Online Raman Spectroscopy for Bioprocess Monitoring

by

Gustavo Adolfo Gil

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Bachelor of Science in Electrical Science and Engineering

and

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2005
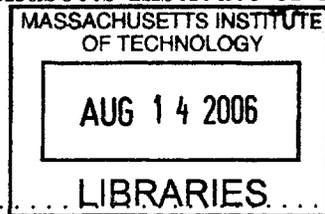
Author ....
Department of Electrical Engineering and Computer Science
August 5, 2005

Certified by ....
Rajeev Ram
Associate Professor
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Chairman, Department Committee on Graduate Theses

Online Raman Spectroscopy for Bioprocess Monitoring

by

Gustavo Adolfo Gil

Submitted to the Department of Electrical Engineering and Computer Science
on August 5, 2005, in partial fulfillment of the
requirements for the degrees of
Bachelor of Science in Electrical Science and Engineering
and
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Online monitoring of bioprocesses is essential to expanding the potential of biotechnology. In this thesis, a system to estimate concentrations of chemical components of an *Escherichia Coli* fermentation growth medium via a remote fiber-optic Raman spectroscopy probe was studied in depth. The system was characterized to determine sources of instability and systematic error. A complete first-order error analysis was conducted to determine the theoretical sensitivity of the instrument. A suite of improvements and new features, including an online estimation of optical density and biomass, a method to correct for wavelength shifts, and a setup to increase repeatability and throughput for offline and calibration methods was developed accordingly. The theoretical and experimental ground work for developing a correction for spectrum distortions caused by elastic scattering, a fundamental problem for many spectroscopic applications, was laid out. In addition, offline Raman spectroscopy was used to estimate concentrations of fructose, glucose, sucrose, and nitrate in an oil palm (*Elais guineensis*) bioreaction. Finally, an expansion of optical techniques into new scale-up applications in plant cell bioprocesses, such as plant call differentiation was explored.

Thesis Supervisor: Rajeev Ram
Title: Associate Professor

Acknowledgments

The good people that have contributed, directly and indirectly, to the creation of this document are too many to mention. I will do my best to thank as many as I can remember, although this is in no way an exclusive list. I hope that in my characteristic bad memory, I do not offend those that I inadvertently omit, but hope that they know that their contributions to this thesis or to my life are appreciated and treasured.

First and foremost, I would like to thank my advisor, Rajeev Ram, without which this project would not even have existed. His love of science and engineering is an inspiration that kept me always wanting to push forward and do more. From the moment I signed on, I don't think I have ever seen him frown. Also instrumental to this project was Harry Lee, who pioneered the Raman work and got me up to speed quickly. His help during the long process of the fermentation alone would be worthy of note, but his constant willingness to help makes him more than a good co-worker, it makes him a good friend. Also contributing directly to this Project were numerous members of the Sinskey lab. Paolo Boccazzi deserves particular note for always nicely and promptly answering my seaming stream of questions about the biology and for helping to run the fermentation. Joerg "Joe" Schoeneit deserves much thanks for running the HPLC samples for both the E. coli and the oil palm samples. Laura Willis and Phil Lessard both deserve thanks for their consultation and help with the oil palm fluorescence stuff. Finally, I'd like to thank Anthony Sinskey for letting me use so much of his valuable staff's time. I'd also like to thank Daniel Choi for preparing the oil palm samples.

Special thanks also go out to all of the POE lab guys whose input helped out in so many ways during my time here. I'd like to thank Peter Mayer for being a huge pillar of moral support by always listening to me when I vented my concerns and worries, and offering sanguine and compassionate advice. Thanks go out to Tom Liptay for his infinite graciousness in letting me hog the spectrometer, as well as for his stimulating conversation about all economics and financial markets. Much thanks to Kevin Lee, who with whom I would bounce ideas off with and crack jokes. Kevin proves a man can be super smart and hardworking and still be funny, upbeat, and full of energy. I'd like also to thank Tauhid Zaman and Xiaoyun Guo for letting me share an office with them.

Of course, I would not even be here if it wasn't for those people who sculpted me into the young man I am today. Thanks go out to my parents for raising me right and always putting Mara and I first. To my father, who taught me that great passion, tempered by integrity are the stuff that leads every Venezuelan man to the three important things in life, salut, dienro, y amor. To my mother, who taught me that balance in life is important for the mind and soul. To my sister Mara, who always teaches me to enjoy the simpler things n life and to never take myself too seriously. To my uncle Darin, who, at a very critical juncture in my life, showed me that I and I alone am the only master of my destiny. Finally, to my darling Katherine, whose love and support give me a reason to exist. And finally, to God, for sending me my darling Katherine.

I'd also like to thank Cindy Garcia for believing in a shy stuttering kid. Her support and help turned that kid into an Eagle Scout. I would also like to thank Kelly "Stewy" Stewart, for always giving me advice and shaping me into a good leader.

Finally, I would like to thank a personal hero. He is the man who taught me the value of a dollar, that hard work is its own reward, and that it's possible to regularly eat four cheeseburgers for dinner and still be thin in your 50s. I remember our trips in his big old car, and I remember his sagely advise on life, his fatherly love, and his energy that kept him going up and down the basketball court in that striped uniform. Although our paths in life have diverged, I will always remember Louis "Lou" Hardy as my own father, and hope that he would consider me as his own son. For that reason I have, and will always have, that red and black jacket.

Thanks Lou.

Contents

List of Figures

11

List of Tables

Chapter 1

Introduction

1.1 Motivation

Bioprocesses are processes that use living cells or their components (e.g. enzymes, chloroplasts) to effect physical or chemical changes, usually for the production of a desired chemical or biological product. In essence, the cells in a bioprocess are genetically engineered microscopic factories that produce chemicals and tissues. These cells are usually bacterial single cell organisms, which are used to produce a large assortment of chemicals for use as components of drugs, food additives, and polymers [1, 2, 3]. However, cell suspensions of plant and animal cells for propagation of clonal material also exist, usually for medical applications. To optimize a bioprocess, the biomass needs to grow in a tightly controlled environment, requiring reliable, online monitoring of all conditions, including pH, temperature, oxygen transfer rate, turbidity, and chemical composition. Previous monitoring technologies required individual chemical or biological sensors for composition monitoring. As the number of chemicals to be monitored increases, this method becomes increasingly difficult to implement. A possible solution to this scaling problem is a system that can measure the concentrations of multiple chemical components in a bioreactor. Such a method should also be non-invasive and contribute little to no risk of biological contamination. These requirements are easily met with optical methods, which can acquire data without causing chemical changes or physical tampering of the bioreactor or peripheral

equipment. Finally, many biological species are optically active in some way, indicating that optical and spectroscopic methods are well suited for bioprocess monitoring [4]. Many different optical techniques have already been implemented. Two-dimensional fluorescence spectroscopy has been demonstrated for bacterial growth [5], but suffers from interference of the fluorescence of medium components with the desired signal. Near-infrared and mid-infrared absorption spectroscopy have likewise been explored [6], but suffer from a strong signal from water, due to the strong IR absorption of O-H bonds. For a chemical composition monitoring application, low interference from both the biomass and water is desired.

Raman Spectroscopy offers a noninvasive, reagentless, cheap, and remote method that exhibits a weak signal from water. Recent advances in holographic filters and Charge Coupled Device (CCD) cameras have made Raman spectroscopy a viable choice for bioprocess monitoring [7, 8]. Previous Raman systems attempted to monitor particle concentrations using algorithms such as partial least squares regression and machine learning techniques using training samples, thereby accounting for other effects implicitly [9, 10]. However, a simple model treating a Raman spectrum as the linear combination of component Raman spectra and using linear algebra techniques to quickly determine the error has recently been devised and demonstrated in *Escherichia coli* bioreactions [11]. Such a system eliminates the need for training the algorithms, thereby lowering operating costs. Furthermore, by using this more explicit method, it is possible to quantify the limits of detection.

Using explicit methods, previous results of the *in situ* performance of Raman spectroscopy were compared to High Performance Liquid Chromatography (HPLC) for two separate bacterial fermentations. These results show estimation errors are not dominated by the noise in the measurement spectra, but by systematic model errors or noise in the calibration spectra, particularly in the case of the online measurements [11]. It is therefore desirable to improve the theoretical error model to include errors in the calibrations and reduce the systematic experimental errors. Particular attention should be focused on the wavelength dependence of scattering from large particles, as it is one of the largest contributors to the error in the online measurements. With all of these corrections, Raman spectroscopy using explicit methods could be an ideal solution

for concentration estimation.

The utility of Raman spectroscopy is not limited to concentration estimation of microbial bioreactions. Some other applications and functions of this technique have yet to be exploited. For microbial bioreactions, where the populations can become very large, it is desirable to get an online measurement of the biomass, and while optical density (OD) probes exist for this application [12], it is desirable to do so with the Raman probe directly. Furthermore, to the author's knowledge, Raman spectroscopy has not been applied as a plant cell culture monitoring solution before. Since these bioreactions generally involve the use of multiple sugars, Raman spectroscopy shows particular promise, since it shows significant contrast between these otherwise very similar molecules. For concentration estimation, there is little difference in the experimental methods used for plant cell bioreactions, but an investigation into Raman spectroscopy's ability to resolve components relevant to these bioprocesses has yet to be conducted. Finally, Raman spectroscopy could be applied to other functions for plant cell cultures, including sorting and differentiation applications. The central aim of this thesis is to characterize the errors and limitations as well as the quantitative utility and growing applications of Raman spectroscopy for bioprocess monitoring.

1.2 Raman Spectroscopy

Although Raman scattering was discovered in 1928, it has become a convenient and available technique only in the last two decades. The weakness of the Raman signal with respect to other optical signals made acquiring spectra a difficult task. The current state of-the-art allows Raman spectroscopy to be not only possible, but affordable and convenient. As mentioned earlier, this has been largely due to technological advances involving all aspects of spectroscopic acquisition. For example, an integrated CCD/dispersive spectrometer available in 1997 provided approximately 50,000 times more signal than a single channel system in 1985 with similar excitation power and integration-time [13]. These technological improvements have allowed for a proliferation of Raman spectroscopic techniques and applications, making the field subject to

thousands of research papers and dozens of monographs. What follows is a brief overview of the fundamentals of dispersive vibrational Stokes-Raman spectroscopy along with some brief information about other techniques. It is by no means comprehensive, and the reader is encouraged to examine the references for further information.

1.2.1   Raman Scattering

Excited   ───────────────────────────────────── v = 3
Electronic   ───────────────────────────────────── v = 2
State   ───────────────────────────────────── v = 1

Virtual
State
      $h\nu_0$ $h\nu_0$ $h\nu_0 - h\nu_1$

Ground   v = 3
State   v = 2
                                                          v = 1

Rayleigh   Stokes   Anti-Stokes   Resonance
Scattering   Raman   Raman   Raman
                Scattering   Scattering   Scattering
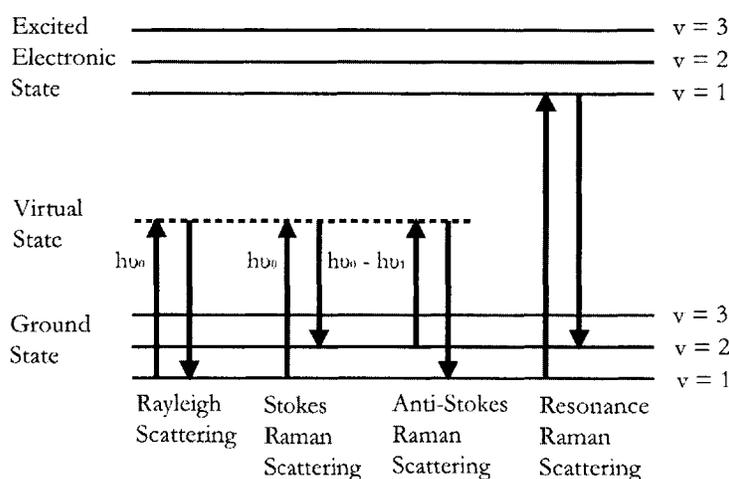
Figure 1-1: Spectroscopic transitions for various kinds of Raman spectroscopy. ν indicates laser frequencies while v is the vibrational quantum number.

Raman spectroscopy takes advantage of Raman scattering, discovered by C. V. Raman in 1928 [14, 15]. Raman scattering, also referred to as inelastic light scattering, is caused by the interaction between the optical oscillations of light with the vibrational motion of molecules. To explain Raman scattering, consider the case of a monochromatic beam of light with energy $h\nu_0$. When the light comes into contact with a group of molecules much smaller than the wavelength of the light, most of the light will scatter elastically (at the same wavelength), an effect called Rayleigh scattering. A much smaller amount of light, anywhere between 10^{-6} and 10^{-10} of the total scattered light, will scatter off of the particle with energy $h\nu_1 \neq h\nu_0$. The incident photon with energy $(h\nu_0)$ excites vibrational motion in the molecule with energy (hf), causing some of the energy to be given to the molecule, while the rest of the energy is scattered off as a new

photon. In this case, the scattered photon has a lower energy, $h\nu_1$:

$$h\nu_1 = h\nu_0 - hf \tag{1.1}$$

As seen in Figure 1-1, when the scattered light has less energy than the incident light ($h\nu_1 < h\nu_0$), it is referred to as Stokes Raman scattering. The case where $h\nu_1 > h\nu_0$ is referred to as anti-Stokes Raman scattering. Since anti-Stokes Raman scattering requires that the molecules begin in an excited state, it is generally much weaker.

In general, all spectroscopies can be viewed as ways of observing a mechanism by which the incident radiation interacts with the molecular energy levels of a sample [16]. In Infrared (IR) spectroscopy, like Raman spectroscopy, the mechanism is molecular vibration. Unlike Raman, however, IR spectroscopy probes the vibrations directly. An incident photon vibrating at a frequency of ν is incident on a molecule vibrating at the same frequency, the photon will be absorbed, increasing the amplitude of the vibration. For this reason, vibrational Raman and IR are called vibrational spectroscopies. For fluorescence spectroscopy, the observed mechanism is the spontaneous emission from electrons settling to the ground state from an excited electronic state. For vibrational Raman spectroscopy, the mechanism is the interaction of radiation with a polarizable electron cloud, modulated by molecular vibrations. Raman scattering can also be modulated by rotational changes in the excited molecules, but these are lower in energy than vibrational transitions. As a consequence, this work will focus on vibrational Raman spectroscopy. More information about the theory of Raman scattering can be found in the literature [17, 18, 19].

A number of factors can affect the intensity of a Raman signal. McCreery [13] summarizes all of the factors with the simple equation:

$$I_R = I_0\sigma_j Ddz \tag{1.2}$$

where I_R is the intensity of the Raman signal, I_0 is the intensity of the excitation light at the sample, D is the number density of scatterers, dz is the path length or depth of field of the

laser in the sample, and σ_j is an empirically determined cross-section of Raman scattering. The cross-section of scattering is used here mostly for cultural reasons, as many other scattering applications use this notation. It is not well suited here, however, because it is the cross-section over all angles, i.e., all 4π steradians around the sample. For Raman spectroscopy it is more useful to define the differential Raman cross-section, β:

$$\beta(cm^2 molecule^{-1} sr^{-1}) = \frac{d\sigma_j}{d\Omega} \tag{1.3}$$

In the literature, Raman intensities are often expressed in terms of the differential cross-sections to facilitate repeatability. Since β is an empirical quantity, a number of observations as to what can affect β have been summarized by various authors [13, 16]. Note that many of these are empirical and should be treated as general rules, not fundamental laws:

1. Stretching vibrations associated with chemical bonds (i.e. changes in the length of the bonds) should be more intense than deformation vibrations (i.e. changes in angle of the bonds relative to each other).

2. Molecules with only single C-H, C-O, and C-C bonds usually have small cross-sections.

3. Molecules containing large or electron-rich atoms, such as sulfur or iodine, will have larger cross-sections. Likewise, molecules with small and electron-poor atoms, such as H_2, CO, and N_2 will have smaller cross-sections.

4. Multiple chemical bonds should create intense stretching modes and thus bigger cross-sections. For instance, a C=C vibration will be more intense than a C-C.

5. Bonds in large, polyatomic molecules, such as the S-S linkages in proteins, can give rise to large stretching modes and thus large βs.

6. Spectra acquired in liquids will be higher by factors of 2 to 4 than those for the same vibration in gases due to field effects [13]. Note that this is independent of the smaller particle density, D, of gases, which will be the dominant limiter of the overall Raman signal.

7. The Raman spectrum of a substance in solution is different than the substance itself as a liquid. This is due to disassociation. Molecules in solution may be separated by the solvent.

8. The Raman spectrum of an acid or base will change with pH. Specifically, it will be disassociated from H^+ or OH^- respectively into a salt.

9. If the energy added by the excitation is large enough to reach the first excited state, resonant effects can cause a massive and selective increase in signal. This is referred to as resonance Raman spectroscopy.

1.2.2 Raman Spectroscopic Technology
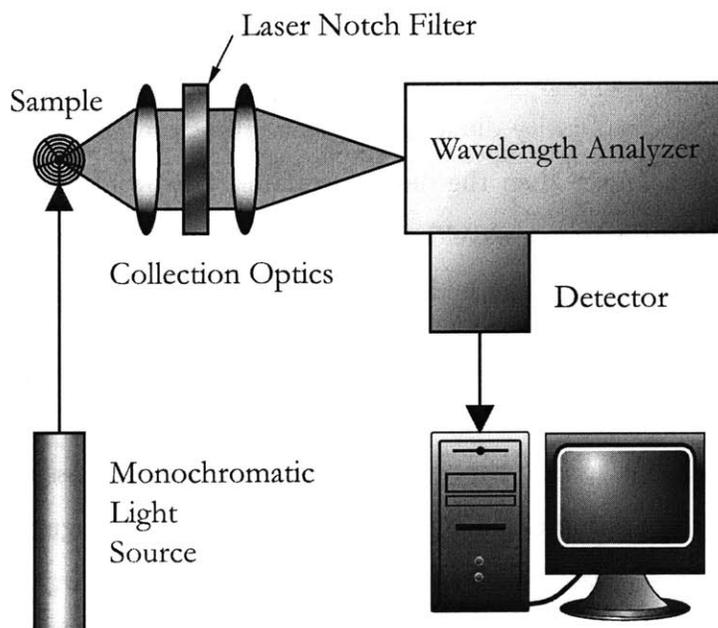
Figure 1-2: Generic Raman spectroscopy system sowing main components: Light source, collection optics, wavelength analyzer, detector, and computer. Collection is at 90°, although many other collection geometries are commonly in use as well.

In order to acquire Raman spectra, it is important to understand the components that make up a typical spectroscopy setup. Figure 1-2 displays a generic Raman spectroscopy setup and

its main components. A radiation source illuminates a sample and the scattered light from the sample is filtered and collected into a wavelength analyzer of some kind, which separates the light into a spectrum that can be captured by a detector and saved on a computer. Other light delivery and collection geometries exist as well, but the principles remain the same.

Since it is the Raman shift relative to the excitation wavelength that is desired, it is important to have a very narrow-band light source. Furthermore, since the Raman signal is extremely weak, even small amounts of sideband light from a laser can interfere. For this reason, Raman spectroscopy commonly uses high-power lasers, such as Ar^+ ion (514.5nm) and Nd:YAG (1064nm) lasers, with significant laser-line filtering. These systems offer very stable excitation, but at high cost. Frequency stabilized diode lasers are also emerging as a low cost alternative. Previously, the temperature dependence of the gain curve as well as the multimode operation and low power made these lasers unsuited for Raman spectroscopy, but with modern thermo-electric coolers (TEC) and external cavity techniques, these problems can be reduced so as to be insignificant [20]. An external cavity diode laser uses a diffraction grating to setup an external resonator which is much longer than the diode resonator, narrowing the line width. Moving the grating can also tune the center wavelength, allowing for more control over laser operation. More information on these devices can be found in the literature [20, 21].

As mentioned above, there are many different collection geometries possible for Raman spectroscopy. Figure 1-2 shows the 90° collection for ease of illustration, but increasingly, 180°, or backscatter, collection is used. This is mostly due to alignment issues. Using the same lens for collection and excitation eliminates the need to separately align the collection optics. One advantage of using separate collection optics, however, is that larger numerical aperture optics can be used, allowing for a greater collection of the scattered light. The ability to use the backscatter geometry is what makes fiber Raman probes possible, and most of these probes made commercially today implement filtering capability in the probes themselves. In both cases, the Rayleigh scattered light must be filtered out so hat it does not interfere with the signal. Rejection of the Rayleigh scattered light was one of the largest experimental issues to overcome before modern holographic filters [7]. Presently, holographic notch filters with high rejection

(OD > 8) and small transition bands are available commercially.

(a) Classic Chezny-Turner Dispersive Spectrome- (b) Michelson Interferometer FT-Spectrometer
ter
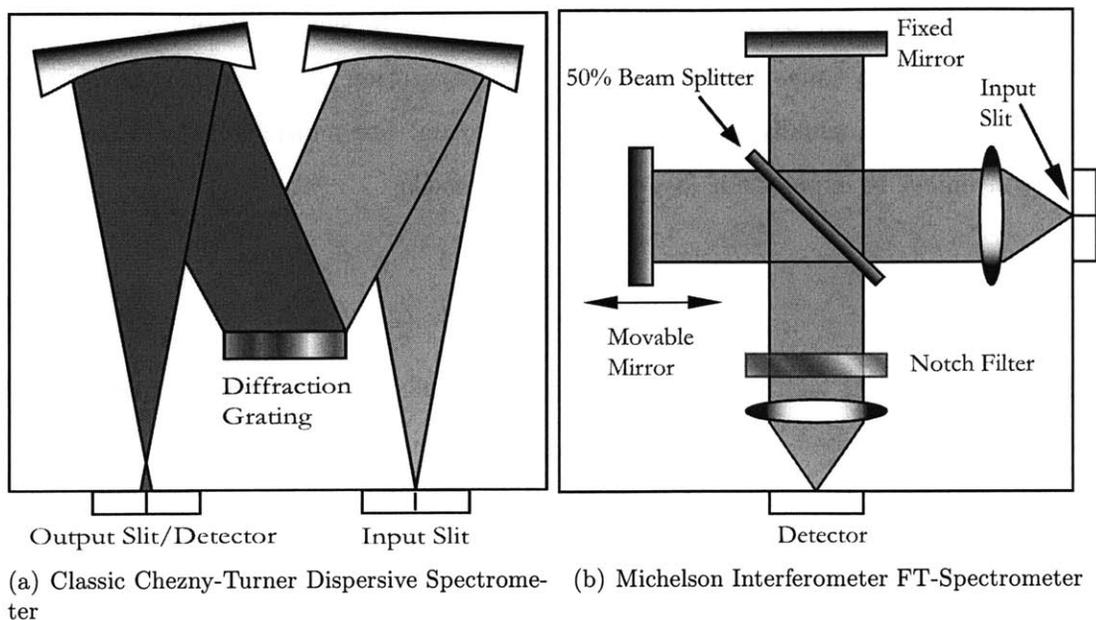
Figure 1-3: Wavelength analyzers for Raman spectroscopy.

Before 1986, the only available spectrometers for Raman spectroscopy were single-channel scanning systems with as many as three diffraction gratings. Starting in 1986, however, Fourier transform (FT) Raman spectrometers became available. In these systems, a Michelson interferometer is used for frequency selectivity. By moving a mirror at a constant velocity, the fringes on the detector will move at a constant rate determined by the speed of the mirror motion and the wavelength of the light, as shown in Figure 1-3(b). This will create a time varying intensity signal on the detector A Fourier transform of this signal will then reveal the Raman spectrum. These systems are primarily used for high resolution and long wavelength applications, but suffer from low sensitivity and generally have poor SNR. Figure 1-3(a) shows a typical dispersive spectrometer in the classic two mirror (Chezny-Turner) configuration. Dispersive spectrometers such as this one spectrally separate light via a diffraction grating that disperses the light according to wavelength. Using a lens or curved mirror, that light can be separated in space. This gives the choice of either single or multichannel operation. If a slit is put at the output plane, a signal wavelength will be resolved. Scanning the wavelength will therefore produce the Raman

spectrum. If a detector with numerous pixels is placed at the output plane, such as a CCD camera, all wavelengths can be observed at once, increasing the sensitivity of the overall system for the same acquisition time. More specifically, this separation of wavelengths is described as linear dispersion, $dl/d\lambda$, usually stated in units of mm/nm. For multichannel devices, however, it is more convenient to express the dispersion in nm/pixels:

$$\frac{nm}{pixel} = \left(\frac{d\lambda}{dl}\right) W_p \tag{1.4}$$

where W_p is the physical width of a pixel along the wavelength axis and the reciprocal of the linear dispersion is described by:

$$\frac{d\lambda}{dl} = \frac{\cos\theta}{pmF_2} \tag{1.5}$$

where p is the pitch of the grating, m is the diffraction order (0,1,2,...), F_2 is the focal length of the focusing mirror, and θ is the angle of the diffracted light leaving the grating relative to the grating surface normal. Since θ is typically very small, it is common to refer to the linear dispersion as a constant in sales literature. Note that this is a property of the grating and spectrometer only, indicating that changing excitation sources will not affect the dispersion over wavelength. However, what is important in Raman spectroscopy is the frequency shift, usually expressed as wavenumbers. Using the same grating and spectrometer will produce a different range of wavenumbers if the grating is moved to a different center wavelength. It is therefore desirable to convert Equation 1.5 to determine the dispersion with respect to the wavenumber, $\bar{\nu}$:

$$\frac{d\bar{\nu}}{dl} = \frac{\cos\theta(\bar{\nu})^2}{pmF_2} \tag{1.6}$$

This sets the resolution per pixel in terms of (cm^{-1}/pixel):

$$resolution = \left(\frac{d\bar{\nu}}{dl}\right) W_p \tag{1.7}$$

Note that this assumes that the resolution is limited by the grating, and not by the width of the input slit of the spectrometer. If the slit width is greater than the pixel width of the detector, the resolution will be the slit width. The resolution is also different for every pixel, since it varies with wavenumber. Thus, multiplying the resolution calculated at a particular pixel by the number of pixels will only yield an approximate solution to the spectral range. The difference in absolute wavenumber is fairly small for most of the common excitation wavelengths used today, so the approximate solution is adequate, especially considering the small finite number of options for spectrometers and gratings:

$$range = resolution(\mu_{mid}) \times N_p \qquad (1.8)$$

Where μ_{mid} is the resolution computed at the wavenumber corresponding to the middle of the desired range, and N_p is the number of pixels.

The weakness of Raman signals made photon counting photomultiplier tubes (PMTs) the dominant detectors for Raman spectroscopy until the introduction of Charged Coupled Device (CCD) cameras. In PMTs, a photon strikes a photocathode, ejecting a photoelectron, which is accelerated by an electric field into a dynode, which releases more electrons and strikes another dynode, and so on, all the way to an anode, amplifying the signal by a large factor along the way (10^4 to 10^6). PMTs are physically large in size, making them ill-suited for multichannel operation. CCDs, however, are multichannel devises that can have small pixel sizes. A single CCD array can have millions of pixels. Since the light from a dispersive spectrometer is dispersed linearly, one dimension of pixels can be binned together to increase sensitivity. CCDs are typically sensitive between 200nm and 1100nm, making them practical for UV, Visible, and NIR excitation. In each pixel, an incident photon will create an electron/hole pair in the semiconductor (usually silicon). A metal plate held at positive potential attracts the electrons and holds them in the region of semiconductor close to it, creating a well of electrons. These wells can store up to 10^6 electrons before they are at full capacity. Thus, the well acts as an integrating detector. After the integration period, the electrons are cleared electronically and the signal is converted to counts by an A/D converter. The gain of a CCD is the number of

stored electrons required to yield one count (e^-1/count). This is an important figure of merit, as all noise statistics must be calculated using the number of electrons, not number of counts.

