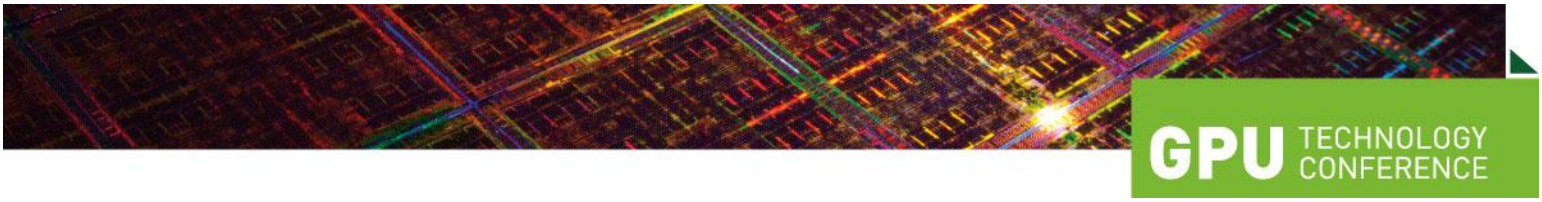

GPU Technology Conference, May 14-17, 2012
McEnery Convention Center, San Jose, California
www.gputechconf.com

Sessions on General Interest (subject to change)

IMPORTANT: Visit http://www.gputechconf.com/page/sessions.html for the most up-to-date schedule.

TUTORIALS

S0005 - Languages, APIs and Development Tools for GPU Computing
Will Ramey (NVIDIA)
Day: Monday, 05/14 | **Time:** 9:00 am - 10:20 am
Topic Areas: General Interest; Development Tools & Libraries; Application Design & Porting Techniques
Session Level: Beginner

Get a head start on the conference with this first-day introduction to key technologies for GPU Computing. This 90-minute tutorial session will cover the key features and differences between the major programming languages, APIs and development tools available today. Attendees will also learn several high level design patterns for consumer, professional and HPC applications, with practical programming considerations for each.

SESSIONS

S0527 - GPUs and the Next-Generation Aerial Surveillance
Nikola Bozinovic (MotionDSP)
Day: Tuesday, 05/15 | **Time:** 9:00 am - 9:25 am
Topic Areas: General Interest
Session Level: Beginner

Graphics processors are already used for computationally intensive video tasks in many ISR (Intelligence, Surveillance, Reconnaissance) applications; GPU-based system for video enhancement and analytics outperforms a similarly priced CPU-based system 5-to-1 at HD resolutions. Our initial tests on 64 megapixel Wide Area Aerial Surveillance (WAAS) data show at least 10x speedup with tasks such as super-resolution or moving target indication. In this talk, we'll discuss unique design and implementation challenges of real-time processing of very large video data sets. We will demonstrate our existing GPU-based software, IKENA ISR, and discuss its video-processing pipeline and innovative processing solutions that are promising to dramatically expand capabilities of emerging aerial surveillance platforms.

S0040 - Introducing CUDA in KBE Applications for Digital Vehicle Development Programs
Avijit Santra (Tata Motors Limited)
Day: Tuesday, 05/15 | **Time:** 9:30 am - 9:55 am
Topic Areas: General Interest
Session Level: Intermediate

Get the latest development in Next Generation Knowledge Based Engineering (KBE) software which provides real results over the traditional design approach. Today there exist numerous KBE applications in the field of vehicle

ergonomics, suspension, NVH, safety, regulations etc which deal with huge number of iterations and mathematical algorithm. With GPU computing and CUDA the KBE kernel is restructured to incorporate parallel programming model which helps the applications run faster and achieving time reduction from hours to seconds. KBE geometry kernel also gets benefited by enabling CUDA in topology based operations which take lot of time when performed on CPU.

S0013 - GPUs for Fast Triggering in NA62 Experiment
Gianluca Lamanna (CERN), Marco Sozzi (Physics Department of Pisa)
Day: Tuesday, 05/15 | **Time:** 10:00 am - 10:25 am
Topic Areas: General Interest
Session Level: Intermediate

We discuss an approach for using commercial graphic processors (GPUs) at the earliest trigger stages in high-energy physics experiments, and study its implementation on a real trigger system in preparation. In particular we focus on the possibility to reconstruct rings in a Cherenkov detector as building block of a selective trigger condition for rare decay search. Latency and processing rate measurements on several state-of-the-art devices are presented, and the potential issues related to processing time jitter and data transfer throughput are discussed.

