

Heterogeneous Processing of Large, Multidimensional Imaging Data Sets Generated by Optical Microscopy Combined with Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging.

David Kissick¹, Ta-Hsuan Ong², Stanislav Rubakhin³, and Jonathan Sweedler⁴

¹ University of Illinois at Urbana-Champaign, Department of Chemistry, Urbana, USA.

² University of Illinois at Urbana-Champaign, Department of Chemistry, Urbana, USA.

³ University of Illinois at Urbana-Champaign, Department of Chemistry, Urbana, USA.

⁴ University of Illinois at Urbana-Champaign, Department of Chemistry, Urbana, USA.

Our long-term goal is to be able to chemically characterize the cells in a population using mass spectrometry to find important rare events. We are using matrix assisted laser desorption/ionization mass spectrometry (MALDI-MS) as it has sufficient sensitivity to probe individual cells.[1,2] This measurement involves spreading the cells on a microscope slides, optically imaging the slide and obtaining thousands of spectra on the cells. Data analysis has become our limitation in this workflow, making efficient data processing essential to understanding these measurements. Currently available software is typically executed as single threads on a computer's central processing unit (CPU). Recently, the development of programming languages like CUDA and OpenCL allowed utilization of graphics processing units (GPUs) in general purpose processing of vast amounts of multidimensional data. A typical GPU is capable of running hundreds of parallel threads, speeding up execution of appropriate algorithms. This parallelization is ideal for processing multidimensional data sets in which the same operation is applied to many subsets. In this work, we present a GPU based toolset for processing large, multidimensional data sets acquired in microscopy and chemical imaging experiments.

These data processing improvements are demonstrated with MALDI MS neuropeptide profiling of neurons. Here we used two separate neuronal populations from the well-used neuronal model *Aplysia californica*, the bag cell neurons and pleural sensory neurons, as both have well defined peptide complements.[3,4] Locations of these individual cells on indium tin oxide-coated microscope slide are then determined for subsequent MALDI-MS analysis so that only cell locations, and not the empty spaces between cells, are used to collect mass spectra. In this process, darkfield and autofluorescence images are acquired simultaneously for the entire cell containing area on a slide at 10x magnification using an inverted microscope (Zeiss, Axiovert 200M), resulting in ~50GB of raw data (see Figure 1). A Python module using AMD's APPML libraries was developed to allow easy access to GPU based basic linear algebra subroutines (BLAS) and fast Fourier transforms (FFTs). The phase correlation method [5] was implemented in this module and used to stitch the greater than 700 individual images. Cell coordinates were extracted from the result using intensity thresholding and used to generate a custom geometry file to direct the MALDI-TOF mass spectrometer (Bruker Daltonics, UltrafleXtreme) to acquire spectra only from cells. Principal component analysis was also implemented for the GPU and used to discriminate between the two cell types based on the chemical profiles.

The use of GPU hardware results in a ten to one hundred fold decrease in execution time, but is strongly dependent on the hardware. Furthermore, the Python based wrapper for these GPU BLAS and FFT libraries will allow these methods to be applied to many other imaging data sets. Using this approach, we plan to implement new data mining algorithms such as artificial neural networks like auto-encoders[6]. The final outcome will be a workflow that enables rapid characterization of large

populations of cells for those with unusual chemical profiles, an important step forward in finding unexpected and rare events.

References:

- [1] Rubakhin SS, Romanova EV, Nemes P, Sweedler JV. *Nat Methods*. 2011 Apr;8(4 Suppl):S20-9.
- [2] Rubakhin SS, Lanni EJ, Sweedler JV. *Curr Opin Biotechnol*. 2013 Feb;24(1):95-104.
- [3] Hatcher NG, Sweedler JV. 2008 Jan;99(1):333-43.
- [4] Romanova EV, Sasaki K, Alexeeva V, Vilim FS, Jing J, Richmond TA, Weiss KR, Sweedler JV. 2012;7(11):e48764.
- [5] Kuglin CD, Hines DC. *Int. Conf. on Cybernetics and Society, Proc. IEEE*. 1975. 163–165.
- [6] Hinton GE, Salakhutdinov RR. *Science*. 2006. 504-507.
- [7] The authors acknowledge funding from the National Institutes of Health through 1R21 MH100704 and the National Science Foundation through CHE-11-11705. DJK acknowledges the support of the Springborn Postdoctoral Fellowship

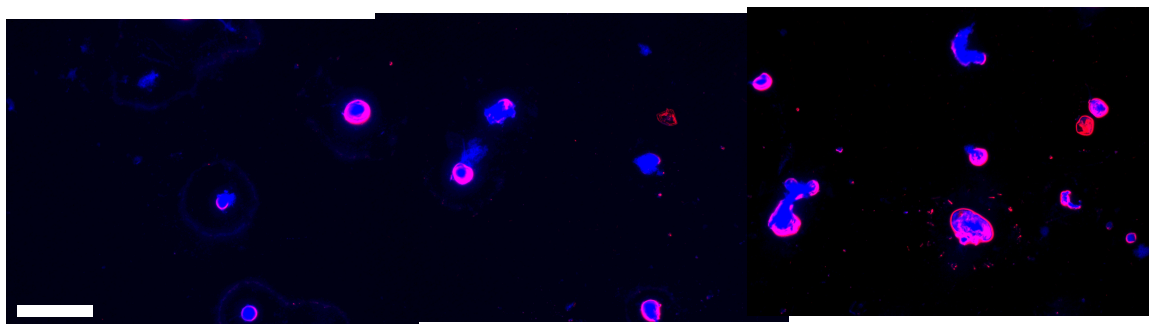


Figure 1. Optical imaging is used to find the cell soma on the slide, and the subsequent coordinates are used to guide mass spectral acquisitions of from these positions. Here, three adjacent optical images were used to locate the cells prior to MALDI mass spectrometry. The offsets show that stitching of these images is required. Blue corresponds to autofluorescence. Red corresponds to edges detected by dark field microscopy. Scale bar is approximately 100 microns.