1.2.3   Noise and SNR in Raman Spectra

The weakness of Raman signals makes noise analysis an important part of any application that is quantitative. Raman spectra are typically analyzed digitally on a computer by software that quantizes the signal's spectrum into bins and the signal's amplitude into counts. To understand the noise contribution, consider an individual bin. An individual instantiation of that bin will have a value, n, that will be a random variable with an average value of μ_n, subject to some noise, defined as σ_n. More specifically, the mean represents the sum of the average signal over the average background:

$$\mu_n = \mu_S + \mu_B \tag{1.9}$$

where μ_S is the average value of the signal and μ_B is the average value of the background. The signal to noise ratio for a particular measurement is defined using these values to be:

$$SNR = \frac{\mu_S}{\sigma_n} \tag{1.10}$$

The average value, μ_n is a function of excitation power, collection efficiency, and integration-time, and is therefore set by the optics and experimental parameters of an experiment. The standard deviation of the noise can be reduced by averaging many spectra, but if the total integration time per acquisition is fixed, it is important to understand the sources of noise in order to reduce them. Table 1.1 shows all of the sources of noise in a typical Raman system and their designations. These noise sources are related to the overall noise by:

$$\sigma_n = (\sigma_S^2 + \sigma_B^2 + \sigma_D^2 + \sigma_F^2 + \sigma_R^2)^{1/2} \tag{1.11}$$

Shot noise due to the background and signal in this thesis will be generally considered together

Noise Sources in Raman Spectroscopy		
Noise Type	Cause or Origin	Designation
Signal shot noise	Poisson distributed photon flux	σ_S
Background shot noise	Poisson distributed photon flux	σ_B
Readout noise	Electronics (A-D conversion)	σ_R
Dark current noise	Thermal generation of e^-	σ_D
Flicker noise	Laser power fluctuations	σ_F

Table 1.1: Sources of noise, their causes, and their designations.

for the sake of convenience as simply shot noise, which varies with a standard deviation of:

$$\sigma_s = (\sigma_S^2 + \sigma_B^2)^{1/2} \tag{1.12}$$

Shot noise in both cases is fundamental to spectroscopy and results from any process governed by Poisson statistics, such as photons from a laser source. Poisson statistics describe processes which involve counting some quantity that arrives at random intervals with an overall average arrival rate. In the case of spectroscopy, we are concerned with counting photons (PMTs) or electrons (CCDs). If n is large, it can be shown that the standard deviation of the Poisson process is the square root of the mean:

$$\sigma_s = \mu_n^{1/2} \tag{1.13}$$

More information about how this value is eerived can be found in Appendix A. It is clear from Equation 1.13 that in the case where shot noise is the dominant error, i.e. $\sigma_n \approx \sigma_s$, the SNR will be completely described by the means:

$$SNR = \frac{\mu_S}{\mu_n^{1/2}} = \frac{\mu_S}{(\mu_S + \mu_B)^{1/2}} \tag{1.14}$$

Interpreting this expression reveals two limiting cases. In the case where the background is negligible compared to the signal, the $SNR = \mu_S^{1/2}$. This suggests a certain theoretical maximum to the error that cannot be exceeded for light sources that obey Poisson statistics. In the limit where the background is much larger than the signal, the $SNR = \mu_S/\mu_B^{1/2}$, indicating that

SNR will be small despite the large overall signal. This case shows that fluorescence and stray excitation light can significantly reduce the SNR of a Raman spectrum. The large background shot-noise limit will play a significant role in determining the lowest possible error for a Raman bioprocess monitoring system.

Readout noise refers to the process of converting electrons from the detector to the digital data that can be manipulated. More specifically, if a bin could be reliable filled with the same number of electrons for multiple iterations, the standard deviation of the digital values that would be produced by the ADC is the readout noise, σ_R. This property of the electronics and is typically very small in the current state-of-the-art and can typically be ignored for situations where the signal is large. Flicker noise refers to the noise created by fluctuations in the laser power and is only important for scanning spectrometers. As the laser fluctuates in power, the amplitude of the Raman spectrum will change. For a scanning spectrometer, this could cause the appearance of noise if the fluctuations are much faster than the scanning speed. For multichannel spectrometers, all of the bins are filled simultaneously, so laser power fluctuations cause all bins to be scaled simultaneously, making flicker noise only an issue when averaging multiple acquisitions. Even this problem can be achieved by normalizing the spectra before analyzing them. Finally, dark current noise arises from thermal noise in the detector. This source of error can be reduced by cooling the detector, and modern liquid nitrogen cooled detectors have made this problem virtually non-existent.

1.3 Bioreactors and Bioprocesses

The design of stirred-tank bioreactors has undergone only very moderate changes over the last 40 years, do in part to the complexity of biological processes, which makes the costs of radical technology changes difficult [22]. Liquid bioreactors vary in size from 500mL for some research reactors to 300,000 L for industrial reactors. The principles of operation remain the same over the size and application range. A schematic of a typical stirred tank bioreactor is given in Figure 1-4. Bioreactor ports are typically on a top-plate secured to a tank with a tight o-ring seal to
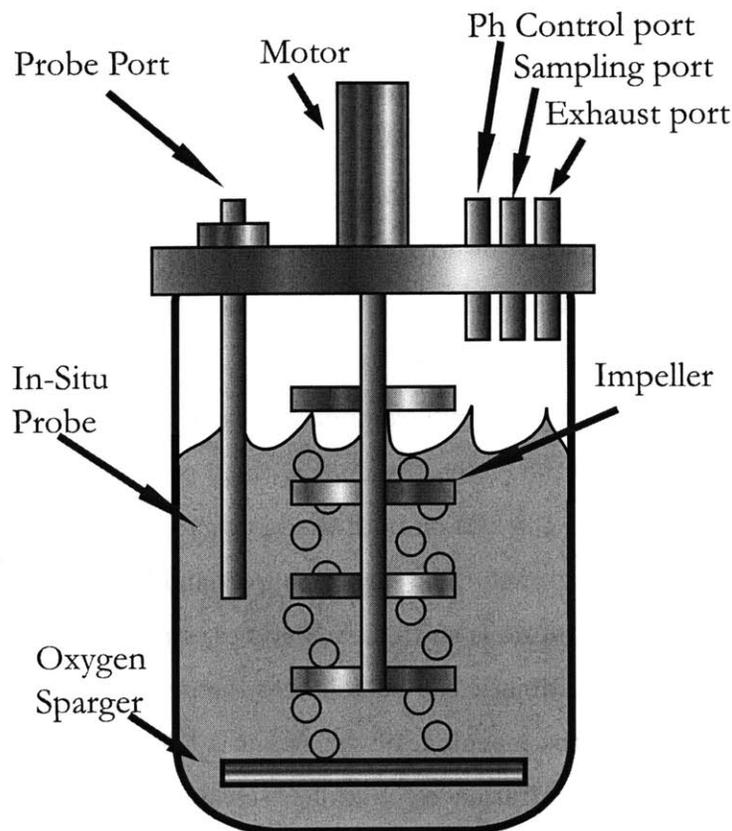
Figure 1-4: Schematic of typical stirred-tank bioreactor.

prevent contamination. Agitation is provided by impellers on a central shaft driven by a motor. Aeration is used to deliver oxygen via a sparger at the bottom of the tank. For small tanks, this combination can result in close to perfect mixing of the bioreactor medium, but as the size of the bioreactor increases, the probability and lifetime of localization of the oxygenation of the medium becomes problematic [23].

The bioreactor must remain a perfectly controlled system for industrial applications. First the tank is typically placed in a heated water bath to heat it as uniformly as possible. Once a bioreaction has begun, all input of materials is done through a finite set of ports. Injection of acid and base have separate dedicated ports; a sampling port is used for extraction; a third port is used for input of antifoam and other materials; and a condenser is placed on a fourth to release the output gas without losing medium and water. This off-gas is sometimes also monitored for

chemical content. In this way, the bioreactor seeks to make the entire system closed, not just for contamination elimination, but also for process efficiency, as it is desirable to create the most product using the least amount of material. With the environment successfully controlled, the conditions of the bioreactor need to be monitored closely to control the production process. A number of *in-situ* probes are typically inserted into pre-made ports for monitoring pH, Dissolved Oxygen (DO), and increasingly, Optical Density (OD) to measure biomass. As we will soon see, these ports are well suited for a Raman probe as well.

To run a fermentation, all of the medium components except for the carbon sources (sugars) and supplements are placed into the bioreactor along with all probes and equipment and sealed. Steam sterilization at 120°C and 138 kPa on both the bioreactor and the carbon sources ensure no other bacteria survive before growth. The carbon sources and filtered supplements are added to the assembly and the fermentation is started. Sometimes, strains resistant to one particular antibiotic are used so that the antibiotic can be added to the medium to further reduce the risk of contamination. Soon after, the inoculum of the strain is added and the bioprocess begins. Under current technology, concentration estimation is conducted by extracting samples and performing HPLC offline after the bioreaction is over. Typically, these same samples are used for OD measurements offline as well. A method to measure both online would therefore be an obvious improvement to the technology and drastically increase the efficiency of all bioprocesses.

1.4 Previous Work

Raman spectroscopy has already been demonstrated to work as an offline monitoring system for microbial bioprocesses. FT-Raman spectroscopy has been used to simultaneously estimate glucose and ethanol in *Saccharomyces cerevisiae* during ethanol fermentation [24] using partial least squares (PLS) regression, which was capable of constraining errors to under 10mM offline. This method was later used to observe lactic acid offline with 8mM of accuracy [6]. Separately, offline concentration estimates of Gibberellic acid 3 (GA3) were attempted using PLS and artificial neural networks as the concentration estimation algorithms [10]. Online concentration

estimates for glucose and ethanol were performed on a shake-flask culture of bakers yeast with PLS regression using a dispersive Raman setup and NIR excitation, and showed estimates within 5% for all measurements [9]. All of these techniques used implicit methods, however, making it impossible to propagate the error due to noise and come to a fundamental understanding to the limits of detection of the instruments.

Bioprocess monitoring via *in situ* Raman spectroscopy using explicit methods has been demonstrated on *Eshericia coli* bioreactions [11]. An algorithm was developed where calibration spectra of all of the pure components were acquired at known concentration before the bioreaction. Cosmic rays were removed, water was subtracted, and the spectra were normalized to create a basis set that, when combined with basis vectors for a fourth order polynomial, composed the columns of a pure spectrum matrix, K. Concentrations, c, were estimated from the measured spectra, s, in a least-squares sense by solving $s = Kc$. To correct for scattering, all estimated concentrations are normalized to the estimated concentration of water. This algorithm relied on a number of assumptions: The total Raman scattering collected was a superposition of pure component Raman spectra; the interference from smoothly varying fluorescence was additive in nature; noise in the spectra is dominated by shot-noise in the detector, there are background subtraction errors in the pure spectra; and scattering from biomass and air bubbles affected primarily the amplitude of the signal.

Results of Previous Work			
		Max Error [mM]	
Component	SNLD [mM]	BR1	BR2
Phenylalanine	0.04	1.0	3.4
Glucose	0.10	34.5	66.5
Lactate	0.19	20.6	48.8
Formate	0.11	59.2	6.6
Acetate	0.15	39.7	87.5

Table 1.2: Error results for offline *in situ* measurements of an anaerobic fermentation (BR1) and an aerobic fermentation (BR2), along with the previously calculated shot noise limit of detection (SNLD).

The results of this previous work presented an error model predicting a noise floor in the

$100\mu M$ range for most of the components. Meanwhile, *in-situ* measurements of the components showed errors in the 10mM range for 2 separate bioreactions. Measurements are *in situ* but not online because of a distortion of the Raman spectrum of the sapphire window after autoclaving, which created the need to acquire a new calibration spectrum for sapphire after the bioreactions. These results show that errors were two orders of magnitude above shot-noise predicted levels, indicating that there were significant systematic errors. A number of possible systematic errors were proposed. First, errors arising from noise in the calibrations were not modeled. Second, additional systematic errors caused by the calibration setup were unknown. Third, the temperature dependence of the fiber Raman probe and the components were unknown. Fourth, while the change in the sapphire spectrum after autoclaving was known, its true effect on the estimates could not be known without further experiments. Last, the wavelength dependence of the scattering off the biomass and air bubbles was ignored. In addition, other effects not suspected at the time were also causing errors. First, errors caused by drift in the laser line in wavelength were unnoticed and thus not corrected. Second, calibration spectra for the organic acids were not pH controlled, resulting in calibration spectra for these components that were erroneous. Finally, errors in the HPLC concentration estimates were not questioned, even though estimated starting concentrations exceeded nominal starting concentrations by a significant amount. Investigating these issues to bring online Raman spectroscopy closer to commercial viability is the subject of this thesis.

1.5   Outline of Presented Work

The research presented in this thesis will describe experiments and theory with the goal of describing the sensitivity, resolution, and future opportunities of Raman spectroscopy as a bioprocess monitoring solution. Chapter 2 addresses the constraints of the current state-of-the-art of Raman spectroscopy monitoring systems by determining, verifying, and correcting possible sources of error from the components. This will also be done to the theoretical model. Section 2.2 covers experiments conducted to characterize the Raman spectroscopy bioprocess monitor-

ing setup to address and discover sources of error and spectral distortion. Section 2.3 outlines a number of experimental improvements to address these instabilities and sources of distortion. Specifically, Section 2.3.2 describes a method to remove the errors caused by wavelength shifts in the excitation laser, and Section 2.3.1 describes in detail the experimental constraints in acquiring repeatable calibration spectra and demonstrates a method that produces such spectra. Section 2.4 describes a complete noise model to determine the contribution of shot-noise in the calibration and measured spectra to the errors in the concentration estimates of a Raman bioprocess monitoring system.

Chapter 3 lays the groundwork for the last step needed to bring the online concentration estimates closer to the offline concentration estimates by describing the effects of scattering from the biomass and the air bubbles in the bioreactor. Section 3.1 is an overview of the rigorous (Mie) scattering solution to scattering by spheres and its applicability to the bioprocess monitoring setup we have implemented. Section 3.2 describes the distortions caused by the air bubbles and the biomass, and proposes a method of correcting them to improve the performance of the device. Finally, Section 3.3 describes a novel technique that takes advantage of the attenuation of the concentration estimates of water caused by scattering to get an online measurement of the biomass. Chapter 4 discusses a series of experiments to extend Raman spectroscopy and other optical methods to plant bioprocess applications. Section 4.2 shows a set of measurements conducted for an oil palm (*Elaeis guineensis*) cell culture. Section 4.3 discusses an exploration into spectroscopic methods for plant calli differentiation. While Raman spectroscopy turns out to be ill-suited for this application, there is promise in fluorescence spectroscopy as a method. Finally, Chapter 5 summarizes all of these findings and proposes future work.

Chapter 2

Online Raman Spectroscopy

To make an online Raman spectroscopy bioprocess monitoring system a practical solution for commercial and research applications, a number of issues must be addressed. The environmental conditions between the online measurements and the calibration spectra can give rise to systematic sources of error that limit the accuracy of the concentration estimates. It is therefore important to determine how the Raman spectra will change when placed under different conditions. If a problem that gives rise to systematic errors cannot be eliminated, then a method to correct for its effects should be implemented. To determine whether systematic sources of error exist, a complete error analysis that can predict the errors caused by noise is needed. In this Chapter, numerous stability tests are outlined and solutions to some of the issues are proposed. A complete error analysis of an *E. coli* fermentation is compared with the observed errors in the concentration estimates for both online and offline estimation.

2.1 Experimental Setup

Figure 2-1 shows a schematic of the Raman spectroscopy based bioprocess monitoring system implemented at the Physical Optics and Electronics group of the Research Laboratory for Electronics at MIT. Unless otherwise stated, this is the setup used to acquire spectra seen in this thesis. Excitation was set at 785nm to reduce the background from fluorescence while staying

(a) Spectrometer and Camera

(b) Laser Assembly

(c) Fiber Raman Probe

(d) Top-Level Setup
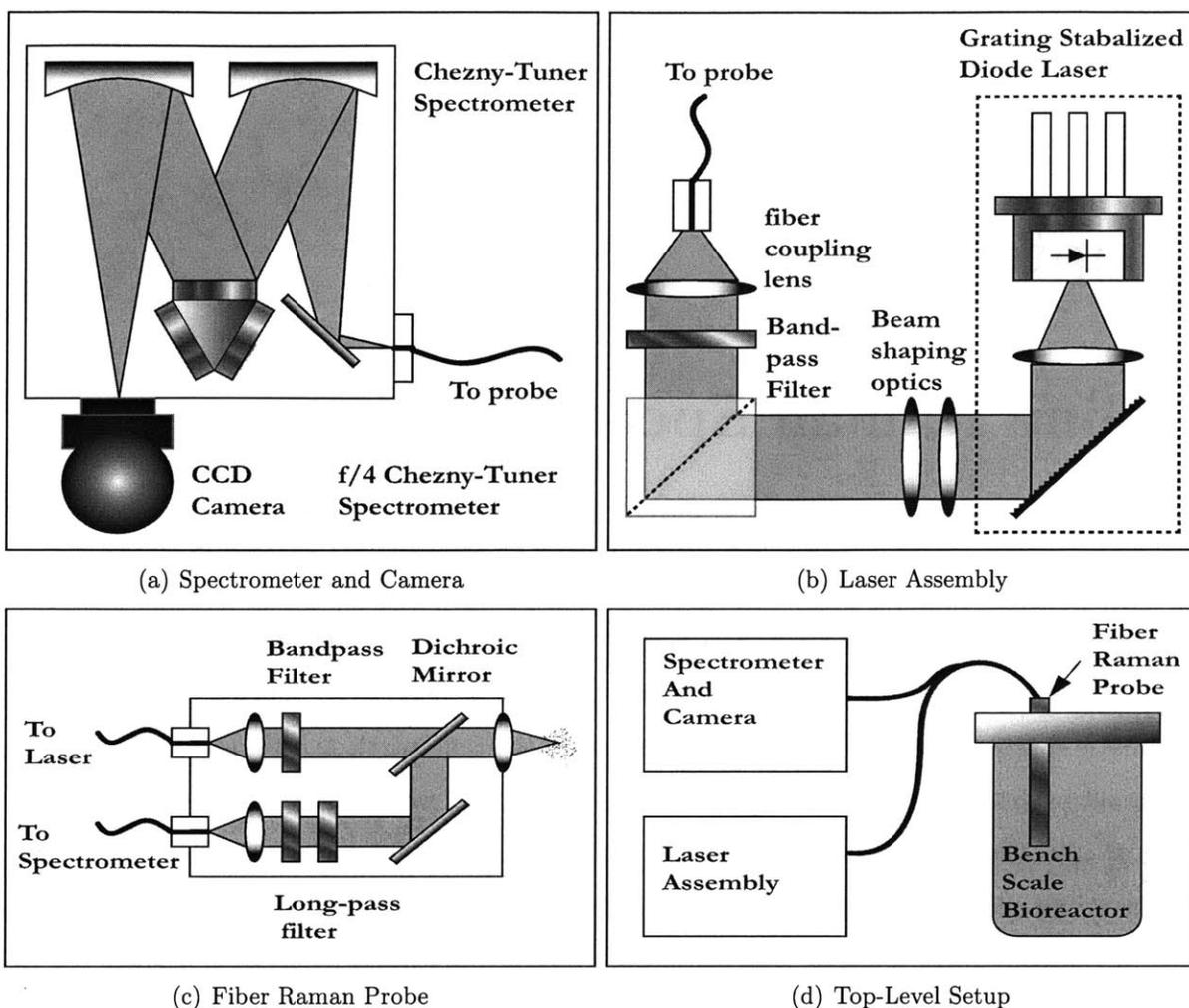
Figure 2-1: Schematics of a dispersive Raman spectroscopy bioprocess monitoring setup.

within the sensitivity of commercial CCDs. The diode laser source was grating stabilized by an external cavity laser, similar in design to [21], as shown in Figure 2-1(b). The light was then filtered by a Kaiser(TM) volumetric holographic grating and coupled into the 105μm excitation fiber of the fiber Raman probe.

Figure 2-1(c) shows the InPhotonics(TM) fiber Raman probe. Excitation light is collimated before going through a bandpass filter and dichroic mirror to remove the Raman scattered light from the excitation fiber. The narrow-band light is focused down by a lens (NA 0.4, $f = 5mm$) to a focal depth of 1mm and spot size of 200μm. Raman scattered light is then collected by

the same lens, resulting in a 180° backscatter geometry. Both the dichroic mirror and the long pass filter in the collection path attenuate the Rayleigh scattered light, resulting in an overall attenuation of OD > 8. The filter cutoff, as observed on the Raman spectra taken, is around 800nm, corresponding to a Raman shift around 200 cm^{-1} for 785nm excitation. The remaining Raman scattered light is focused into a 200μm collection fiber.

Composition of Escherichia coli Medium					
Component	Formula	Conc. (g/L)	Conc. (mM)	Raman Activity	Purpose
Glucose	$C_6H_{12}O_6$	30.0	167.0	Strong	Carbon Source
Potassium Phosphate	K_2HPO_4	10.6	61.0	Strong	Buffer/Phosphorous Source
Sodium Phosphate	NaH_2PO_2	4.2	35.0	Strong	Buffer/Phosphorous Source
Ammonium Sulfate	NH_4SO_4	2.0	15.0	Strong	Buffer/Nitrogen Source
Ammonium Chloride	NH_4Cl	3.74	70.0	Undetectable	Nitrogen Source
Magnesium Sulfate	$MgSO_4 \cdot 7H_2O$	0.2	0.76	Weak	Buffer
Calcium Nitrate	$Ca(NO_3)_2 \cdot 4H_2O$	0.0144	0.06	Undetectable	Nitrogen Source
Ammonium Molybdate	$(NH_4)_6Mo_7O_{24}$	0.0153	0.003	Undetectable	Trace Element
Boric Acid	H_3BO_3	0.025	0.4	Undetectable	Trace Element
Manganese Chloride	$McCl_2$	0.0163	0.08	Undetectable	Trace Element
Zinc Sulfate	$ZnSO_4$	0.0029	0.01	Undetectable	Trace Element
Ferric Chloride	$FeCl_3$	0.0234	0.14	Undetectable	Supplement
Tryptophan	$C_{11}H_{12}N_2O_2$	0.041	0.2	Weak	Supplement
Tyrosine	$C_9H_{11}NO_3$	0.0906	0.5	Weak	Supplement
Thiamine	$C_{12}H_{17}ClN_4OS$	0.0006	0.002	Undetectable	Supplement
Ampicilin	$C_{16}H_{19}N_3O_4S$	0.115	0.285	Undetectable	Supplement
Acetic Acid	CH_3COOH	0.0	0.0	Undetectable	Fermentation Product
Formic Acid	$HCOOH$	0.0	0.0	Undetectable	Fermentation Product
Lactic Acid	$C_3H_6O_3$	0.0	0.0	Undetectable	Fermentation Product
Succinic Acid	$C_4H_6O_4$	0.0	0.0	Undetectable	Fermentation Product
Phenylalanine	$C_9H_{11}NO_2$	0.0	0.0	Undetectable	Fermentation Product

Table 2.1: Composition of *Escherichia coli* medium.

Figure 2-1(c) shows the Acton SpectraPro 300i spectrometer. This is a single grating Chezny-Turner design, image corrected and optimized for CCD binning. The triple-grating turret of the spectrometer was equipped with a 600 lines/mm grating, yielding a theoretical resolution of 0.18nm/pixel, with a 1μm blaze for improved collection efficiency close to the region of interest. This f/4 spectrometer represents a good compromise between the high collection efficiency of low f/# and the reduced aberrational effects of large f/#. Light from the fiber probe collection fiber is refocused with a coupling lens into the entrance slit and collimated by the collimating

mirror. The collimated light is separated spectrally by the grating and refocused onto the exit aperture and dispersed over the width of the CCD camera. The CCD camera is a Princeton Instruments Spec10:400BR back-illuminated, deep-depletion camera. In addition, the camera was liquid nitrogen cooled to -85°C (as measured by a temperature sensor interfaced to the camera software), reducing the dark counts to an average of 2 counts/pixel/hr, or essentially zero. The spectrometer was run in low-noise mode, giving it a gain of about $5.19e^-$/count. All spectra were acquired with WinSpec(TM) software provided by Princeton Instruments. Signal processing was done in MATLAB.

Figure 2-1(d) shows how the pieces fit together. The Raman probe was adapted for insertion into a standard 19mm bioreactor port using a custom made assembly. The assembly consisted of an outer anodized aluminum tube, an inner brass tube, a window holder, and a window, made of either sapphire or quartz. For more information on the windows see Section 2.2.3. The inner brass tube (1.43cm O.D., 1.28cm I.D) was threaded at one end to facilitate precise axial adjustment of the probe focal point. The outer tube (1.88cm O.D., 1.43cm I.D.) was sealed with a single o-ring at the port side and threaded to accept the window holder on the other. Two window caps were designed for holding a sapphire and quartz window respectively, which were bonded with epoxy onto the caps. This housing, once assembled, was placed into a 2.5L stirred tank bioreactor (Chemap CMF100), using a typical 19mm port nut.

A strain of phenylalanine producing *E. coli*, ATCC31883, was inoculated into the bioreactor culture at 37°C. HCl and NaOH were added to control the pH at 7.0 ±0.05. A list of the medium components and products of the mixed acid fermentation are shown in Table 2.1. The impeller was set to agitate the mixture at 800rpm with filtered aeration at 1 VVM (volume air per volume liquid per minute). The strain was grown for 26 hours and samples were taken at intervals of 30 to 90 minutes, with readings of dissolved oxygen taken at these points. Offline OD_{650} measurements were taken on these samples along with HPLC measurements (Agilent 1100 series) for glucose, phenylalanine, acetate, formate, and lactate. Samples for the HPLC were first filtered by 1μm filters to remove biomass. The online Raman measurements were taken between 5 to 75 minutes apart.

2.2 Stability Tests

Previous and current results show a disparity between the accuracy of offline and online measurements. One possible explanation of this disparity is that environmental changes could cause changes in the equipment transfer functions. Alternatively, these environmental changes could be affecting the Raman cross-sections directly. It is therefore necessary to conduct tests to determine the stability of the components and identify the sources of error and thereby quantify their affects on concentration estimates.

2.2.1 Spectrum Variability

(a) Raman spectrum of a cuvette

(b) Raman spectrum of the probe optics
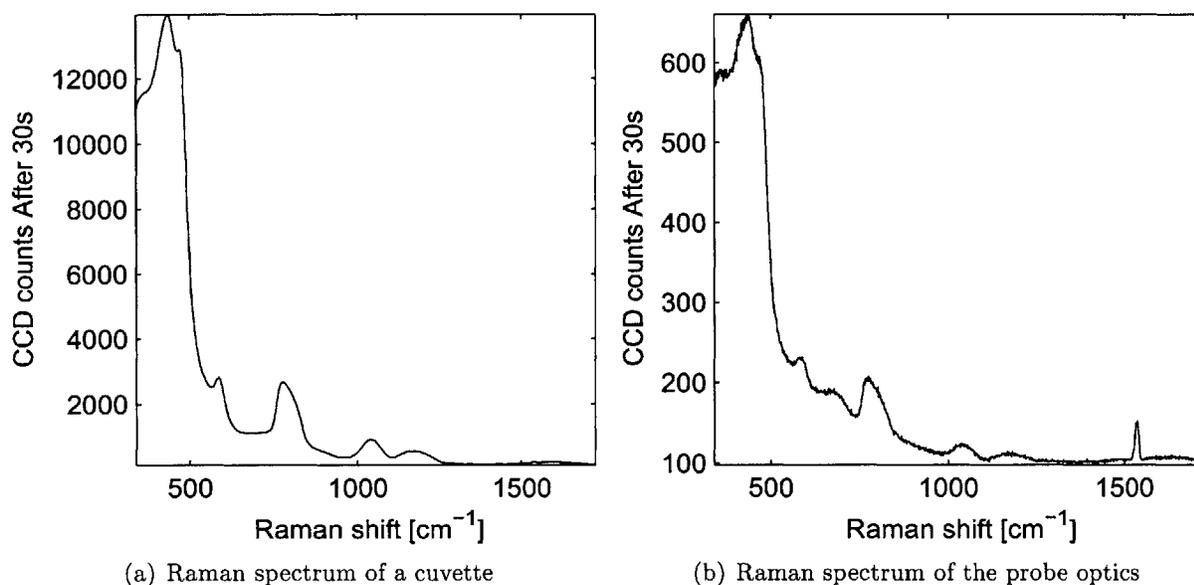
Figure 2-2: Schematics of single and multiple scattering situations.

