as a one-line change to working Python code.
17:30	**Improving Engineering Productivity with HPC and GPU-Accelerated Simulation**	**Ray Browell**	Lead Product Manager	ANSYS	High Performance Computing has been a mainstay of increased productivity for years now. But recently, GPUs have enabled another level of performance without the significant purchase and power consumption required by additional nodes. ANSYS, Inc. continues to develop customer focused HPC solutions incorporating the latest hardware technologies including NVIDIA GPUs.
Wednesday, November 14, 2012					
10:30	**New Features in CUDA 5**	**Mark Harris**	Chief Technologist, GPU Computing Software	NVIDIA	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. In this talk you will learn about powerful new features of CUDA 5 and the Kepler GPU architecture, including CUDA Dynamic Parallelism, CUDA device code linking, and the new NSight Eclipse Edition.
11:00	**Fast Simulations of Fast Processes - How GPUs Give Insight Into the Interaction of Light and Matter**	**Michael Bussmann**	Group Leader Computational Radiation Physics	Helmholtz-Zentrum Dresden-Rossendorf	Laser-driven radiation sources can potentially help us to cure cancer or understand the dynamics of matter on the atomistic scale. With GPUs we today can simulate these sources at a frames-per-second rate. This in turn enables us to make them affordable to more users than ever before.
11:30	**Quarks, GPUs and Exotic Nuclear Matter**	**Balint Joo**	Staff Computer Scientist	Thomas Jefferson National Accelerator Laboratory	Hadronic matter, such as protons and neutrons, is composed of quarks bound together by gluons whose interactions are described by Quantum Chromodynamics (QCD). In this talk I will describe the use of GPUs for computing the spectrum of QCD. Of special interest are exotic states, such as those which will be sought by the Glue-X experiment of the Jefferson Lab 12 GeV upgrade, since these states can elucidate the role of the gluons. The calculations range from capacity work on small partitions with a few GPUs to capability sized partitions such as will be available in the Titan system at the Oak Ridge Leadership Facility (OLCF) and I will discuss the work on scaling our application (the Chroma code combined with the QUDA library for QCD on GPUs) to such large systems.

GPU Technology Theater @ SC12 (Booth #2217)

