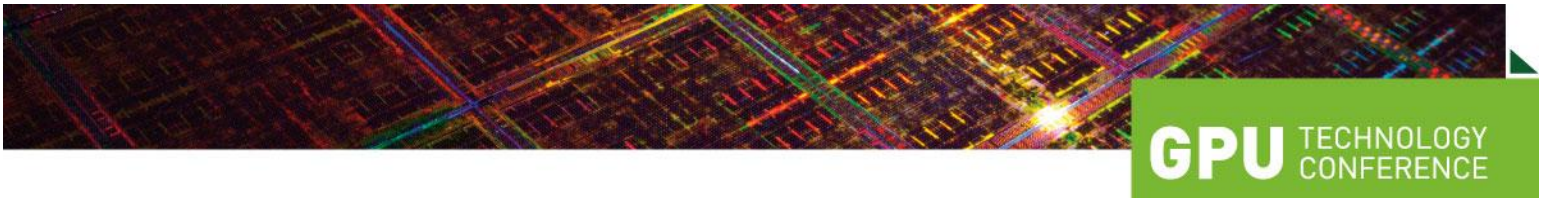


GPU Technology Conference, May 14-17, 2012
McEnery Convention Center, San Jose, California
www.gputechconf.com

Sessions on Bioinformatics (subject to change)

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

S0337 - High-Throughput Epistasis Screening Using GPUs
Mark Seligman (Insilicos LLC)
Day: Tuesday, 05/15 | **Time:** 9:00 am - 9:25 am
Topic Areas: Bioinformatics; Life Sciences; Supercomputing; Cloud Computing
Session Level: Intermediate

Epistasis is the interaction of two or more genes in coding for a biological property. Epistasis is believed to be an important factor in an individual's susceptibility to disease, and the search for epistasis is a major component in the development of personalized approaches to genomic medicine. Statistical tests for epistasis are typically confounded by the multiple-testing problem, that is, the aggregated loss of precision incurred through repeated hypothesis testing. One way to circumvent this problem is to simulate a false-discovery rate via resampling. We report success in using GPUs to accelerate these highly compute-intensive resampling techniques.

S0083 - Swift: A GPU-based Smith-Waterman Sequence Alignment Program
Pankaj Gupta (St Jude Children's Research Hospital)
Day: Tuesday, 05/15 | **Time:** 9:30 am - 9:55 am
Topic Areas: Bioinformatics
Session Level: Beginner

This session describes Swift, a GPU-based Smith-Waterman implementation for aligning short DNA sequences to large genomes. Swift has been designed to reduce computation time and lower hardware cost. Also, unlike other leading GPU-based Smith-Waterman sequence alignment programs like CUDASW++ and SWCUDA which focus on protein sequence alignment, Swift has been developed for DNA sequence alignment. Swift performs 200x faster than CUDASW++ using a test data set containing 1000 reads (100 bases each) and 1000 references (1000 bases each), and it performs 11x faster than the CPU-based implementation of Smith-Waterman using 24 million reads (100 bases each) and human chromosome 1.

S0376 - Dynamic Programming on CUDA: Finding the Most Similar DNA Sequence
Grzegorz Kokosiński (IBM Poland) , Krzysztof Zarzycki (IBM Poland)
Day: Tuesday, 05/15 | **Time:** 10:00 am - 10:25 am
Topic Areas: Bioinformatics; Algorithms & Numerical Techniques
Session Level: Intermediate

Learn a couple of techniques to speed up compute-heavy Dynamic Programming algorithms on the GPU. Our particular problem regarded DNA sequences: given a reference sequence, how to find the one most similar to it among a large database? The sequences are millions characters long, and their similarity is calculated with a (quadratic) DP algorithm, which makes the problem very tough even for the GPUs. We speed up both the theoretical and practical side: we present programming techniques that enable Dynamic Programming to be

performed at the hardware speed, and improvements to the algorithm itself that drastically lower the execution time.