S0379 - GPU-based High-Performance Simulations for Spintronics
Jan Jacob (University of Hamburg - Institute of Applied Physics and Microstructure Research Center)
Day: Tuesday, 05/15 | **Time:** 2:30 pm - 2:55 pm
Topic Areas: General Interest; Computational Physics; Application Design & Porting Techniques
Session Level: Intermediate

The joint utilization of the electron's charge and spin in "spintronics" represents a promising technology for data processing and storage in nanostructures. The complex quantum effects like the spin-Hall effect in these devices require demanding numerical simulations providing a convenient link between idealized analytical models to often very complex results from measurements. The simulations involving multiplications and inversions of large matrices provide an ideal showcase for performance gain by employing GPGPUs in the execution of the algebraic routines on these matrices in computing environments with shared execution of algorithms on multiple nodes with multiple GPGPUs and CPU cores.

S0050 - High Performance Logic Simulation with GPUs
Yangdong Deng (Tsinghua University)
Day: Tuesday, 05/15 | **Time:** 4:00 pm - 4:50 pm
Topic Areas: General Interest; Algorithms & Numerical Techniques
Session Level: Advanced

Verification has become the bottleneck of IC design process due to its fast increasing complexity. The fundamental means of verifying digital circuits is logic simulation, which can be performed at both register-transfer level (RTL) and gate level. In this work, we developed GPU based logic simulation solutions. We implemented a Chandy-Misra-Bryant parallel simulation protocol on GPUs for sufficient parallelism. A dynamic GPU memory allocator was introduced to efficiently manage GPU memory resources. RTL simulation is performed in a compiled-code scheme by translating Verilog code into equivalent CUDA code. Experimental results proved that the GPU simulators significantly outperform their CPU counterparts.

S0253 - Sensor Processing with Rugged Next-Generation NVIDIA GPUs (Presented by GE Intelligent Platforms)
Dustin Franklin (GE Intelligent Platforms)
Day: Wednesday, 05/16 | **Time:** 9:00 am - 9:50 am
Topic Areas: Audio, Image and Video Processing; General Interest; Machine Vision; Computer Vision
Session Level: Intermediate

Swimming in sensors and drowning in data? Turn the tide on high-bandwidth sensors with rugged next-generation GPUs from NVIDIA. See how we deploy NVIDIA GPUs into the most extreme of environments, providing GPGPU capabilities onboard platforms where SWaP and GFLOPS/watt is key. Dig into four realtime CUDA sensor processing applications - Hyperspectral Imaging, Wide-Area Surveillance, 360Â° Situational Awareness, and GSM cellular SIGINT. Discuss the CUDA algorithms, interconnects, and rugged platforms behind each. Learn how we utilize GPUDirect and realtime Linux for improved latency and determinism.

S0016 - NVIDIA Grad Fellowship Fast Forward
David Luebke (NVIDIA Research)
Day: Wednesday, 05/16 | **Time:** 10:00 am - 10:50 am
Topic Areas: General Interest
Session Level: Beginner

We invite you to a special presentation from our 2011-2012 Graduate Fellowship recipients to learn "what's next" in the world of research and academia. The NVIDIA Graduate Fellowship recipients were selected from 200 applications in 27 countries. Sponsored projects involve a variety of technical challenges, including computer architecture, computer vision, programmability and optimization for heterogeneous systems, automotive computing and much more. We believe that these minds lead the future in our industry and we are proud to support the 2011-2012 NVIDIA Graduate Fellows. For more information on the 2011-2012 NVIDIA Graduate Fellows, please visit www.NVIDIA.com/fellowship.

S0352 - GPU-Accelerated Parallel Computing for Simulation of Seismic Wave Propagation
Taro Okamoto (Department of Earth and Planetary Sciences, Tokyo Institute of Technology)
Day: Wednesday, 05/16 | **Time:** 10:30 am - 10:55 am
Topic Areas: Computational Physics; General Interest
Session Level: Advanced

We adopted GPU to accelerate large-scale, parallel finite-difference (FDTD) simulation of seismic wave propagation. Effective parallel implementation is needed because the size of the memory of a single GPU is too small for real applications. Thus we describe the memory optimization, the three-dimensional domain decomposition, and overlapping the communication and computation adopted in our program. We achieved so far a high performance (single-precision) of about 61 TFlops by using 1200 GPUs of TSUBAME-2.0, the GPU supercomputer in Tokyo Institute of Technology, Japan. As an important application, we show the results of the simulation of the 2011 Tohoku-Oki mega-quake.