To perform a controlled experiment, it is necessary to make a large number of assumptions. Many of these assumptions are made implicitly, such as the assumption that the optical table will not shake, or that such vibrations will not cause a change in the Raman spectra acquired. The large number assumptions make it necessary to empirically determine the variability of the setup. More specifically, it is desirable to know how much the Raman spectra taken under

(a) Mean

(b) Standard Deviation
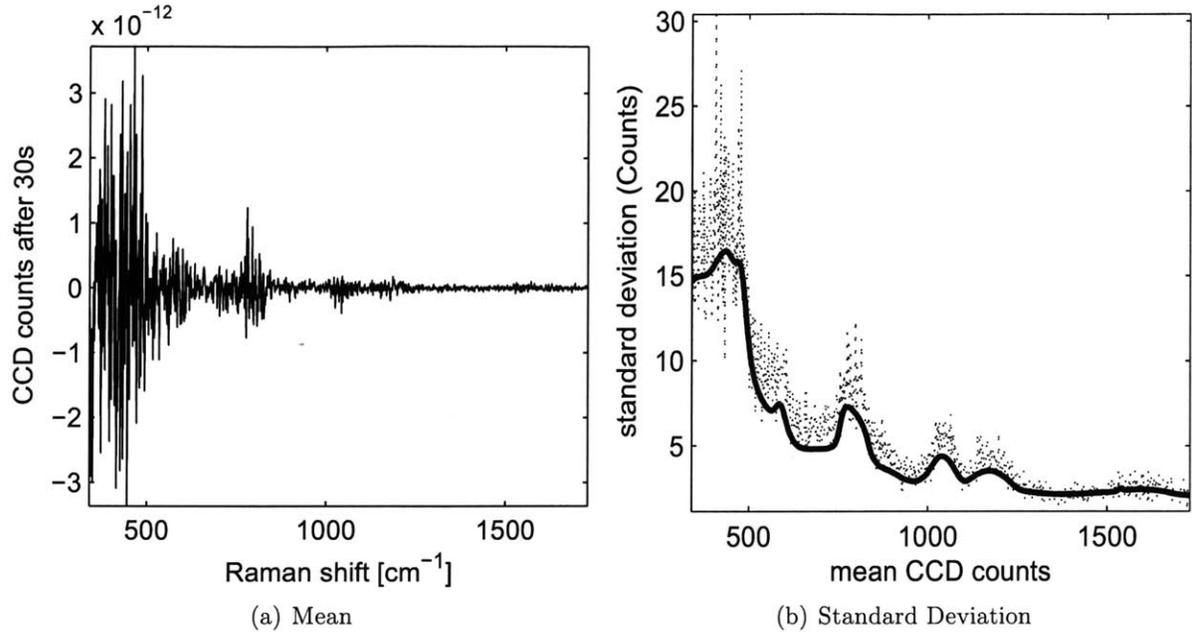
Figure 2-3: Mean and standard deviation of the error in each pixel of the Raman spectrum of UV fused silica.

nearly identical conditions will be different. The Raman spectra of UV fused silica from the side of a cuvette and polycarbonate were acquired many times and compared. Figure 2-2 shows the Raman spectrum of the fused silica (a) and the Raman spectrum of the background (b) taken with 30s acquisition time. The lens and optical components of the probe are made out of fused silica and thus give rise to a similar spectrum to the cuvette wall. To identify errors, the many spectra of the cuvette wall were subtracted from the mean of the spectra. The mean of the errors, shown in Figure 2-3(a), indicates that the noise is centered around zero, validating the assumption that there are no distortions. Furthermore, the standard deviations of the error, shown in Figure 2-3(b), are approximately the same as theory relying on shot-noise as the sole source of error.
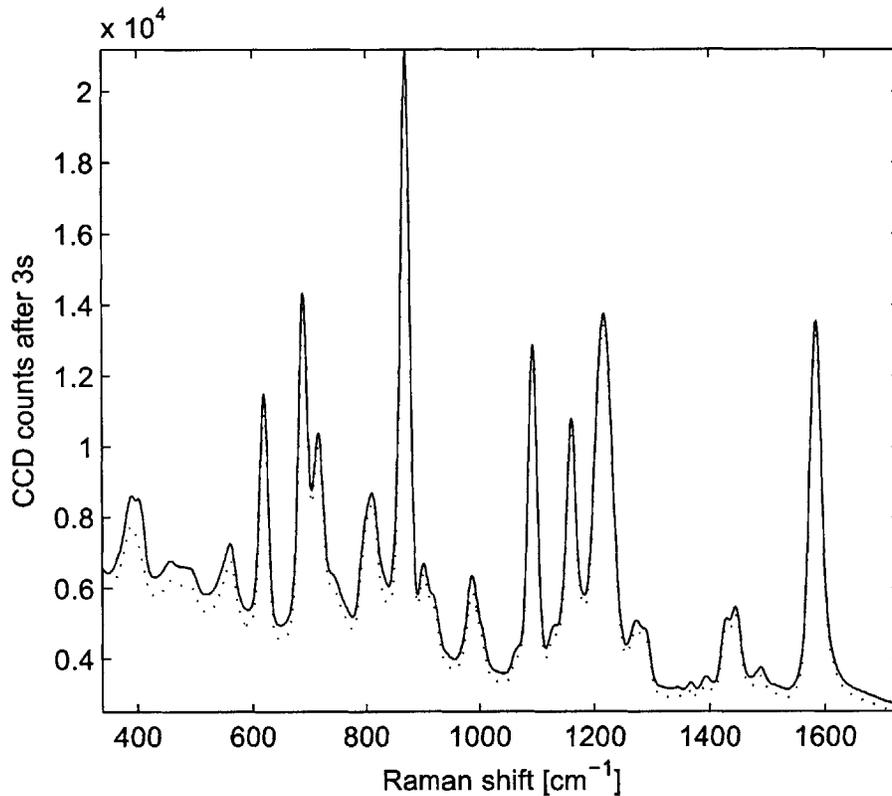
Figure 2-4: The Raman spectrum of polycarbonate.

2.2.2 Fiber Optic Cable Stability

All fiber-optic cables exhibit bending losses that are inversely proportional to the radius of the bend. The stability of the Raman spectra of two substances, UV fused silica and polycarbonate, were tested. To do this, multiple acquisitions were taken for two configurations of the fiber optic cable that connects the probe to the laser and spectrometer: a tightly coiled fiber, and an uncoiled, straight fiber. The two substances were used because of their different Raman cross-section properties. While the fused silica has a smoothly varying and weak Raman signal, the polycarbonate has a strong Raman signal and sharper peaks. To identify additive errors, the averaged spectrum of the coiled fiber configuration was subtracted from the straight fiber configuration, which, if the error is noise, should show noise centered around a mean of zero. Instead, the result, as shown in Figure 2-5, shows that for both substances, there is a defined

structure to the residual. To identify multiplicative errors, the averaged spectrum of the straight fiber configuration was divided from the coiled fiber configuration, which, if the error is noise, should show noise centered around a mean of one. There is instead a definitive structure in this result for both substances, as shown in Figure 2-6.

(a) Fused Silica   (b) Polycarbonate
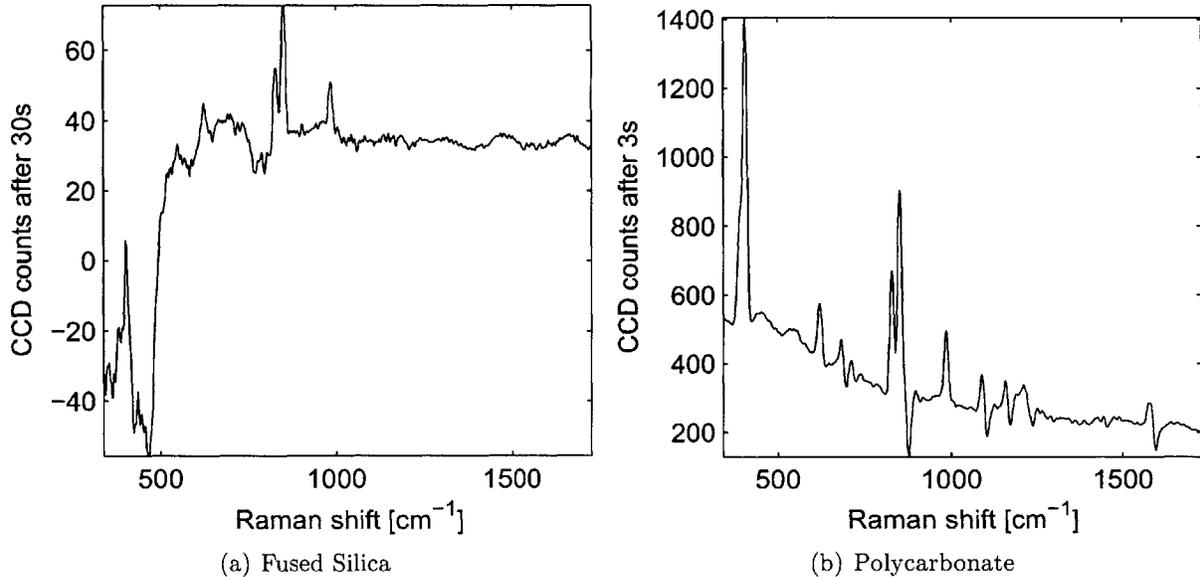
Figure 2-5: Difference of Raman spectra acquired with an uncoiled and tightly coiled fiber optic cable.

The structure shown in the difference spectrum in 2-5(b) shows that the strain causes a peak shift. This is apparent by the difference peaks, which are a peaks followed dips or vice versa. To ensure that the difference peaks are real and not a result of the mode-hopping of the laser, the wavelength correction described in Section 2.3.2 was used to adjust the spectra via observation of the Rayleigh scattering. In addition, the quotient spectra show an inverse dependence of the intensity to the shift. This is seen best in Figure 2-6(a). The higher intensity peaks are only slightly attenuated, while the lower intensity wavenumbers show greater distortions. This is also apparent in Figure 2-6(b), as the peaks are closer to unity than the troughs. For both materials, there exist sharp peaks at 400, 830, 850, and 980 cm^{-1} for the straight fiber measurements. These are caused by the room lights and are therefore not real distortions. It is important to include here that similar tests were conducted for slight changes in the fiber optic cable and

(a) Fused Silica

(b) Polycarbonate
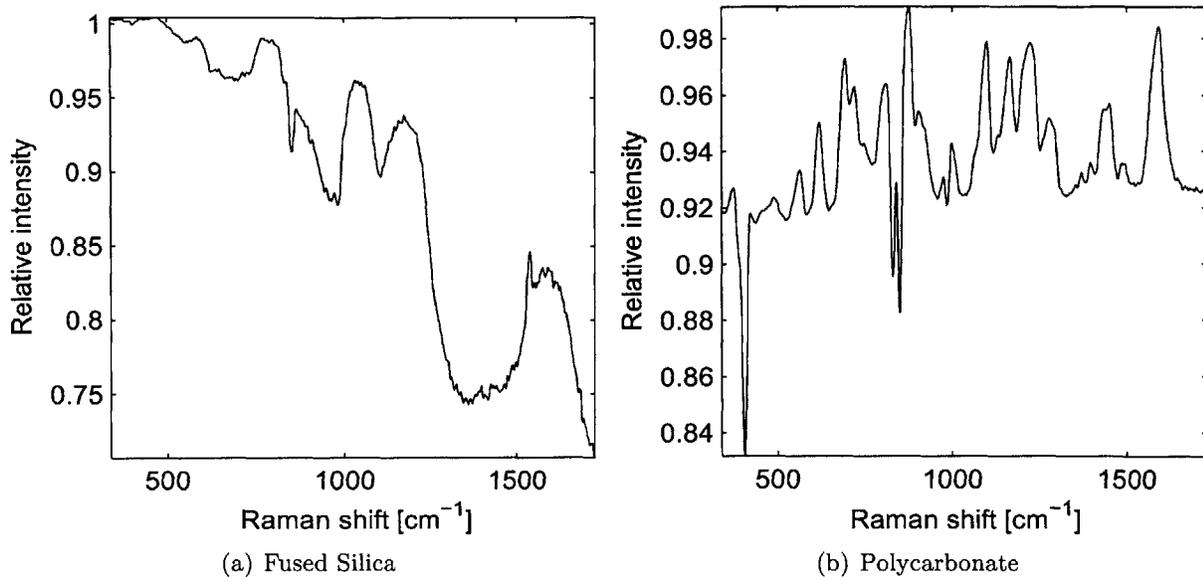
Figure 2-6: Quotient of Raman spectra acquired with an uncoiled and tightly coiled fiber optic cable.

no differences were observed, indicating that as long as the fiber optic cable is in a particular configuration, it is insensitive to slight changes. To conclude, it is clear that bending the fiber causes distortions in Raman spectra both in terms of intensity and in terms of wavelength. For the offline and calibration measurements presented throughout this thesis, the fiber was coiled in approximately the same configuration. However, we expect this to be one systematic source of error when comparing measurements over the course of several days.

2.2.3 Probe Window Stability

Previous online results were found to have been distorted by a change in the Raman spectrum of the sapphire window used to shield the probe from the bioreaction chamber. This problem led to significant errors in the online measurements, requiring an offline calibration after the fermentation. To investigate the hypothesis that the sapphire window was changed by the process of autoclaving, or steam sterilization at 120°C, a sapphire window was placed in an autoclave in a sealed glass jar submerged in the same growth medium used in previous fermentations, LB.

Raman spectra of the sapphire window were acquired before and after this process and verified with a control window to ensure that any changes were not due to other unseen changes in the setup. The results, as shown in Figure 2-7, show the relative peak intensities of sapphire were distorted by steam sterilization. In addition, two different sapphire windows do not have the same Raman spectra, as the control window is different from the other window. Furthermore, it was found that distortions were localized on the sapphire window.
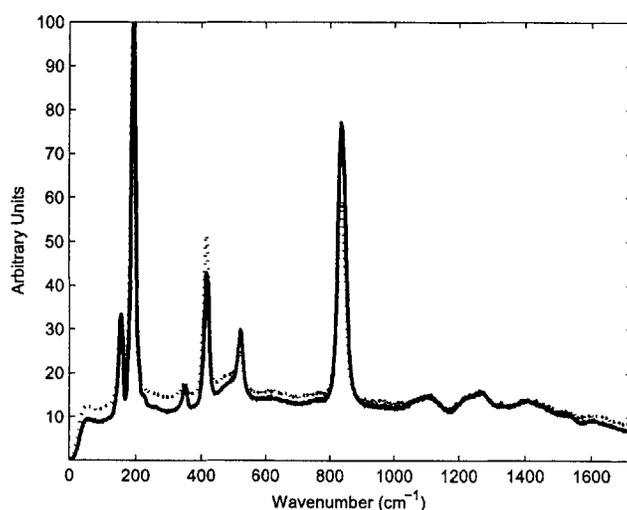
Figure 2-7: The Raman spectrum of a sapphire window before (solid) and after (dotted) autoclaving.

A rigorous study on the Raman spectrum of sapphire in the literature was conducted for a sapphire window at room temperature as well as a window that was cooled to liquid nitrogen temperatures in a cryostat [25]. This experiment also showed a localization problem in the Raman spectrum of the previously cooled sample. This rigorous study showed that the localization portrayed radial symmetry, with the center of the window being most similar to the original spectrum. They hypothesized that this change was due either to impurities in the production of the sapphire or to a phase change in the crystalline structure of the sapphire. There are two water-free crystalline phases of the sapphire crystal, known as $\alpha - AL_2O_3$, and $\gamma - AL_2O_3$. Synthetic sapphire is produced by a process that converts amorphous alumina to

$\gamma - AL_2O_3$ which is in turn converted into $\alpha - AL_2O_3$ (corundum) which is the most stable [26]. The hypothesis is that the stress caused by the temperature extreme could be causing a slight reversal of this process. This theory is supported by a similar study showing a change in the Raman spectrum due to a localized disordered phase of an $\alpha - AL_2O_3$ crystal caused by weight stress [27]. Current conventional wisdom states that sapphire is a good choice of stress-resilient window material for Raman probes, and thus for bioprocess monitoring. The findings here as well as in [25] and [27] put this idea into doubt.
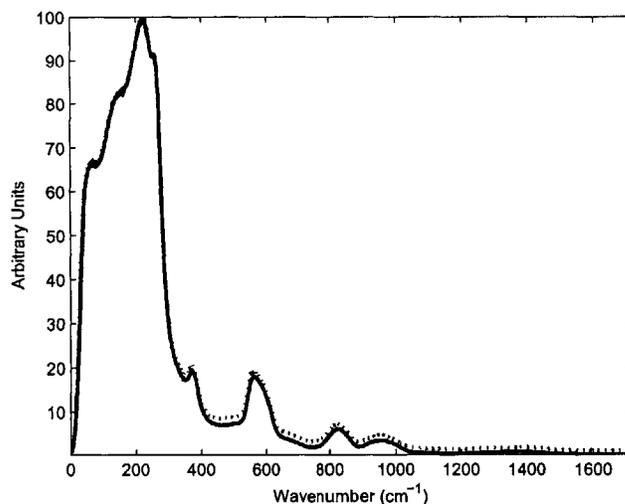
Figure 2-8: The Raman spectrum of a quartz (Supracil 300) window before (solid) and after (dotted) autoclaving.

A Supracil 300, or IR-quartz, window was concurrently tested in the same manner described for the sapphire window. Unlike the sapphire, the window was found to be stable before and after autoclaving, with no localized defects, as shown in Figure 2-8. Also unlike the sapphire, the Raman spectrum of quartz is uniform over all of the windows. As a consequence, the Raman spectra of the control window and the sample window before and after autoclaving are indistinguishable with our instrument. The full effect of the new window on concentration estimation was tested by conducting a fermentation. The results of the fermentation, as described in detail in Section 2.4.2, show that other systematic errors are still limiting the accuracy of the

concentration estimates.

2.2.4   Probe Temperature Stability
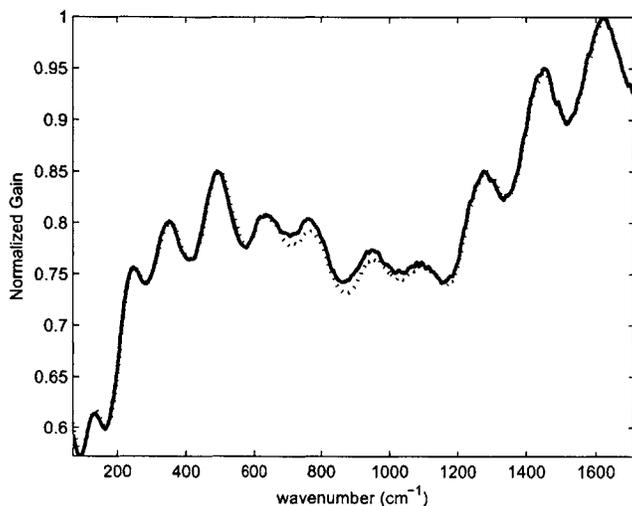
Figure 2-9: Probe transfer function at 22°C (solid) and 37°C (dotted).

A test for temperature stability of the InPhotonics Raman probe transfer function was performed for a temperature range of 22°C to 52°C, with measurements 1°C apart. Previously pure component spectra were taken with the fiber Raman probe at the temperature of the room, which is controlled to 22°C, while measurements in the bioreactor were taken at the reaction temperature of 37°C. The temperature stability experiment was performed by enclosing the fiber Raman probe in an aluminum jacket mounted on a hot plate, as shown in Figure 2-10. Measurements of the probe temperature were taken by a thermocouple and white light was collected by the probe from a fiber coupled to an Ocean Optics Tungsten lamp. These measurements revealed no detectable difference in the probe transfer function over the range of interest. The probe transfer functions for the cases where the probe is at 22°C and 37°C is shown in Figure 2-9. The InPhotonics probe uses thin-film dielectric filters, which are sensitive to the index of refraction. Specifically, the change in center wavelength of such a filter is related to the change in refractive index by:

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta n}{n} \qquad\qquad (2.1)$$
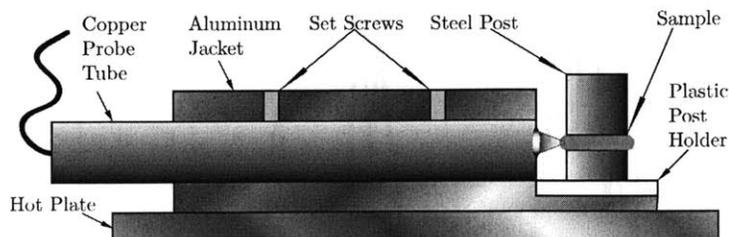
Figure 2-10: Graphic representation of experimental setup for the probe temperature stability measurements.

Assuming that the filters are made of fused silica, which is a good assumption since it is the only zero-field spectrum when the excitation laser is on, and realizing that the worst case will be for the edge of the field of interest at 927nm, we can solve for the expected peak shift. Since, for fused silica, $n_{22c} = 1.4514$ and $n_{37C} = 1.45159$ [28], then $\Delta n = 1.9 \times 10^{-4}$, which would mean that we expect a $\Delta\lambda = 0.12$nm. Although the CCD camera and spectrometer can achieve a resolution of approximately 0.18nm, the large slit width used in this setup limits the resolution further, indicating that this sub-pixel shift will certainly be indistinguishable with this instrument. This explains the temperature stability of the probe.

2.2.5 Component Temperature Stability

Pure component spectra were also measured for temperature stability. Raman spectra of water, glucose, phenylalanine, lactic acid, formic acid, and acetic acid were taken at 22oC, 37oC, and 52°C and relative peak intensities were compared. The experiment was conducted by placing aqueous samples in IR-Quartz cuvettes on a hot plate with temperature readings verified with thermo-couple submerged in the cuvettes. Measurements of the Raman spectra were performed and showed no detectable change in relative peak intensity, as seen for glucose in Figure 2-11. This is consistent with no structural changes (i.e. phase changes, denaturing, etc.) occuring in the molecules of interest.
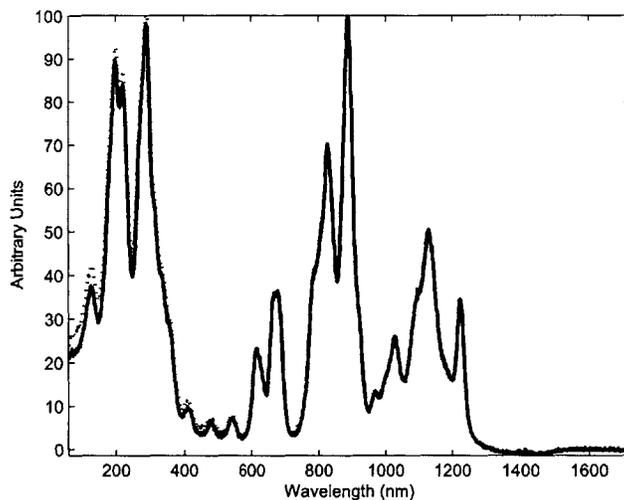
Figure 2-11: Raman spectrum of glucose at 22°C (solid) and 37°C (dotted).

2.2.6   Air Bubble Stability

Typical bioreactions produce many moving air bubbles to deliver oxygen to the cells. These bubbles will scatter both the excitation and the collection light of the Raman signal, causing a possible distortion. It is important to determine whether or not a correction can be made by determining the stability of this behavior. Four separate experiments were performed in the bench-top bioreactor. One experiment used water as the sole component in the bioreactor, while the others used a 1M glucose solution. The effect of scattering from large particles is assumed to be multiplicative and is described at length in Chapter 3. In both experiments, acquisitions of the Raman spectra were taken with the solution still and with the solution moving at various speeds. For the first glucose experiment, no oxygen was added and the impeller speed was set to 600RPM. For the second glucose experiment, the impeller speed was adjusted to 1200RPM. For the third glucose experiment, an air intake of approximately 1 VVM (volume air per volume liquid per minute) was added while maintaining the impeller speed at 1200 RPM. For the water experiment, the impeller speed was set to 800RPM with an air intake rate of approximately 1 VVM, much like the fermentation that was later performed. To remove the spectrum of the quartz, concentration estimation was performed on the still solution spectra

(a) 1M Glucose, 600RPM

(b) 1M Glucose, 1200RPM

(c) 1M Glucose, 1200RPM, air

(d) Pure Water, 800RPM, air
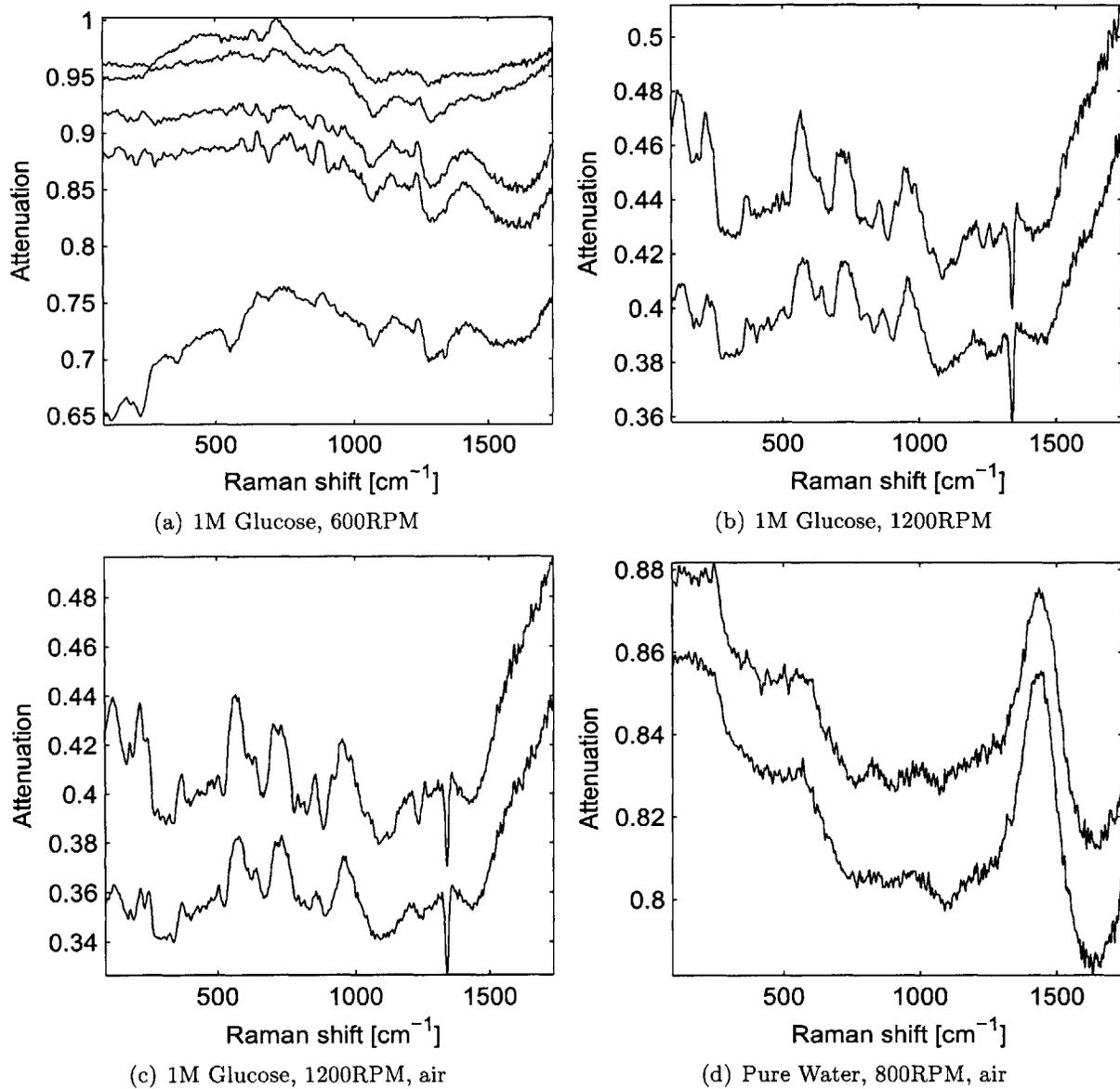
Figure 2-12: Plots of the attenuation of Raman spectra by air bubbles in a stirred tank bioreactor. Parts (a)-(c) show the attenuation of a 1M glucose solution. Part (a) shows four instantiations of the attenuation caused by an impeller speed of 600RPM. Part (b) shows two instantiations of the attenuation caused by an impeller speed of 1200RPM. Part (c) shows two instantiations of the attenuation caused by an impeller speed of 1200RPM with an air intake of 1 VVM. Part (d) shows two instantiations of the attenuation of pure water caused by an impeller speed of 800RPM with an air intake of 1 VVM.

and the concentration estimated quartz was subtracted from both. This technique required the assumption that the contribution to the spectra from quartz is unaffected by the scattering.