S0519 - GPU Accelerated Bioinformatics Research at BGI
Bing Qiang Wang (BGI)
Day: Tuesday, 05/15 | **Time:** 2:00 pm - 2:25 pm
Topic Areas: Bioinformatics; Life Sciences; Algorithms & Numerical Techniques; Supercomputing
Session Level: Intermediate

After digitizing DNA double helix by sequencing, computation is the key connecting raw sequences with life science discoveries. As massive data is generate, how to process and analysis as well as storage them in an efficiently manner turns out to be a major challenge. By developing GPU accelerated bioinformatics tools and integrate them into pipelines, BGI researchers now run analysis pipelines in several hours instead of several days. These tools include SOAP3 aligner, SNP calling and tool for population genomics. The speed up is generally around 10-50x comparing with traditional counterparts.

S0084 - CUMACH - A Fast GPU-based Genotype Imputation Tool
Hu Xiaohan (NVIDIA)
Day: Tuesday, 05/15 | **Time:** 2:30 pm - 2:55 pm
Topic Areas: Bioinformatics
Session Level: Intermediate

The goal of this session is to introduce a GPU-implemented tool in bioinformatics. Genotype imputation is method which extrapolates genetic correlations from a densely characterized reference panel to a sparsely typed study sample. There have already been lots of CPU-based tools, but they all cost lots of time for large data-set. In this session, we try to implement a GPU-based imputation tool which can get relatively good result and fast speed. There will be three main parts for the session: 1) Introduce the background and its HMM based algorithm, 2) GPU implementation and optimization, 3) Results.

S0109 - SOAP3: GPU-based Compressed Indexing and Ultra-fast Parallel Alignment of Short Reads
Tak-Wah Lam (The University of Hong Kong), Chi-Man Liu (The University of Hong Kong)
Day: Tuesday, 05/15 | **Time:** 3:00 pm - 3:25 pm
Topic Areas: Bioinformatics
Session Level: Advanced

We give the fi_x000C_rst implementation of a compressed index (Burrows-Wheeler Transform) on the GPU, supporting very efficient parallel alignment of short patterns (reads) onto the human genome. The new alignment software SOAP3 is tens of times faster than existing ones and can catch up the throughput (Giga to Tera bp) of next generation DNA sequencer. It takes 2.4 seconds to perform exact matching for one million length-100 reads (tens of seconds for small-error approximate matching). Technically, we show how to minimize memory accesses to the index from individual threads and to control the branching and divergence of the threads.

S0152 - Accurate Sequence Alignment using Distributed Filtering on GPU Clusters
Reza Farivar (University of Illinois at Urbana-Champaign), Shivaram Venkataraman (University of Illinois at Urbana Champaign)
Day: Tuesday, 05/15 | **Time:** 3:30 pm - 3:55 pm
Topic Areas: Bioinformatics; Algorithms & Numerical Techniques
Session Level: Intermediate

Learn how GPUs enable new ways to rethink a complex bioinformatics problem: Accurate sequence alignment. What was once prohibitive to compute can become the basic block of novel GPU-based algorithms. Modern DNA sequencing machines generate enormous amounts of short sequences within minutes, and they should be aligned to a reference genome in real time. Most solutions only find a few locations that match a short sequence. We introduce a new technique to find all matching locations inside a reference sequence for a given number of mismatches. Our technique is based on a distributed filtering scheme and GPU based processing.

S0008 - Algorithms and Tools for Bioinformatics on GPUs
Bertil Schmidt (Nanyang Technological University)
Day: Tuesday, 05/15 | **Time:** 4:00 pm - 4:50 pm
Topic Areas: Bioinformatics; Life Sciences
Session Level: Intermediate

Learn how to use GPUs to accelerate compute- and data-intensive applications and algorithms Bioinformatics. High-throughput techniques for DNA sequencing and gene expression analysis with microarrays have led to a rapid growth in the amount of digital biological data, e.g. the NCBI Sequence Read Archive (SRA) houses raw sequence data generated by next-generation sequencing (NGS) technologies which succeeds 25 trillion base-pairs. Therefore, modern bioinformatics tools need to be scalable; i.e. they need to deal with an ever growing amount of data. GPUs and CUDA provide the opportunity to significantly reduce the runtime of many biological algorithms on inexpensive hardware.

