

# GPU accelerated signal processing in Ion Torrent analysis pipeline

Mohit Gupta  
Ion Torrent by Life Technologies

## ABSTRACT

The area of DNA sequencing has always been at the forefront of high performance computing due to vast amounts of computing power needed to complete the journey from the raw data produced by the sequencers to DNA sequences comprising nucleotides (T,A,C,G). In an attempt to reduce the time to discovery in life sciences, new technologies providing high performance computing are always explored and adopted. Cloud computing has been explored as a viable option since data analysis jobs can be delegated to compute farms equipped with tremendous computing power available through cheap commodity hardware. But, this is not an option for all due to high bandwidth network connection required to the cloud. A more practical solution is to scale the computing power tremendously by using a coprocessor like GPU alongside the traditional multicore CPU. We, at Ion Torrent, have adopted this solution to provide fast analysis results to our customers by accelerating our signal processing pipeline using Tesla C2050 GPU. This poster presents a high level view of GPU application to our processing pipeline.

## CONTACT

Mohit Gupta  
Ion Torrent by Life Technologies  
Email: [mohit.gupta@lifetech.com](mailto:mohit.gupta@lifetech.com)  
Website: [www.lifetechologies.com](http://www.lifetechologies.com)

Poster Design & Printing by GenGraphics® - 800.790.4301

## INTRODUCTION

Ion Torrent pairs proprietary semiconductor technology with a simple sequencing chemistry that is based on a well-characterized biochemical process. In nature, when a nucleotide is incorporated into a strand of DNA, a hydrogen ion is released as a byproduct as shown in Figure 1. Ion Torrent uses a high density of micro machine wells etched on a semiconductor chip to perform this biochemical process in a massively parallel way. These hydrogen ions are sensed by ion sensors laid under these wells and a corresponding voltage signal is generated. These voltage signals are processed by the primary analysis pipeline to generate sequencing data. There is a strong background signal riding on these voltage signals and signal processing is all about separating this noise from the actual voltage signals. The number of wells on Ion chips (314, 316, 318) varies from 1.3 million to 12 million.

## SIGNAL PROCESSING

Divide and conquer strategy is adopted in processing the complete chip.

- Chip area is divided into small regions which can undergo signal processing independently to highly parallelize the algorithm.
- Each chip region is submitted as a processing job to a shared queue.
- There is a Proton Flux model and Incorporation Signal model at play to model the net hydrogen ion generation, nucleotide diffusion into the well and reaction with polymerase.
- Custom Levenberg-Marquardt non-linear regression algorithm is used to fit the model to the measured data.
- Some parameters are fitted for each well and others are fitted for the region.
- The number of wells need to undergo the model fitting and the iterative nature of the Levenberg-Marquardt algorithm makes it a computational bottleneck in the analysis pipeline.

## GPU Implementation

GPU has been seamlessly integrated into the signal processing pipeline as a coprocessor. The implementation of the signal processing pipeline on GPU takes advantage of several powerful features in cuda 4.0 to maximize the utilization of the GPU.

- Multiple threads share the same GPU since they have the same context.
- Page locked host memory has been used for a lot of host allocations to have faster transfer time across the PCI-Express bus.
- Look up table for Poisson CDF approximation is used in the fitting model. Constant memory has been used to store these tables for faster access times.
- Matrix data is over allocated in Levenberg-Marquardt algorithm causing memory access to happen in multiples of 64 which increases the effective memory bandwidth.
- Cuda Math functions like erf() and exp() are used which are highly optimized for GPU.
- Access patterns for the data buffers doesn't permit coalesced reads and writes as the algorithm progresses. Shared memory has been used to counter this bottleneck.

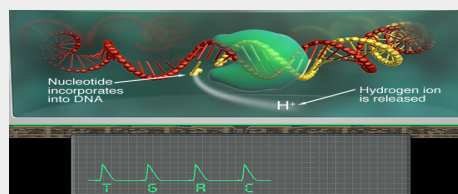


Figure 1. Incorporation Reaction

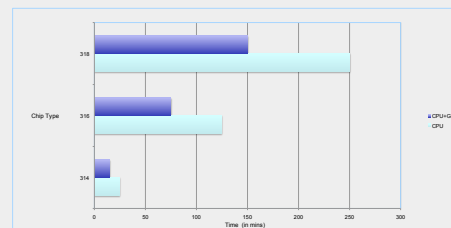


Figure 2. Run time for signal processing for different chip type

## RESULTS AND DISCUSSION

The signal processing pipeline runs on a dual hex-core processor. We have been able to achieve 10X speedup over a single CPU core (2.6 GHz) using a Tesla C2050 GPU. The overall speed boost to the signal processing pipeline is about 40% since a single GPU is competing with 12 CPU cores. The region with higher number of wells takes more time to process than a region with a lower number of wells. The shared queue is sorted such that the regions with higher number of wells are at the front. GPU threads finish faster than CPU threads and are able to process heavier workloads better, therefore, they consume work from the front of the queue. GPU is not being utilized to its full potential as the algorithm progresses in the fit and less number of wells have to be processed with each successive iteration. This phenomena interferes with data access pattern in the GPU kernels and there is an increased amount of uncoalesced reads and writes. We have found that it is beneficial to delegate the task back to the CPU when the number of wells to process reaches below a threshold. At this point, GPU takes more time than CPU because the memory access latency becomes huge in GPU kernels.

## CONCLUSIONS

GPU has considerably sped up our signal processing pipeline. The iterative nature of the fitting algorithm doesn't allow the most efficient utilization of this hardware. The GPU kernels performance is limited by the global memory latency due to uncoalesced reads and writes. New ways of laying out the data buffers have to be explored to realize order of magnitude improvement from this massively parallel computing device.

## REFERENCES

1. [www.lifetechologies.com](http://www.lifetechologies.com)
2. [www.iontorrent.com](http://www.iontorrent.com)
3. [www.nvidia.com](http://www.nvidia.com)

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.

© 2011 Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation and/or its affiliate(s) or their respective owners.