Figure 2-12 shows the results of the multiplicative attenuation caused by the bubbles. The results are different for each case, although in each case, a definite pattern emerges, indicating that an experimental correction for the scattering from bubbles is possible, but that a software correction is unlikely given the strong dependence of the bubbles on the parameters such as impeller speed and relative peak intensity. For the case where the impeller speed is low, shown in Figure 2-12(a), the wavelength dependence is weaker, which was matched by the personal observation that the bubbles were large, often over 1mm in diameter, indicating that the scattering was in the geometric limit. Since the large bubbles would divert all light away from the collection geometry, this result matches the intuition that geometric scattering has a weak wavelength dependence. The dips at low wavenumbers in some of the spectra are from the quartz subtraction. When the impeller is set to higher RPM, the bubbles are more plentiful and smaller. The smaller bubbles create a stronger wavelength dependence, which is described in more detail in Chapter 3, while the increase in the number of bubbles causes greater overall attenuation, as seen in Figure 2-12(b). The oxygen sparger produces very large bubbles, as much as 5mm in diameter. These bubbles scatter light in the geometric limit. Thus, these larger bubbles have little effect on the distortion caused by the air bubbles, as supported by Figure 2-12(c). With the impeller speed kept constant, there is little difference between the attenuation caused by the bubbles with the air intake on or off. It would be best to re-design the setup in such a way as to remove the bubble problem altogether. Finally, using different solutions produces different results. The peaks in the attenuation appear to be correlated to the peaks in glucose, and the same is true for the case of pure water, as shown in Figure 2-12(d). It has been shown that the attenuation of the bubbles is a function of the relative intensities of the peaks of the spectrum of the medium as well as the speed of the impellers. The former reason would seem to indicate that the effect is not entirely multiplicative.

Although the impeller speed is generally kept constant during a fermentation, the composition of the medium changes with time, indicating that an online calibration method is necessary.

Alternatively, the spectra should be taken in an enclosed chamber, which is continually pumped with solution from the tank. Such a design for online Raman spectroscopy has already been presented in the literature for yeast cells in shake flasks [9]. Greater understanding about the source of the attenuation is also desirable before more experiments can be conducted. This is the topic of Chapter 3 of this thesis.

2.3 Calibration Techniques

In addition to the stability tests above, other sources of instability were identified and corrected. The ultimate goal of these corrections is to create universal calibration spectra. This means that once a calibration spectrum for a particular component is acquired, it should never have to be acquired again. To achieve this end, the calibration setup must be rigid and well defined to avoid systematic errors caused by changing conditions. Second, it must accommodate for any conditions that cannot be controlled, such as the laser power and excitation wavelength. This will allow the operator to acquire a library of Raman spectra that can be re-used for any bioprocess, vastly increasing the scale-up potential of a Raman spectroscopy bioprocess monitoring system.

2.3.1 Universal Calibration Setup

Previously, two different setups were used for acquiring calibration spectra. In the first method, Raman spectra from pure samples of the components and products were measured by filling a 5mL cuvette to the brim such that the meniscus was flat, after which the Raman probe was positioned above the meniscus and Raman spectra acquired. This method required a large amount of substance (5mL) that was difficult to measure out to exact proportions repeatably, causing scalar errors, and wasted product. In the second method, a $150\mu L$ solution of sample was held by its surface tension in a 0.1 inch diameter hole drilled through a stainless steel rod, and the Raman probe was positioned to acquire spectra laterally. This method, in fact, was modified previously for the probe stability measurements, and is diagramed in Figure 2-10. This method used a small amount of product that was easily inserted via a micro-pipetter. This
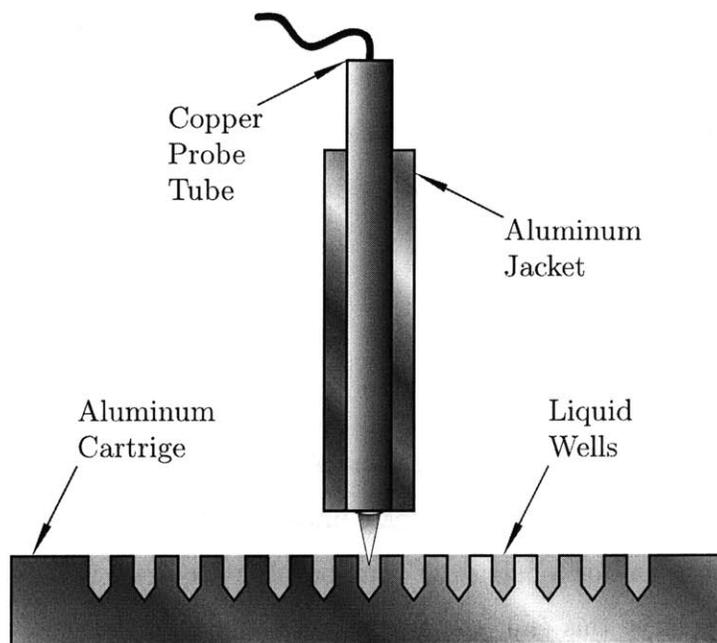
Figure 2-13: Stable, high-throughput, universal calibration and offline measurement setup.

measurement, unfortunately, suffered from the fact that the shape and curvature of the meniscus was very sensitive to the volume, making $1\mu L$ perturbations in the inserted volume result in distortions in the Raman spectra due to the meniscus lensing. Furthermore, to reduce noise, component spectra are usually acquired with highly concentrated samples. These solutions are at such high concentrations typically that they cause a change in the surface tension of water, causing some samples to not hold inside the cavity. It is therefore desirable to make a setup that uses a small sample, does not rely on surface tension, and is rigid. In the mean time, the fundamental advantage of both previous setups, that there is no window to interfere with the signal, should be maintained.

A calibration setup that meets all of the requirements set forth and eliminates all of the drawbacks of the previous two setups was developed and is diagramed in Figure 2-13. Small wells are drilled into a large aluminum cartridge to a defined volume. Aluminum and stainless steel are good materials because they do not exhibit significant Raman spectra. The Raman probe is positioned above the well held in a metal jacket and secured by plastic set screws. The
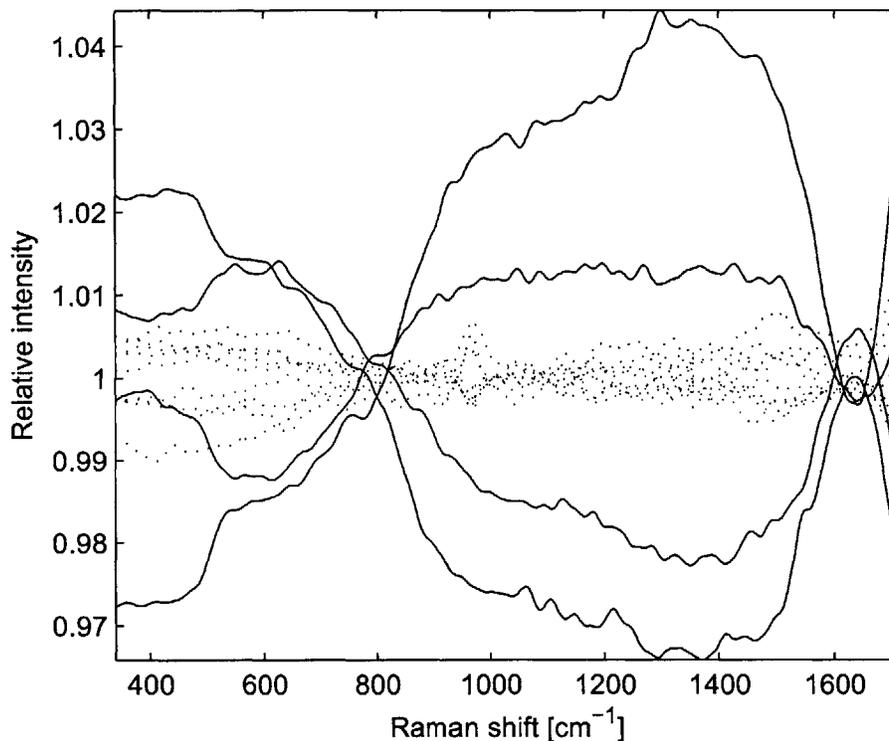
Figure 2-14: Deviation of successive Raman spectra of water from the mean. Nine instantiations of the new calibration setup (dotted) are compared with four instantiations of the old (solid) setup.

Aluminum cartridge was inserted into a tight slot to reduce all other degrees of freedom. Multiple sizes were tried. Smaller wells caused the interference from the otherwise weak aluminum signal to become slightly noticeable at high acquisition times. The smallest well that was found to have negligible aluminum contribution was a $220\mu L$ well with a diameter of $3/16$ inches. At $220\mu L$, the meniscus is totally flat and insensitive to slight perturbations in volume. Furthermore, the sample may be reliably inserted with a micro-pipetter, and liquids with a low surface tension can be used. To validate the superior repeatability and reliability of this setup, numerous spectra of water were taken over multiple days with this setup and the lateral setup that relied on surface tension. The spectra were then subtracted and divided by the overall average spectrum of water to reveal any distortions. The results, shown in Figure 2-14, shows that there are no detectable distortions in the Raman spectrum of water for the new setup, showing errors

within the shot noise limit. However, there are noticeable distortions for the previous setup. The new setup has only been known to cause distortions for cases where the fiber optic cable was changed significantly, as shown in Section 2.2.2. It is important to note that the fiber optic cable was assumed moved everyday, indicating a certain immunity to slight changes in the fiber configuration.

2.3.2   Rayleigh Line Auto-Calibration

(a) Before Correction   (b) After Correction
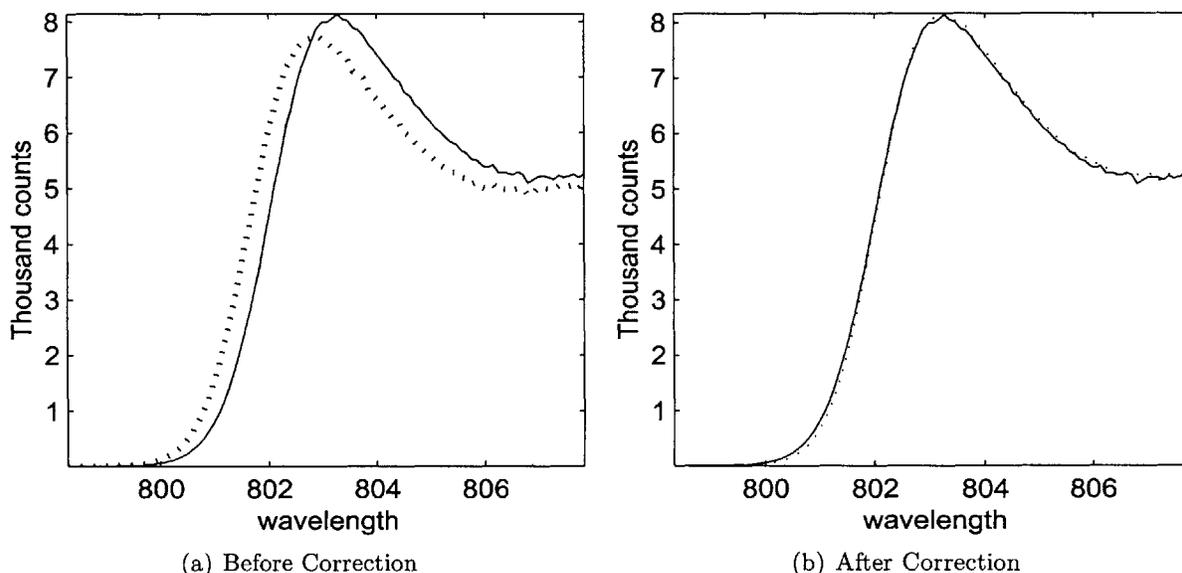
Figure 2-15: The difference in a water peak before and after wavelength correction.

Even with an external cavity and TE cooling, diode lasers can still fluctuate slightly in power and wavelength. This will cause a change in the Raman spectra acquired, since a Raman spectrum is a measure of the frequency shift relative to the excitation. The common solution to such a problem is to acquire Raman spectra of substances with well-defined Raman shift peaks, such as acetone. This method reduces throughput and is ill-suited for *in-situ* measurements because the probe cannot be removed. It is therefore desirable to calibrate the Raman shift *in situ*. To achieve this, the grating in the spectrometer was adjusted to include the Rayleigh line so that fluctuations in the laser line wavelength could be captured. Figure 2-15 illustrates
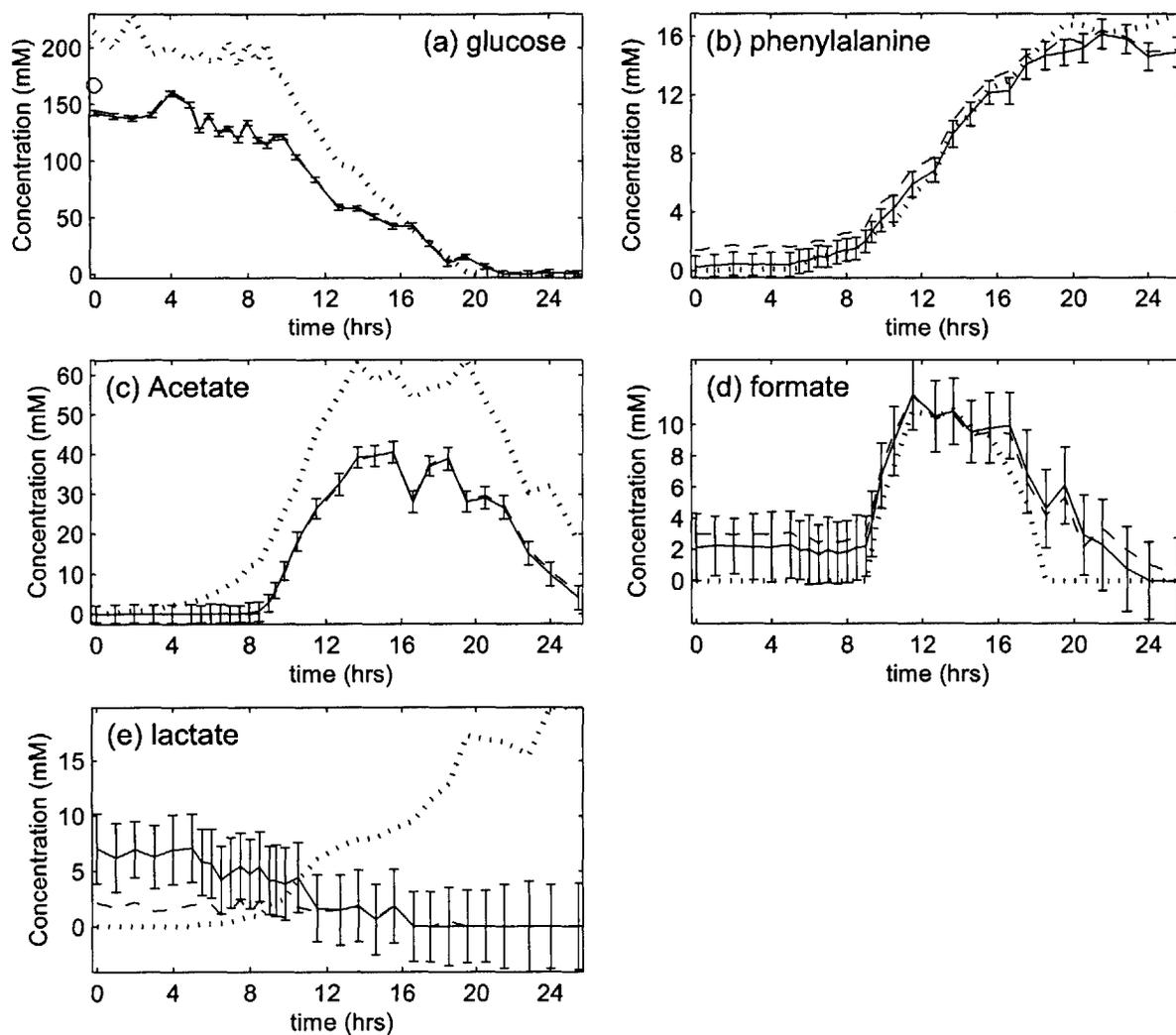
Figure 2-16: Concentration estimation of six analytes using offline Raman spectroscopy with (solid) and without (dashed) wavelength correction. Both results are compared using HPLC (dotted).

this by showing selected peaks of two Raman spectra of water before correction (a) and after correction (b). This line can also be used to measure the differences in laser power, which, when applied to the online measurements, could help differentiate the attenuation from laser power fluctuations from the attenuation due to scattering. To see the difference, offline measurements of samples taken from the bioreaction described in Section 2.1 were acquired using the same setup developed for the calibration spectra, described in Section 2.3.1. Table 2.2 shows the effect this technique had on the concentration estimates between uncorrected and corrected spectra of offline measurements. This is graphically visualized in Figure 2-16. The estimates and error bars in Figure 2-16 were created using the algorithm described in Section 2.4. The offline estimates with (solid) and without (dashed) the wavelength correction, show similar agreement to the HPLC data (dotted) for most of the components. This shows that the effect of wavelength shifts in the laser excitation do not significantly affect the estimates. This method combined with the rigid calibration method above still fail to remove all of the systematic error in the offline concentration estimates, indicating that calibration spectrum repeatability is not the dominant source of the systematic errors.

Component	Max. (mM)	Avg. (mM)
Glucose	2.08	1.28
Phenylalanine	1.17	0.83
Acetate	1.01	0.31
Formate	1.29	0.57
Lactate	4.73	1.95

Table 2.2: Differences in the concentration estimates, c, caused by the wavelength correction.

2.4   Error Analysis

Once spectra are acquired using the setup described in Section 2.1, they are processed digitally using MATLAB to generate the concentration estimates. The algorithm to generate the concentration estimates is as follows. Assuming that the Raman spectrum of an aggregate is the linear combination of the components, then the concentration of the components can be determined

by least squares fitting of a set of calibration spectra to the measured spectrum. To enable this, we have assumed that background fluorescence results in additive error, that scattering from biomass and bubbles only cause a scalar multiplicative effect, and that the wavelength dependence of scattering is negligible. Under these conditions, shot-noise from the calibration spectra and the measured spectra should be the dominant source of error. This will allow the operator to acquire a library of Raman spectra that can be re-used for any bioprocess, vastly increasing the scale-up potential of a Raman spectroscopy bioprocess monitoring system. Calibration spectra are normalized to 1mM concentration. Included with these calibration Raman spectra is a fourth order polynomial to model the additive background fluorescence, together they form the pure component spectrum matrix, K. For a given measured spectrum vector, s, the vector concentrations, c, can be acquired by solving $s = Kc$ using pseudo-inverse methods [29]. The scalar amplitude correction for scattering and laser drift is done by normalizing all component concentrations to the concentration of water. In this way, water is used as an internal standard, as described by [30]. Unlike the implicit algorithms described in Section 1.4, this method makes it possible to propagate the error caused by noise which shall presently be discussed.

2.4.1 Error Propagation Matrices

The algorithm for predicting the contributions of the shot-noise limited measurement noise uses a first order error analysis of the pseudoinverse. The matrices one experimentally has available are $\tilde{K}$ and $\tilde{s}$, which will produce $\tilde{c}$ using the pseudoinverse, where the tilde represents that these values are the actual matrix with some noise added:

$$\tilde{K} = K + \Delta K \tag{2.2}$$

$$\tilde{s} = s + \Delta s \tag{2.3}$$

$$\tilde{c} = c + \Delta c \tag{2.4}$$

where ΔK, Δs, and Δc are the noise, or perturbations, on K, s, and c respectively. To estimate

Δc, we use the expression:

$$\Delta c = \tilde{c} - c = \tilde{K}^+ \tilde{s} - K^+ s = \underbrace{(\tilde{K}^+ - K^+)\tilde{s}}_{\Delta c_K} + \underbrace{K^+ \Delta s}_{\Delta c_s} \qquad (2.5)$$

where K^+ is the pseudoinverse of K and $\tilde{K}^+$ is the pseudoinverse of $\tilde{K}^+$. However, we would ideally like to represent Δc in terms of the unperturbed quantities. Given that K is an over-determined matrix, i.e. that it has more independent rows than columns, then from Equation 2.5 using the proof provided in [31] (and restated in appendix B for convenience), we get:

$$\Delta c = \underbrace{-K^+ \Delta K \tilde{c}}_{\Delta c_{K,1}} + \underbrace{(K^+ K)^+ (\Delta K)^T \tilde{r}}_{\Delta c_{K,2}} + \underbrace{K^+ \Delta s}_{\Delta c_s} \qquad (2.6)$$

where $\tilde{r}$ is the perturbed residual, defined as $\tilde{r} = \tilde{s} - \tilde{K}\tilde{c}$. Assuming that the perturbations are small enough, than we can linearize Equation 2.6 by noting that $\tilde{c} \approx c$ and $\tilde{r} \approx r$ to show:

$$dc = \underbrace{-K^+ dK c}_{dc_{K,1}} + \underbrace{(K^+ K)^+ (dK)^T r}_{dc_{K,2}} + \underbrace{K^+ ds}_{dc_s} \qquad (2.7)$$

where the d denotes small perturbations. In truth, if the perturbations are small enough, ds, dK, and dc are differentials. Since this is a rather large and complicated relationship between dc and dK, the terms can be rewritten in terms of error propagation matrices, $E_{K,1}^+$, $E_{K,2}^+$, and E_b^+:

$$
\begin{aligned}
dc &= & dc_{K,1} & + & dc_{K,2} & + & dc_b \\
&= & E_{K,1}^+ dK(:) & + & E_{K,2}^+ dK(:) & + & E_s^+ ds \\
&= & & & E_K^+ dK(:) & + & E_s^+ ds
\end{aligned}
$$

where the modifier $dK(:)$ is a vector consisting of dK's columns. It has been shown in the literature that these matrices can be described using the SVD of K [32, 31], and are described by:

$$E_{A,1}^+ = V[\Sigma^{-1}\|x\|]W^T \tag{2.8}$$

$$E_{A,2}^+ = V[\Sigma^{-2}\|r\|]Y^T \tag{2.9}$$

$$E_s^+ = V\Sigma^{-1}U^T \tag{2.10}$$

$$= K^+ \tag{2.11}$$

where W has the columns

$$w_i = -(u_i \frac{x^T}{\|x\|})(:), i = 1, \ldots, n \tag{2.12}$$

and where Y has the columns

$$y_i = (\frac{r}{\|r\|}v_i^T)(:), i = 1, \ldots, n \tag{2.13}$$

It is therefore shown that the error contribution due to noise in the component spectra can be considered independently from the error due to noise in the measurement spectra by use of the error propagation matrices.

2.4.2 Error Modeling Results

In order to use the error propagation matrices to propagate errors due to the noise, a model for how the noise behaves is necessary. We can assume that the photons striking the surface of the CCD camera create the electrons with some quantum efficiency $Q < 1$. This means that the noise bottleneck, the part with the least number of randomly created quanta, is the charge created by the electrons in the camera. If we assume that successive electrons are independent and arriving at random intervals, then the electron arrival is dictated by Poisson statistics. The independent electron arrival rate is caused by independent photon arrival rate, which is a good assumption for laser sources. To determine the probability, $P(n)$, that n electrons are counted in a pixel with an integration time of t, divide t into N time bins such that no more than one

electron occupies a time bin. This means that there will be n filled bins and $(N - n)$ empty bins. If N is large, then the probability that there will be n filled bins is the Poisson distribution [33]:

$$P(n) = \frac{\mu_n^n}{N!} e^{-\mu_n} \qquad (2.14)$$

which has a mean of μ_n and a variance of

$$\sigma_n^2 = Np(1 - p) \qquad (2.15)$$

where p is the probability that a bin is filled. Since $p = \mu_n/N \ll 1$, the standard deviation will therefore be:

$$\sigma_n \approx \sqrt{\mu_n} \qquad (2.16)$$

When n is large, the distribution created is similar to a Gaussian distribution with the same mean and standard deviation of the Poisson distribution. Thus, it is possible to simulate the shot noise by modeling it as independent Gaussian noise with the statistical parameters defined above.

Pure	Online Measurements				Offline Measurements			
Component	σ_s	σ_k	σ_n	Max. Error	σ_s	σ_k	σ_n	Max. Error
Glucose	0.295	0.254	0.390	113.15	0.200	0.160	0.256	92.32
Phenylalanine	0.093	0.074	0.119	3.09	0.071	0.133	0.089	2.60
Acetate	0.289	0.235	0.371	24.25	0.207	0.160	0.263	35.74
Formate	0.216	0.163	0.271	5.30	0.192	0.133	0.235	6.09
Lactate	0.375	0.310	0.481	19.80	0.267	0.209	0.340	19.84

Table 2.3: Theoretical error due to measurement noise and maximum measured errors for online and offline results. All values are in units of mM.

