May 2015

Scattering Correction Methods of Infrared Spectra Using Graphics Processing Units

Asher Imtiaz
University of Wisconsin-Milwaukee

SCATTERING CORRECTION METHODS OF INFRARED SPECTRA USING GRAPHICS PROCESSING UNITS

by

Asher Imtiaz

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Computer Science

at

The University of Wisconsin-Milwaukee

May 2015

ABSTRACT
SCATTERING CORRECTION METHODS OF INFRARED SPECTRA USING GRAPHICS PROCESSING UNITS

by

Asher Imtiaz

The University of Wisconsin-Milwaukee, 2015
Under the Supervision of Professor Roshan M. D'Souza

Fourier transform infrared (FTIR) microspectroscopy has been used for many years as a technique that provides distinctive structure-specific infrared spectra for a wide range of materials (e.g., biological (tissues, cells, bacteria, viruses), polymers, energy related, composites, minerals). The mid-infrared radiation can strongly scatter from distinct particles, with diameters ranging between 2-20 micrometer. Transmission measurements of samples (approximately 100 micrometers x 100 micrometers x 10 micrometers) with distinct particles. will be dominated by this scattering (Mie scattering). The scattering distorts the measured spectra, and the absorption spectra appear different from pure absorbance spectra. This thesis presents development and implementation of two algorithms for processing of FTIR spectra and evaluating the resulting mid-FTIR images. The first procedure removes Mie scattering spectral features, and shows resulting spectra and images to confirm that scattering intensity has been minimized, and the second procedure is a spatial deconvolution algorithm which is used to improve the contrast and fidelity of the imaging data.

Both the algorithms discussed in this thesis were implemented using Graphics Processing Units (GPUs) for fast hyperspectral processing by exploiting the parallelism in

distributed computational environment. 30x speedup was achieved in spatial deconvolution algorithm implementation as compared with MATLAB implementation of the same problem specifications. Scattering correction implementation on GPU achieved 10x speedup for single iteration as compared with previous MATLAB implementation. Next, some tests were run on real datasets and its' GPU implementation time is compared with previous implementation on CPUs. In the end some future directions and prospects are mentioned.

TABLE OF CONTENTS

1. INTRODUCTION 1

1.1. BACKGROUND 1

1.2. CONTRIBUTIONS 6

2. RELATED WORK 7

3. IMPLEMENTATION 11

3.1. GRAPHICS PROCESSING UNITS (GPUS) 11

4. DECONVOLUTION 16

4.1.1. USER INTERFACE FOR SPATIAL DECONVOLUTION 20

5. SCATTERING 22

5.1.1. USER INTERFACE FOR SCATTERING CORRECTION 25

6. SIMULATED DATA 27

7. DATA 30

7.1. SPIDER SILK 30

7.2. DRG NEURONS.....31

8. SCATTERING AND DECONVOLUTION.....33

8.1. RESULTS.....33

9. FUTURE WORK.....37

9.1. TOMOGRAPHY.....37

9.2. FTIR BIOSPECTROSCOPY.....40

9.3. GPU TECHNOLOGY.....41

10. REFERENCES.....43

LIST OF FIGURES

Figure 1: FTIR spectroscopy work flow for imaging and diagnosis. Reproduced from reference [17]. **3**

Figure 2: Overview of IRENI beamline. Reproduced from reference [10]. **5**

Figure 3: Hyperspectral image cube. Reproduced from reference [23]. **5**

Figure 4: Flowchart of the algorithm proposed by Bassan, et al., in [4] **9**

Figure 5: Growth of floating-point operations per second for the CPU and GPU. Reproduced from reference [21]. **12**

Figure 6: GPU Fermi architecture [29]. **16 multiprocessors are positioned around a common L2cache. Each orange portion is scheduler and dispatch, green portions are execution units and light blue portions are memory for register file and L1 cache** **13**

Figure 7: CUDA Model [21]. **Threads, blocks, and grids, with corresponding memory spaces for private per-threads, shared per-block, and global per-applications.** **15**

*Figure 8***: *Raw and deconvolved chemical images; Extracted single spectra from the original (left-hand side) and deconvolved data sets (right-hand side) taken along the white line in the respective images. Reproduced from reference [10].*** **18**

Figure 9: User interface for spatial deconvolution algorithm **21**

Figure 10: Reference Spectrum **24**

Figure 11: Real refractive index **24**

*Figure 12***: *User interface for scattering correction algorithm.*** **26**

Figure 13: Simulated data result with reference spectra as input Z_Raw. 27

Figure 14: Simulated data result 1. 28

Figure 15: Simulated data result 2. 28

Figure 16: Simulated data result 3. 29

Figure 17: Chemigrams of DRG Neuron dataset. Original raw data (left) and deconvolved (right) with 74x absorbtion option. 33

Figure 18: Chemigrams of spider silk image data 34

Figure 19: Average spectra of original image 35

Figure 20: Average spectra after scatter correction 35

Figure 21: Average spectra - Scatter correction and then deconvolution 36

Figure 22: Single pixel spectra of spider silk tomography dataset. Original data spectra (left), Scatter corrected then deconvolved (right) 36

Figure 23: FTIR spectro-microtomographic imaging of Zinnia. Reproduced from reference [8]. 39

ACKNOWLEDGEMENTS

I would like to begin by thanking my supervisor, Dr. Roshan M. D'Souza, to whom I am humbly indebted; I am grateful for the consistent guidance he has offered during the course of my research and studies. I also express my deepest appreciation and gratitude for my co-supervisor Dr. Carol Hirschmugl, for her generous support and invaluable encouragement. Additionally, I am thankful to my friends and colleagues, Ivan Komarov and Ebrahim Aboualizadeh, who were helpful not only during the implementation process, but also in the understanding of several parts of project. Likewise, I extend my gratitude to Eric Mattson for the insights he shared during the preliminary stages of the project. Finally, I am grateful to my masters thesis committee members, Professors Dr. Susan McRoy and Dr. Hossein Hosseini, for not only their time contributions, but also their helpful feedback in regards to this thesis. I would like to thank my friends Muhammad Aurangzeb, Muhammad Latif, Laeeq Aslam, Muhammad Ali, Sarmad Abbasi and Ali Dashti for their direct and indirect contributions at different stages of my academic life. In the end I would like to thank Dr. John Boyland for his continuous support throughout the course of my studies.

I would like to dedicate this thesis to my parents for their unconditional love.

1. Introduction

This thesis presents development and implementation of two algorithms for processing of Fourier transform infrared spectra and evaluating the resulting mid-FTIR images. The first procedure removes Mie scattering spectral features, and shows resulting spectra and images to confirm that scattering intensity has been minimized. The scattering correction algorithm is modified from the one presented in [4] and is implemented using Graphics Processing Unit. The second procedure is a spatial deconvolution algorithm which is used to improve the contrast and fidelity of the imaging data. In the first section of the thesis, some subject background of the FTIR imaging is summarized and contributions of this thesis are mentioned. In the next section, previous related work in the area is presented. In the end future directions of the work is mentioned specially highlighting microtomography for which these two processes are important to be computed efficiently.

1.1. Background

Spectroscopy that deals with the infrared region of the electromagnetic spectrum is known as Infrared spectroscopy or IR spectroscopy. It is based on absorption spectroscopy and covers a wide range of techniques. This technique can be used to identify and study chemicals just like with all other spectroscopic techniques.

In absorption spectroscopy technique we measure the absorption of radiation, as a function of frequency or wavelength, due to its interaction with a sample.

FTIR stands for Fourier Transform InfraRed and is considered as the preferred method of infrared spectroscopy. In infrared spectroscopy, infrared radiation is passed through a sample. Some of the infrared radiation is absorbed by the sample and some of it is passed through or transmitted. The resulting spectrum represents the molecular absorption and transmission, creating a molecular fingerprint of the sample. Like a fingerprint no two unique molecular structures produce the same infrared spectrum. This makes infrared spectroscopy useful for several types of analysis. FTIR can provide useful information such as it can identify unknown materials, it can determine the quality or consistency of a sample and it can also determine the amount of components in a mixture.

During the last decade, there has been a significant increase in the use of infrared micro-spectroscopy to study biological and biomedical samples. In the field of cancer diagnosis, several papers demonstrate that infrared micro-spectroscopy can be used to analyse biopsy samples, and in some cases more reliably than a trained histopathologist, eliminating subjectivity and offering the possibility of high-throughput screening. In addition to studying tissue samples at sub-cellular resolution there has been increasing interest in the study of single biological cells.

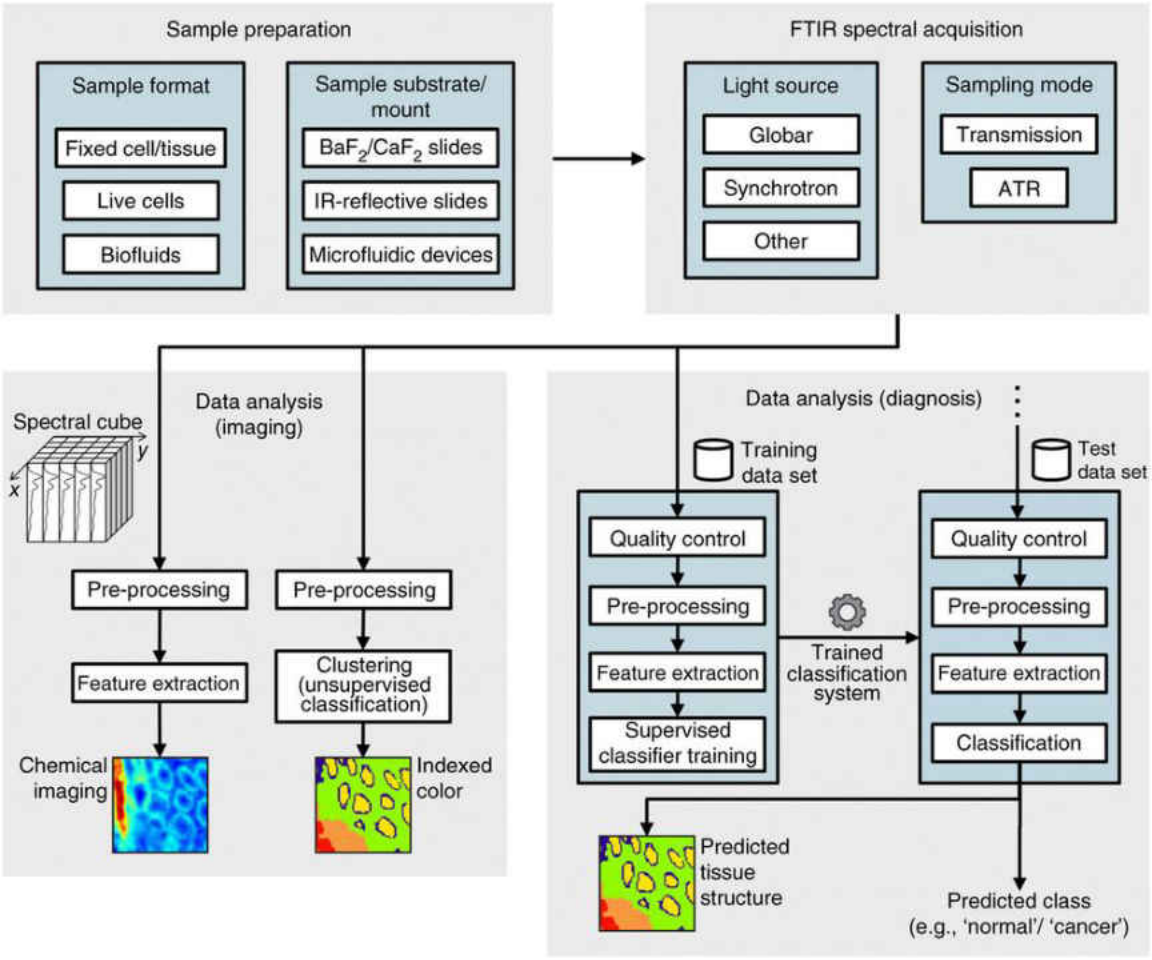


Figure 1: FTIR spectroscopy work flow for imaging and diagnosis. Reproduced from reference [17].

The application of Fourier-transform infrared (FTIR) spectroscopy in biomedical research is relatively novel. Although the physical and mathematical principles behind FTIR spectroscopy have been laid for over a century, it is only in the last fifty years or so that the required technology – both instrumental and computational – has been fully developed to make FTIR spectroscopy instruments available. [1] summarizes in detail the major applications of FTIR spectroscopy seen over the past decades which includes a diverse range of investigations. In the survey [1] gives the following examples and

applications of FTIR spectroscopy; the technique is used to study the presence and development of cancer in cervical cytology; in tissue samples such as brain, colon, lymph, colorectal, prostate, skin, liver, endometrium, blood vessels, bone marrow, cervix, breast, larynx, and stomach; toxicological studies such as the Syrian hamster embryo (SHE) assay; environmental studies; presence of disease in biofluids such as serum; taxonomic identification; detection of cellular mechanisms; and, detection of stem cells in cornea and intestine. [1]

Synchrotron-based multiple-beam Fourier transform infrared (FTIR) chemical imaging known as IRENI (infrared environmental imaging) has recently been developed at the Synchrotron Radiation Center (SRC) in Madison, WI, USA [6]. The infrared imaging data used in the work presented in this thesis is all collected on IRENI at SRC. Overview of the IRENI beamline is illustrated in Figure 2 and explained in details in [6].