SCHEDULE SUBJECT TO CHANGE - **All times are Mountain Time Zone (GMT-7).**

Time	Talk Title	Presenter Name	Presenter Title	Presenter Organization	Talk Abstract
12:00	**Recent Progress on "Path to Exascale" Computational Challenges in Fusion Energy Sciences**	**William Tang**	Head, Fusion Simulation Program	Princeton Plasma Physics Laboratory	Advanced computing is recognized as a vital tool for accelerating progress in scientific research in the 21st Century. The fusion energy research community has made excellent progress in developing advanced codes with associated programming models for which computer runtime and problem size scale well with the number of processors on massively parallel supercomputers. Come see examples of algorithmic progress from the Fusion Energy Sciences area.
12:30	**Molecular Dynamics with LAMMPS on a Hybrid Cray Supercomputer**	**W. Michael Brown**	Computational Scientist	National Center for Computational Sciences, Oak Ridge National Laboratory	We present software development efforts in LAMMPS that allow for acceleration with GPUs on supercomputers. We present benchmark results for solid-state, biological and mesoscopic systems along with results from simulation of liposomes, polyelectrolyte brushes, and copper nanostructures on graphite. We present methods for efficient simulation with GPUs at larger node counts.
13:00	**Productive Performance on the Cray XK System Using OpenACC Compilers and Tools**	**Luiz DeRose**	Programming Environment Director	Cray Inc.	The Cray XK high level parallel programming environment was developed to help the widespread adoption of GPUs in HPC. Ease of use is possible with OpenACC compilers, making it feasible for users to write applications in Fortran, C, or C++, and tools and libraries to help users port, debug, and optimize for hyrid systems.
13:30	**Accelerating Compute-Intensive Processing with Hybrid GPU Parallelization and Cloud Computing**	**Ravi Kunju**	Senior Director, Strategy and Marketing, Enterprise Solutions	Altair Engineering	In this presentation, Altair will discuss how innovative hybrid parallelization using multiple GPUs and MPI dramatically reduces runtime for certain classes of compute-intensive workloads. Offloading intensive computations on the GPU and using heterogeneous computing with optimized workload management improves performance; users also benefit from simplified, accelerated access to compute resources via cloud portals.
14:00	**CUDA: Past, Present and Future**	**Ian Buck**	General Manager, GPU Computing	NVIDIA	The GPU evolved from its humble beginnings as a "VGA Accelerator" to become a massively parallel general processor for heterogeneous computing systems. Once the opportunity became obvious, the challenge was how to best develop a general purpose programming model to preserve the GPU's architectural advantage. Learn how CUDA as a parallel computing platform and programming model came to be, how it's being leveraged in a wide range of fields, and get an exciting preview of where it's going.
14:30	**Preparing S3D for Titan from a Domain Scientist's Perspective**	**Ray Grout**	HPC Applications Researcher	National Renewable Energy Laboratory (NREL)	Readying S3D, an explicit solver for the compressible reacting Navier-Stokes equations, for Titan took place in conjunction with an effort to move the code from an MPI-everywhere design to a hybrid MPI+x design. The design trade offs and considerations in this process will be described that led to a code ready for large scale GPU computing.
15:00	**NVIDIA Application Lab at Jülich**	**Dirk Pleiter**	Group Leader "Application Oriented Technology Development"	Jülich Supercomputing Centre	The NVIDIA Application Lab at Julich, established by JSC and NVIDIA in June 2012, aims on enabling scientific applications for GPU-based architectures. Selected applications and their performance characteristics will be presented. Strategies for multi-GPU parallelizations (necessary to meet computing demands) will be discussed.
15:30	**Accelerating Science and Engineering with Kepler GPUs in Blue Waters**	**Wen-mei Hwu**	Professor	University of Illinois at Urbana-Champaign	The Blue Waters supercomputer at the University of Illinois is the fastest production machine for the US NSF community, with over 11PF in peak performance. I will give an overview of the use of Kepler GPUs in Blue Waters and how we expect them to accelerate science and engineering work by more than 32 peta-scale application teams.
16:00	**A Large-scale LES Wind Simulation using Lattice Boltzmann Method for a 10km x 10km Area in Tokyo on TSUBAME 2.0**	**Takayuki Aoki**	Professor / Deputy Director	Tokyo Institute of Technology	Lattice Boltzmann Method describes airflow around complicated building structures with newly-developed LES (Large-Eddy Simulation) model without taking spatial average to specify a constant value. We carry out a 4000-GPU run to study a turbulent flow for 10km x 10km Area in Tokyo with high Reynolds number on TSUBAME 2.0
16:30	**The PGI Accelerator Compilers and OpenACC**	**Michael Wolfe**	Compiler Engineer	The Portland Group, Inc.	OpenACC uses directives and API library routines to program GPUs and accelerators using standard C, C++ and Fortran. We discuss the status of OpenACC and the PGI Accelerator compilers, and present advanced features that will soon be available.
17:00	**Technical Challenges on the Road to Exascale**	**Bill Dally**	Chief Scientist and SVP of Reseach	NVIDIA	From cell phones to supercomputers the scaling of future information systems depends on dramatic improvements in energy efficiency and parallel programming. To build an ExaScale supercomputer this decade with reasonable power, we must reduce operation energy from about 2nJ today to less than 20pJ. Cell phone, server, and embedded processors require similar improvements. Only 3-4x of this 100x reduction is expected to come from improved semiconductor technology. Because most of the energy is spent moving data, architecture, applications, and programming systems must work together to exploit all available locality. Energy efficient, throughput-optimized cores are also required. Future performance improvements will come from increased parallelism. Single thread performance is no longer increasing. Incremental approaches to parallel programming are problematic. However, with appropriate hardware mechanisms, locality-aware, fine-grain parallel programming can be made both efficient and productive.

