# Accelerating Mass Spectrometry-Based Protein Identification Using GPUs

You Li [1], Leihao Xia [1], Hao Chi [2,3], Xiaowen Chu [1,4,*]

[1]: Department of Computer Science, Hong Kong Baptist University.

[2]: Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China..

[3]: Graduate University of Chinese Academy of Sciences, Beijing 100049, China.

[4]: Institute of Computational and Theoretical Studies, Hong Kong Baptist University.

[*]: To whom correspondence should be addressed.

Emails: YL (youli@comp.hkbu.edu.hk); LX (10050760@life.hkbu.edu.hk); HC (hchi@jdl.ac.cn); XC (chxw@comp.hkbu.edu.hk)

Tandem mass spectrometry-based database searching is currently the principal method for protein identification in shotgun proteomics. The explosive growth of protein and peptide databases due to genome translations, enzymatic digestions, and post-translational modifications (PTMs), is making computational efficiency in database searching a serious challenge. Profile analysis shows that most search engines spend 50%-90% of their total time on the scoring module, and that the spectrum dot product (SDP) based scoring module is the most widely used. As a general purpose and high performance parallel hardware, graphics processing units (GPUs) are promising platforms for speeding up many bioinformatics tools [1] [2].

In this poster, we show our design and implementation of a parallel SDP-based scoring module on GPUs that exploits the efficient use of GPU registers and shared memory. Compared with the CPU-based version, we achieved a 30 to 60 times speedup using a single GPU. We also implemented our algorithm on a GPU cluster and achieved an approximately linear speedup.

Our GPU-based SDP algorithm can significantly improve the speed of the scoring module in mass spectrometry-based protein identification. The algorithm can be easily implemented in many database search engines such as X!Tandem, SEQUEST, and pFind.

More details about this work can be found in [3]. A software tool implementing this algorithm is freely available at [4].

## Acknowledgement

This work was supported by research grants FRG2/10-11/099 and FRG2/11-12/158 from Hong Kong Baptist University.

## References

1. C. M. Liu et. al. SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. Bioinformatics, 28(6), 878-879, 2012.
2. K. Zhao and X.-W. Chu. GPU-BLASTN: Accelerating Nucleotide Sequence Alignment by GPUs. Poster at RECOMB 2013.
3. Y. Li and X.-W. Chu. Speeding up Scoring Module of Mass Spectrometry Based Protein Identification by GPU, The Fifth International Symposium on Advances of High Performance Computing and Networking, Liverpool, UK, June 2012.
4. http://www.comp.hkbu.edu.hk/~chxw/ProteinByGPU.html