In FTIR based imaging, the data is collected as a hyperspectral cube (x,y, Abs(v)), which convolutes the sample information with the point spread function (PSF) of the instrument. Hyperspectral cube is illustrated in Figure 3.

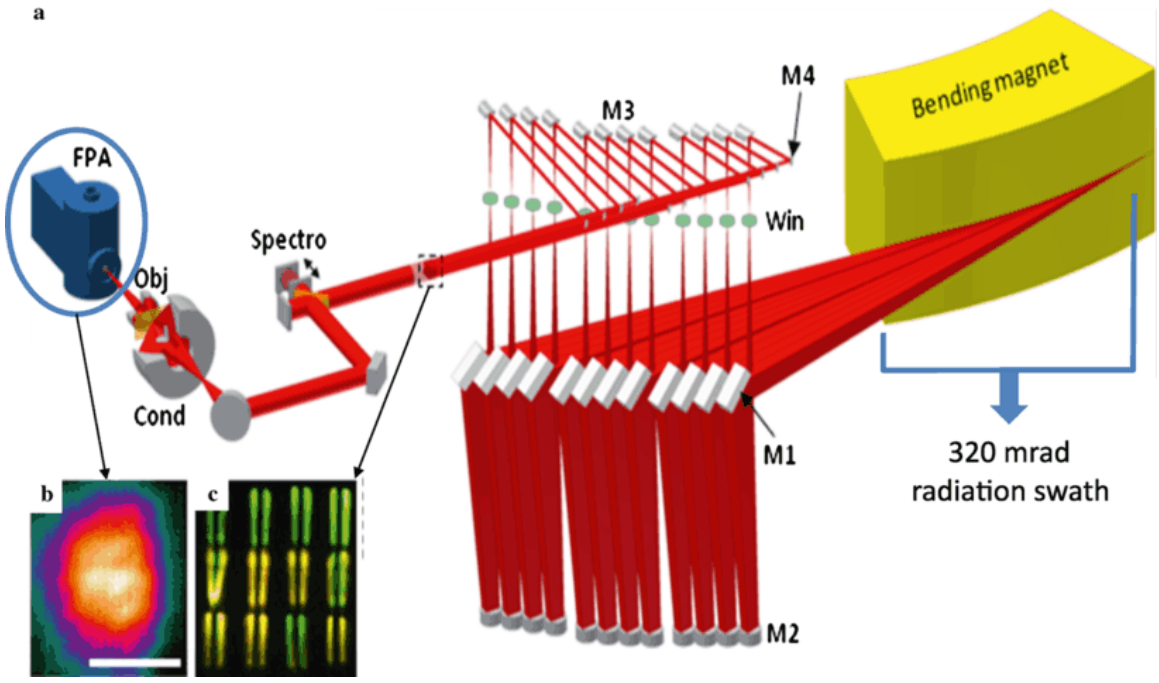


Figure 2: Overview of IRENI beamline. Reproduced from reference [10].

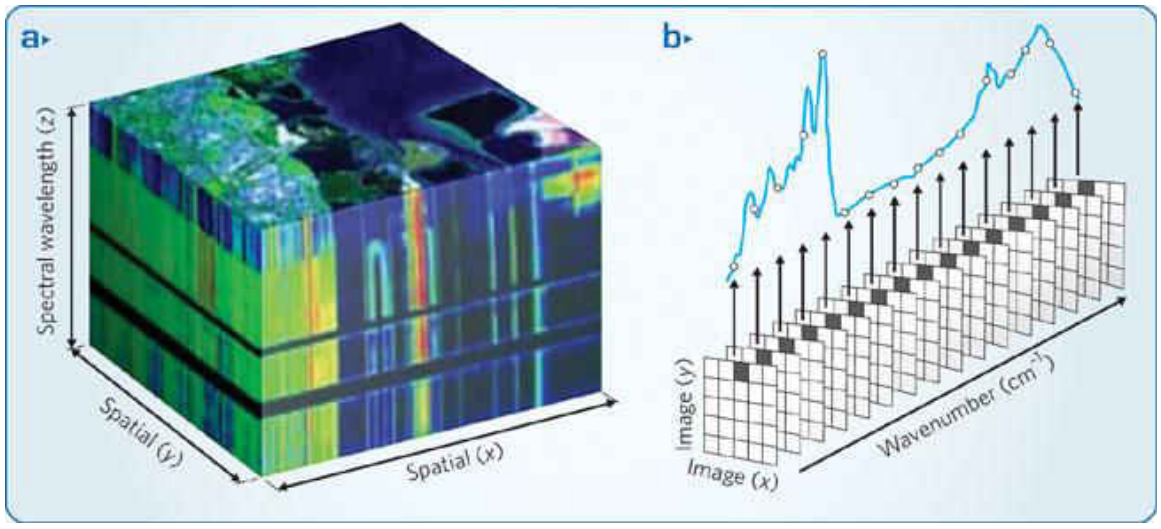


Figure 3: Hyperspectral image cube. Reproduced from reference [23].

1.2. Contributions

In this thesis, two algorithms for processing of FTIR spectra are implemented using Graphics Processing Units. The first removes Mie scattering spectral features and for the first time all the main steps of algorithm pipeline were executed on GPUs as compared to previous implementations where only a small portion of algorithm was implemented on GPU. The algorithm is based on the preliminary algorithm presented in [4] and some modifications have been done in the algorithm. One of the major changes is the way principle components were found. The second procedure, a spatial deconvolution algorithm was completely implemented on GPUs which is used to improve the contrast and fidelity of the imaging data. The unique contribution of this thesis is reporting the results of applying combination of deconvolution and scattering correction on the same FTIR imaging data. The implementations presented in this thesis achieved 10x and 30x speedup, respectively, as compared with an existing MATLAB implementation of the same problem specifications. The other main contribution was the development of the reference spectrum through experimentation which is used as an input for scatter correction algorithm. The reference spectrum is the pure absorbance spectrum which is used to correct the scattered spectrum. Software with very intuitive user interface was also developed to parse input image data, process image data and generate output image data in different formats. Many parameters to control algorithms are also provided to user to play with for different image data.

2. Related Work

In the recently published survey of the area [17], the authors have grouped the spectral correction methods into mainly two groups which researchers are using currently. Two groups are 'physics based' and 'model based'. In physics based approach, explicit optical image–formation modeling from first principles is used to predict and correct data. Here each sample effect (boundary scattering, scattering centers in the sample and substrate) needs to be explicitly accounted for. The theory has been shown to be generally valid and there are methods mentioned in [17] which exist now for correcting the same for films, spheres and fibers. In case of model based methods, a model is assumed to explain all sample effects, typically, Mie scattering. Subsequently, rigorous theory is used to recover spectra, e.g., including extended multiplicative scattering correction (EMSC), resonant Mie scattering correction (RMieSC) and rubber band baseline correction. Model-based methods will generally be faster than explicit modeling methods and may prove to be broadly useful but need to be validated in each case. This thesis is based on the model based work to correct spectra. Bhargava, et al., in [19, 20] provides the first principles calculations that is the basis of the physics behind the scattering calculations implemented in this thesis. They focused on identifying, understanding, and removing Mie scattering effects in spectra, i.e., light scattering from particles that are comparable in size to the wavelength of the incident radiation. The scatter is characterized by large baseline variations; strong absorption bands frequently exhibit dispersive band shapes, so that apparent peak maxima are shifted and absorption intensities are changed. These effects arise from fundamental interactions between light and materials and have been

successfully modeled on the basis of Maxwell's equations in [19, 20], but they require computationally expensive calculation.

Bassan, et al., in [4] proposed scattering correction algorithm for infrared spectra as shown in Figure 4. His algorithm uses full Mie theory and he also used GPU to compute one part of the algorithm.

Research has been going on understanding the spectral features for many years and in 2009 the observed distorted spectral features were resolved theoretically by the Manchester (UK) research group centered around P. Gardner. In these recent investigations it was reported that Mie scattering and the "dispersion artefact" were in fact closely related, and they indicated that the use of the term "dispersion artefact" should be discontinued, since the distorted band shapes are due to a real physical effect, rather than an instrumental or computational artefact. Figure 4 shows the flow chart of algorithm and details of the algorithm are presented in [4].

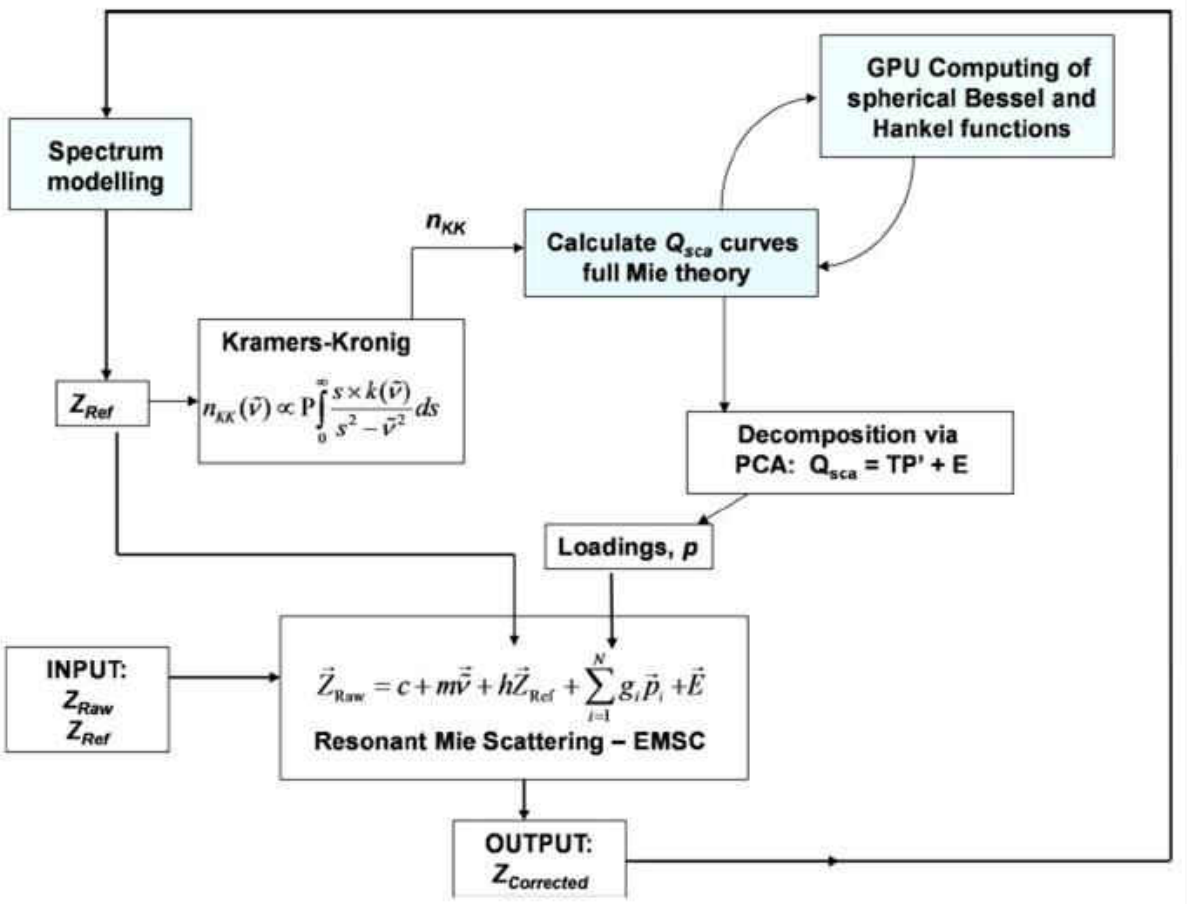


Figure 4: Flowchart of the algorithm proposed by Bassan, et al., in [4]

Bird, et al., also in [5] have also proposed a two step resonant Mie scattering correction based on the work of [4]. In [5], the authors have applied a different approach than that proposed by Bassan et al in [4]. They utilized a set of standard spectra for each class type that are devoid of dispersive line shapes instead of using an iterative process for correction which was presented in [4]. These standard spectra are extracted from class type libraries built from Hierarchical Cluster Analysis (HCA) imaging results. These standard spectra are converted to pure reflective spectra via numerical Kramers-Kronig

transform, and are used as interference spectra in an EMSC based correction procedure. The average of all standard spectra was used as the reference spectrum in this EMSC calculation. They report that the approach is very and hence advantageous. They also mentioned that the standard spectra are transferable for samples that contain the same cell or tissue type. This last point attests to the power and flexibility of the method [5].