A simulation using 10,000 iterations on each time point of both the online and offline samples of the *E. coli* bioreaction described in Section 2.1 was conducted. The simulation assumed that all of the other sources of noise other than the shot noise were negligible, and used the error propagation matrices above to estimate the errors in the concentrations. The result of the
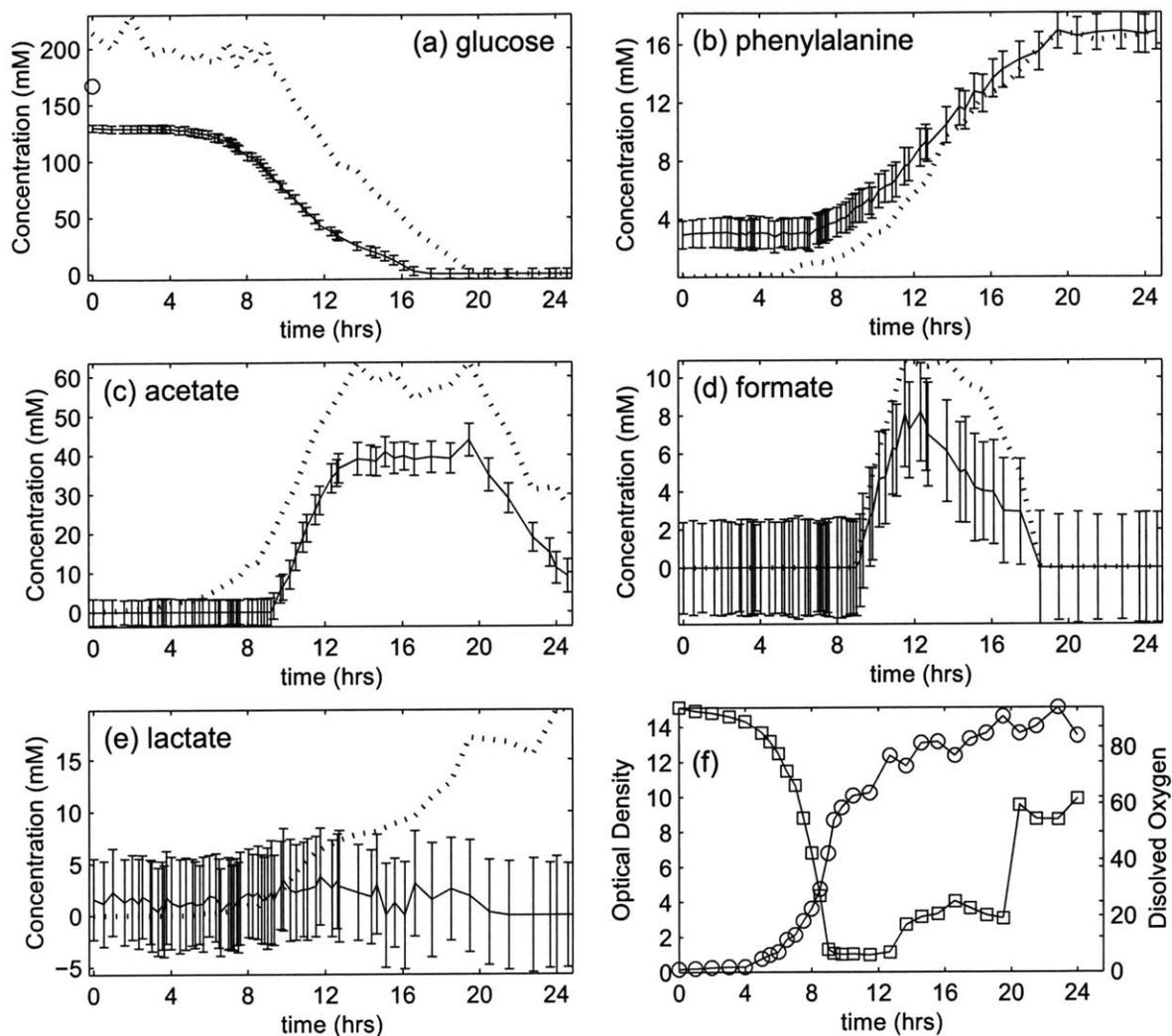
Figure 2-17: Concentration estimation of six analytes using online Raman spectroscopy (solid) compared using HPLC (dotted). A plot of dissolved oxygen and OD are also provided for reference.

simulation is shown in Table 2.3 along with the maximum observed errors for both online and offline measurements. The standard deviation of the concentration estimates due to the shot noise is separated into two components:

$$\sigma_n^2 = \sigma_K^2 + \sigma_s^2 \tag{2.17}$$

where σ_K and σ_s are the standard deviations in the concentration estimates due to shot noise in K and s respectively. This table shows that systematic errors still dominate the errors in the system. Figure 2-17 shows the online results with confidence intervals set at 3σ. This means that if the system functions in a shot-noise limited way, 98.7% of the measured points will fall within the error bars. The fact that the errors are well outside the error bars for many of the data points indicates that there are still unresolved systematic errors. For the case of the online measurements, changes in the optical fiber as well as scattering could be causing significant errors. For both cases, however, the acetate and the glucose are consistently underestimated. This could be a problem associated with HPLC. The nominal starting concentration of glucose, for instance, is 167mM, indicating that the HPLC results overestimate glucose. In addition, the Raman estimates for lactate are in direct violation of HPLC measurements, and is likely caused by a binding to NADH. In *E. coli* fermentations, lactate is made to facilitate electron transport between NAD- to NADH [34], indicating that there could be high quantities of this substance and that it could be binding to the lactate, thereby changing the Raman spectrum of it. Further biological tests are required to test this hypothesis.

2.5 Summary

An existing Raman spectroscopy bioprocess monitoring system has undergone extensive characterization. The system was found to produce repeatable spectra over time when there is no change in the fiber geometry. There was found to be a significant difference in the collected Raman spectra, however, for different radii of curvature of the fiber optic cable. While this should not cause much difference in the offline and calibration spectra, it is expected to cause

alterations in the online measurements, due to the fact that the cable is uncoiled in order to reach the bioreactor. The fiber optic probe was found to be stable within the temperatures of interest (22°-37°). In addition, it was shown that the Raman spectra of the components are largely unaltered by increases in temperature as well. A previously observed alteration of the Raman spectrum of sapphire after autoclaving was confirmed, and the window was replaced with a UV fused silica window which was shown to be stable in the temperatures of interest. The attenuation due to the air bubbles was found to be independent of the oxygen intake rate, but exhibit a strong dependence on both the impeller speed and the relative peak heights of the spectrum being attenuated. The dependence on impeller speed was expected, as this will change the size distribution of the bubbles. The dependence on the relative intensities of peaks and troughs indicates that a software correction based solely on the size and nature of the bubbles will not be sufficient to correct for the scattering. Either an experimental correction or a software correction that includes the measured spectrum as a parameter is required.

A number of improvements were made to reduce systematic errors. A novel approach to solving the errors caused by an excitation source that varies in wavelength and intensity was devised. This approach involved moving the spectrometer grating to include the Rayleigh scattered line, allowing for an automatic reference line for each acquisition. This method thereby removes one of the fundamental experimental challenges to quantitative Raman spectroscopy. In addition, a new design for taking offline and calibration spectra has been shown to produce repeatable spectra over many days of operation. This setup removes interference from a window or cuvette wall by exciting and collecting light in a vertical geometry. This setup does this with a small $220\mu L$ volume and is insensitive to small perturbations in the volume size by using wells in an aluminum cartridge of the defined volume to produce a flat meniscus. This device also has scale-up potential. Making a cartridge with multiple wells will allow for many calibration spectra to be taken between well cleanings. With a stationary fiber optic cable, these two techniques make offline concentration estimation nearly free of systematic errors due to the optics.

The theoretical model to determine the detection limits due to shot-noise has also been improved. The new model accounts for noise in the calibration spectra as well as the measured

data. An algorithm to estimate the error contributions separately using error propagation matrices has been outlined. Using MATLAB simulations it is possible to get limits of detection for a bioreaction before actually running a fermentation, allowing for the operator to determine if concentration estimation on a component is feasible before use. With this algorithm, confidence intervals on all online and offline measurements of an *E. coli* bioreaction were generated. Both online and offline results showed that most of the error was caused by systematic error, not noise. To improve the error in the online results to be within the error in the offline results, a correction of distortions caused by the changes in the optical fiber configuration and a correction for the scattering is necessary. To improve the offline measurements, it is likely that the errors are caused by problems in the biological model instead of the optical system. More tests of components such as NADH need to be conducted, and erroneous HPLC results need to be corrected.

Chapter 3

Wavelength Dependant Scattering

Previously, we have discussed that Raman scattering is accompanied by Rayleigh scattering. For an aqueous suspension of individual molecules, this description is sufficient, as Rayleigh scattering is the limit of elastic light scattering for particles much smaller than the wavelength of excitation ($\lambda \gg d$), such as individual molecules. Bioreactors, however, also contain large particles, usually in the form of air bubbles or biomass. These particle sizes can be too large for Rayleigh scattering approximations and too small to for geometric ray tracing ($\lambda \ll d$). *E. Coli.* cells, for example, have dimensions on the scale of 1μm, which is on the order of the observed wavelengths of a typical Raman bioprocess monitoring setup. Elastic scattering has previously been modeled as a wavelength invariant effect on a Raman signal. Using this assumption, a scalar correction was implemented. A more rigorous look at the elastic scattering of large particles shows that there is, in fact, a wavelength dependence. This effect will manifest as a multiplicative effect on the Raman spectrum of the solution and could give rise to some of the systematic error in the online measurements. In Chapter 2, many of the unexplained differences between the online and offline results were believed to be attributed to this wavelength dependence. This chapter will discuss the theory that describes the errors caused by scattering as well as experiments to confirm them.

3.1   Elastic Scattering Theory
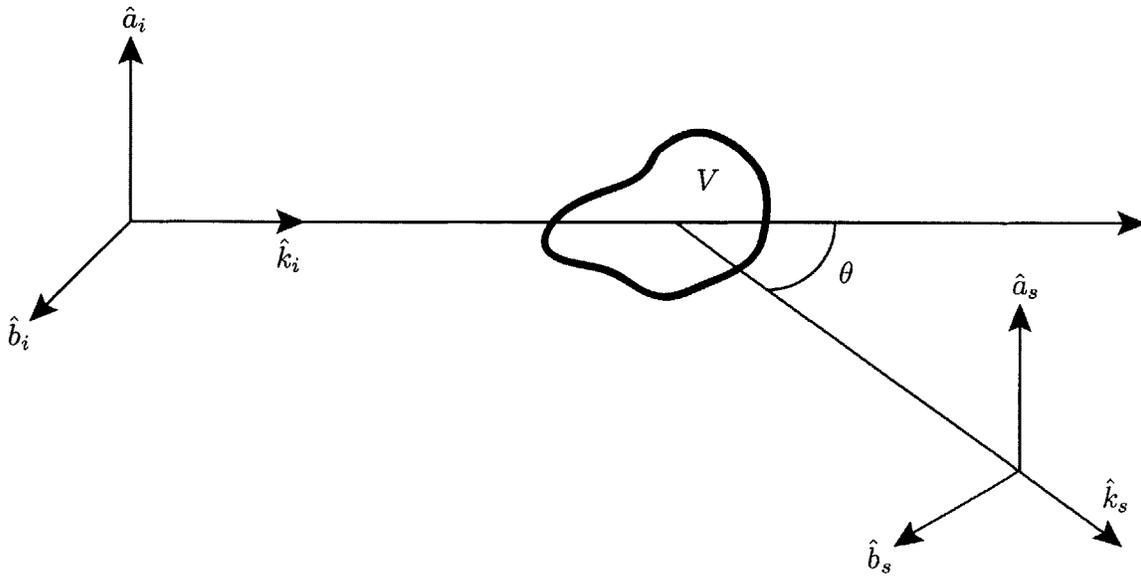
Figure 3-1: Geometry for defining the orthonormal unit system used for elastic scattering. The angle between $\hat{k}_i$ and $\hat{k}_i$ is θ. The plane containing $\hat{k}_i$ and $\hat{k}_s$ is the scattering plane.

Elastic scattering of electromagnetic waves has been studied and explored at length due to its applicability to almost every subfield of electromagnetics. Readers unacquainted with the theory are recommended to read the commonly referenced monographs as a starting point [35, 36, 37], as well as a review of recent work in scattering theory [38]. The basic problem of light scattering is defined in Figure 3-1. An incident electromagnetic wave traveling in the direction of $\hat{k}_i$ is incident on a particle, which scatters the light in some direction $\hat{k}_s$ an angle θ from $\hat{k}_i$. More specifically, an incident plane wave with an electric field described by:

$$\bar{E}_i = \hat{e}_i E_o e^{jk\hat{k}_i \bar{r}} \tag{3.1}$$

where $\bar{r}$ is the position vector, k is the wavenumber, and time dependence $e^{-j\omega t}$ has been suppressed for convenience. In the far field, the scattered light will resemble a spherical wave:

$$\bar{E}_s = \hat{e}_s f(\hat{k}_s, \hat{k}_i) E_o \frac{e^{jk\hat{k}_i \bar{r}}}{r} \tag{3.2}$$

where $\hat{e}_s$ is the polarization vector indicating the direction of the electric field and $f(\hat{k}_s, \hat{k}_i)$ is the scattering amplitude function from direction $\hat{k}_i$ to direction $\hat{k}_s$. This amplitude function is at the core of scattering theory. Given an input wave, a location, and an $f(\hat{k}_s, \hat{k}_i)$, the entire wave resulting from the scattering event can be described as a coherent superposition of the incident field and the scattered field. Furthermore, since the information is still in wave form, the field resultant from the elastic scattering from any collection of particles can be computed from a superposition of the fields from all of the particles. Of course, for any practical application, the complexity of the scattering amplitude functions and the number of particles combine to make the math impossible to compute in this way. At their core, all scattering methods and algorithms are ways of simplifying the problems of acquiring $f(\hat{k}_s, \hat{k}_i)$ and combining the fields.

3.1.1   Beer's Law for a Homogenous Slab of Scatterers

(a) Forward scattering from a single particle   (b) Forward scattering from a volume of particles
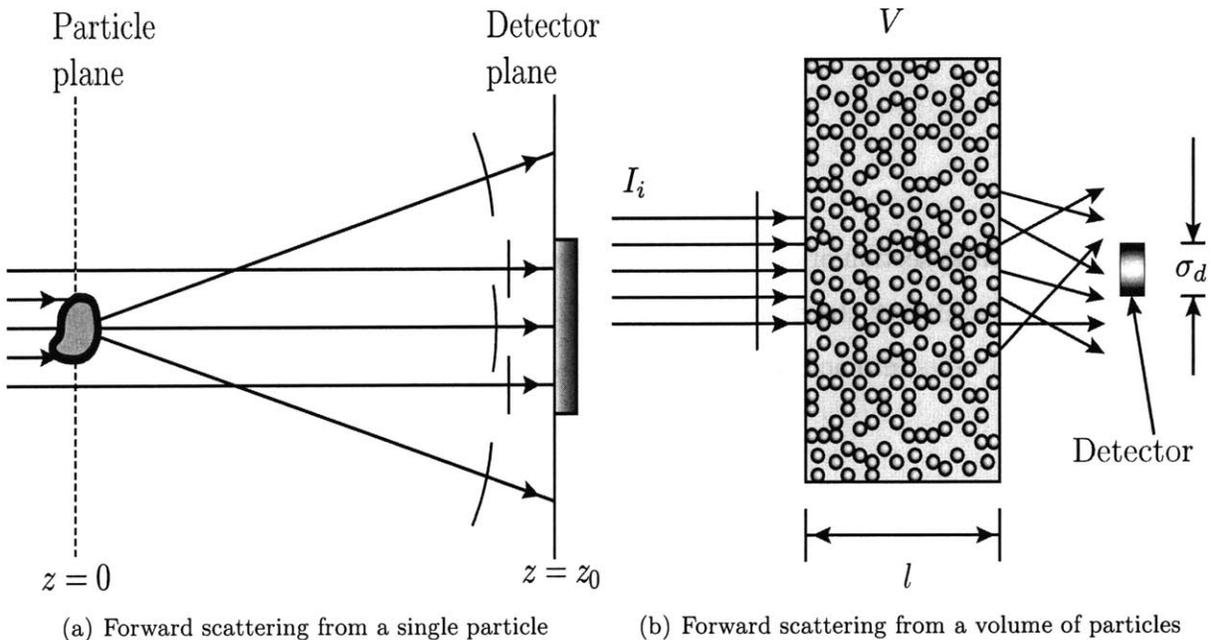
Figure 3-2: Schematics of single and multiple scattering situations.

For Raman spectroscopy of a substance in the presence of scatterers, a model for describing how much attenuation the Raman scattered light will undergo as a function of wavelength is

desired. First, consider the forward scattering ($\theta = 0$) case for a single arbitrarily shaped particle, as shown in Figure 3-2(a). A plane wave of intensity I_{inc} illuminates a particle which scatters some of the light. If a detector of diameter d is placed at a distance z from the scatterer in the far field, ($z \gg d$), then the contribution of the scattered light will be infinitesimal since it is a modified spherical wave. Therefore, the power of the light on the detector, P will be:

$$P = \sigma_d I_{inc} - P_{ext} \tag{3.3}$$

where σ_d is the area of the detector, I_{inc} is the intensity of the incident radiation and P_{ext} is the power scattered or absorbed by the particle. This total power extinguished will be a function of the morphology of the particle and the incident radiation, leading to:

$$P_{ext} = \sigma_{ext} I_{inc} \tag{3.4}$$

and therefore:

$$P = \sigma_d I_{inc} - \sigma_{ext} I_{inc} \tag{3.5}$$

where σ_{ext} is the cross-section of extinction. Intuitively, σ_{ext} can be seen as the size of a perfectly absorbing disk placed over the detector that would extinguish the same amount of light [35]. This cross-section takes into account both scattering and absorption, and is related to them by:

$$\sigma_{ext} = \sigma_{sca} + \sigma_{abs} \tag{3.6}$$

where σ_{abs} is the cross-section of absorption, and σ_{sca} is the cross-section of scattering. Naturally, for weakly absorbing particles, $\sigma_{ext} \approx \sigma_{sca}$. Determination of σ_{ext} will be described later.

Now consider if, instead of one particle, there are many inside a volume, as shown in Figure 3-2(b). If the particle concentration is dense, then they cannot be seen as independent particles, but as a single diffuse body that scatters fields in all directions isotropically. For these suspensions, radiative transfer theory must be applied. For dilute suspensions of scatterers, there is no

field coupling between spheres and the events can be looked at independently. In independent, or single scattering, the particles are far enough that the scattered fields from other particles are negligible compared to the excitation field. Stated another way, the probability that a scattered photon hits another scatterer is very low. In the absence of field coupling, it is clear that the power lost by extinction in the slab will be the sum of the power lost by extinction of each particle:

$$P_{ext} = \sum_i P_{ext,i} \tag{3.7}$$

which means that we can do the same with their cross sections:

$$\sigma_{ext} = \sum_i \sigma_{ext,i} \tag{3.8}$$

We would like to convert this summation into a description of the extinction based on the macroscopic parameters. If we initially expect that there are ρ identical particles per unit volume then the total change in power scattered per unit volume is:

$$\frac{dP_{ext,tot}}{dV} = \rho P_{ext} = \rho \sigma_{ext} I_{inc} = K_{ext} I_{inc} \tag{3.9}$$

where K_{ext} is called the extinction coefficient. K_{ext} represents how much power is extinguished from the excitation per unit distance. Furthermore, K_{ext} can be split into scattering and absorption as well:

$$K_{ext} = K_{sca} + K_{abs} \tag{3.10}$$

Furthermore, K can be generalized. Specifically, for a solution that contains particles of j different morphologies, each with a radius of a and a size distribution of $n_j(a)$ per unit volume, then K_{ext} will be:

$$K_{ext} = \sum_j \int_{a_{min}}^{a_{max}} n_j(a) \sigma_{ext,j} da \tag{3.11}$$

Since a change in the scattered power causes a change in the detected power, and since the area of interest remains constant, Equation 3.9 leads to an expression for the change in intensity at the detector per unit thickness of the slab:

$$\frac{dP_{ext,tot}}{dV} = -\frac{dI_{out}}{dl} = \rho K_{ext} I_{out} \tag{3.12}$$

Integrating this linear 1st-order ordinary differential equation, we get:

$$I_{out} = I_{inc} e^{-\int_0^L K_{ext} dz} \tag{3.13}$$

$$= I_{inc} e^{-K_{ext} L} \tag{3.14}$$

$$= I_{inc} e^{-\tau} \tag{3.15}$$

where τ is called the turbidity or optical thickness of the slab. This relationship is known as Beer's law. In the literature, it is also common to refer to the mean free path, l_p, defined as the inverse of the slab extinction parameter:

$$\tau = K_{ext} L = \frac{L}{l_p} \tag{3.16}$$

Where l_p is the mean distance traveled by a photon before a scattering event occurs. If $L \ll l_p$, then the chances that a photon is scattered twice is very small, indicating that the density of the suspension is in the single scattering regime. Thus, we expect this to be true for $\tau \ll 1$. In truth, if the aperture of the detector is properly contained, it is possible to get good agreement up to $\tau \approx 10$. This will be discussed later in section 3.2.1.

3.1.2   Derivation of the Rigorous Solution for a Single Sphere

It has now been shown that forward scattering from a slab of scatterers can be predicted with the proper choice of scattering cross-section. Determining the scattering cross-section depends on the size and morphology of the particle. As mentioned above, since $\lambda \approx d$, neither the

Rayleigh nor geometric scattering limits can be used for the application at hand, so a rigorous elastic scattering solution is necessary. The most widely used and important solution for the scattering problem is that for a sphere, derived by Mie in 1908 [39], and remains the cornerstone of rigorous scattering theory. The Mie formulation is a solution to the scalar wave equation for a homogenous medium using the appropriate spherical boundary conditions. Since air bubbles will be spheres, it should accurately describe their scattering behavior. Furthermore, many cell types are at least spheroidal and therefore approximate a sphere. It is therefore from Mie theory that the σ_{ext} will be acquired. In addition to the material presented here, many of the details of the derivation can be found in Appendix C.

Remembering the scattered field from an incident plane wave is a spherical wave modified by the function $f(\hat{k}_s, \hat{k}_i)$, we can create a 2 port model that correctly describes the input and output behavior:

$$\begin{bmatrix} E_{as} \\ E_{as} \end{bmatrix} = \begin{bmatrix} f_{aa}(\hat{k}_s, \hat{k}_i) & f_{ab}(\hat{k}_s, \hat{k}_i) \\ f_{ba}(\hat{k}_s, \hat{k}_i) & f_{bb}(\hat{k}_s, \hat{k}_i) \end{bmatrix} \begin{bmatrix} E_{ai} \\ E_{bi} \end{bmatrix} \qquad (3.17)$$

where the subscripts a and b indicate the orthogonal components of polarization in the relative axes, as shown in Figure 3-1. The input and output fields are described by:

$$\vec{E}_i = \left(\hat{a}_i E_{ai} + \hat{b}_i E_{bi} \right) e^{j\bar{k}_i \bar{r}} \qquad (3.18)$$

$$\vec{E}_i = \left(\hat{a}_s E_{as} + \hat{b}_s E_{bs} \right) \frac{e^{j\bar{k}_s \bar{r}}}{r} \qquad (3.19)$$

This allows for a solution for an arbitrary polarization by splitting the two components into linearly polarized plane waves. Since the sphere is isotropic and radially symmetric, we can assume that there is no coupling between the two components of the E-Field, indicating that the scattering does not rotate the polarization state of the light. Using this fact and placing the sphere at the origin then creates the following relations:

$$f_{aa}(\hat{k}_s, \hat{k}_i) = \frac{j}{k} S_1(\theta) \tag{3.20}$$

$$f_{ab}(\hat{k}_s, \hat{k}_i) = 0 \tag{3.21}$$

$$f_{ba}(\hat{k}_s, \hat{k}_i) = 0 \tag{3.22}$$

$$f_{bb}(\hat{k}_s, \hat{k}_i) = \frac{j}{k} S_2(\theta) \tag{3.23}$$

Leading to:

$$\begin{bmatrix} E_{as} \\ E_{bs} \end{bmatrix} = \begin{bmatrix} \frac{j}{k} S_1(\theta) & 0 \\ 0 & \frac{j}{k} S_2(\theta) \end{bmatrix} \begin{bmatrix} E_{ai} \\ E_{bi} \end{bmatrix} \tag{3.24}$$

Which indicates that the scattered light will be symmetric about $\hat{k}_i$. To solve for these two functions, we solve the scalar wave equation:

$$\nabla^2 \Psi + k^2 m^2 \vec{E} = 0 \tag{3.25}$$

with the appropriate boundary conditions:

$$\hat{n} \times (\vec{E}_1 - \vec{E}_2) = 0 \tag{3.26}$$

$$\hat{n} \times (\vec{H}_1 - \vec{H}_2) = 0 \tag{3.27}$$

where k is the wavenumber of the light in the suspension, m is the relative index of refraction between the particle and the suspension, $m = n_p/n$, and $\vec{E}_1$, $\vec{E}_2$, $\vec{H}_1$, and $\vec{H}_2$ are the fields on either side of the boundary. Solving will result in expressions for $S_1(\theta)$ and $S_2(\theta)$:

$$S_1(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \{ a_n \csc \theta P_n^1(\cos \theta) + b_n \frac{d}{d\theta} P_n^1(\cos \theta) \} \tag{3.28}$$

$$S_2(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \{ b_n \csc \theta P_n^1(\cos \theta) + a_n \frac{d}{d\theta} P_n^1(\cos \theta) \} \tag{3.29}$$

Where P_n^1 is an associated Legendre function and the coefficients a_n and b_n are given by:

$$a_n = \frac{\psi_n' \psi_n(mx) - m\psi_n(x)\psi_n'(mx)}{\zeta_n'(x)\psi_n(mx) - m\zeta_n(x)\psi_n'(mx)} \tag{3.30}$$

$$b_n = \frac{\psi_n \psi_n'(mx) - m\psi_n'(x)\psi_n(mx)}{\zeta_n(x)\psi_n'(mx) - m\zeta_n'(x)\psi_n(mx)} \tag{3.31}$$

Where ψ_n and ζ_n are Riccati-Bessel functions of the 1st and 3rd kind respectively, $x = d/\lambda$ is the size factor relating the diameter of the particle to the excitation wavelength, and m is the quotient of refractive index of the medium and the particle [36].

The last step is to compute the scattering cross-section for $S_1(\theta)$ and $S_2(\theta)$. Given an arbitrary input polarization and amplitude, $E_0\hat{e}_i$, where $\hat{e}_i$ is at an angle ϕ from the scattering plane

$$I_s(\theta) = \langle S \rangle_t = \frac{1}{2\eta k^2 r^2} E_0^2 \left[|S_1(\theta)|^2 \sin^2 \phi + |S_2(\theta)|^2 \cos^2 \phi \right] \tag{3.32}$$

where η is the impedance of the suspension. As discussed above, the scattering cross-section will be the total power scattered divided by the incident field, where the power is the intensity function integrated over all solid angles:

$$\sigma_{sca} = \frac{P_s}{I_i} \tag{3.33}$$

$$= \frac{2\eta}{E_0^2} \int_{4\pi} I_s(\theta) d\Omega \tag{3.34}$$

$$= \frac{1}{k^2} \int_0^{2\pi} \int_0^{\pi} \sin\theta \left[|S_1(\theta)|^2 \sin^2 \phi + |S_2(\theta)|^2 \cos^2 \phi \right] d\theta d\phi \tag{3.35}$$

$$= \frac{\pi}{k^2} \int_0^{\pi} \sin\theta \left[|S_1(\theta)|^2 + |S_2(\theta)|^2 \right] d\theta \tag{3.36}$$

From Equation 3.36, the value of σ_s can be determined numerically, and this method will become useful later in determining alternative cross-sections. For the simple Mie solution at $\theta = 0$, however, we can take advantage of orthogonality relations of Legendre polynomials to

get an analytical expression for the cross-section:

$$\sigma_{sca} = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \left(|a_n|^2 + |b_n|^2 \right) \tag{3.37}$$

For the case where there is very little absorption, the scattering cross-section is adequate for determining the extinction ($\sigma_{ext} \approx \sigma_{sca}$). There are cases when the scatterers also absorb, however, resulting complex refractive indices of the particles. This information can be retrieved from the Mie coefficients for the forward scattering case as well, using the co-called optical theorem [37, 40], which states that the total power loss is related to the imaginary part of the of the scattering amplitude in the forward direction, which in this case is:

$$\sigma_{ext} = \frac{2\pi}{k^2} Re \left\{ \sum_{n=1}^{\infty} (2m+1)(a_n + b_n) \right\} \tag{3.38}$$

For the complete derivation of the optical theorem, please see Appendix C. From this point on, the remaining scattering treatment will concern itself with particles that are very weak absorbers, and so the cross-section of scattering will be used interchangeably with the cross-section of extinction unless otherwise noted.