S0259 - A High Performance Platform for Real-Time X-Ray Imaging
Suren Chilingaryan (Karlsruhe Institute of Technology)
Day: Wednesday, 05/16 | **Time:** 3:00 pm - 3:25 pm
Topic Areas: General Interest; Supercomputing; Audio, Image and Video Processing; Algorithms & Numerical Techniques
Session Level: Intermediate

We will share our experience on development of the GPU-based platform for synchrotron-based X-ray imaging aimed to analysis of dynamic processes. The complete data flow from the camera to the data storage will be

discussed with a special focus on I/O issues, hardware platform, and ways to utilize the available system resources. An efficient GPU-implementation of filtered back projection will be presented highlighting differences of implementations for GT200, Fermi, and AMD Cypress architectures. We will introduce our software platform used to abstract current configuration of the imaging station and to simplify the development of parallel image processing algorithms.

S0311 - Teaching Applied Parallel Computing with GPUs
Chris Lupo (California Polytechnic State University)
Day: Wednesday, 05/16 | **Time:** 5:30 pm - 5:55 pm
Topic Areas: General Interest; Ray Tracing
Session Level: Intermediate

Learn how the next generation of HPC developers are learning hands-on skills with GPUs, and how GPU computing is being incorporated into Computer Science courses. We will discuss how GPUs are being used to enhance student learning of parallel computing concepts through a cross-teaching approach, where students with different domain expertise are grouped into teams and tasked with parallelizing an application such as ray tracing. We'll show that student projects that emphasize optimization of architectural resources and performance tuning allow students with no prior experience to parallelize a large-scale application with significant performance improvement in as little as six weeks.

S0290 - Algorithm Acceleration for Geospatial Analysis
James Goodman (HySpeed Computing LLC), Matthew Sellitto (Northeastern University)
Day: Thursday, 05/17 | **Time:** 9:30 am - 9:55 am
Topic Areas: Algorithms & Numerical Techniques; General Interest
Session Level: Intermediate

Learn how the power of GPU computing is being leveraged to accelerate algorithms in the field of geospatial image analysis. The data volume and computation requirements associated with geospatial imagery are rapidly expanding as a result of the increasing number of satellite and airborne sensors, greater data accessibility, and expanded utilization of data intensive technologies. This equates to a growing need for high-performance computing in this field. We demonstrate the capacity for GPU computing to meet this need by accelerating a complex non-linear optimization algorithm used for the mapping and assessment of coral reef ecosystems.

S0354 - Bcl::ChemInfo Suite Enables Machine Learning-Based Drug Discovery Using GPUs
Edward Lowe (Vanderbilt University), Nils Woetzel (Vanderbilt University)
Day: Thursday, 05/17 | **Time:** 9:30 am - 9:55 am
Topic Areas: General Interest
Session Level: Intermediate

High-throughput screening data allows the training of machine learning quantitative structure activity relationship models which can be used for in silico drug discovery screening. Here, we present a GPU-accelerated suite for descriptor generation, model training, feature selection, and data set similarity analysis, bcl::ChemInfo. The suite provides functionality for the analysis of constructed models as well as for screening external libraries of compounds. We examine case studies illustrating how this workflow can now be completed in a single day on a Tesla equipped workstation with speedups reaching 300x providing a complete GPU-accelerated cheminformatics framework for drug discovery.

S0360 - Set GPUs Free: Integrating a File System with CUDA Programs
Mark Silberstein (UT Austin), Emmet Witchel (UT Austin)
Day: Thursday, 05/17 | **Time:** 9:30 am - 9:55 am
Topic Areas: General Interest
Session Level: Intermediate

This session seeks the answer to the question: “Can we simplify and speed up CUDA programs by allowing them to access files residing on a host?" To prove our affirmative answer, we demonstrate how the concept of a file system enables programs with non-trivial CPU-GPU and GPU-GPU interactions to be efficiently and easily implemented on top of a new GPU file-system layer. We also show that such a file system enables implementation of fully stand-alone GPU programs without any CPU wrapper code. Finally we outline the details of the file system design which contributed to scalability, data consistency and performance.