S0037 - SeqNFind™: Application Of CUDA GPU Technologies To Sequence Alignment Techniques
D. Andrew Carr (Accelerated Technology Laboratories, Inc.)
Day: Tuesday, 05/15 | **Time:** 5:00 pm - 5:25 pm
Topic Areas: Bioinformatics; Algorithms & Numerical Techniques
Session Level: Advanced

Explosive growth in the amount of genomic data has created a need for faster systems that align and compare nucleotide sequences. With the development of tools for leveraging the massively parallel architecture of NVIDIA GPUs it is a logical next step to construct algorithms for genomic analysis on GPU clouds/clusters. Although a seemingly simple task, there are a number of challenges to deploying the current algorithms. Every algorithm from Smith-Waterman to BLAST has its own unique set of barriers. Presented here some of the lessons learned and how ongoing genomic research projects have benefitted from the increased speed and accuracy.

S0156 - Towards Computing the Cure for Cancer
Wu Feng (Virginia Tech)
Day: Tuesday, 05/15 | **Time:** 5:00 pm - 5:50 pm
Topic Areas: Bioinformatics; Life Sciences; Supercomputing; Algorithms & Numerical Techniques
Session Level: Intermediate

Learn about how to create "designer" genomic analysis pipelines as part of the "Compute the Cure" for cancer initiative from NVIDIA Foundation. Get an overview of an open-source framework that enables the creation of customized genomic analysis pipelines. Discover how different plug-ins from the “mapping/realignment/discovery" repositories, respectively, can be composed to form a genomic analysis pipeline. Learn to use next-generation sequencing data to characterize previously undetectable genetic changes between normal and malignant cells. Find out how you can contribute to the "Compute the Cure" cause.

S0072 - GPU-Enabled Spatiotemporal Model of Stochastic Cardiac Calcium Dynamics and Arrhythmias
Mohsin Jafri (George Mason University), Hoang-Tron Minh Tuan (George Mason University)
Day: Wednesday, 05/16 | **Time:** 9:00 am - 9:50 am
Topic Areas: Life Sciences; Bioinformatics
Session Level: Beginner

Calcium ions play a central role controlling the contraction of the heart to pump blood. This requires tight regulation of cellular calcium dynamics which depends upon over 1,000,000 calcium channels that open and close stochastically and have a very specific spatial arrangement. In the School of Systems Biology at George Mason University, CUDA technology coupled to novel algorithms for Monte Carlo simulation have made possible this computationally expensive spatiotemporal model of calcium dynamics in the heart muscle cell to study the regulation of calcium dynamics and what aberrations leads to cardiac arrhythmia.

S0262 - GPU-Accelerated Model-Based Drug Development
Chee Ng (Children Hospital of Philadelphia/University of Pennsylvania)
Day: Wednesday, 05/16 | **Time:** 10:00 am - 10:25 am
Topic Areas: Life Sciences; Algorithms & Numerical Techniques; Bioinformatics
Session Level: Beginner

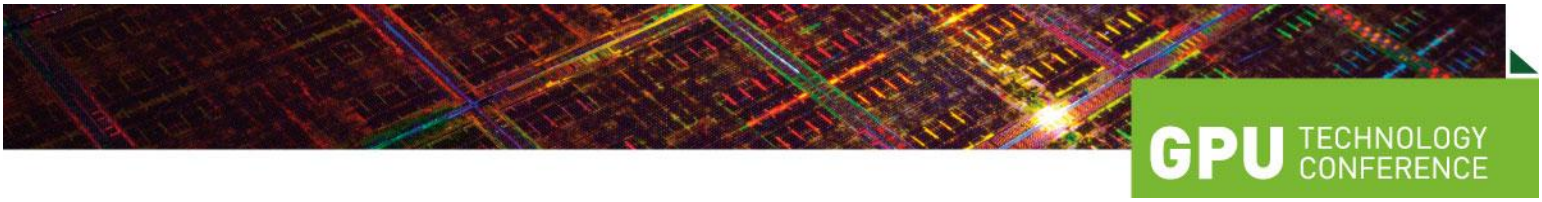
Explore how GPUs can be used to improve the efficiency of drug development. Drug development is a very time-consuming, complex and expensive process that has low successful rate. A model-based drug development paradigm has been proposed as a possible solution to overcome these problems. A key challenge is to develop computational intensive drug and disease-specific models from a large quantity of highly complicated preclinical and clinical data. This session will describe how GPUs can and will play a key role in shortening the model development times and improving the efficiency of model-based drug development.