3.1.3 Generalized Lorenz Mie Theory

The theory developed thus far is derived only for the case of a distant plane wave exciting a slab of scatterers. This is a restrictive set of assumptions that does not meet the experimental conditions of a Raman bioprocess monitoring system shown in Figure 3-3. In this setup, excitation light will undergo forward scattering before exciting Raman scattering. The Raman scattered light will then be forward scattered on its way back to the detector. This indicates a number of flaws in the theory. First, the excitation is not a plane wave, but a focused Gaussian beam. Second, the collection optic is not a small detector at infinity, but a large lens with high numerical aperture. Finally, the scatterers will not generally be monodisperse solutions of perfect spheres, but polydisperse solutions of non-spherical bodies. The theory therefore needs to be adjusted to fit these constraints.
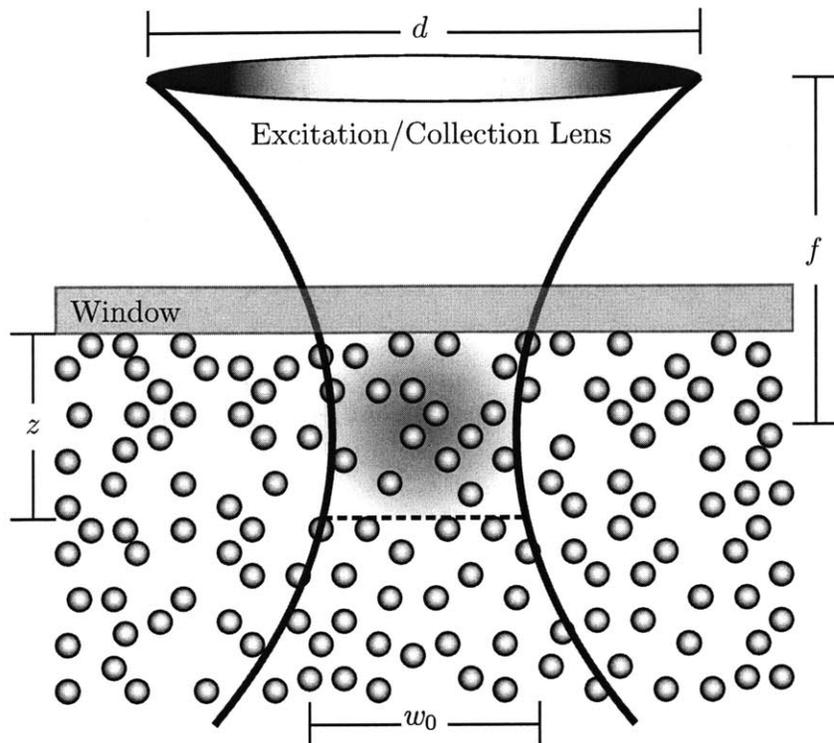
Figure 3-3: Actual geometry for an online Raman spectroscopy bioprocess monitoring setup. Most of the collected Raman scattered light comes from inside the cylinder of length z and diameter w_0.

The solution for elastic scattering of a focused Gaussian beam from a single homogenous sphere has been derived be Gouesbet [41, 42]. This theoretical development, known as Generalized Lorenz-Mie Theory (GLMT), aims to use the same formalism used in Mie theory; Namely, it is a method to solve for the coefficients of the scattering amplitude matrix, $S_1(\theta)$ and $S_2(\theta)$. This is method gives rise to a set of beam shaping coefficients, $g_{n,TM}^m$ and $g_{n,TE}^m$, which, when the particle is placed on axis, simplify to g_n:

$$g_n = \frac{2n+1}{\pi n(n+1)} \cdot \frac{1}{(-1)^n j^n} \int_0^\pi \int_0^\infty jkr \sin^2 \theta \cdot f(r,\theta) \cdot e^{jkr \cos \theta} \cdot G_n^1(kr) P_n^1(\cos \theta) d\theta d(kr) \quad (3.39)$$

where $f(r,\theta)$ is rigorously defined as [43]:

$$f(r, \theta) = G_0 \left(1 - \frac{2}{jl_d + 2z} r \cos \theta \right) \tag{3.40}$$

where G_0 is the fundamental mode solution and l_d is the diffraction or spreading length. Using g_n, 3.28 and 3.29 can be altered:

$$S_1(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left[a_n \csc \theta P_n^1(\cos \theta) + b_n \frac{d}{d\theta} P_n^1(\cos \theta) \right] g_{n,TE}^m \tag{3.41}$$

$$S_2(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left[b_n \csc \theta P_n^1(\cos \theta) + a_n \frac{d}{d\theta} P_n^1(\cos \theta) \right] g_{n,TM}^m \tag{3.42}$$

Because of the computationally impractical nature of these expressions, much of the literature relating to GLMT focuses on approximation methods to determine the beam shaping coefficients. All of these methods relay on decomposing the incident wave into an infinite series of elementary constituents, such as spherical vector wave functions [44, 45], or plane waves [46, 47]. The fastest and simplest algorithm for computing the beam spectrum coefficients was given for the case of a particle on axis using a localized approximation [43, 48, 49]:

$$g_n = exp - \left[\frac{\left(n + \frac{1}{2}\lambda \right)}{2\pi w_0} \right]^2 \tag{3.43}$$

where w_0 is the Gaussian beam waist. When w_0 is large, g_n tends to one, indicating a similarity to a plane wave, which is the proper limiting behavior. Using the fact that a typical Raman excitation is radially symmetric and the spheres are isotropic, it can be assumed that $g_n = g_{n,TM}^m$ and $g_{n,TE}^m$. Therefore, Equation 3.43 can be inserted into Equations 3.41 and 3.42 to fully compute the scattering coefficients for this setup. The effect of the focused beam on the forward scattering of the excitation is therefore modeled.

3.2 Spectral Distortion of Raman Spectra

Spectral distortion of Raman spectra due to elastic scattering has been hypothesized as a source of systematic errors in online concentration estimates. To properly apply the theory from Section 3.1 to the setup described in Figure 3-3, a number of experimental verifications and explanations need to be accomplished. First, knowing that Beer's law is an effective tool for the single scattering case (i.e. low τ), an experiment to determine where optical thickness begins to fail in predicting the attenuation must be done. This is important since bioreactions of bacterial cells can grow to high optical densities, and so it is desirable to show that this method will hold for all cases. Second, experimental verification that theory can predict the attenuation caused by the elastic scattering from a plane wave excitation to verify the basic theory is necessary. Finally, experiments involving the actual probe geometry are necessary to determine if an amplitude correction can be computed.

3.2.1 Multiple Scattering Regime

Use of Beer's law for the case of single scattering of homogenous spherical particles has been theoretically assumed valid at $\tau \ll 1$, but if it is to be practically applied to the application at hand, the limits of this first order scattering model need to be tested. An experiment showing that Beer's law is adaptable for multiple scattering has been demonstrated for a single wavelength in the literature [50, 51]. The experiment shows that for a turbid solution of water and polystyrene microspheres of known size, attenuation of the forward scattered light could be predicted for values of $1 < \tau < 10$ using the Mie formulation. This regime corresponds to multiple scattering from non-interacting particles by restricting the solid angle at the detector. In this regime, each scattering object is far enough from the others to prevent inter-particle interactions and wave coupling. Thus, if the solid angle of the output is restricted, then we can assume that most scattered photons will not scatter back into the un-scattered beam path, as shown in Figure 3-4(b), extending the applicability of Beer's law into this regime of multiple scattering.

Experimentally this is done by the placement of two apertures between a detector and the

(a) Multiple Scattering Experimental Setup   (b) Ray Illustration of the Role of the Apertures
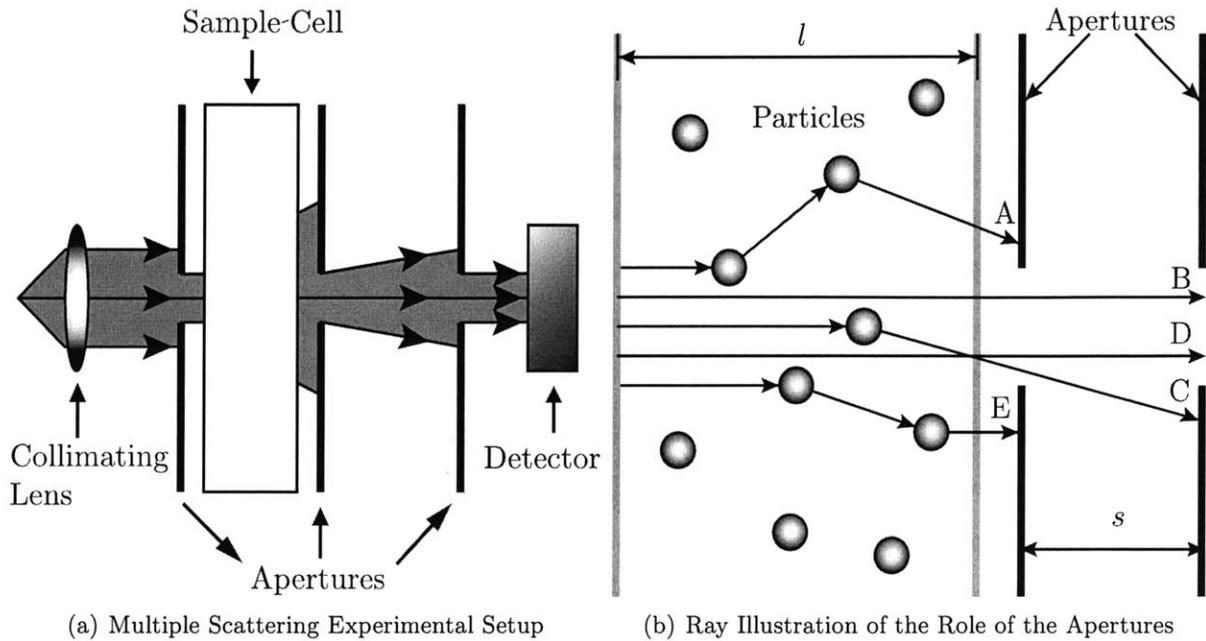
Figure 3-4: The experimental setup for observation of forward scattering in the multiple scattering regime. The role of the apertures is to remove multiply scattered light. The leftmost aperture removes light that was scattered back to $\theta = 0$, while the far aperture removes singly and multiply scattered off axis ($\theta \neq 0$) light.

scattering cell, as shown in Figure 3-4(a). The first aperture, placed directly in front of the cell, prevents off-axis multiply scattered light from entering the output beam. The second aperture, placed directly in front of a detector in the far field from the scattering cell, will restrict the light that exits the cell at an angle greater than the desired acceptance angle. In short, we are only concerning ourselves with the first scattering event. Quantitatively, the literature suggests that the half-angle of acceptance should be less than one tenth of the first angular minimum in the Fraunhofer diffraction pattern of a disk equal to the particle in the projected area [52]:

$$\theta_{1/2} \leq 0.122\frac{\lambda}{d} \tag{3.44}$$

where λ is the wavelength in the medium and d is the diameter of the particle diameter.

This experiment was replicated with the spectrograph described in Section 2.1 to show that this experiment could be done for all wavelengths of interest simultaneously, thereby generating

(a) Sphere diameter = 750nm   (b) Sphere diameter = 1000nm   (c) Sphere diameter = 1500nm
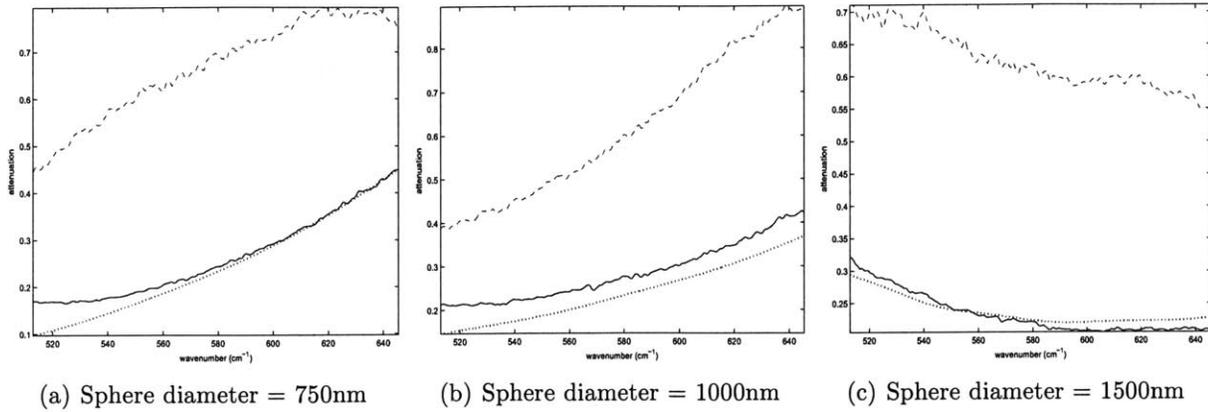
Figure 3-5: Attenuation of white light by a 1 cm, $\tau = 5$, slab of polystyrene spheres. The theory (dotted) is well matched by the results where the solid angle was restricted (solid), while the unrestricted solid angle (dashed) failed to eliminate all of the scattered light.

a wavelength dependant amplitude correction from Mie scattering. Instead of a laser source, the sample was illuminated with a tungsten halogen white light source (Ocean Optics LS-1). The white light was delivered by a 1 mm fiber butt-coupled to the lamp. Output light was collimated and passed through samples of polystyrene microspheres of $0.535\mu m$, $0.771\mu m$, $1.072\mu m$, $1.53\mu m$, and $2.061\mu m$ diameters (Polysciences Polybead(TM)). For the restricted solid angle case, the forward scattered light was restricted by an aperture close to the slab and coupled into a $62.5\mu m$ fiber set approximately 10cm away. The fiber was agitated mechanically to scramble the modes. The small core diameter of the bare fiber made it a natural limited aperture at this distance. For the unrestricted solid angle case, the first aperture was removed. Figure 3-4(b) illustrates the difference. If the first aperture is removed, then some of the multiply scattered light, such as ray A in Figure 3-4(b), can still enter the restricted aperture in the distance. A 600lines/mm grating with a 500nm blaze was set to a center wavelength of 580nm for a spectral range of 512nm to 646nm. For data prediction, a suite of MATLAB functions and scripts were developed that numerically solved for the attenuation of each pixel. The new suite was developed because existing scripts used slow algorithms to compute multiple wavelengths, did not take into account complex indices of refraction, and often did not make use of existing fast MATLAB functions. The source code of this suite is given in Appendix D. The solutions from the new suite were

(a) Sphere diameter = 535nm

(b) Sphere diameter = 771nm

(c) Sphere diameter = 1072nm

(d) Sphere diameter = 1530nm
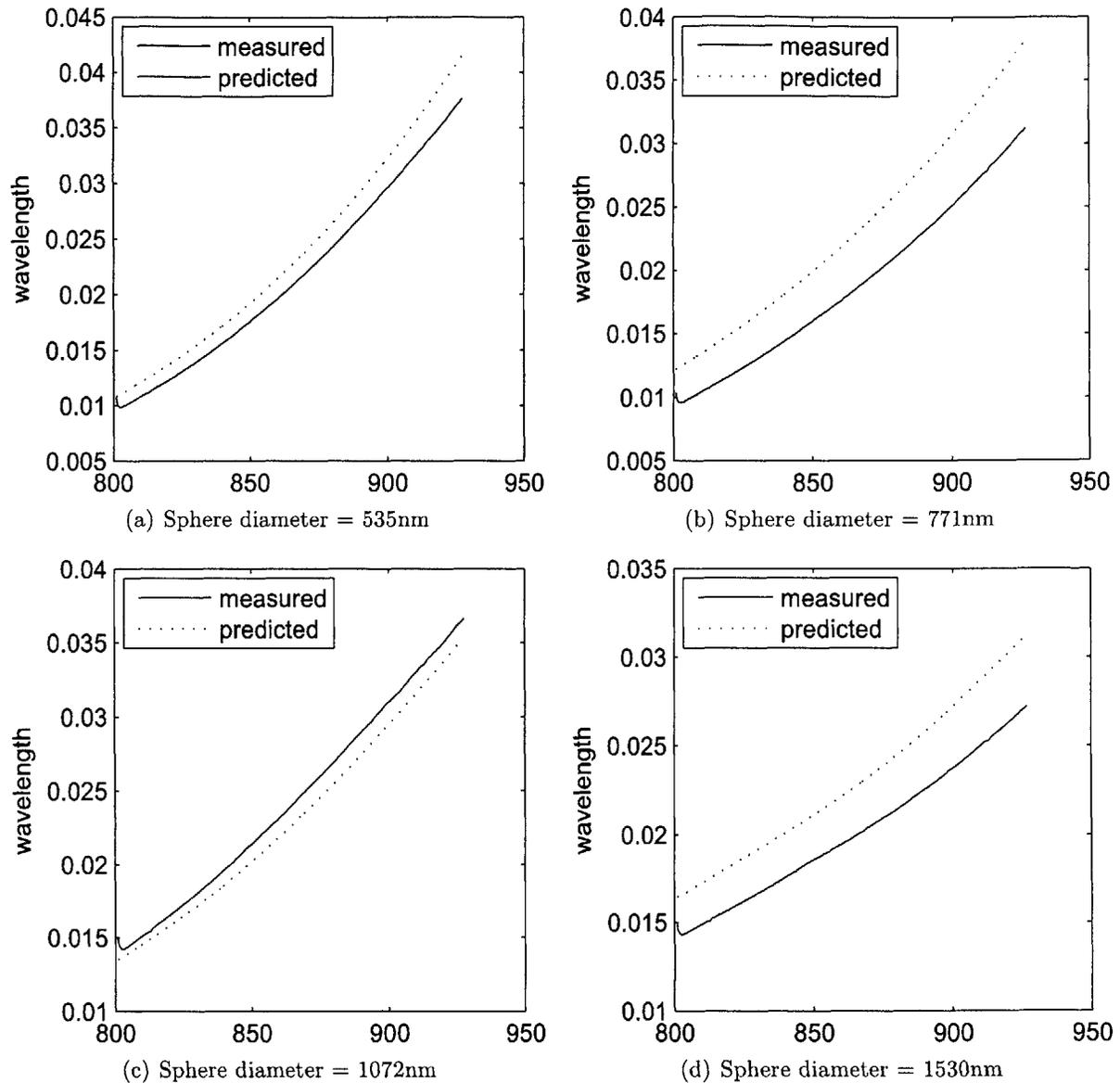
Figure 3-6: Attenuation of white light by a 1 cm, $\tau = 5$, slab of polystyrene spheres using the fiber Raman probe. The theory (dotted) is well matched by the measured values (solid).

compared to the output of the existing slower MATLAB scripts to test for correctness. As expected, the data matches the predicted values well for a restricted angle of acceptance. The data for varying particle sizes with turbidities of $\tau = 5$ is shown in Figure 3-5. It shows that when the solid angle is restricted, the wavelength attenuation can be predicted with our model

of using visible wavelengths and similar optics.

The experiment was replicated with a few optical changes more relevant to remote Raman spectroscopy. Namely, the fiber Raman probe was used as the collection optics instead of the raw fiber. To couple the light into the focusing optics of the fiber, a lens was placed confocal to the fiber probe in the place of the detector in Figure 3-3. The results, shown in Figure 3-6, indicate that there is still good agreement between the theoretical Mie solution and the measured values. The errors in this final optical setup show a maximum transfer function error of less than 0.69% of total attenuation and less than 45% of the transfer function. The overall error analysis in Table 3.1 shows the percent error between theory and the measured values for various particle sizes at $\tau = 5$. More important to note, however, is that most of the error is scalar. That is to say that the evolution of attenuation with wavelength is matched almost perfectly. This indicates that if the elastic scattering of the Raman light can be properly modeled, then a scattering correction is possible with the optical components already in use.

d (μm)	$A_m - A_t$ (%)	$(A_m - A_t)/A_t$ (%)
0.535	0.390	12.7
0.771	0.694	30.1
1.072	0.154	10.2
1.530	0.406	16.3
2.061	0.669	44.5

Table 3.1: Error analysis of the measured forward attenuation. For each sphere diameter, d, the percent error of the total attenuation, A, is given. In addition, the percent error with respect to the curve values is also given.

3.2.2 Elastic Scattering of Raman Spectral Signals

To confirm that the elastic scattering of Raman spectra can be predicted, an experiment was devised again involving a collimating lens. This time, laser excitation at 785nm was applied through the excitation fiber and collimated by a UV fused silica lens. Samples of water and polystyrene microspheres in cuvettes were illuminated from above for maximum working distance. Since the geometry of this setup is not a strict forward scattering experiment, the theory
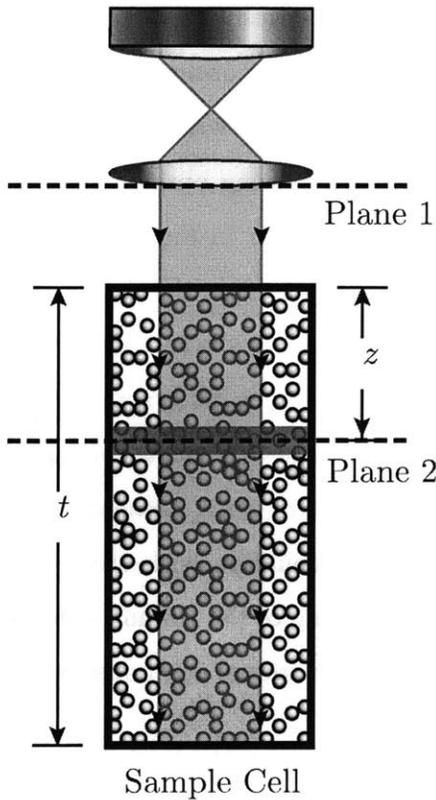
Figure 3-7: Schematic of the geometry of a Raman and elastic scattering experiment.

must be re-examined. Figure 3-7 shows a schematic of the geometry. Consider then, a slim sheet of the sample, located a distance z from the surface. Without elastic scatterers, the incident light with intensity $I_{inc,1}$ will pass virtually unattenuated to the sheet and excite a sheet of Raman scatterers with quantum efficiency Q. The spherical waves from each scatterer has random phase, so we will assume the all the scatterers in a sheet will combine to create a sheet of Raman scattered light of intensity $I_{r,2,j}$ which will travel back virtually unattenuated to the detector. Since this is just the Raman scattered light of a single infinitely small sheet, the total Raman scattered light will be the total integrated light from all of the sample:

$$I_{r,1} = \int_0^t I_{r,1,j} = Q \int_0^t I_{r,2} = Qt I_{inc,1} \qquad (3.45)$$

which properly shows the linear dependence of the overall Raman scattered light to the path

length. Now assume that large scattering bodies exist. The intensity of the excitation light at the sheet will therefore be reduced by forward scattering to:

$$I'_{inc,2} = I'_{inc,1} e^{-\sigma_i \rho z} \tag{3.46}$$

where σ_i is the cross-section of scattering for the incident light and ρ is the number density of the particles. Thus, at the jth sheet will again create some Raman scattering:

$$I'_{r,2,j} = Q I'_{inc,2} = Q I'_{inc,1} e^{-\sigma_i \rho z} \tag{3.47}$$

which will travel through the same field of scatterers, but at a different wavelength:

$$I_{r,1,j} = I'_{r,2,j} e^{\sigma_r \rho z} = Q I'_{inc,1} \exp\left[-(\sigma_i + \sigma_r)\rho z\right] \tag{3.48}$$

where σ_r is the cross-section of scattering for the Raman scattered light. Integrating this expression over z:

$$
\begin{align}
I'_{r,1} &= \int_0^t I'_{r,1,j} dz \tag{3.49} \\
&= \int_0^t Q I'_{inc,1} dz \exp\left\{-(\sigma_i + \sigma_r)\rho z\right\} \tag{3.50} \\
&= \frac{Q I'_{inc,1}}{(\sigma_i + \sigma_j)\rho} \left[1 - \exp\left\{-(\sigma_i + \sigma_r)\rho t\right\}\right] \tag{3.51} \\
&= I_{r,1} \left[\frac{1}{(\sigma_i + \sigma_r)\rho t} \left(1 - \exp\left\{-(\sigma_i + \sigma_r)\rho t\right\}\right)\right] \tag{3.52}
\end{align}
$$

Thus, the scattered light is modeled, to the first order, to be dependant solely on the sample properties and the cross-sections as defined by the Mie scattering formulation. To test this theory, serial dilutions of 1530nm polystyrene microspheres were analyzed with the setup described above. The low sphere concentration made the contribution to the Raman signal from the spheres insignificant in comparison to the signal from water. Thus, the Raman signal of water was the spectrum under test. Table 3.2 shows the concentrations of the spheres tested

(a) $\tau_f = 0.1$ (b) $\tau_f = 0.2$

(c) $\tau_f = 0.4$ (d) $\tau_f = 0.5$
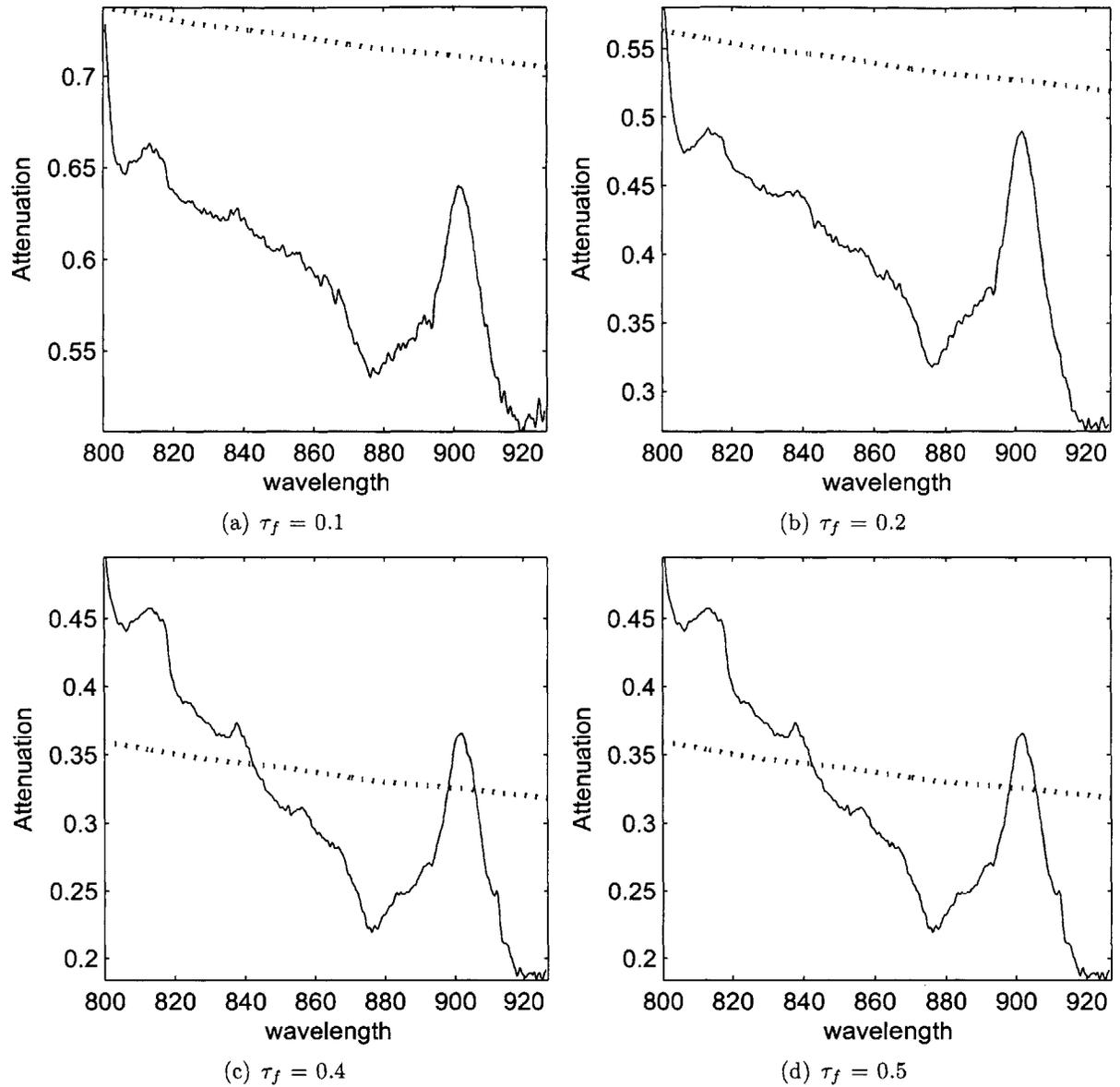
Figure 3-8: Theoretical (dotted) and measured (solid) attenuation of the Raman spectrum of water.

along with other parameters. τ_f is the τ that would have been experienced if with a volume 4.572cm thick ($t = z = 4.572$cm) was inserted into the forward scattering setups in Section 3.2.1. In addition, τ_i corresponds to the attenuation of the excitation beam, while τ_r corresponds to the average attenuation of the Raman scattered light in this setup.