S0521 - Desktop Supercomputing in the Soft-Matter Physics Laboratory
Peter Lu (Harvard University)
Day: Thursday, 05/17 | **Time:** 10:00 am - 10:50 am
Topic Areas: General Interest
Session Level: Beginner

While many GPGPU applications reside on large clusters, in many laboratories the time to move data to an external cluster would exceed the time to analyze it upon arrival. By bringing high-throughput computational power to the data in the laboratory, GPUs offer new capabilities in doing science. This session offers a number of ways in which GPUs are making a significant impact on our research in experimental physics, biology and chemistry, from designing and building apparatus (Quadro and Tesla), to collecting data on portable devices (Tegra), to high-throughput analysis of large data sets (Tesla). It also presents results from studies investigating the motion of diffusing and aggregating colloidal particles and swimming bacteria, observing liquid-gas phase separation onboard the International Space Station, applying high dynamic-range techniques to optical tomography, and using low-cost devices to detect chemical and microbial contamination in the third world.

S0320 - PTask: OS Support for GPU Dataflow Programming
Christopher Rossbach (Microsoft Research Silicon Valley), Jon Currey (Microsoft Research Silicon Valley)
Day: Thursday, 05/17 | **Time:** 2:00 pm - 2:50 pm
Topic Areas: Development Tools & Libraries; General Interest; Parallel Programming Languages & Compilers
Session Level: Advanced

This session considers the PTask API, OS-level abstractions that support GPUs as first-class computing resources, and supports a dataflow programming model. With PTask, the programmer specifies where data goes, rather than how and when it should get there, allowing the system to provide fairness and isolation guarantees, streamline data movement in ways that currently require direct programmer involvement, and enable code portabality across diverse GPU-based platforms. Our experience building the PTask APIs shows that PTask can provide important system-wide guarantees and can enable significant performance benefits, for example improving the throughput of hand-tuned CUDA programs by up to 2x.

S0032 - Teraflop GPU Acceleration of Large Matrix Algebra
Ronald Young (Multipath Corporation)
Day: Thursday, 05/17 | **Time:** 2:30 pm - 2:55 pm
Topic Areas: Development Tools & Libraries; General Interest
Session Level: Beginner

Learn how Multipath's Fast Matrix Solver (FMS) is setting performance records using multiple GPU's solving large matrices in production applications. By (1) leveraging NVIDIA's CUBLAS library, (2) operating multiple GPUs in parallel and (3) overlapping data transfers with computation, FMS averages over 2 teraflops of performance, even on jobs lasting for days. The presentation also includes a description of what problems FMS solves and how it is incorporated into applications programs.

S0392 - Large-Scale First Principle Pseudopotential DFT Calculations on GPU Clusters
WeiLe Jia (Supercomputing Center of CNIC, Chinese Academy of Sciences), Long Wang (Supercomputing Center of CNIC, Chinese Academy of Sciences)
Topic Areas: Quantum Chemistry; General Interest
Session Level: Advanced
Day: Thursday, 05/17 | **Time:** 3:30 pm - 4:20 pm

In this session, we will present a series of work on density functional theory (DFT) plane wave pseudopotential(PWP) calculations on GPU clusters. The GPU version is developed based on a CPU DFT-PWP code: PEtot, which can calculate ~1000 atoms on thousands of processors. Our test indicates that the GPU version can have a ~20 times speedup over CPU code. A detail analysis of the speed-up and the scaling on the number of CPU/GPU(up to 256) will be presented.As far as we know, this is the first GPU DFT-PWP code scalable to large number of CPU/GPU.

S0065 - Satellite HUB Communication System GPU Based
Gaetano Mendola (MBI srl), Francesco Basile (MBI srl)
Day: Thursday, 05/17 | **Time:** 4:30 pm - 5:20 pm
Topic Areas: General Interest
Session Level: Intermediate

In the last few years the increasing GPU computational power has opened new perspectives in telecommunication fields trough SDR (software defined radio) approach. Some tasks, such as the one we had to deal with, do not offer negotiation margins with the execution speed due to the real-time analysis of a radio signal. We coped with the implementation of the lowest layer in the protocol stack for a land mobile satellite communication system, and we were able to deliver a product with a reduced time to market with respect to traditional FPGA approach.

S0134 - On the Integration of OpenCL into a Software Defined Radio
Michael Dickens (University of Notre Dame)
Topic Areas: General Interest
Session Level: Intermediate
Day: Thursday, 05/17 | **Time:** 5:30 pm - 5:55 pm

We will present a software defined radio system that allows for heterogeneous processing using a host computer's CPUs and GPUs, via dynamic runtime resource allocation provided by our Surfer framework and extensions to it using OpenCL. This system collects runtime statistics including samples / second throughput for each signal processing block, data transfer latency between different processors, and the host CPU cores' loads. Using this information, a supervisor can move computations between processors during runtime, without interrupting data processing. We will demonstrate an OFDM transmitter, graphing the system throughput and CPU loads while selecting where processing occurs for each block.