3. Implementation

Implementation on Graphical Processing Units

The need for development of algorithms and implementation on massive parallel processors such as GPUs arises because of the size of FTIR microspectroscopy imaging data. A typical dataset has around 800 images of 128 x 128 pixels. Gardner group in 2010 reported in [2, 4] that to correct an IR image with 10 iterations used to take approximately 2 days. In [4] they improved and optimized the image correction to 20 minutes compared to 2 days. Our implementation corrects the whole spectra for a single iteration in around 12 seconds. This higher speed was achieved by implementing all main steps of the algorithm pipeline using GPU.

Parallel processing on GPU

The algorithms implemented in this thesis use parallel processing with NVidia GPUs. CUDA (Compute Unified Device Architecture) is the parallel computing platform and programming model which is implemented by NVidia's GPUs. In this section the various tools, techniques and programming models used in this implementation are described.

3.1. Graphics Processing Units (GPUs)

Graphics Processing Units were initially developed to handle computations related to graphics rendering. The need for specialized rendering routines led GPU vendors to provide user-defined functionality through shader programming [29]. Subsequently,

researchers used shaders to essentially trick GPUs into performing scientific computations. This further led GPU vendors to develop extensions of the C language in order to directly access the parallel processing power for general purpose scientific computing. Examples of these APIs include CUDA [25] and OpenCL [26]. Driven mostly by 3D graphics, GPU has evolved into a highly parallel processor with tremendous computational horsepower [21]. Figure 5 shows the growth over the last decade or so.

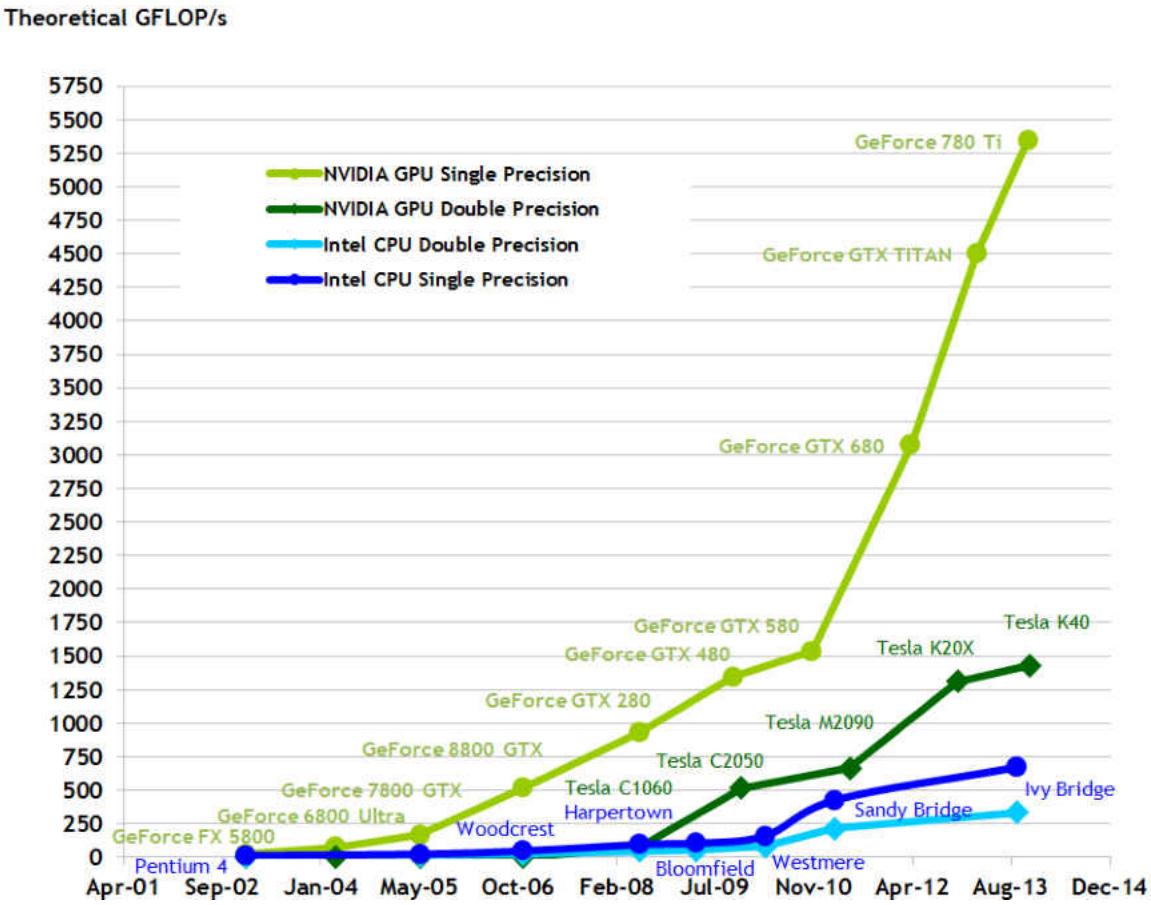


Figure 5: Growth of floating-point operations per second for the CPU and GPU. Reproduced from reference [21].

GPUs are based on single instruction and multiple data (SIMD) architecture. The programming model created by NVIDIA and implemented by the GPUs is called CUDA (Compute Unified Device Architecture). CUDA provide both low level and high level API for developers. In GPUs, more transistors are dedicated to computation. GPU architecture has two main components, global memory and streaming multiprocessors (SMs). The implementation for the work presented in this thesis use Fermi architecture by NVIDIA as shown in Figure 6. Each SM in Fermi has 32 CUDA cores, cache and shared memory.

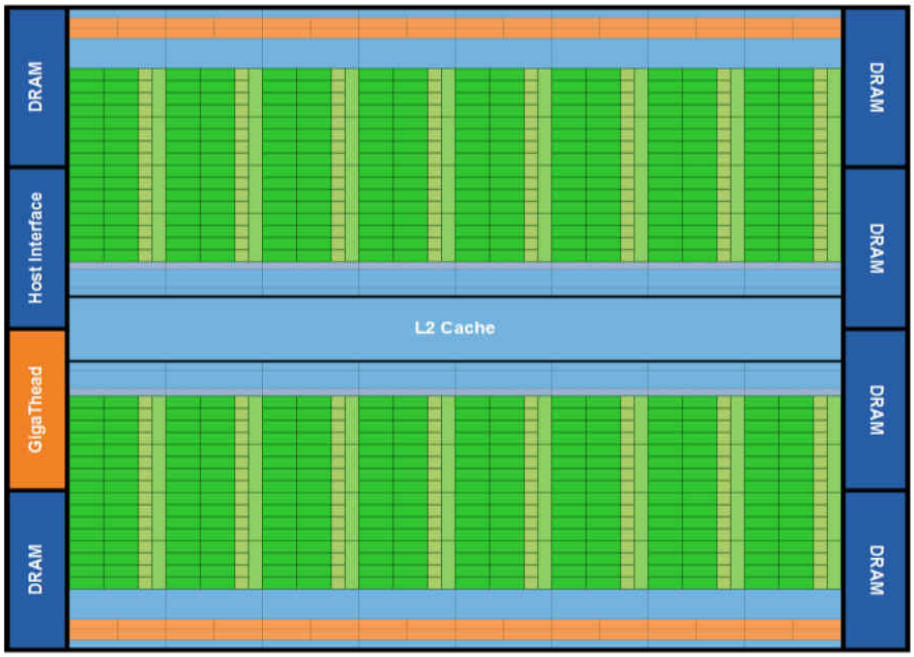


Figure 6: GPU Fermi architecture [29]. 16 multiprocessors are positioned around a common L2cache. Each orange portion is scheduler and dispatch, green portions are execution units and light blue portions are memory for register file and L1 cache

The basic execution unit is a thread. Every thread executes the same program called a kernel. Threads are organized into logical partitions called thread blocks (TBs). During execution, all threads in a TB are assigned to a single MP. Threads in a TB can communicate via shared memory and can be synchronized. In a typical execution, the number of TBs far exceeds the number of MPs (Figure 7). At the hardware level, threads are organized into warps. All threads in a warp execute in lock-step fashion. Warps are equivalent to threads in the symmetric multi-process context. It is therefore advisable to avoid thread divergence with a warp. If two or more threads in a warp access the same shared memory bank, this operation will cause bank conflicts and serialization. However, if all threads access a single shared memory location, then an efficient broadcast mechanism is used. In accessing global memory, all threads in a warp must access memory within a contiguous 128b segment. Non-compliant memory accesses are called un-coalesced accesses and are serialized.

Many high performance numerical libraries like CUBLAS and CULA provide BLAS and LAPACK implementations on GPUs as well as hybrid computations involving both, CPUs and GPUs. CUDA 4.2 API based on Fermi architecture was used to implement both algorithms presented in this paper.

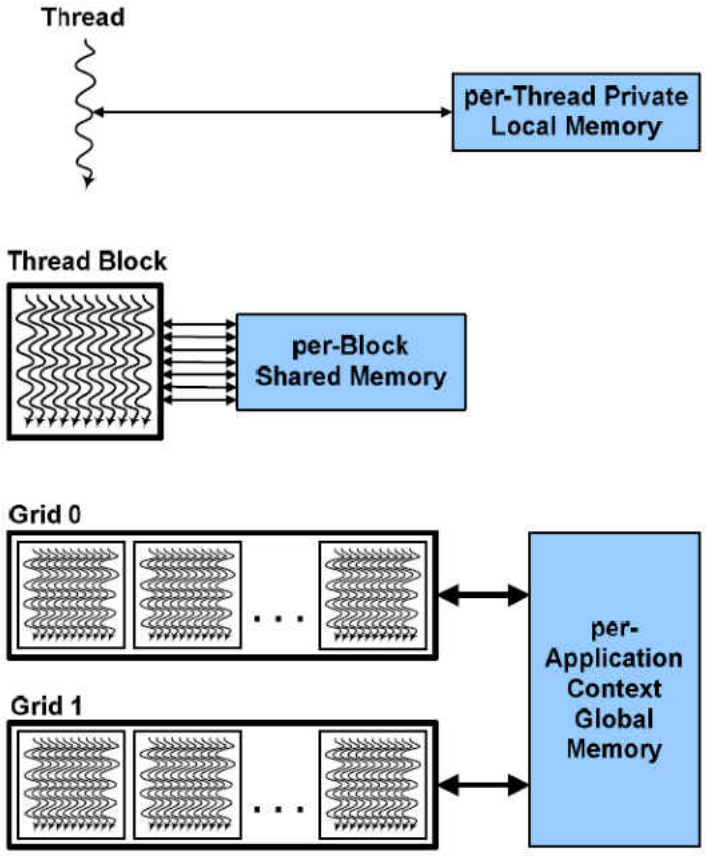


Figure 7: CUDA Model [21]. Threads, blocks, and grids, with corresponding memory spaces for private per-threads, shared per-block, and global per-applications.

Following GPUs were used for the development and all the tests.

GPU 1: NVIDIA Tesla C2050, 2.6 GB Global Memory

GPU 2: NVIDIA GeForce GTX 580, 3 GB Global Memory

4. Deconvolution

The first part of the project was to implement deconvolution algorithm on GPU as first part of spatial image restoration of chemical images. A GPU based program has been implemented and rapid deconvolution has been applied to Fourier transform infrared imaging data collected on IRENI at the Synchrotron Radiation Center (SRC). A deconvolution approach to remove the PSF has been described in detail in a recent publication [9], and improves spatial and spectral fidelity of the data sets for a well-known test sample.

The point spread function (PSF) of an optical imaging system provides a measure of diffraction effects, chromatic aberrations and other factors that distort the object being imaged. The relationships between an observed image, the true form of the object being imaged and the PSF that represents the optical response of the system to a point light source are described within the degradation model of image formation.

Mathematically, these three quantities are related by the convolution operation [9]

$$\boldsymbol{I}_{\text{meas}} = \boldsymbol{I}_{\text{true}} * \boldsymbol{PSF} \qquad (1)$$

where * denotes convolution. If the PSF for a given optical system is known, one can remove the blurring because of diffraction effects from the measured images, to restore images with enhanced resolution and contrast.

Here a summary is provided for the computational method presented in [9] for spatial deconvolution.

For the IR imaging community, it has long been a goal to perform spatial and spectral deconvolution of hyperspectral cubes. However, to perform spatial deconvolution properly, an appropriate characterization of the PSF of the optical imaging system must be performed and implemented. In doing this, we must account for the fact that the PSF is intrinsically wavelength-dependent, yet we need to apply it to the broad mid-IR wavelength range (2.5−10 μm, for IRENI). Other approaches to deconvolution of IR hyperspectral data sets have applied the well-known methods of Fourier Self-Deconvolution (FSD) to the spectral domain as well as the spatial domain. These methods, implemented in commercial software packages, deconvolute Lorentzian lineshapes from the spectra, as well as from the two-dimensional images in the hyperspectral cube. The fundamental difference in the method presented in [9] is that the work has focused on deconvolving a PSF which truly represents the response of the imaging system to a point light source and thus correctly describes the mechanism of image formation. For our current version of software, two different psf for two different optics (36x, 0.5 NA and 74x, 0.65 NA objectives) are implemented. By employing the convolution theorem, (1) may be inverted and the solution for the deconvoluted image may be expressed as

$$I_{\text{true}} = \text{FT}^{-1}\left[\frac{\text{FT}\left(I_{\text{meas}}\right)}{\text{FT(PSF)}}\right]$$

Results in [9] have shown that upon comparison of the raw and deconvoluted data from both objectives, it is clear that in all cases the deconvolution produces images showing sharper, more well- defined features with enhanced contrast.

