

# CS267 Assignment 0: GPU-BLAST

Yinghui Huang  
02-06-2013

## 1. Bio

I am Yinghui Huang (English name is Kris), an exchange student learning computer science in UC Berkeley for this semester. My major is Mathematics & Applied Mathematics and this is the fourth year of my undergraduate study. I am fascinated by parallel computing since I was a member of a project aimed at speeding up BLAST by using graphics processors while generating the same result with NCBI-BLAST, where I obtained my first learning about parallel programming in my home university.

In this course, I plan to learn parallel-computing systematically and know more about parallel algorithm design. My interests are including but not limited to high performance computing and algorithms in numerical computing and machine learning. I am a fast learner, new knowledge is always welcomed.

## 2. BLAST and GPU-BLAST

BLAST, short for Basic Local Alignment Search Tool, is developed by National Center for Biotechnology Information (NCBI), based on BLAST algorithm delivered by Stephen Altschul et al. in 1990. It “finds regions of local similarity between sequences.” “The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.”<sup>[1]</sup> The result from NCBI-BLAST has been always regarded authoritative.

GPU-BLAST, developed by Carnegie Mellon University in 2010, is “an accelerated version of the NCBI-BLAST.” “The implementation is based on the source code of NCBI-BLAST, thus maintaining the same input and output interface while producing identical results. In comparison to the sequential NCBI-BLAST, the speedups achieved by GPU-BLAST range mostly between 3 and 4.”<sup>[2]</sup>

## 3. GPU programming

Theoretically speaking, GPU, which has a great number of arithmetic units and share memory within blocks, is a satisfactory environment for alignment comparison. Every subject in database is almost mutually exclusive finding similarity between input-sequence. The size of input-sequence is usually small to put into share memory. Thus, you can separate sequences of database for comparison in different arithmetic units without too much communication.

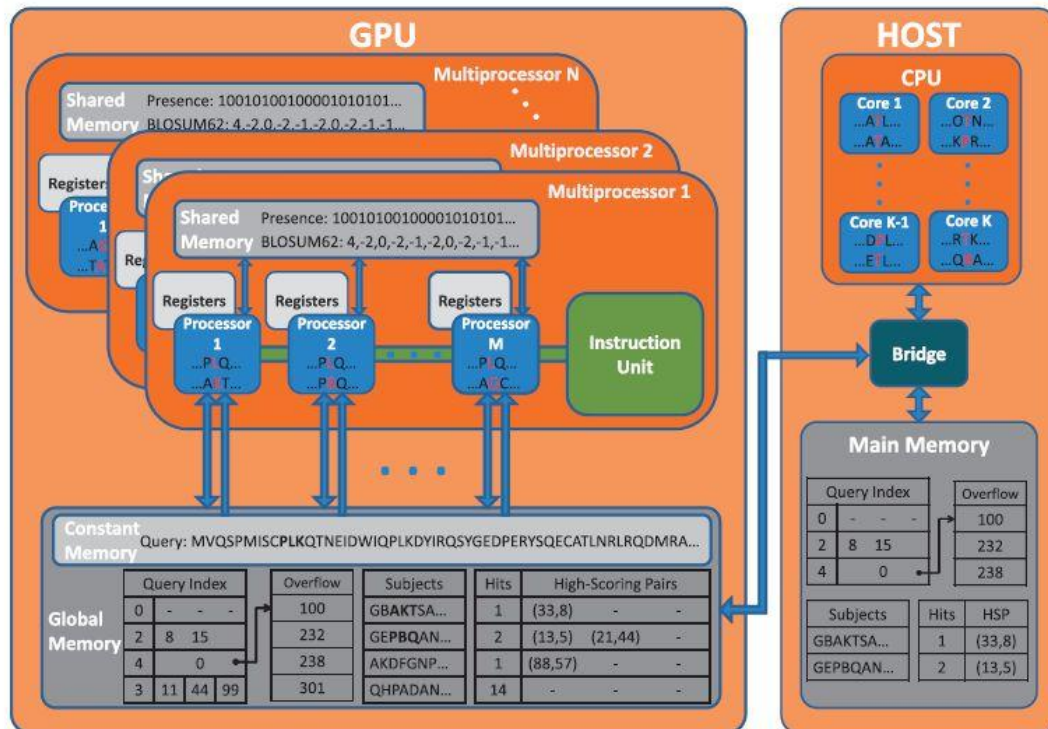


Figure 1<sup>[2]</sup>: the architecture and memory organization of NVIDIA's GPU executing GPU-BLAST

GPU-BLAST is writing on NVIDIA's CUDA (Compute Unified Device Architecture), which is a parallel computing platform and programming model facilitates GPU programming with C compiler.<sup>[3]</sup>

#### 4. Difficulties and Future Improvement

##### A. Load balance

This difficulty is on account of data itself. GPU-BLAST distributes workload of every thread by the number of sequences. However, the sequences in database are not equal-size. And even when there are equal-size sequences, the time every thread spends on is unpredictable because more similar one costs more time to store information and further study. Hence, it is difficult to balance thread's workload.

##### B. Small size of share memory

According to the GPU-BLAST execution flow, gapped alignment-matches need to be re-calculated on CPU, which blocks the speed-up of this part which is usually required to be executed since gene mutation is common in biology. I think it is because the algorithm of gapped part requires more communication and more space to store intermediate results.

#### Reference

[1] <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

- [2] Panagiotis D. Vouzis and Nikolaos V. Sahinidis, *GPU-BLAST: using graphics processors to accelerate protein sequence alignment*. In *Bioinformatics*, 2010.
- [3] <https://developer.nvidia.com/what-cuda>