ρ (m^{-3})	τ_f	τ_i	τ_s
0.4135×10^{12}	0.1	0.3090	0.3882
0.8271×10^{12}	0.2	0.6179	0.7764
1.2406×10^{12}	0.3	0.9269	1.1646
1.6524×10^{12}	0.4	1.2358	1.5528
2.0677×10^{12}	0.5	1.5448	1.9410

Table 3.2: The concentrations and relevant optical thicknesses for observing the attenuation of the Raman signal of water using dilutions of 1.53μm spheres.

Figure 3-8 shows the results of the experiments. The overall attenuation of the Raman spectrum of water appears to scale down in accordance with the theory. In all of the cases shown in Figure 3-8, the observed attenuation (solid) is slightly below the predicted value (dotted). Unlike, the theory, however, the wavelength dependence of the attenuation is not monotonic. This data therefore indicates that the distortions in the Raman spectra caused by scattering are not entirely described by the theory presented. The attenuation appears to be dependant on the signal intensity. This is similar behavior to the observed attenuation caused by elastic scattering from air bubbles, as discussed in Section 2.2.6. Spectral distortions caused by changing fiber positions of Raman signals of polycarbonate and fused silica have been shown to be intensity dependant as well, as discussed in Section 2.2.6. These other results, when combined with the results presented in this section, point to a possible single source of error related to the fiber optics of the system. It is believed by the author that this source of error, whatever it is, should be the first priority to isolate and fix.

3.2.3 Amplitude Correction Function

Finally, the desired product of this scattering formulation is an amplitude correction function which could correct for scattering via software. To do this, it is important to attempt an experiment using the correct probe geometry to test the theory. First, the theory so far formulated needs one more step. The probe geometry shown in Figure 3-3 also shows that the lens collects a large solid angle. Specifically, for a focal length f and diameter d, the half angle of acceptance will be:

$$\theta_{1/2} = \frac{d}{2f} \qquad (3.53)$$

instead of almost zero. This indicates that some of the scattered light will still be accepted into the detector. Note that f is still much larger than the typical sphere diameter, a, indicating that all other assumptions used about the far-field should still hold. Therefore, to predict the total power attenuated, P_a, the solid angles being collected must be removed:

$$P_a = P_s - P_c \qquad (3.54)$$

where P_c is the scattered power collected. Furthermore, using 3.4, a relation for the cross-sections can be determined:

$$P_{atten} = P_s - P_c \qquad (3.55)$$

which when combined with Equation 3.36 yields a new useful cross-section:

$$\sigma_{atten} = \frac{\pi}{k^2} \int_{\theta_{1/2}}^{\pi} \sin\theta \left[|S_1(\theta)|^2 + |S_2(\theta)|^2 \right] d\theta \qquad (3.56)$$

which can be solved numerically by doing this integral explicitly. Using σ_{atten} instead of σ_{sca} in Equation 3.15 should therefore account for the reduced attenuation due to the increased numerical aperture of detection. The results of an initial experiment are shown in Figure 3-9. The measured attenuation is plotted along with the theory with (dashed) and without (dotted) the cross-section computation being restricted to angles not included within the numerical aperture of the lens. The line created using the total scattering cross-section appears to be in better agreement to the data than the line created using the restricted scattering cross-section. This result indicates that there is another mechanism that is restricting the collection of light scattered with small scattering angles. More data is required, however, before this hypothesis is ruled out. The sharp drop at high wavenumbers is real (i.e. not a result of filtering) and seems to support the hypothesis that the scatterers could have sharp absorption bands previously
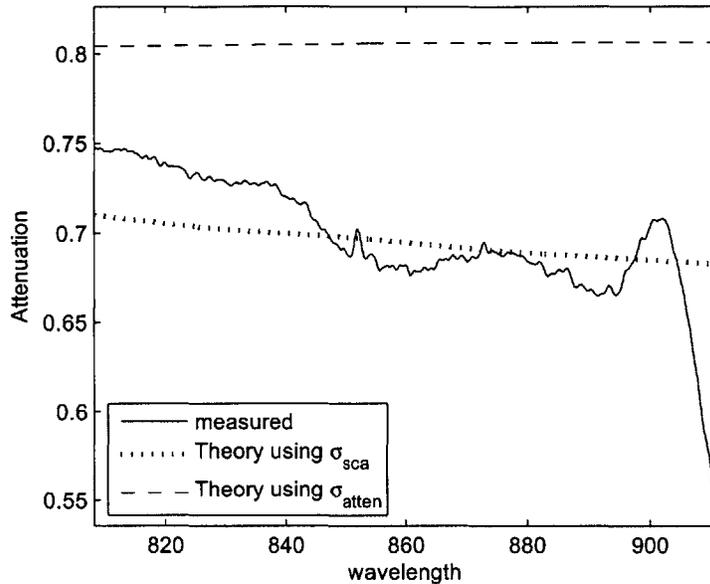
Figure 3-9: Results of an initial scattering experiment using 1530nm polystyrene spheres with ρ $= 2.063 \times 10^{13}$ m^{-3}, $\tau_f = .1$, and $t = 1$mm. This result indicates that restricting the cross-section is unnecessary.

unknown to the author. The complex index of refraction of polystyrene used in the author's MATLAB simulations is based by the data in [53]. This data set only has one sample every 30nm, making it impossible to know if there are sharp absorption peaks in the NIR here. This should be investigated along with the other issues mentioned in Section 3.2.2. The general trend of the line is captured, however, indicating that if these issues are fixed, there is promise for a software scattering correction for online Raman spectroscopy. Even more likely, however, the source of the unexplained distortion could be some higher order problem with the fiber optics, as previously discussed.

3.3 Online Biomass Estimation from Scattering

While measurements of pH, dissolved oxygen, and component concentrations all give useful information about the cell growth implicitly by monitoring conditions, it is important to monitor the cell growth of the biomass directly for bioprocess optimization. The typical method used for

measuring biomass is a measurement of the optical density (OD) at a particular wavelength and path length. For offline samples, the technique is similar to the scattering technique described in Section 3.1.1, except that a monochromatic light source is used, and the attenuation law has changed:

$$I_{out} = I_{inc}e^{-\tau} = I_{inc}10^{-OD} \tag{3.57}$$

where I_{out}, I_{inc}, and τ have already been defined in Section 3.1.1, and $OD = \tau/ln(10)$. To keep the OD linear with biomass, samples are diluted by a defined amount and measured in the single scattering regime. This induces noisy behavior when the dilution ratios are high (above 10:1). This measurement can later be calibrated with offline dry cell weight measurements after the bioreaction is over. If the calibration is known for a particular strain, then all subsequent estimates of OD can result in a measurement of total biomass.

Ideally, this measurement would be done online for reasons stated above. In fact, probes for measuring OD in a bioreactor already exist [12, 54], and are now commercially available. The most common technology is a linear extinction measurement, with the probe holding a small LED and detector, as shown in Figure 3-10. Since the probes are in situ, there can be no dilution and the response of extinction with biomass will be non-linear. Furthermore, this method is sensitive to scattering from bubbles, not unlike a Raman probe. This means that an additional calibration step is needed to fit the actual extinction to linear OD. Usually, that step is to run a test fermentation using similar growth conditions and impeller speed, then fit the curve of linear extinction to offline OD with a non-linear function [54]. Bioreactors have a limited number of ports, however, so integrating online OD measurements into an existing probe would be ideal.

Since it is also an optical method that is affected by the presence of biomass, information about the OD is already contained in the measured Raman spectra. It is possible to integrate this function into the fiber Raman probe entirely via signal processing, simultaneously obtaining online estimates of the biomass. One way of extracting this information from the spectra is by examining the concentration estimate of water before normalization. Since water is most of the bioreactor medium solution, in the absence of laser power drift and scattering, the concentration
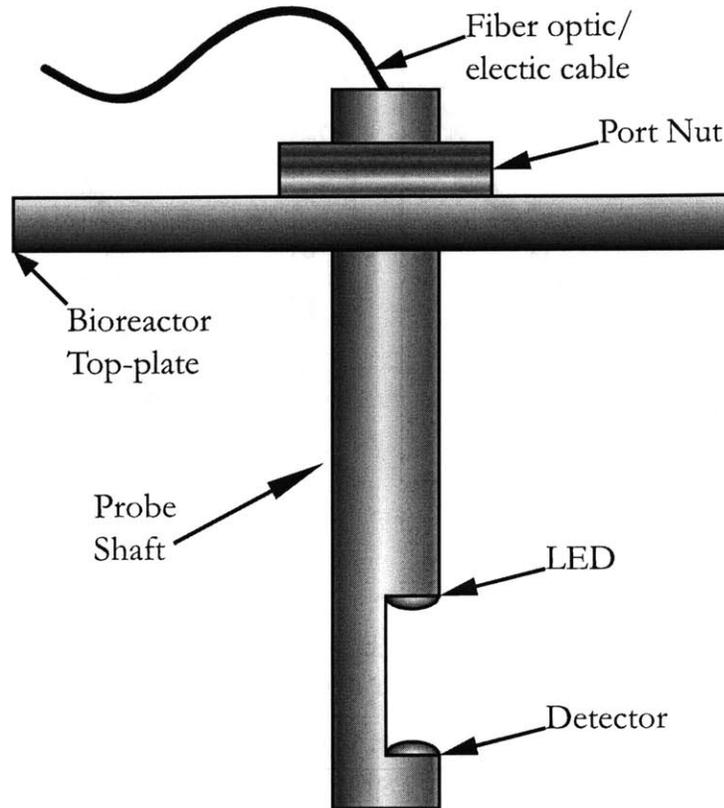
Figure 3-10: Schematic of a state-of-the-art linear extinction probe for online monitoring of OD in a bioreactor.

estimate of water should remain within the shot-noise limited values. Assuming that the effects of laser power drift are small and that scattering causes mostly a scalar change in the overall Raman spectrum, the change in the concentration estimate of water will be a result of the extinction caused by the biomass and air bubbles. Further assuming that the attenuation contribution from the bubbles remains fixed as the cells grow, normalizing the concentration estimates of water, c_w, to the concentration estimate of the first point, $c_w(t = 0)$, should reveal the attenuation due to the biomass, A_b:

$$A_b(t) = 1 - \frac{c_w(t)}{c_w(0)} \tag{3.58}$$

Figure 3-11(a) is a plot of the Offline OD versus $A_b \times 100\%$. The actual OD and the OD that

(a) OD vs. attenuation   (b) OD vs. time
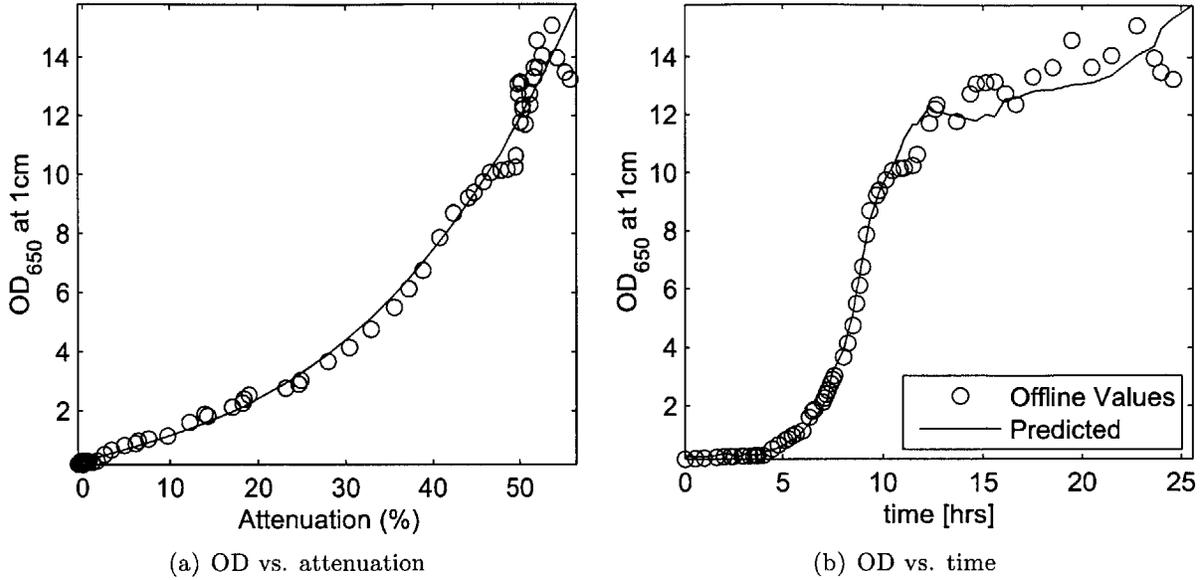
Figure 3-11: Cubic fit of OD to attenuation of the estimated concentration of water (a) and the resultant prediction of OD using the fit on the same fermentation (b).

Coeff.	Value
a_0	2.8074×10^{-1}
a_1	7.9094×10^{-2}
a_2	1.9911×10^{-4}
a_3	5.7445×10^{-5}

Table 3.3: Coefficients for the polynomial fit in Figure 3-11, where $y = a_3 x^3 + a_2 x^2 + a_1 x^1 + a_0$.

would be estimated by this line fit versus time for the same fermentation are shown in Figure 3-11(b). Since offline data was taken at different points than online data, the points shown were made by linearly interpolating between the OD measurements. As expected, the attenuation of the water signal grows nonlinearly with the OD due to multiple scattering at high OD. The solid line shown is a third order polynomial fit, with coefficients defined in Table 3.3. The two methods are in agreement to within 0.43 OD for OD < 10 and within 2.1 OD for OD > 10. The fact that the offline OD measurements get significantly noisier after the dilution was changed to 40:1 indicates that the dilution process could be the dominant source of error for high offline OD measurements. Likewise, the increased noisiness in the line fit at high OD could be due to errors in the concentration estimate of water caused by the non-scalar effect of scattering discussed
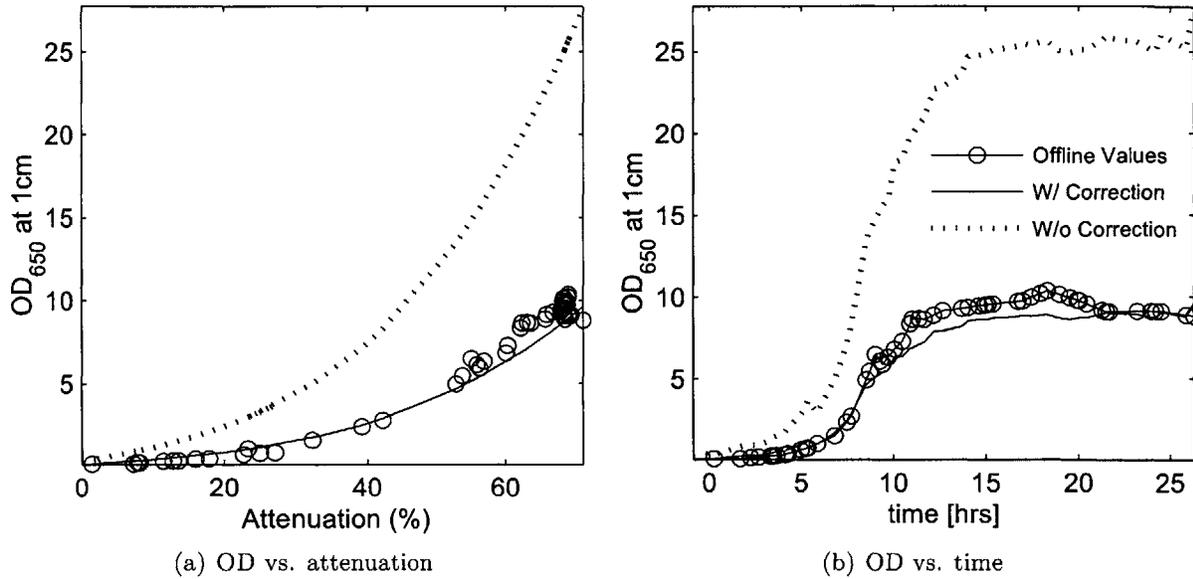
Figure 3-12: The predicted OD on another fermentation before (dotted) and after (solid) a correction for different focal depths. Both predictions are compared with the offline measured OD values (circles).

earlier in this chapter. Data from a previous fermentation of the same strain of *E. Coli* using the same conditions was obtained to test the OD prediction. The only difference between the two fermentations was that the older fermentation, described in [11], placed the focal point of the lens deeper inside the tank from the window. This caused the effect of the scattering to be stronger. Assuming that the scattering is a multiplicative effect, and that scattering from air bubbles is affected by the same amount as scattering from biomass, a correction can be made for this difference. To do this, it is possible to use the ratio of the attenuation of the concentration estimate of water at time 0 from the nominal value for each fermentation to get the value of a correction constant, C_c:

$$C_c = \frac{c_{nom,t} - c_{w,t}(0)}{c_{nom,m} - c_{w,m}(0)} \tag{3.59}$$

where $c_{nom,t}$ is the nominal concentration of water for the training fermentation, $c_{nom,m}$ is the nominal concentration of water for the measured fermentation, $c_{w,t}(0)$ is the starting estimated

concentration of water for the training fermentation, and $c_{w,m}(0)$ is the starting estimated concentration of water for the measured fermentation. The OD prediction curve can then be multiplied by the correction coefficient to yield a proper OD estimate. Figure 3-12 shows the results of the OD estimation using both the uncorrected (dotted) and corrected (solid) OD predictions. It is clear that with the correction in place, the OD prediction is in good agreement with the measured OD values for the second fermentation. The maximum error observed is 1.73 OD, while the average error observed is 0.66 OD. This result proves that the OD of a fermentation can be predicted from the attenuation in the concentration of water after a calibration fermentation has been conducted.

3.4 Summary

In this chapter, Mie scattering cross-sections have been shown to be effective in predicting the wavelength dependence of elastic scattering for the forward scattering collimated geometry. This has been effectively demonstrated into the multiple scattering regime for turbidity values of τ > 5 (OD > 2.17) using the same optical components used for the Raman bioprocess monitoring setup. Using GLMT, the perturbations in the theory caused by the new Gaussian shape have been accommodated. Using a quadrature method, it is possible to determine the actual amount of light scattered into a lens, allowing for large solid angle collection in forward scattering theory. Spectral distortion of Raman spectra by elastic scattering has also been demonstrated, although our model fails to describe some of the wavelength dependence in these geometries. One hypothesis is that the polystyrene spheres may have absorption peaks in the NIR. The lack of available absorption spectra in the NIR for polystyrene makes it difficult to know whether of not this is true without conducting an experiment. More likely, however, is that the additional distortions in the spectra of both experiments are related to how light is coupled into the fiber optic cable. More experiments must be conducted to fully establish a firm conclusion about the theoretical treatment here. First, absorption spectra of polystyrene should be acquired to ensure that there are not unknown peaks. Second, the spectral distortions caused by manipulating the

fiber should be more carefully modeled, as these distortions exhibit intensity dependence as well. Third, experiments involving complex mixtures with scattering spheres should be conducted to quantify the effect the scatterers have on concentration estimates. Fourth, samples of cells should be acquired to determine the effect they have on concentration estimates offline. This would allow for a more complete picture of the scope of the problem.

We have also shown that a monotonic relationship exists between the attenuation in the concentration estimate of water and OD, indicating that the Raman probe could be used as an online OD monitor. In fact, with a scattering correction and other systematic error removals, it is expected that the online Raman measurements will be less noisy than the offline OD estimates. A third order polynomial was proposed as a possible fit and shows good agreement when used to predict OD on the same fermentation. Using this fit, OD from a previous fermentation was predicted with a maximum error of 1.73 OD and an average error of 0.66 OD, proving the effectiveness of this technique. If a scattering correction is also introduced before such a fermentation, it is likely that smooth, repeatable estimates will be achieved.

Chapter 4

Optical Spectroscopies for Plant Cell Cultures

4.1 Plant Cell Cultures

As mentioned in Chapter 1, plant cell culture bioreactions are now being used in a number of medical and industrial applications. Plant cell cultures differ from their microbial counterparts in that, while they can be used to create biochemicals metabolically in the suspension culture directly [55], they can also be use for somatic clonal propagation of certain plant species [56]. In both cases, the bioreactor is setup and operated in a similar manner to bacterial fermentations. Some principle differences, however, are lifetimes and particle sizes. For plant cell cultures, lifecycles are on the time scale of months, instead of the time scale of days seen by microbial bioreactions, and can result in increased necessity for proper sterile techniques, as contamination from bacteria or fungi can overwhelm the plant cells. In addition, plant cells must grow into multi-cellular aggregates, or calli, to survive. These calli can grow over 1mm in size and are fragile, making them susceptible to the shear stresses caused by the aeration and impeller agitation necessary in stirred-tank bioreactor cultivation [57].

While this last difference adds some constraints to both types of plant cell bioreactions, it can be particularly complicating for embryo production cultures. For these cultures, the calli are

being grown for eventual propagation outside of the stirred-tank reactor. Thus, it is desirable for the calli to be embryogenic clones that properly replicate desired phenotypes. Even in optimal growth conditions, shear stress can reduce yields in the following ways. First, some calli will die, due mostly to lysis of the aggregates and member cells from the shear stress. Second, some calli will enter a non-embryogenic growth path. The dominant cause of this is imperfect growth conditions, although reactor stress has been shown to affect the embryogenic yield [58]. Lastly, some of the embryogenic calli may display an undesired phenotype, despite being a genetic clone. This problem is caused by DNA hyper and hypo-methylation of the calli [59]. When the plant cells are under significant oxidative stress, oxygen radicals can damage the enzymes in charge of methylating DNA segments [60]. This can cause a malfunction in the methylation of DNA, ultimately preventing some proteins from being transcribed, resulting in phenotypic changes.

Once plant cells are removed from cell cultures, they must be grown on solid media to test for embryogenesis. Later, they must be planted and allowed to grow before the desired phenotype will manifest itself. For some kinds of plants, this process is on the time scale of years. It is therefore of critical importance to carefully monitor bioprocesses to get the maximum yield of cells that are both embryogenic and free of epigenetic alterations. It is therefore desirable to implement a bioprocess monitoring system that could help control the conditions inside the bioreactor. Furthermore, there is a demand for a method to determine embryogenesis and epigenetic changes in plant calli.

4.2 Offline Raman Spectroscopy for Plant Cell Bioreactions

One plant that has been grown in bioreactors, and suffers from some of the adverse affects mentioned, is oil palm (*Elaeis guineensis Jacq.*). This plant is of particular interest as it is at the core of a multi-billion dollar palm oil industry. Furthermore, the successful culture of oil palm suspension cells (strain:E90L1) has been demonstrated in a stirred-tank Sixfors(TM) bioreactor, with the goal of growing cultures to develop embryogenic clones of high yield oil

Composition of Oil Palm Medium			
Components	Stock conc. (mg/L)	Starting conc. (mg/L)	Raman Activity
Ammonium nitrate	16500.0	1650.0	Degenerate
Calcium chloride anhydrous	3322.0	332.2	Undetectable
Magnesium sulfate	1807.0	180.7	Strong
Potassium nitrate	19000.0	1900.0	Degenerate
Potassium phosphate monobasic	1700.0	170.0	Strong
Boric acid	6.2	0.62	Undetectable
Cobalt chloride $6H_2O$	0.025	0.0025	Undetectable
Cupric sulfate $5H_2O$	0.025	0.0025	Undetectable
Na_2 EDTA	37.3	3.73	Undetectable
Ferrous sulfate $7H_2O$	27.8	2.78	Undetectable
Manganese sulfate H_2O	16.9	1.69	Undetectable
Molybdic acid $2H_2O$	0.25	0.025	Undetectable
Potassium iodide	0.83	0.083	Undetectable
Zinc sulfate $7H_2O$	8.6	0.86	Undetectable
glycine	2.0	0.002	Undetectable
myo-inositol	100.0	0.1	Undetectable
nicotinic acid	0.5	0.0005	Undetectable
pyroxidine hydrochloride	0.5	0.0005	Undetectable
thiamine hydrochloride	0.1	0.0001	Undetectable
Sodium 2,4-D monohydrate	1.0	0.001	Undetectable
NAA potassium salt	1.0	0.001	Undetectable
Glutamine	100.0	0.0	Undetectable
Sucrose	30000.0	30000.0	Strong
Fructose	0.0	0.0	Strong
Glucose	0.0	0.0	Strong

Table 4.1: Starting composition of oil palm medium along with comparative Raman activity. Raman activity was tested at the nominal beginning concentrations for all components except for glucose and fructose, which were tested at the beginning concentration of sucrose.

production [61]. For reasons given above, it is desirable to monitor the growth conditions of these reactions closely and online. An investigation into the effectiveness of Raman spectroscopy for online concentration estimation has subsequently been conducted.

The first step of this process involved analyzing all constituent components in the oil palm fermentation for Raman activity. Some constituents that are monitored include sugars such as fructose, sucrose, and glucose. The concentrations of these constituents vary in the range of approximately $0 - 30g/L$ [61]. The sensitivity of glucose has already been demonstrated at $0.11g/L$, indicating the promise for sucrose and fructose concentrations [11]. Other constituents of interest include ammonia and glutamine, which have been demonstrated to be Raman active in the near-infrared in the literature [62], although at much higher concentrations.