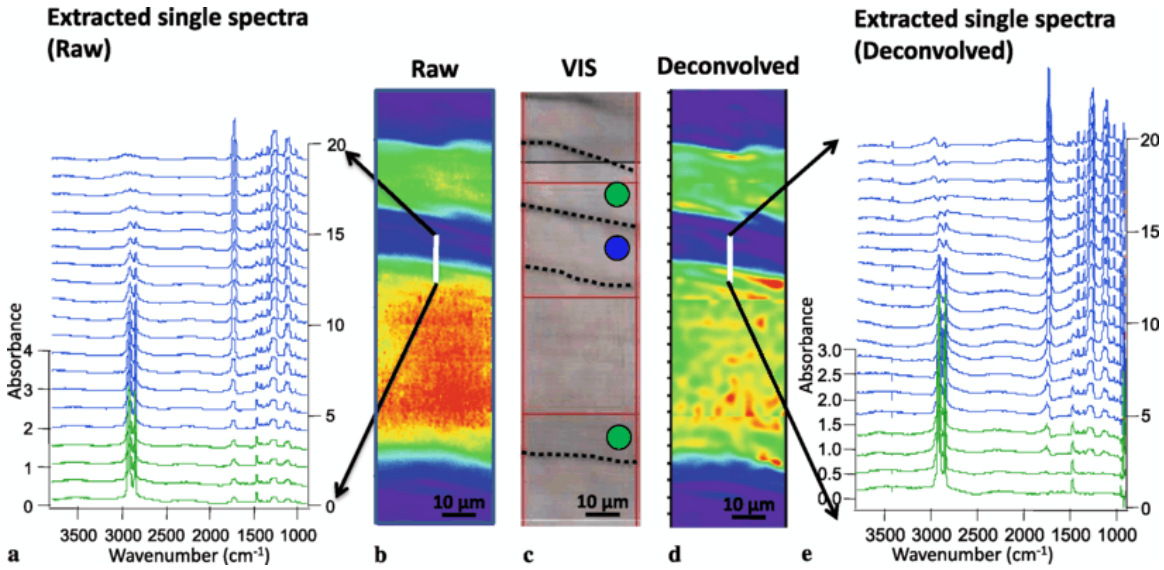


***Figure 8*: Raw and deconvolved chemical images; Extracted single spectra from the original (left-hand side) and deconvolved data sets (right-hand side) taken along the white line in the respective images. Reproduced from reference [10].**

The tasks that are accomplished within GPU for spatial deconvolution algorithm include the following:

- PSF Schwartzchild from spectrum for all wavenumbers
- Parallel reduction to find max for each PSF and scaling the calculated PSF
- Shift PSF data for FFT
- Execute FFT all the PSFs in a batch
- Mirroring image data for FFT
- Shift the image data before FFT
- Executing FFT for all the images in a batch
- Shift back the image data after FFT
- Deconvolve all the images with calculated PSF
- Shift the image data before FFT
- Executing FFT for all the deconvolved images in a batch
- Shift back the image data after FFT
- Get data from mirrored image

Input to the program was image data cube, wave numbers at which each image is taken and a Hanning apodization kernel for deconvolution.

The first step of the algorithm was to measure the fitted wavelength-dependent point-spread-functions (PSFs) which were done for all wave numbers in one GPU call.

Second step was to deconvolve the image data after Fourier with PSF and calculating apodization kernel on the fly on the GPU. For deconvolution, the input image data needs

to be prepared before taking its FFT. For pre-processing image data shift and mirroring operations were also implemented efficiently on GPU. To scale image data, minimum and maximum of image data at each wave number was required. That means calling reduction method multiple times. Thrust library was used efficiently to scale data [27]. As compared to previous implementation of the same algorithm, the time to process similar data sets reduced from around 30-40 seconds to around 1 second.

4.1.1. User Interface for Spatial Deconvolution

The GUI takes Opus file as input which contains the information about image data and also the image data itself. In the first step Opus file is read and data is extracted and converted into raw binary image data. A parser was written in C++ to read the Opus file format and being integrated to work with the GUI. Image data was then passed to a custom library created for deconvolving data. QT which is widely used for application development is used to create the UI. Further adding some imaging capabilities and integrating with CUDA based API. The user interface with the sample run is shown in Figure 9.

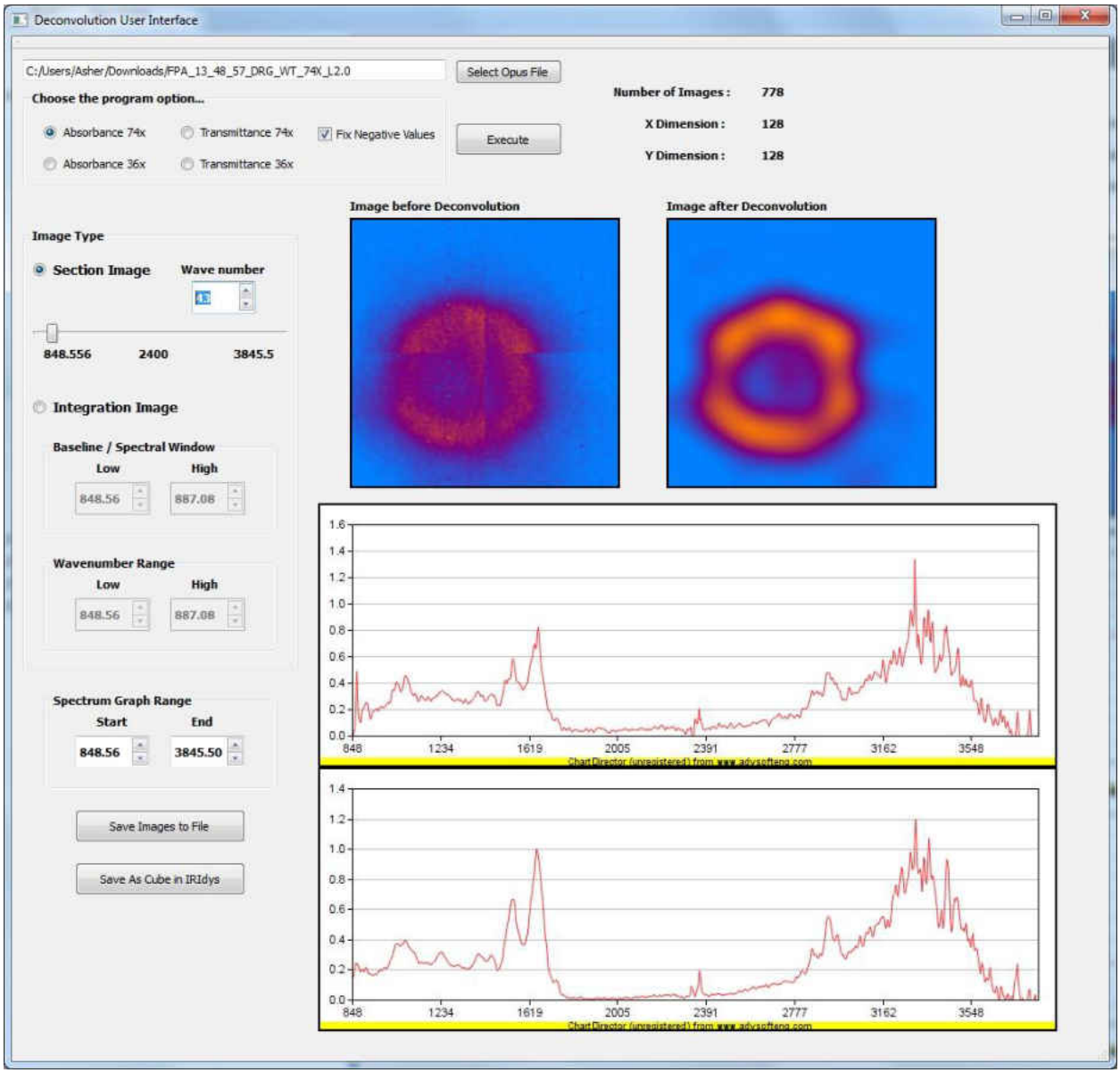


Figure 9: User interface for spatial deconvolution algorithm

5. Scattering

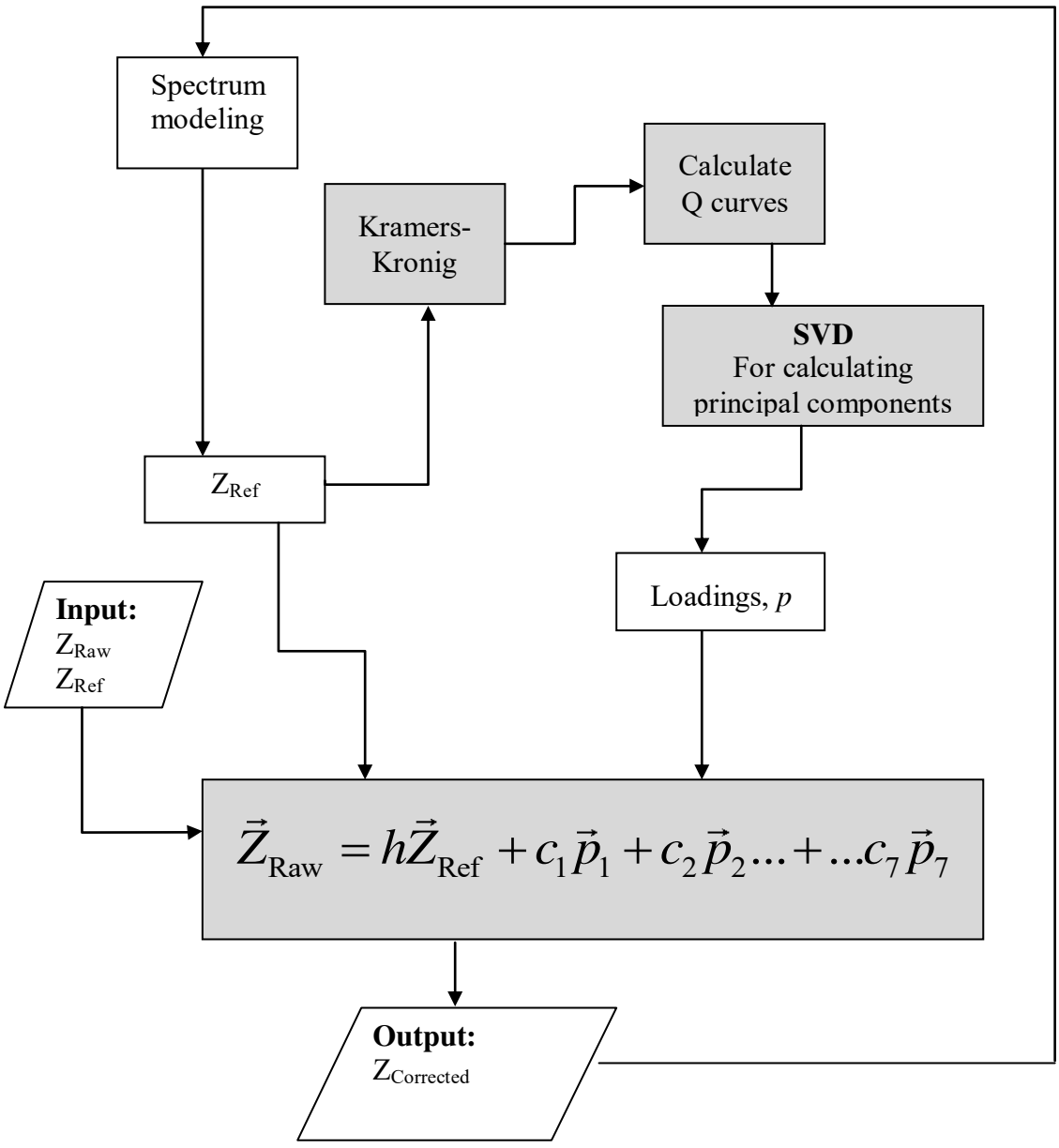
Fourier transform infrared (FTIR) spectroscopy has been demonstrated as a potential tool for disease diagnostics in tissue and in its microscopical form, for cellular analysis. The use of synchrotron sources and attenuated total reflection imaging have achieved subcellular resolution thus promising a powerful technique for the analysis of the biochemical origin of disease. However, both tissue and cellular material are extremely spatially inhomogeneous on a micrometre scale in terms of physical morphology and chemical content and such inhomogeneity can give rise to elastic scattering of the incident radiation which is the origin of a broad undulating background to spectra which have been attributed to Mie scattering. [15]

The algorithm presented by Bassan et al. in [4] is used here with modifications and implemented using GPUs.