S0272 - GPU GWAS - CUDA Based Genome Wide Association Studies
Tim Bi (Johns Hopkins University / George Mason University)
Day: Wednesday, 05/16 | **Time:** 10:30 am - 10:55 am
Topic Areas: Life Sciences; Bioinformatics
Session Level: Intermediate

We have developed a CUDA based GWAS analyzer that has achieved a 10x analysis speed-up per GPU. Genome wide association studies scans through millions of SNP markers across the human genome seeking the genetic basis of life threatening diseases such as coronary artery disease and prostate cancer. The prospect of the $1,000 genome heralds a potential new scale of GWAS involving hundreds of thousands of patients. We will discuss how we utilized the Python, R, and C languages to produce a robust GWAS algorithm that can be extended to multiple GPUs and GPU clusters.

S0327 - Large and Sparse - Mass Spectrometry Data Processing in the GPU
Jose de Corral (Waters Corporation)
Day: Wednesday, 05/16 | **Time:** 2:00 pm - 2:50 pm
Topic Areas: Life Sciences; Bioinformatics
Session Level: Intermediate

Learn how the GPU helps identify millions of ions in datasets of several billion points of four-dimensional sparse data. The data is first reduced to 3D to locate regions of dense data, and then only those regions are processed in 4D. Processing involves combining several steps of convolution filters in three axes, finding local maximums in volumes of data, and extracting information from the data around each local maximum.



S0103 - Accelerating Protein Sequences and Classification using GPU-HMMER Search
Mahesh Khadtare (I2IT, Pune University)
Day: Wednesday, 05/16 | **Time:** 3:30 pm - 3:55 pm
Topic Areas: Life Sciences; Bioinformatics
Session Level: Intermediate

In this paper we present the results of parallelizing HMMer, which is a widely used tool for protein sequence homology detection, as well as functional annotation of homologous protein sequences, and protein family classification. The HMMer program is based upon a Viterbi algorithm coded in C, and is quite time consuming. We modify the Viterbi algorithmic logically to port it on GPGPU. We test multiple enhancements in our GPU kernels in order to demonstrate the effectiveness of each strategy. Our implementation cuda_hmmsearch achieves overall up to 30x speedup over intel single core CPU.

S0516 - The Advantage of GPU Computation for Analyzing Complex Traits
Jun Zhu (Zhejiang University)
Day: Wednesday, 05/16 | **Time:** 4:00 pm - 4:50 pm
Topic Areas: Bioinformatics; Life Sciences
Session Level: Intermediate

Most import agriculture traits and human diseases are complex traits which are controlled by gene network with gene by gene interaction (epistasis) and gene by environment interaction (GE). New statistic methods and software are developed for analyzing genetic architecture for complex traits based on genome-wide association study (GWAS). When deal with large mapping population and huge amount of molecular information, GPU computation has an advantage over CPU computation. We will demonstrate the newly developed GPU based software QTLNetwork V3.0 and GWAS-GMDR for mapping genes with epistasis and GE interaction for complex traits of human, crops, and mouse.

S0303 - GPU Acceleration for Threshold Based Region Growth Algorithms.
Supratik Moulik (University of Pennsylvania), Jason Walsh (University of Pennsylvania 3D lab)
Day: Thursday, 05/17 | **Time:** 9:00 am - 9:50 am
Topic Areas: Medical Imaging & Visualization; Bioinformatics
Session Level: Beginner

Come learn how the massively parallel computing power of modern GPUs help to create faster and more accurate volume rendered images for the medical imaging community. Attendees of this session will gain insight into how GPUs can accelerate region growth algorithms and how these algorithms can be optimized for the latest generation of NVIDIA hardware. Topics covered will include fundamental of region growth, GPU implementations, and practical examples of vessel tracking algorithms based on GPU accelerated algorithms.