GPU Technology Theater @ SC12 (Booth #2217)

SCHEDULE SUBJECT TO CHANGE - **All times are Mountain Time Zone (GMT-7).**

Time	Talk Title	Presenter Name	Presenter Title	Presenter Organization	Talk Abstract
17:30	**Using the Titan Supercomputer for Transformational Materials Science Calculations**	**Markus Eisenbach**	Director of Science	Oak Ridge National Laboratory	Coming soon.
Thursday, November 15, 2012					
10:30	**Tesla Cluster Monitoring and Management APIs**	**Robert Alexander**	Software Engineer	NVIDIA	Learn more about cluster management and monitoring of NVIDIA GPUs. This includes a detailed description of the NVIDIA Management Library (NVML) and user-facing third party software. Additionally, the nvidia-healthmon GPU health check tool will be covered.
11:00	**CARMA: CUDA on ARM Architecture – Developments in Power-Efficient Computing**	**Don Becker**	Systems Architect	NVIDIA	The world's largest computer systems are increasingly being limited by power and thermal limitations. This previous generation's focus on cost-effective and long-term-viable compute platforms has been expanded with the additional requirement for power-efficient computing. CARMA is the introductory system for developing the software ecosystem for the next generation of power-focused, high performance computing. It joins the power efficient high compute performance of a GPGPU, with a power-efficient ARM host processor. Both elements leverage the commodity market cost-effectiveness and leading edge development. This talk will introduce the CARMA hardware, outline its performance characteristics, describe the initial software system, and demonstrate the operational system.
11:30	**CUDA Fortran 2013**	**Brent Leback**	Engineering Manager	The Portland Group, Inc.	We will present an overview of features added to CUDA Fortran in the last year, demonstrate support for CUDA 5.0 features in the upcoming PGI 2013 Release of CUDA Fortran, show how CUDA Fortran can interoperate with other languages and GPU programming models, and preview some new ideas for multi-gpu support.
12:00	**Hybrid CPU-GPU Solutions for Weather and Cloud Resolving Climate Simulations**	**Thomas Schulthess**	Professor of Computational Physics and Director	Swiss National Supercomputing Center	Reliable weather prediction for the Alpine region and cloud resolving climate modeling require simulations that run at 1-2 km resolution. Additionally, since the largest possible ensembles are needed, high fidelity models have to run on the most economical resource in a given time to solution. In this presentation we will give an update on the refactoring of COSMO, a widely used production code in academia as well as seven European weather services, and discuss the performance experience on hybrid CPU-GPU systems.
12:30	**Accelerated ANSYS Fluent: Algebraic Multigrid on a GPU**	**Robert Strzodka**	NVAMG Project Lead	NVIDIA	NVIDIA's NVAMG library is a sophisticated suite of multi-level linear solvers. I will present an overview of our approach to parallelizing all phases of algebraic multigrid, including hierarchy construction and ILU factorization and solve. I will also describe how NVAMG provides GPU acceleration to the coupled incompressible solver in ANSYS Fluent 14.5.
13:00	**Using CAPS Compilers on NVIDIA Kepler and CARMA Systems**	**Francois Bodin**	CTO	CAPS	CAPS compilers provide directive based programming for the Kepler and CARMA systems. They support OpenACC and OpenHMPP directive styles. CAPS compilers are based on a source-to-source specific technology that allows to leverage accelerator's native compilers as well as host ones. This talk is focused on achieving code portability with other accelerator technologies.
13:30	**Introduction to CUDA C/C++**	**Mark Ebersole**	CUDA Educator	NVIDIA	Learn how to access the massively parallel processing power of NVIDIA GPUs using CUDA C and C++. We'll start with a simple "Hello Parallelism!" program and progress on to something a little more complicated. You will see what actually happens when you compile & run and how to add GPU+CPU hybrid computing concepts to accelerate your applications.
14:00	**New Features in CUDA 5**	**Mark Harris**	Chief Technologist, GPU Computing Software	NVIDIA	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. In this talk you will learn about powerful new features of CUDA 5 and the Kepler GPU architecture, including CUDA Dynamic Parallelism, CUDA device code linking, and the new NSight Eclipse Edition.
14:30	**Inside the Kepler Architecture**	**Stephen Jones**	CUDA Software Engineer	NVIDIA	This presentation looks into the features of NVIDIA's latest Kepler GPU architecture. Join us as one of CUDA's language architects explains what's new, why it's exciting, and demonstrates the power of Kepler GPU accelerators with a real-time cosmology simulation in full 3D.