The following are the steps involved in scatter correction algorithm

- Input a scatter free reference spectrum (ZRef)
- Calculate 1000 scattering possibilities using difference particle diameters (QCurves)
- Getting 7 principal components using singular value decomposition (SVD)
- Linear regression to estimate scatter contributions
- Subtract the scatter spectrum

Following is the flow chart of the scatter correction algorithm with grey highlighted parts showing steps performed on GPU. Other than implementation of several parts of algorithm on GPU, one main difference is how principal components are found. We have used singular value decomposition to find the 7 principal components.



The following reference spectrum was used as input for the spider silk datasets which shows a lot of scattering features.

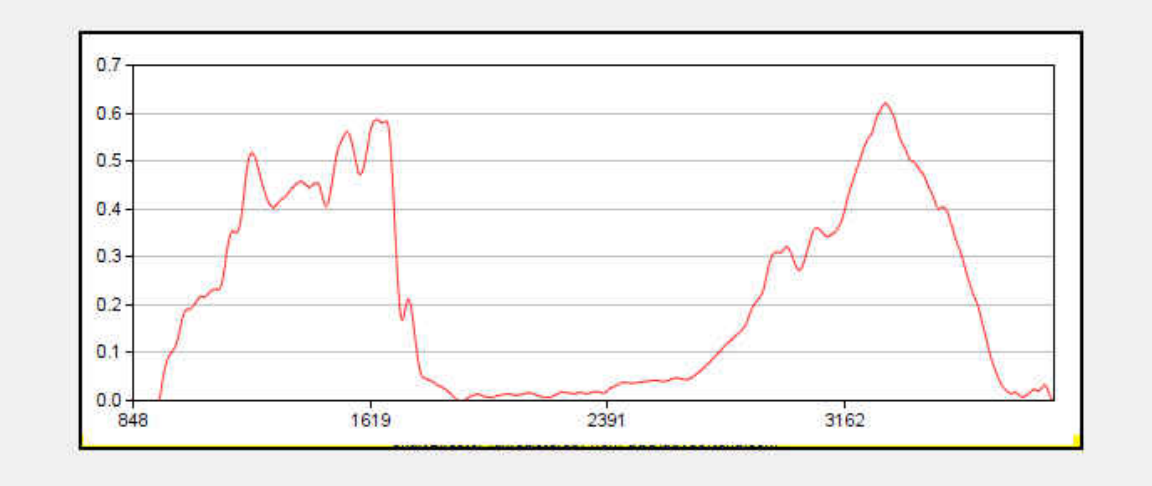


Figure 10: Reference Spectrum

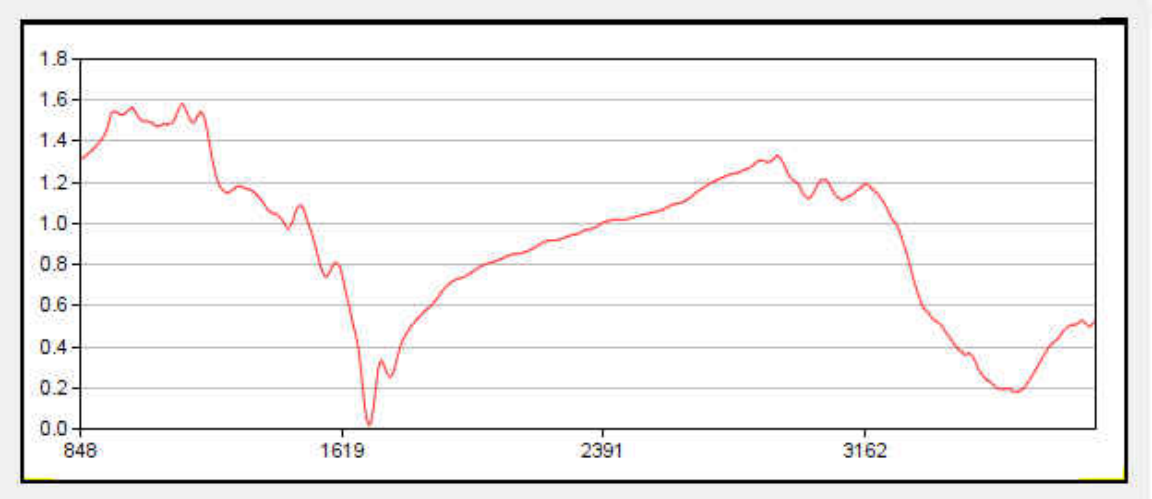


Figure 11: Real refractive index

Overall our scattering correction implementation produces results of a dataset containing around 800 images of 128x128 in around 12 seconds on average.

Process	Matlab version	CUDA version	Speedup
Kramers-Kronig	63.15 ms	2.91 ms	21x
Generating 1000 Q Curves	46.57 ms	2.69 ms	17x

Table 1: Execution time of two computationally expensive tasks on the GPU

5.1.1. User Interface for Scattering Correction

The GUI for scattering correction takes Opus file or comma separated DPT file as input which contains the information about image data and also the image data itself. In the first step input file is read and data is extracted and converted into raw binary image data. Similar to the architecture of previous software mentioned before, image data was then passed to a custom library created for handling all the processing for scatter correction. After the correction, the software generates a binary output file which can be used to generate Chemigrams and further used as input for deconvolution code. The user interface with the sample run is shown in Figure 12.

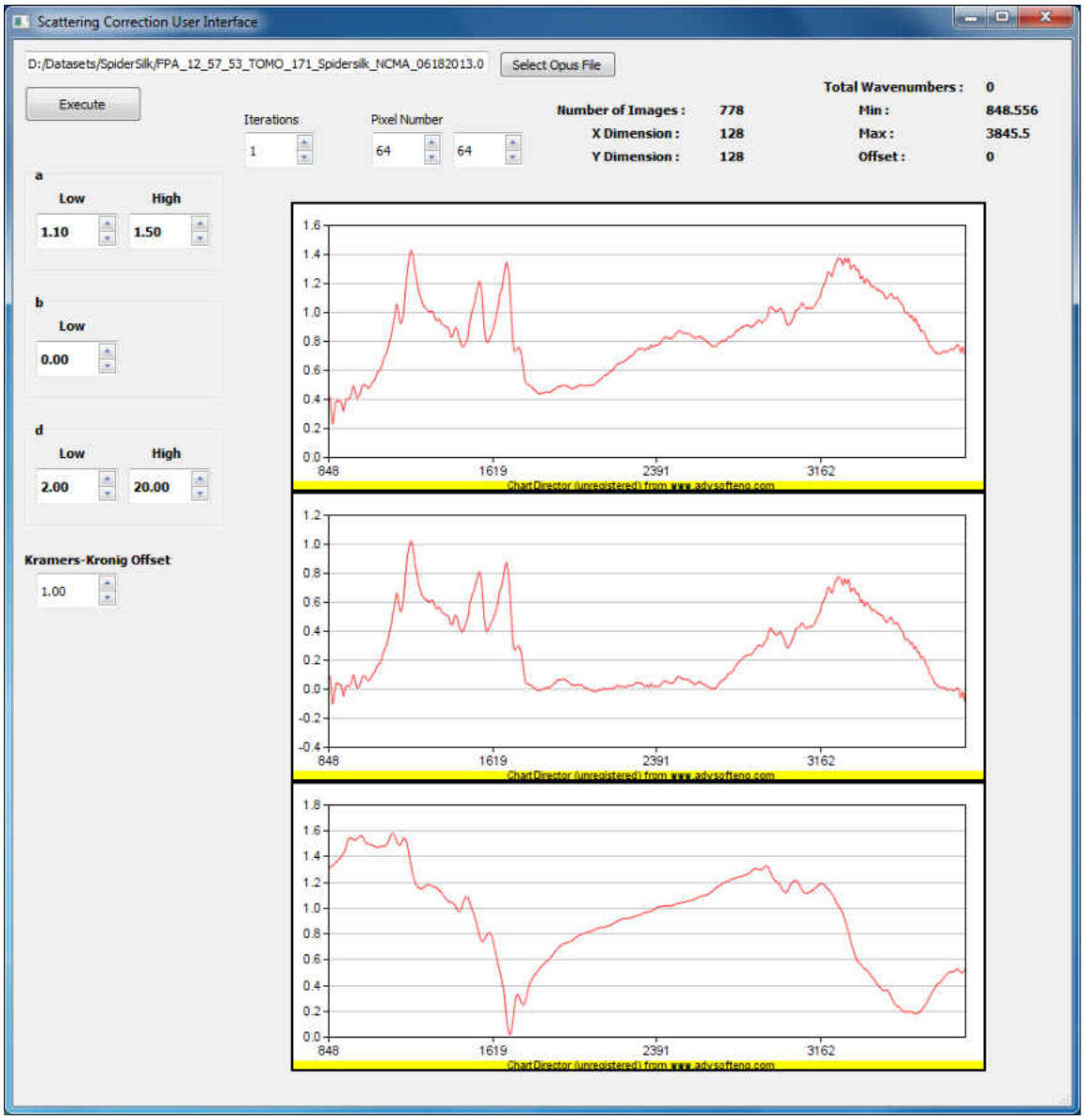


Figure 12*: *User interface for scattering correction algorithm.

6. Simulated Data

To check the fidelity of the scatter correction method, we prepared simulated data sets. The simulated data set was prepared by taking a reference spectrum presented in figure 10. In the first test, reference spectrum was given as ZRaw to the correction algorithm. Since there was no scatter present so the corrected spectra was almost similar to the input as shown in figure 13. Then using van de Hulst approximation equation, 100 different scatter curves were calculated with varying parameters ***a*** (average refractive index), ***b*** (resonance parameter) and diameter values. Then these known scatter curves were added to the reference spectrum to get ZRaw. Then this simulated ZRaw was provided as input to the scattering correction to get corrected spectrum. Selected results are shown in figure 14, 15 and 16.

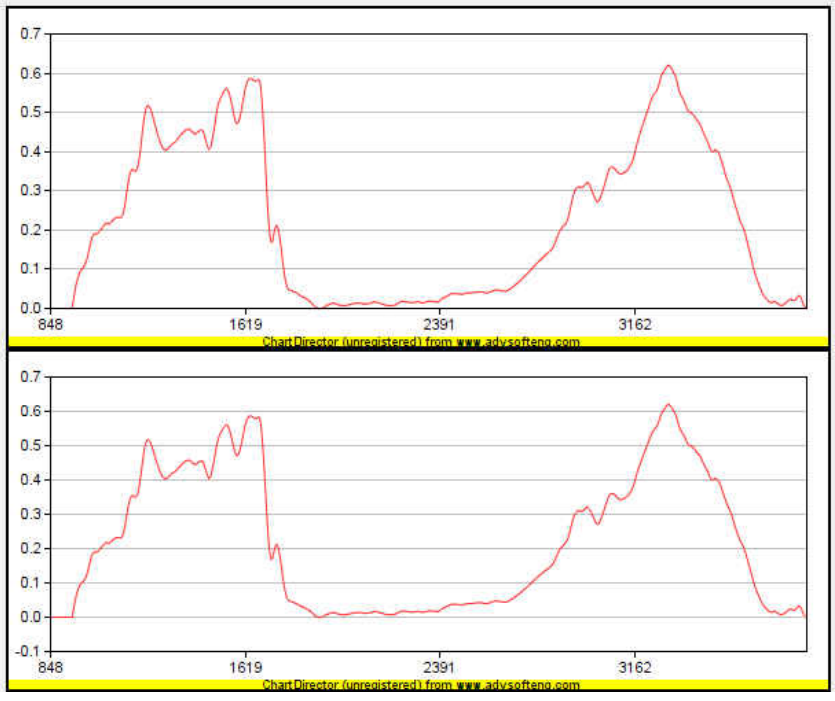


Figure 13: Simulated data result with reference spectra as input Z_Raw.

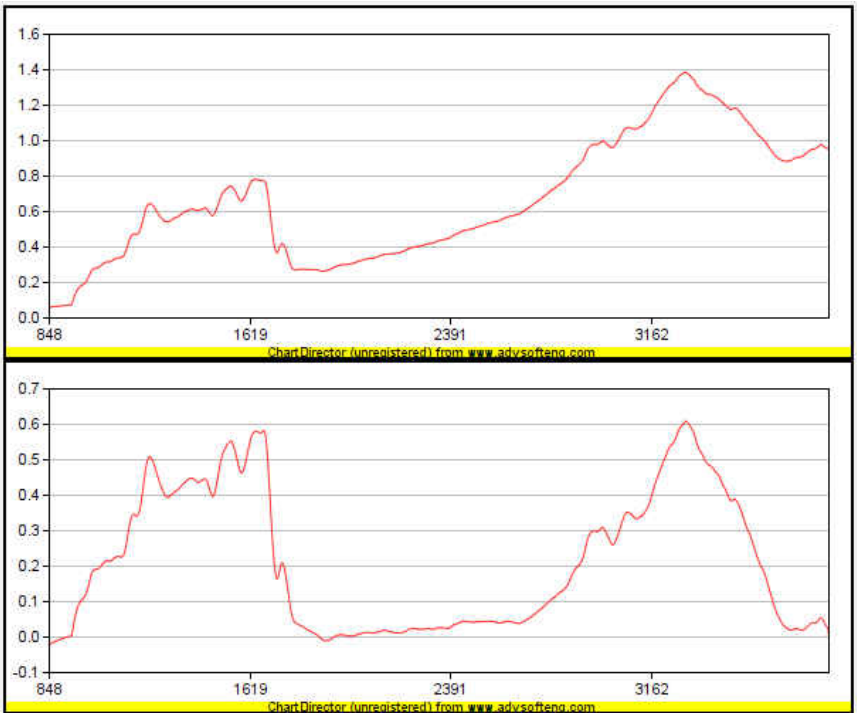


Figure 14: Simulated data result 1.

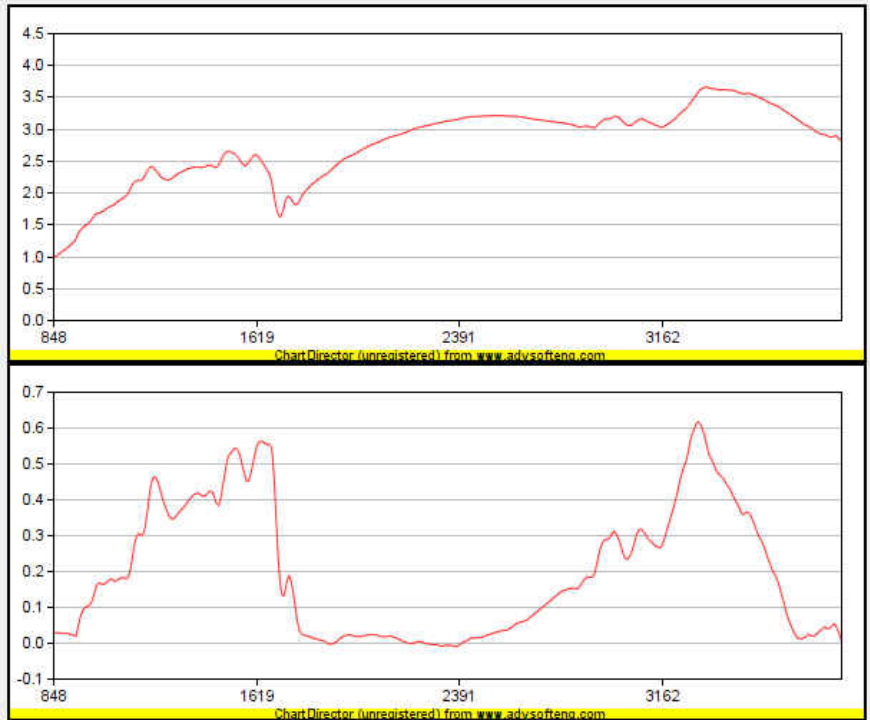


Figure 15: Simulated data result 2.

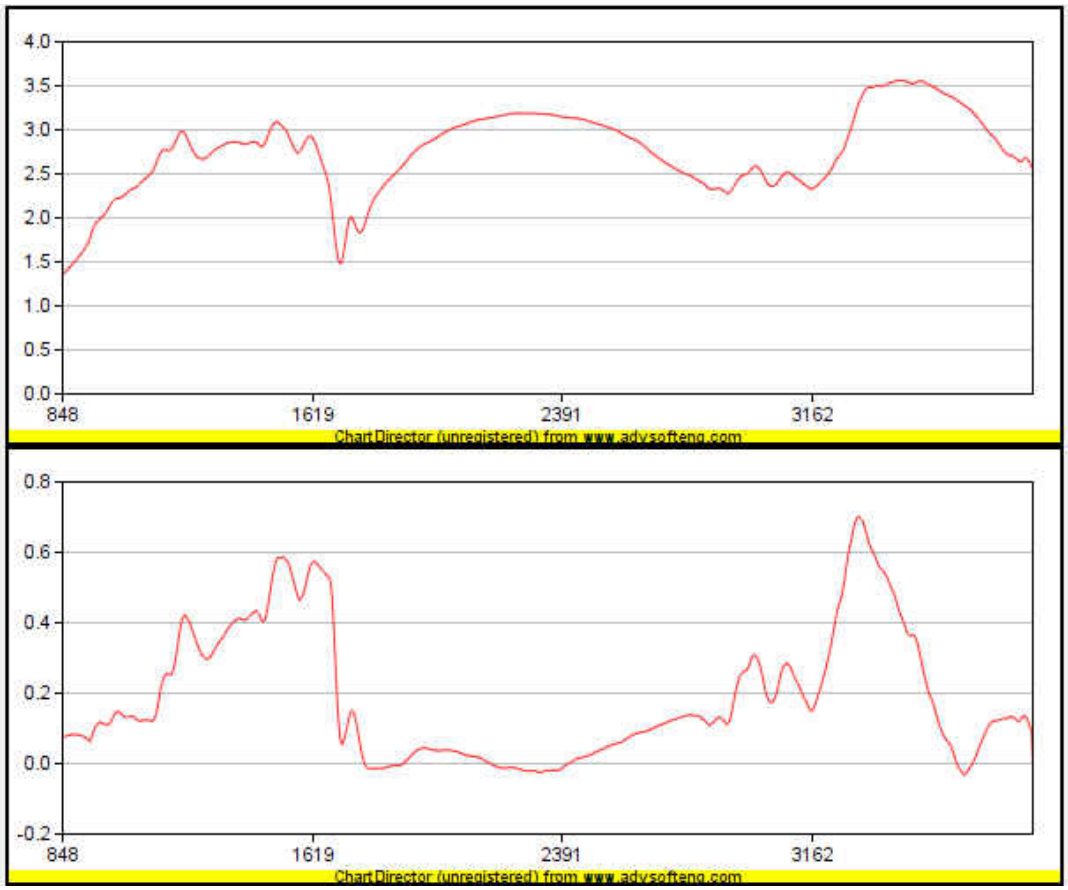


Figure 16: Simulated data result 3.

7. Data

The following datasets have been used to see the results of the code for scattering correction and spatial deconvolution algorithm implementation on GPU.

7.1. Spider Silk

Spider Silk dataset has been provided by Sujatha et al. [12]. The data recorded show exaggerated scattering features and correcting scattering is important in the further data analysis. Spider silk is a remarkable biopolymer, a lightweight protein fiber that spiders produce for diverse purposes including web structure, prey immobilization, making egg cases and which is a key to their survival. With a unique combination of high elasticity, mechanical strength and toughness, its superior material properties surpass those of any currently known man-made fiber. In addition, these fibers are biocompatible and biodegradable. It is one of the best examples of block copolymers in Nature in which repeating structural motifs are linked together covalently resulting in polymers that organize themselves into a variety of nano-structures with exceptional material properties. The extraordinary mechanical properties have been attributed to the hierarchical structuring at the nanometre length scales, and are designed to perform different functions at each level, effectively linking nano-scale constituents to properties at the macro level, a strategy often used in Nature (silk, bone, wood). These features coupled with potential diverse applications render spider silk as an ideal model for

biomimetic research. Over the last few years there has been a great deal of interest in understanding how this nano-structuring works, both from a theoretical and experimental perspective, made feasible with the advent of more advanced experimental and computational facilities. Recent advances in bioengineering have made considerable progress in the production of synthetic silks, however, to date synthetic fibers do not exhibit the same physical or mechanical properties of natural spider silk. The viable production and use of these proteins as synthetic materials requires greater knowledge of the secondary, tertiary, and quaternary structure of the proteins that comprise spider silk.

7.2. DRG Neurons

DRG neurons (dorsal root ganglia) data was collected at the synchrotron radiation center (SRC, Stoughton, WI) using the IRENI beamline [6]. Barabas et al. in [13] hypothesized that heterogeneous populations of sensory neurons likely differ in chemical features. Also they conducted FTIR spectromicroscopy to identify DRG subpopulations based on chemistry. [13] is the first report to use FTIR spectromicroscopy to assess the biochemical composition of somatosensory neurons. They employed synchrotron radiation at the InfraRed ENvironmental Imaging (IRENI) mid-infrared beamline at the Synchrotron Radiation Center (SRC; University of Wisconsin-Madison). The use of this beamline enables rapid, diffraction-limited spatial resolution imaging of individual neurons [6]. Importantly, this experimental approach can probe length scales that are on par with subcellular structures.

Using FTIR spectromicroscopy in [13], the authors were able to visualize the distribution of chemical features and identify changes in chemistry on a single-cell basis. The authors also identified and characterized particularly unusual architectures of chemical structures that were created by an annular "ring" distribution of carbohydrate and/or lipid enrichment in the cytoplasm in sensory neuron subpopulations, a feature that has never before been reported in native cells of any type.

8. Scattering and Deconvolution

This thesis also presents results of applying deconvolution and then correcting scatter of deconvolved FTIR images. To the best of our knowledge such analysis has never been published before. Several studies have employed signal processing techniques to improve image quality. This technique of deconvolution and then scatter correction or scatter correction and then deconvolution is currently being studied and we are looking at results to further highlight its importance. In this section, we used our implementation to run on datasets mentioned earlier and some results are reported.

8.1. Results

The following images are chemigrams of DRG Neuron dataset.

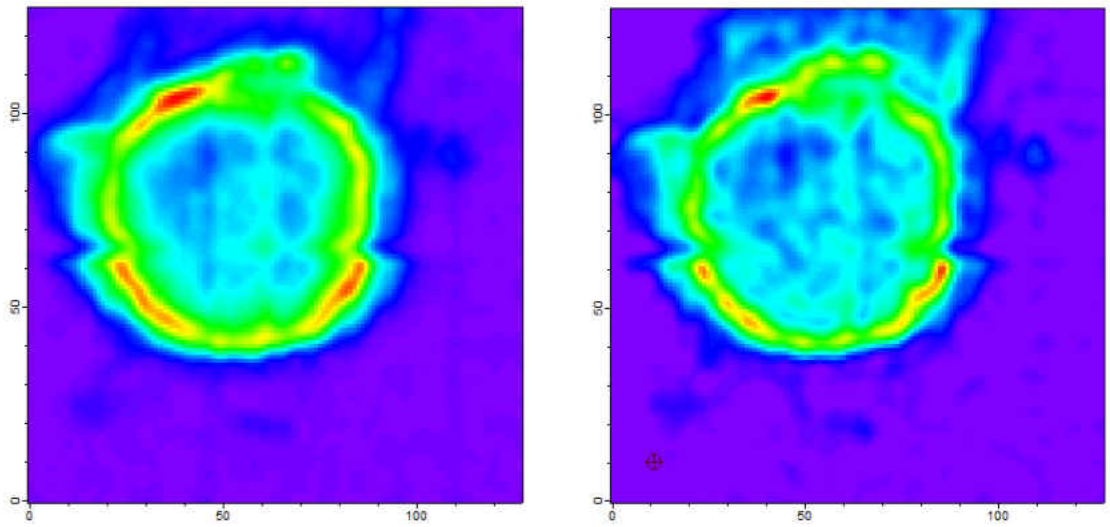


Figure 17: Chemigrams of DRG Neuron dataset. Original raw data (left) and deconvolved (right) with 74x absorbtion option.

The following images are chemigrams of spider silk tomography dataset.

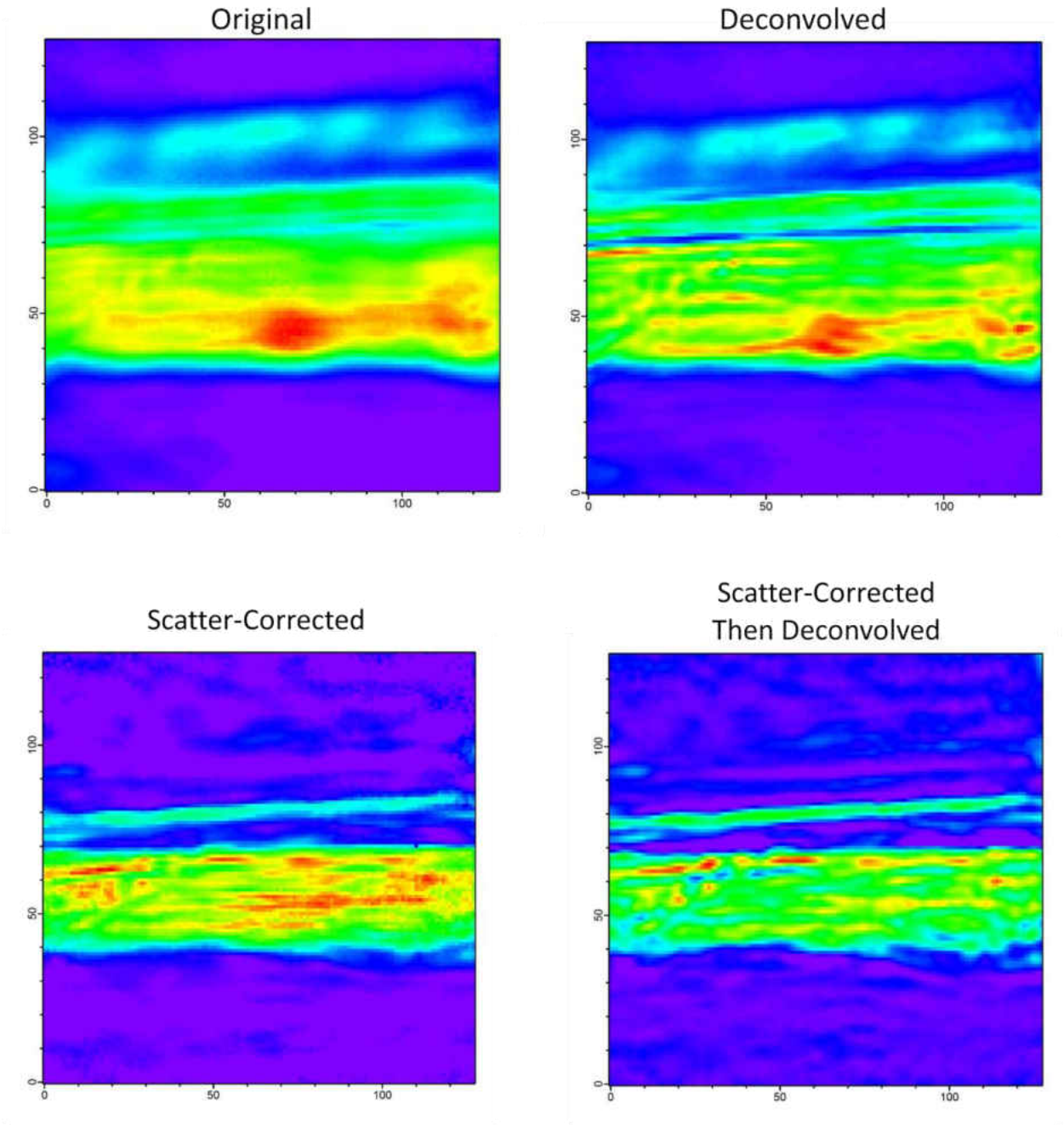


Figure 18*: *Chemigrams of spider silk image data

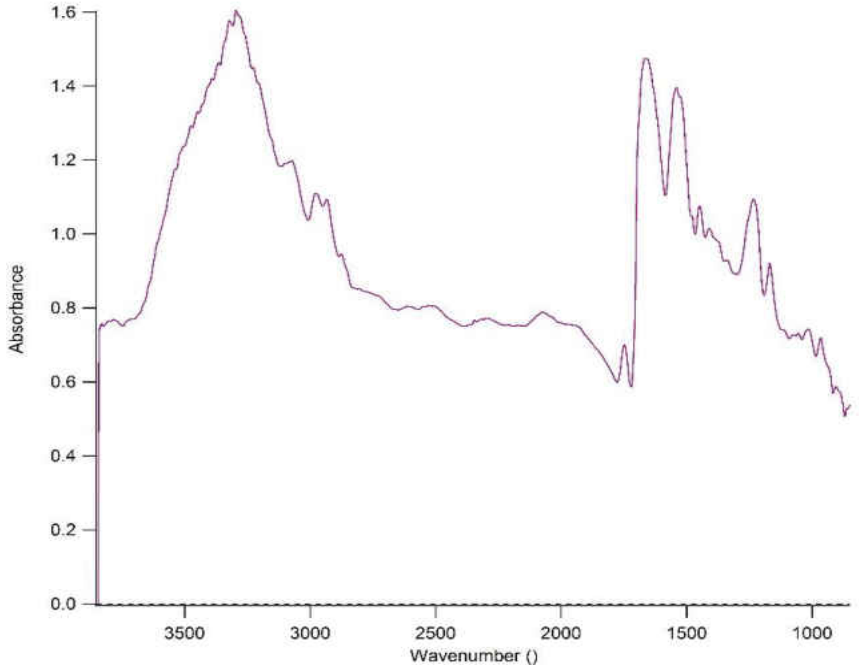


Figure 19: Average spectra of original image

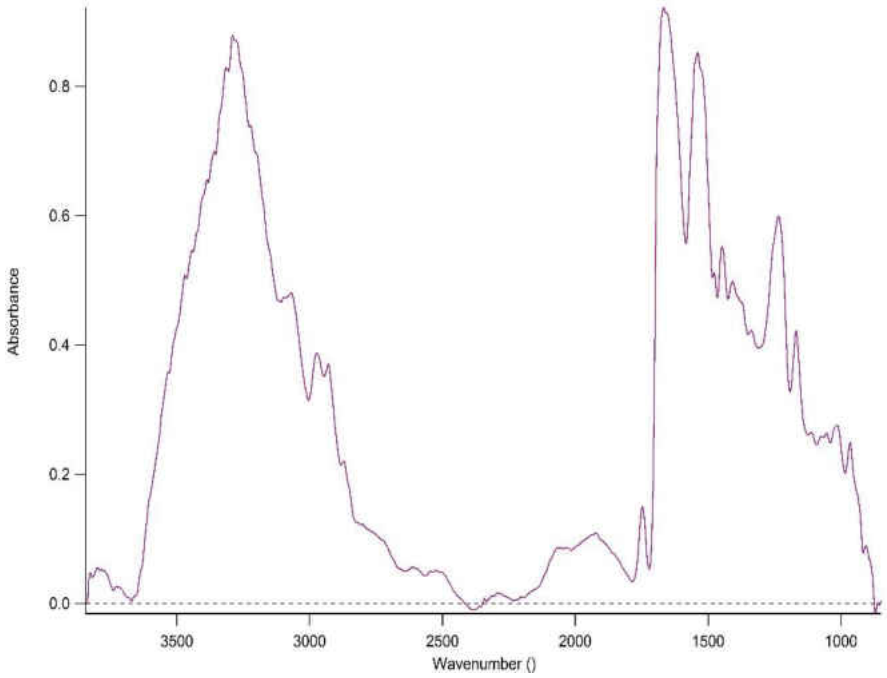


Figure 20: Average spectra after scatter correction

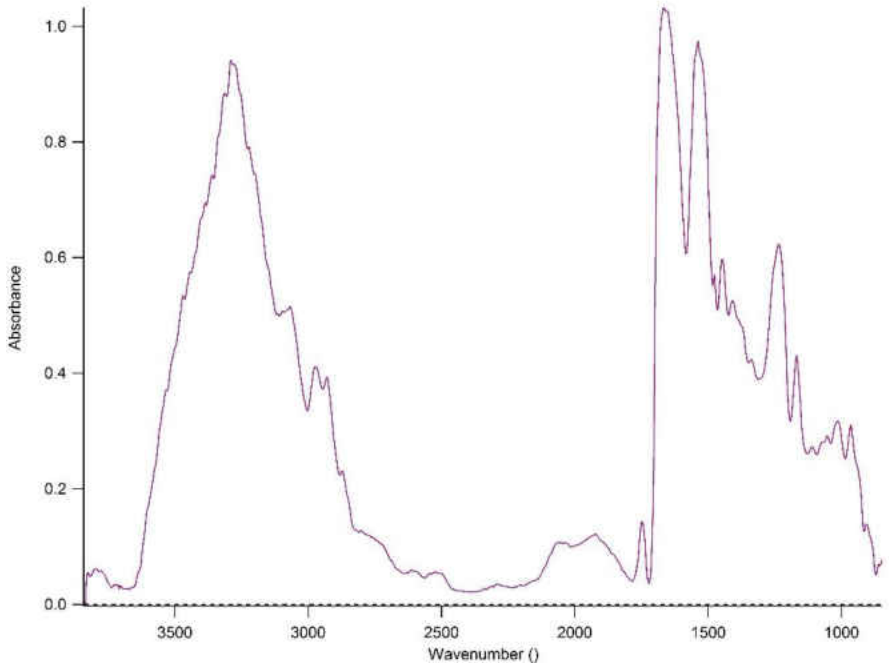


Figure 21: Average spectra - Scatter correction and then deconvolution

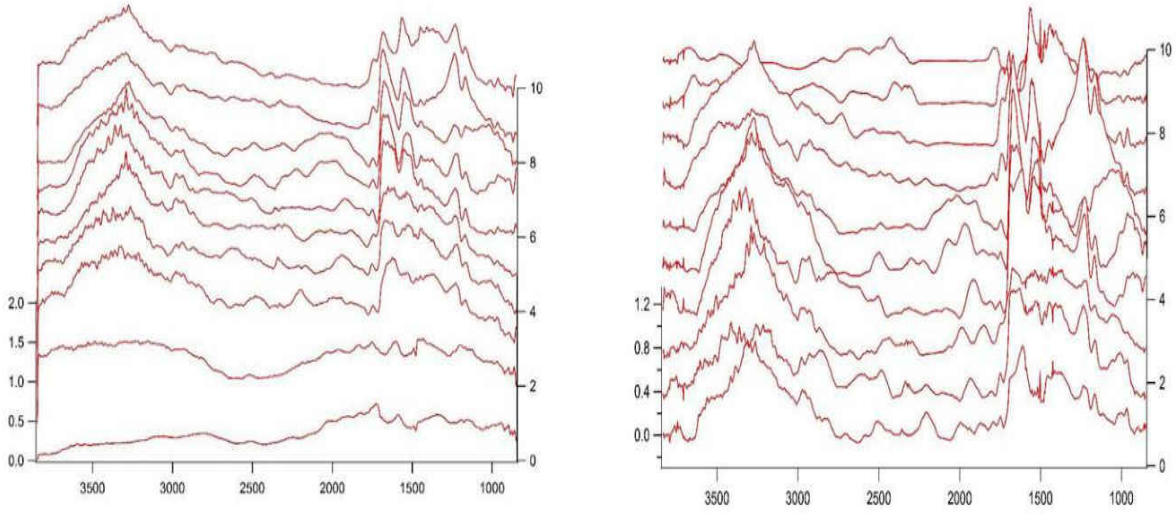


Figure 22: Single pixel spectra of spider silk tomography dataset. Original data spectra (left), Scatter corrected then deconvolved (right)

9. Future Work

Future work has been identified in different areas based on the work presented in this thesis. Our future work will continues in three different areas. The first is the development and implementation of 3D microtomography reconstruction on GPU that outperforms all state of the art algorithms and that can be easily accessible and applicable to huge applications in the area. The second possible target of future work in this area is to release user friendly software for proposed algorithm. The third area lies on the application side. In this work, though we have achieved better speedups and computation time for both the algorithms. There have been many technological advances in GPU hardware and software configurations in the last year or two. New architecture provides more promise to achieve even higher efficiency and algorithm development to improve execution time will be needed in the foreseeable future.

9.1. Tomography

FTIR spectro-microtomography provides spectrally rich, label-free, nondestructive 3D visualizations of biological and materials samples. Several groups are working to create efficient data collection and analysis facilities with the promise of quantitative analysis, which will incorporate advances from protein crystallography, X-ray tomography, medical tomography, FTIR spectroscopy, and computational control and analysis to automate the collection, processing and storage of large spectral tomographic data sets.

Continued advances in mid-IR array detector technologies, synchrotron IR beamlines and computational tomography tool sets are expected to improve FTIR spectro-microtomography. In [8], the methods proposed are expected to facilitate a wide variety of scientific, industrial, materials, energy and medical applications.

Reconstruction of the transmission image of one wavelength is similar to traditional X-ray tomography. Martin et al. in [8] have written scripts that allow reconstruction of any (or all) of the mid-IR wavelengths to compare absorption features, spectral integrations of wavelength regions corresponding to absorption feature(s), ratios of two absorption band intensities or any of these spectral functions with an automated baseline subtraction. By reconstructing all wavelengths, they have shown that they can reassemble the reconstructed spectrum for any or every voxel within the 3D space (an example is shown in Fig. 19). Having complete spectra for every voxel can help users to interpret chemical and structural variations in different regions of the sample and/or enable more advanced spectral comparisons and allow the use of sorting tools such as principal-component analysis, cluster analysis and other statistical analysis techniques. However due to the size of data, this task is time consuming. Using GPUs to do most of the processing, we can get reconstructed spectrum faster and will be one of the next steps in this direction.

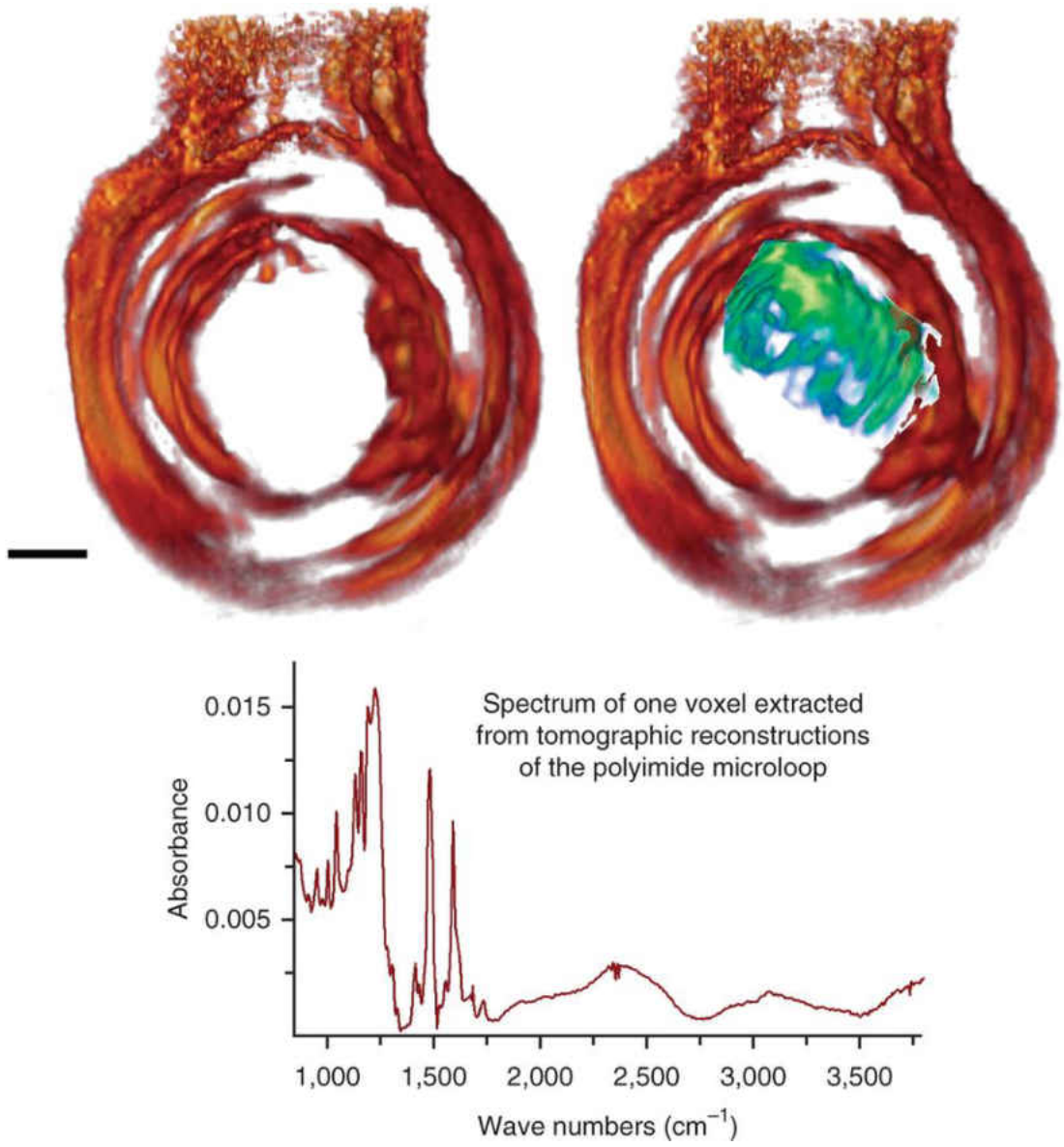


Figure 23: FTIR spectro-microtomographic imaging of Zinnia. Reproduced from reference [8].

9.2. FTIR Biospectroscopy

Biospectroscopy is defined in [1] as applying Fourier-transform infrared (FTIR) spectroscopy to biological questions which is relatively novel. This area of study is growing fast. [1] has identifed some potential fields of application which include cytological, histological and microbial studies. This potentially provides a rapid and non-destructive approach to clinical diagnosis. Its increase in application is primarily a consequence of developing instrumentation along with computational techniques. In the coming decades, biospectroscopy is likely to become a common tool in the screening or diagnostic laboratory, or even in the general practitioner's clinic. Despite many advances in the biological application of FTIR spectroscopy, there remain challenges in sample preparation, instrumentation and data handling. The focus of our work is mainly in data analysis and handling and [1] identify four main goals in this area: Pattern Finding; Biomarker Identification; Imaging; and, Diagnosis. [1] grouped these into two frameworks: Exploratory; and, Diagnostic. They claim that in the near future research should be focused on the challenges of dataset standardization; building information systems; development and validation of data analysis tools; and, technology transfer. While the existing large amount of biological studies that use FTIR biospectroscopy are of remarkable value to validate a novel approach to diagnosis, FTIR biospectroscopy lacks research that is mainly focused on modelling and validation of data processing methods. [1] has summarized some specific challenges in FTIR biospectroscopy data handling.

9.3. GPU Technology

NVIDIA has released two new high performance computing architectures called 'Kepler' and 'Maxwell' since we started this development. GPUs based on Kepler are now widely available and in use. The following are the main improvements and features in the new Kepler based GPUs as compared to the previous Fermi architecture [30].

- Over 2400 stream processors as compared to around 500 processors in Fermi GPUs.
- For single precision around 3.5 TFLOPS and for double precision 1.1 TFLOPS on Kepler GPUs as compared to 1.1 TFLOPS and 500 GFLOPS for single and double precision on Fermi GPUs
- Maximum registers available per thread increased from 63 to 255

The important two new features of Kepler GPUs are ***Dynamic Parallelism*** and ***HyperQ***. With the capability of Dynamic Parallelism, GPU can generate new work for itself. This feature will provide more flexibility to design parallel algorithms. GPUs will be more efficiently used as computation evolves. The code is easier to write for most applications and more control is now inside GPU kernels to adapt to different parallel program structures.

One of the other interesting features HyperQ will allow Kepler GPUs to simultaneously make 32 connections with the CPU which will be managed at hardware level. In Fermi there was just single connection available between CPU and GPU. This will increase GPU utilization dramatically.

Porting the code and further implementation with these new features will surely improve the running time on Kepler and Maxwell. This will result in more speed up and will help improve overall running time of applications which use these algorithms as a building block.

10. References

[1] Trevisan, J., Angelov, P. P., Carmichael, P. L., Scott, A. D., & Martin, F. L. (2012). Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst*, *137*(14), 3202-3215.

[2] Bassan, P., Kohler, A., Martens, H., Lee, J., Byrne, H. J., Dumas, P., ... & Gardner, P. (2010). Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples. *Analyst*, *135*(2), 268-277.

[3] Bassan, P., Byrne, H. J., Lee, J., Bonnier, F., Clarke, C., Dumas, P., ... & Gardner, P. (2009). Reflection contributions to the dispersion artefact in FTIR spectra of single biological cells. *Analyst*, *134*(6), 1171-1175.

[4] Bassan, P., Kohler, A., Martens, H., Lee, J., Jackson, E., Lockyer, N., ... & Gardner, P. (2010). RMieS-EMSC correction for infrared spectra of biological cells: Extension using full Mie theory and GPU computing. *Journal of biophotonics*, *3*(8-9), 609-620.

[5] Bird, B., Miljković, M., & Diem, M. (2010). Two step resonant Mie scattering correction of infrared micro-spectral data: human lymph node tissue. *Journal of biophotonics*, *3*(8-9), 597-608.

[6] Nasse, M. J., Walsh, M. J., Mattson, E. C., Reininger, R., Kajdacsy-Balla, A., Macias, V., ... & Hirschmugl, C. J. (2011). High-resolution Fourier-transform infrared chemical imaging with multiple synchrotron beams. *Nature methods*,*8*(5), 413-416.

[7] Mattson, E. C., Unger, M., Clède, S., Lambert, F., Policar, C., Imtiaz, A., ... & Hirschmugl, C. J. (2013). Toward optimal spatial and spectral quality in widefield infrared spectromicroscopy of IR labelled single cells. *Analyst*,*138*(19), 5610-5618.

[8] Martin, M. C., Dabat-Blondeau, C., Unger, M., Sedlmair, J., Parkinson, D. Y., Bechtel, H. A., ... & Hirschmugl, C. J. (2013). 3D spectral imaging with synchrotron Fourier transform infrared spectro-microtomography. *Nature methods*, *10*(9), 861-864.

[9] Mattson, E. C., Nasse, M. J., Rak, M., Gough, K. M., & Hirschmugl, C. J. (2012). Restoration and spectral recovery of mid-infrared chemical images.*Analytical chemistry*, *84*(14), 6173-6180.

[10] Unger, M., Mattson, E., Patterson, C. S., Alavi, Z., Carson, D., & Hirschmugl, C. J. (2013). Synchrotron-based multiple-beam FTIR chemical imaging of a multi-layered

polymer in transmission and reflection: towards cultural heritage applications. *Applied Physics A*, *111*(1), 135-145.

[11] Mattson, E. C., Unger, M., Manandhar, B., Alavi, Z., & Hirschmugl, C. J. (2013, March). Multi-beam synchrotron FTIR chemical imaging: Impacts of schwarzschild objective and spatial oversampling on spatial resolution. In*Journal of Physics: Conference Series* (Vol. 425, No. 14, p. 142001). IOP Publishing.

[12] Sampath, S., Isdebski, T., Jenkins, J. E., Ayon, J. V., Henning, R. W., Orgel, J. P., ... & Yarger, J. L. (2012). X-ray diffraction study of nanocrystalline and amorphous structure within major and minor ampullate dragline spider silks.*Soft matter*, *8*(25), 6713-6722.

[13] Barabas, M. E., Mattson, E. C., Aboualizadeh, E., Hirschmugl, C. J., & Stucky, C. L. (2014). Chemical structure and morphology of dorsal root ganglion neurons from naive and inflamed mice. *Journal of Biological Chemistry*, *289*(49), 34241-34249.

[14] Bambery, K. R., Wood, B. R., & McNaughton, D. (2012). Resonant Mie scattering (RMieS) correction applied to FTIR images of biological tissue samples. *Analyst*, *137*(1), 126-132.

[15] Kohler, A., Sule-Suso, J., Sockalingum, G. D., Tobin, M., Bahrami, F., Yang, Y., ... & Martens, H. (2008). Estimating and correcting Mie scattering in synchrotron-based microscopic Fourier transform infrared spectra by extended multiplicative signal correction. *Applied spectroscopy*, *62*(3), 259-266.

[16] Finta, C., Teppola, P., Juuti, M., & Toivanen, P. (2013). Feasibility of parallel algorithms and graphics processing unit acceleration in chemical imaging.*Chemometrics and Intelligent Laboratory Systems*, *127*, 132-138.

[17] Baker, M. J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H. J., Dorling, K. M., ... & Martin, F. L. (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nature protocols*, *9*(8), 1771-1791.

[18] Diem, M., Mazur, A., Lenau, K., Schubert, J., Bird, B., Miljković, M., ... & Popp, J. (2013). Molecular pathology via IR and Raman spectral imaging.*Journal of biophotonics*, *6*(11-12), 855-886.

[19] Davis, B. J., Carney, P. S., & Bhargava, R. (2010). Theory of mid-infrared absorption microspectroscopy: II. Heterogeneous samples. *Analytical chemistry*, *82*(9), 3487-3499.

[20] Davis, B. J., Carney, P. S., & Bhargava, R. (2010). Theory of midinfrared absorption microspectroscopy: I. Homogeneous samples. *Analytical chemistry*,*82*(9), 3474-3486.

[21] NVIDIA CUDA Compute Unified Device Architecture Programming Guide, 2014, http://docs.nvidia.com/cuda/cuda-c-programming-guide

[22] CUDA Toolkit 4.2 CUFFT Library Programming Guide, March 2012, http://docs.nvidia.com/cuda/cufft/

[23] Bannon, D. (2009). Hyperspectral imaging: cubes and slices. *Nature photonics*,*3*(11), 627.

[25] Sanders, J., & Kandrot, E. (2010). *CUDA by example: an introduction to general-purpose GPU programming*. Addison-Wesley Professional.

[26] Khronos OpenCL Working Group. (2008). The opencl specification. version, 1(29), 8.

[27] NVIDIA Thrust quick start guide v5.0, October 2012, http://docs.nvidia.com/cuda/thrust/index.html

[28] CULA Dense Free Edition, http://www.culatools.com/

[29] Luebke, D., Harris, M., Govindaraju, N., Lefohn, A., Houston, M., Owens, J., ... & Buck, I. (2006, November). GPGPU: general-purpose computation on graphics hardware. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing* (p. 208). ACM.

[30] NVIDIA Kepler GK110 white paper: http://www.nvidia.com/content/PDF/kepler/NVIDIA-Kepler-GK110-Architecture-Whitepaper.pdf