4.2.1   Component Analysis

Before sample measurements were done, it was necessary to investigate the Raman activity of the constituent components of the oil palm medium. The tests were conducted using the rigid universal calibration setup described in Section 2.3.1. All spectra were the average of ten 30s acquisitions and underwent all the amplitude and wavelength corrections described in Section 2.3. Table 4.1 shows the results of the tests along with their initial concentrations. The fructose and glucose are not added initially; they are the natural byproducts of sucrose and thus appear in large quantities later in the cell culture life cycle. All other components can be assumed to be either consumed in the bioreaction or remain constant. Table 4.1 also shows the results of tests of the Raman activity of the components at the nominal starting concentrations. Fructose and glucose are products of sucrose and so were tested at the nominal beginning concentration of sucrose. The carbon sources (sugars) and the nitrogen sources (nitrates) exhibit comparatively strong Raman spectra, although the nitrate salts have identical Raman spectra. Since nitric acid (HNO_3)is a strong acid, it disassociates completely into the nitrate ion (H_3O^-), and the hydronium ion (H_3NO^+) [63]. Likewise, a nitrate salt will disassociate in solution into the nitrate ion and the cation. It is therefore logical that the nitrate solutions have identical Raman spectra, since the ammonium ion has a very weak Raman cross-section and a single atom has
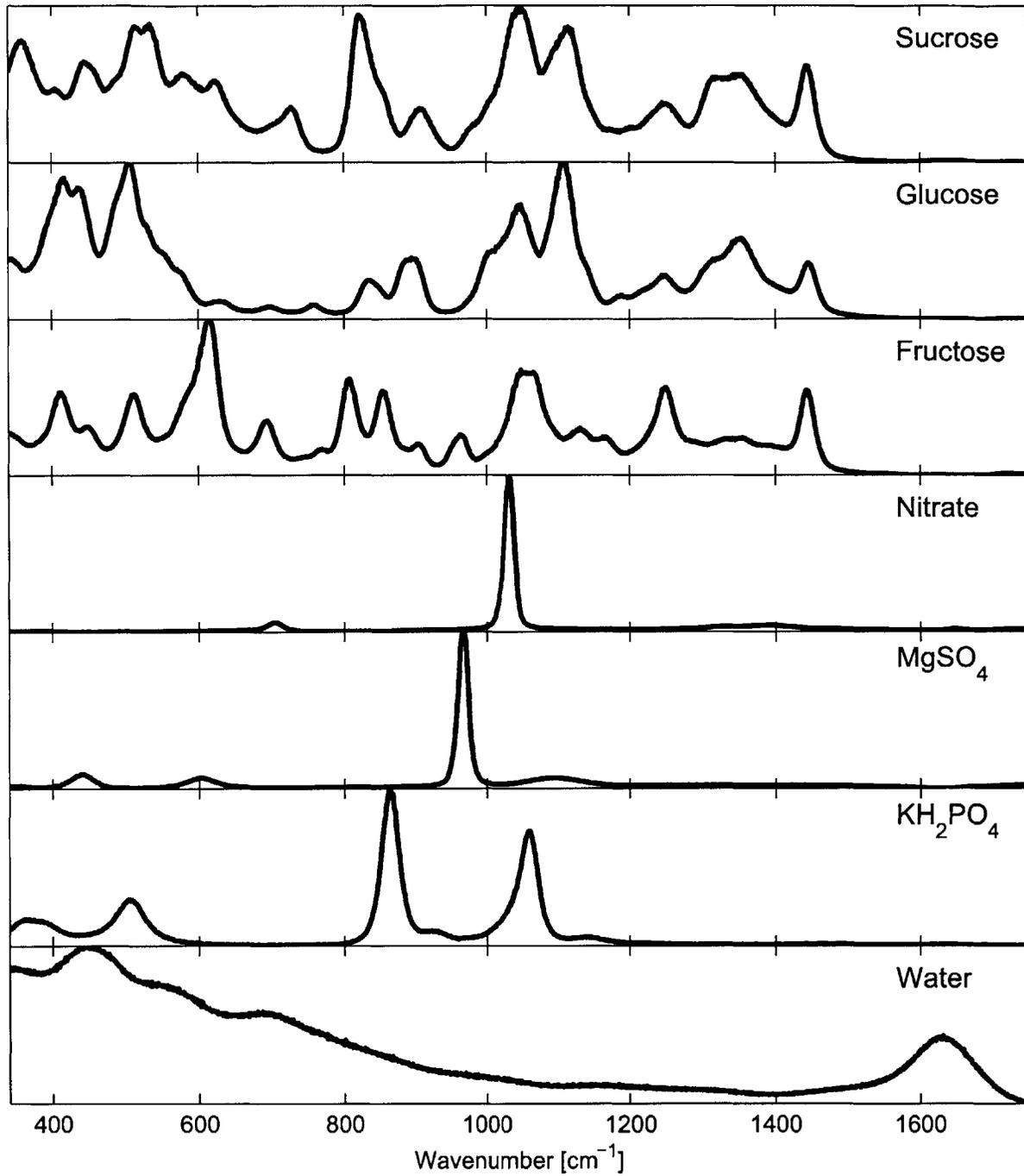
Figure 4-1: Raman spectra of oil palm medium components.
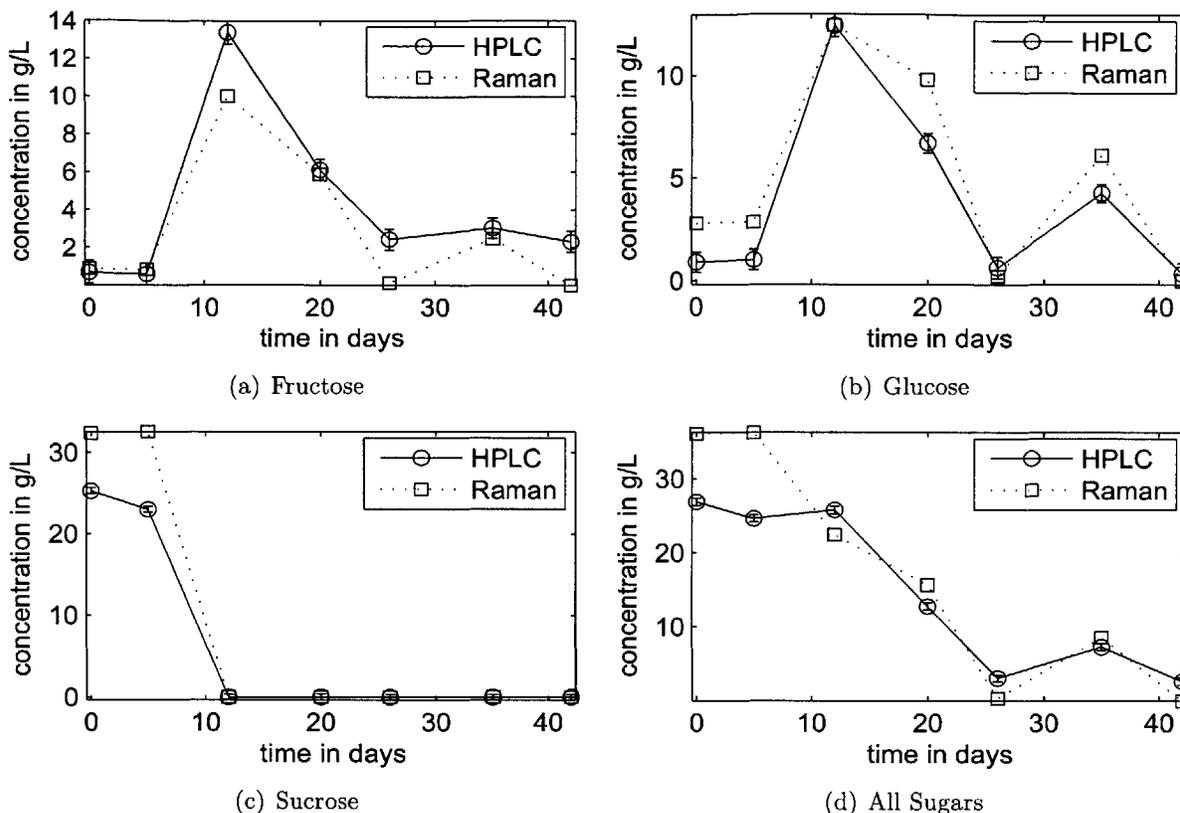
Figure 4-2: Offline Raman measurements of sugars in an oil palm bioreaction compared with HPLC results.

none at all. The medium buffer components were also Raman active, and much like the *E. coli* buffer components, must be included in any calibration set. Figure 4-1 shows all of the Raman spectra.

4.2.2   Offline Raman Measurements

The comparative strength of the Raman spectra of the sugars, nitrates, and buffer components, as shown in Table 4.1, indicate that they will be detected by a Raman spectroscopy based concentration estimation setup. The other components, however, are likely to be in quantities too insufficient to be detectable. Phenolics and other components added or created by the oil palm cells, are only produced to levels up to the 1mM range and are therefore below the

(a) Potassium Phosphate   (b) Magnesium Sulfate   (c) Nitrate
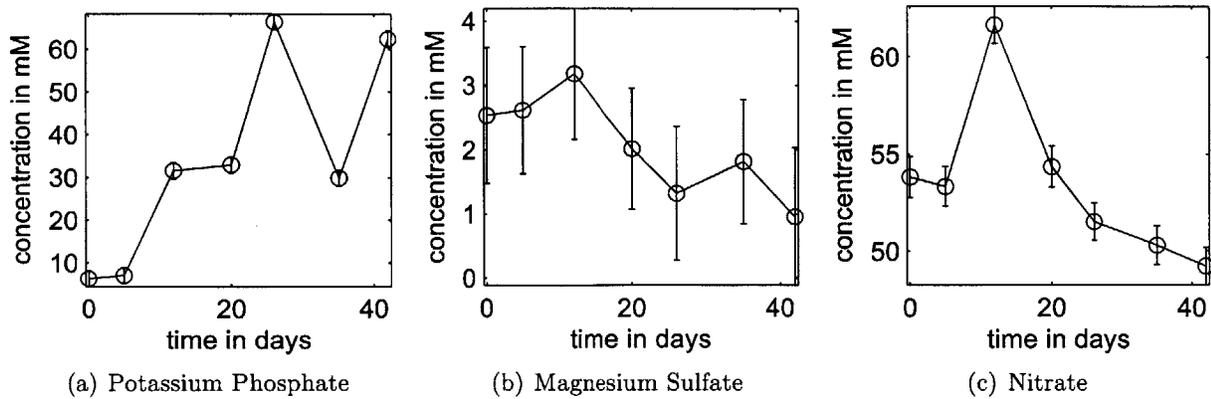
Figure 4-3: Offline Raman measurements of other components in an oil palm bioreaction.

noise floor of the instrument, as the error analysis in Chapter 2.4 illustrates. To confirm these findings, filtered samples of an oil palm cell culture conducted in a modified Sixfors (Infors, AG CH-4103 Bottmingen/Switzerland) laboratory bioreactor were analyzed with the offline Raman instrument used for the calibrations, in a similar manner to the offline bacterial cell samples in Chapter 2. Figure 4-2 shows the results for one of the bioreactors for fructose (a), glucose (b), sucrose (c), and all sugars (d) in g/L. Incomplete and missing HPLC data for nitrate prevented a similar comparison for nitrate. A brief analysis of the errors in the measurements is shown in table 4.2.

Error Analysis for Oil Palm Estimates				
	Theory (3σ)		Maximum Error	
Component	mM	g/L	mM	g/L
Sucrose	2.71	0.93	27.9	9.56
Glucose	3.07	0.55	17.2	3.11
Fructose	3.40	0.61	18.8	3.39
All sugars	9.18	2.09	39.4	11.64

Table 4.2: Error results for offline concentration estimates of sugars in an oil palm cell culture growth medium.

The theoretical values for the error are derived using the same error propagation techniques described in Section 2.4.1. This data shows similar error performance to previous offline measurements for *Escherichia coli* bioreactions. The beginning concentration of sucrose was overes-
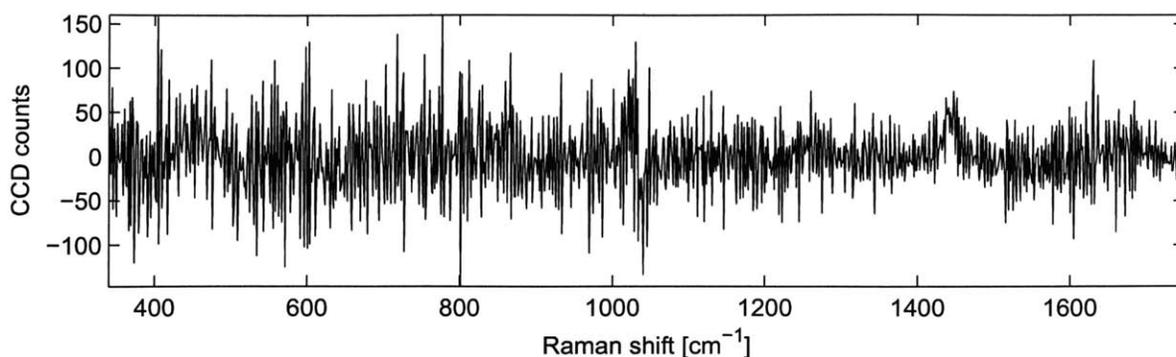
(a) Residual of the concentration estimate on day 0.
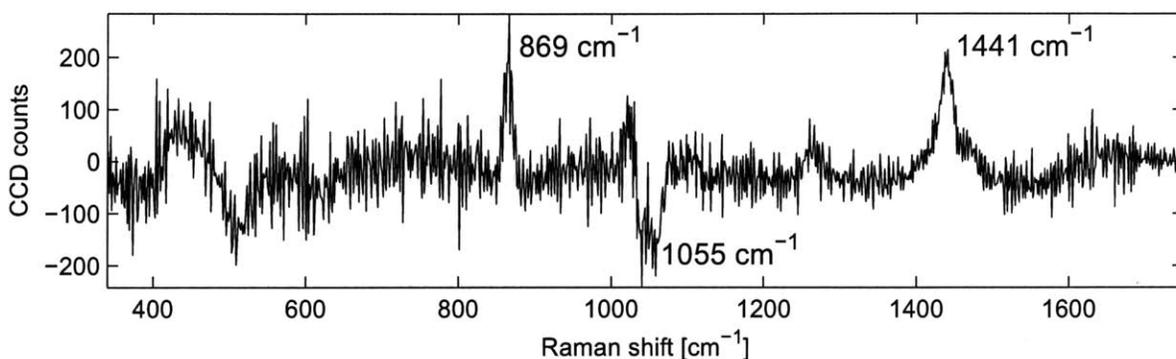
(b) Residual of the concentration estimate on day 12.

Figure 4-4: The residual of the concentration estimation on two different days. The second one shows well-defined peaks, indicating an inadequacy in the physical model.

timated by HPLC, which accounts for some of the error between HLC and Raman measurements. The error between the HPLC concentration estimate and the nominal beginning concentration of sucrose was 7.0mM. Finally, the lack of pH control in this bioreaction causes errors in the concentration estimates. A previous bioreaction of the same strain showed that the pH in the bioreactor dropped dramatically in the first ten days and then gradually over time [61]. As Figure 4-4 illustrates, the residual of all time points after the 5th day of the reaction have well defined peaks. In particular, The peak at 1441 cm^{-1} is shared by all of the sugars, while both of the other peaks correspond to the two peaks of the phosphate buffer. The pH drop is a sign of the failure of the phosphate (pKa = 7.21) to regulate it, causing the potassium and phosphate to associate with hydroxide and hydronium respectively. This could cause the Raman

spectrum due to potassium phosphate to be different at low pH. Furthermore, the concentration estimates of the potassium phosphate, nitrate, and magnesium sulfate, behave erratically, as shown in Figure 4-3, indicating that this could be the case.

The Raman data for nitrate was consistent with what was expected, except for the erroneous data point at 12 days. This spike in nitrate estimate is probably caused by fitting errors, as the only peak in the Raman spectrum of nitrate overlaps a peak in the Raman spectrum of fructose, which is underestimated at that data point. These fitting errors are also likely results of the pH changes, since fructose also shares a peak with potassium phosphate. To correct for these problems, a new calibration set comprised of these components at varying pH should be constructed. More data is needed before these questions can be conclusively answered. In general, however, the oil palm medium is sufficiently simple enough to make Raman spectroscopy using explicit methods a promising method for online monitoring of these bioprocesses.

4.3 Fluorescence Spectroscopy for Calli Differentiation

Plant cell cultures primarily serve the purpose of creating many clones of a plant with a desired phenotype. Even in the most controlled conditions, somaclonal variation and methylation of DNA caused by the unnatural environment can trigger mechanisms that prevent the calli from growing into plants. These non-embryogenic calli are essentially waste and, with current technology, account for most of the calli in the bioreactor. It is therefore useful to determine as early as possible in a growth process whether a callus is embryogenic, i.e. whether it can develop into a plant. Current methods for differentiating embryogenic calli from non-embryogenic calli involve inexact trained observations of callus morphology. These morphological differences are also correlated with differences in chemical composition. For *Medicago Arborea L.* calli, for example, non-embryogenic calli were found to contain less starch [64], implicating that a non-invasive test for starch could be used as a possible differentiation mechanism for future growths.
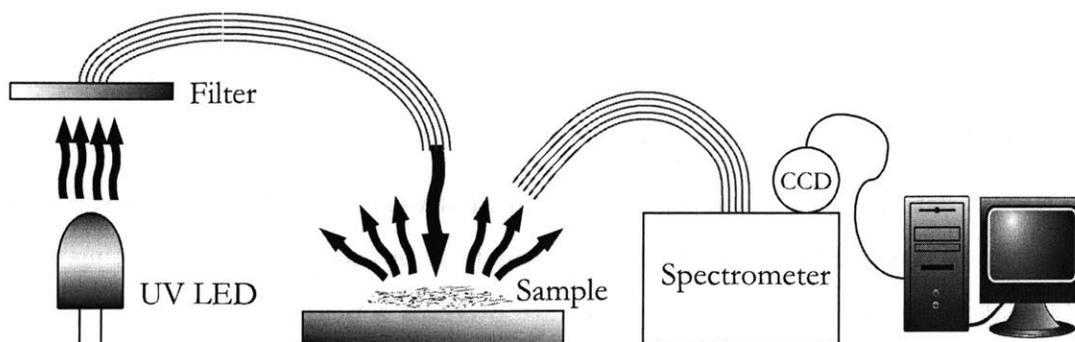
Figure 4-5: Fluorescence spectroscopy setup.

4.3.1   Fluorescence Spectroscopy Setup

Several optical techniques were explored in the examination of oil palm calli. Due to the inexact morphological and phenotypic information given, direct tests for embryogenesis could not be conducted. There is a possible correlation, however, between the friability of the calli and embryogenesis. For the liquid culture phase of propagation, the author was informed that calli of the friable phenotype are more likely to be embryogenic [65]. For solid culture, however, the non-friable calli with small friable nodules appear to have greater embryogenic viability [66]. Therefore, all experiments were conducted to determine the chemical contrast between calli with a friable phenotype and calli with an non-friable, or hard, phenotype. An experiment using Raman spectroscopy as the analysis technique was conducted. Figure 4-6 displays photographs of the calli before and after sample preparation. The complexity of the cells, combined with strong fluorescence, prevented any meaningful spectra from being acquired, and no contrast was observed.

Since the fluorescence signal was strong enough in the NIR to dominate a Raman signal, it was expected that the overall fluorescence signature would be very large in the visible, so a fluorescence spectroscopy setup was assembled as shown in Figure 4-5. The calli samples were differentiated by hand on a glass slide, dried, and granulated to prevent differences in callus geometry from affecting signal amplitudes. Excitation from a 380nm LED was filtered with a shortpass filter with cutoff wavelength of 480nm (3rd Millennium 480sp 40564) butt-coupled to

(a) All callus sizes - whole

(b) All callus sizes - ground

(c) Small and friable whole calli

(d) Mid-sized whole callus, hard with friable nodules

(e) Large, hard, whole callus

(f) Small and friable ground calli

(g) Mid-sized ground callus, hard with friable nodules

(h) Large, hard, ground callus
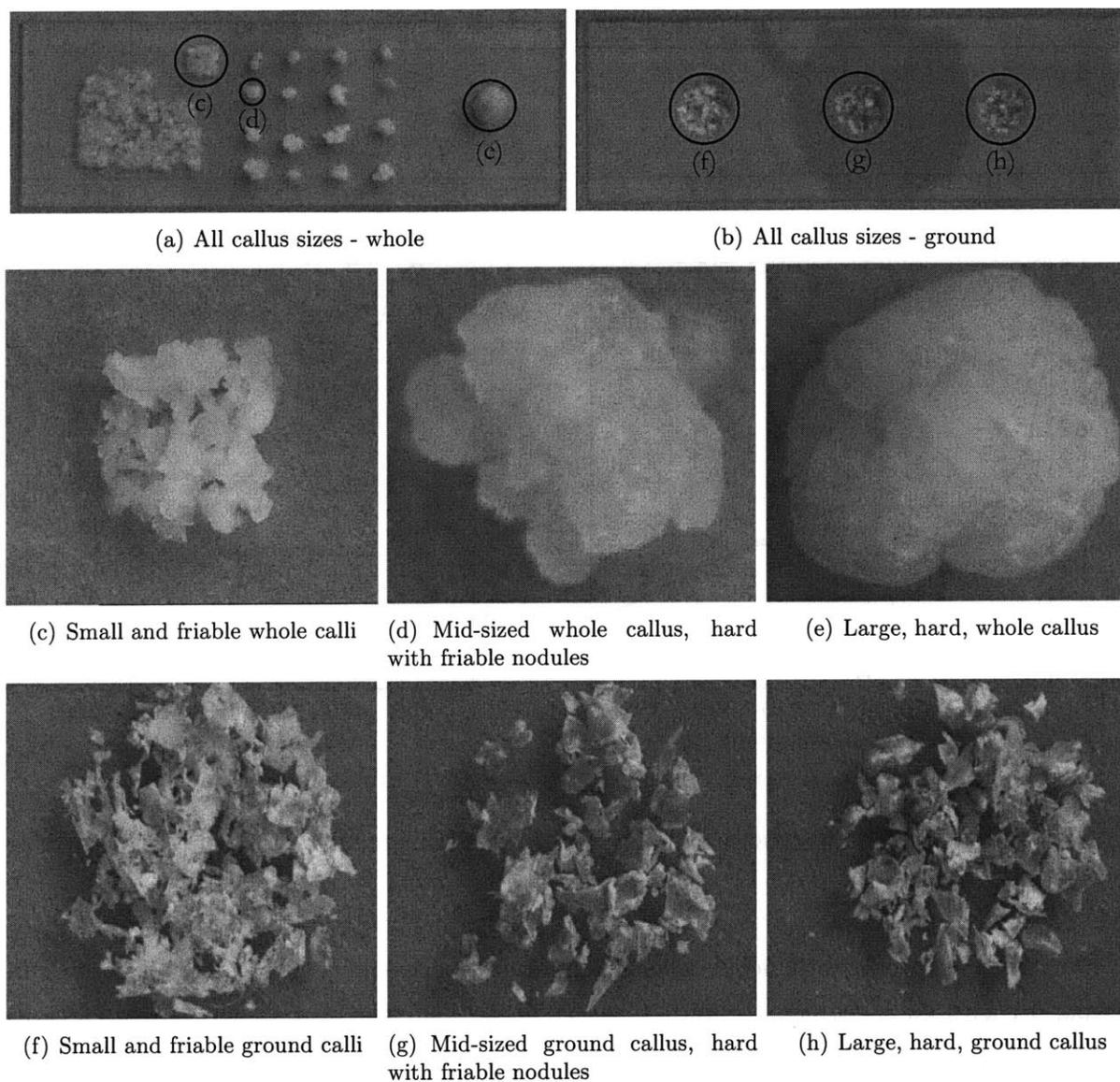
Figure 4-6: Pictures of oil palm calli. All of the whole calli are photographed together (a), and separately (c)-(e), while still wet. The calli are then dried, ground, and photographed again together (b) and apart (f)-(h).

a 20-fiber bundle of 500μm core optical fibers. The excitation was then delivered at normal incidence 1cm above the samples. The fluorescence was collected by a 20-fiber bundle-to-slit that was connected directly into the spectrometer slit. Descriptions of the spectrometer can be found in Section 2.1.

4.3.2 Fluorescence Results

(a) Fluorescence spectra

(b) Contrast ratio
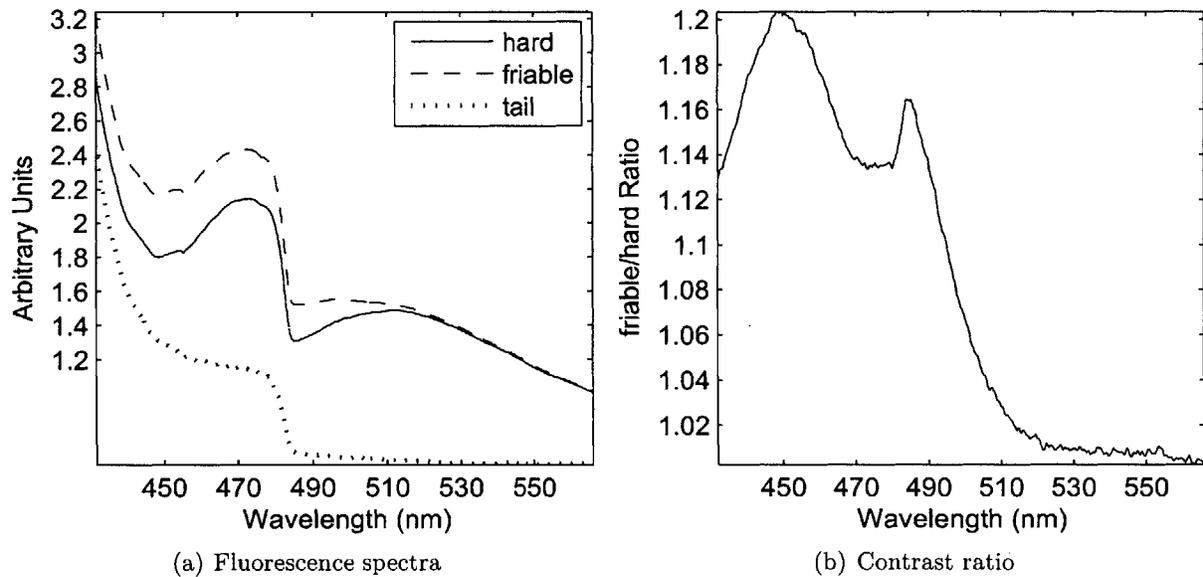
Figure 4-7: Results of a fluorescence experiment conducted on granulated oil palm calli of friable and hard phenotypes.

Fluorescence spectra were acquired using the same spectrometer and camera described in Section 2.1 with an 600lines/inch grating, blazed at 500nm. The grating was moved to three separate wavelength ranges between 400nm and 800nm, with approximate ranges of 400nm to 530nm, 530nm to 660nm, and 660nm to 800nm. Contrast was discovered in the first range, and the grating was adjusted to the 430nm to 560nm range shown in Figure 4-7 to reduce the amount of signal from the tail of the excitation light. All shown results are averages of ten frames taken with 100s integration times. Contrast between the friable and non-friable phenotypes was discovered in the range of 430nm to 560nm and is shown in Figure 4-7(a), normalized to the signal

at 560nm. The contrast region actually extends into shorter wavelengths, but the large signal from the tail of the excitation light dominates the total signal at these wavelengths, effectively masking the fluorescence. A narrow band excitation source could therefore presumably extend this contrast region. The contrast was resolved more explicitly by taking the ratio of the relative fluorescence spectra, as shown in Figure 4-7(a). The non-friable phenotype was seen to emit 20% more light in this blue-green region. This result makes sense because the non-friable phenotype appears yellow, while the friable appears white. This seems to indicate that the contrast could be due to a difference in absorption at these wavelengths, although more work is required to confirm the source of the contrast.

4.4 Summary

Raman spectroscopy has been explored for two separate clonal propagation bioprocess applications. Component analysis of the medium of an oil palm cell culture indicates that glucose, fructose, and sucrose can be differentiated, and that overall nitrate content can be observed. Offline concentration estimation of filtered samples from a batch liquid culture confirmed these findings with estimation errors in the 10mM range for fructose, glucose, and sucrose. Incomplete HPLC data prevented comparison of nitrate with HPLC. Error estimation indicated that 98.7% (3σ) of all sugar concentration estimates should fall within 3.4mM of the actual value for shot-noise limited detection. Peak analysis of the residuals shows that interference of the phosphate buffer could be a major source of error. Research into oil palm cell culture shows that these errors correspond in time with changes in the pH of the medium [61].

These results show promise for the future of Raman spectroscopy as an online bioprocess monitoring system for plant cell cultures. In order to bring the results online, current systematic errors should first be improved. Changes in the pH currently invalidate some calibration spectra, adversely affecting concentration estimates. Acquiring calibration spectra at multiple pH levels should solve this problem. Adding these calibration spectra as separate components to the pure component matrix, or implementing an algorithm that interfaces with pH measurements to

dynamically select the proper pure component spectra, are possible extensions of this solution. In addition, since the plant cell bioreactions grow at a slower rate, online measurements can use longer acquisition times to reduce noise-related errors. Finally, the slow impeller speed, large particle size, and relaxed time requirements of plant cell cultures makes a system for removing scatterers easy to envision. Filtering particles through a sterile filter or dialysis membrane into a chamber surrounding the Raman probe is a possibility, and has already been performed for a shake-flask culture of baker's yeast [9]. With these steps implemented, it should be possible to bring concentration estimation of oil palm online using Raman spectroscopy.

Experiments probing oil-palm calli via Raman spectroscopy revealed that there is too much interference from fluorescence to resolve Raman spectra for 785nm excitation. Fluorescence measurements were then conducted on calli of two separate phenotypes to determine contrast. Fluorescence spectra of oil palm calli of a hard and yellow phenotype were compared with calli of friable and white phenotype. These phenotypes were tested due to a supposed correlation between a callus's friability to its embryogenic properties. Contrast was discovered in wavelengths shorter than 510nm. The tail of the 380nm diode excitation was too strong to below 440 nm to determine where the contrast region ended. Specifically, white phenotype calli were observed to emit up to 20% more than the yellow phenotype for these wavelengths. This contrast is believed to be primarily due to differences in the re-absorption of the fluorescence, rather than the fluorescence itself, as absorption of blue and green wavelengths can give rise to an object looking yellow in color.

These experiments indicate that identification and separation of oil palm calli may be possible using optical methods. First, explicit phenotypic description of which calli are embryogenic is needed before committing any further resources. If the color and friability do not correlate to embryogenesis, then this experiment should be repeated with the proper phenotypic criteria. If the criteria used in the experiment conducted are correct, the next step is to conduct fluorescence experiments using a narrower band of near UV excitation. Another possibility is to switch to another optical method entirely and attempt a diffuse reflectance measurement of a laser line in the wavelengths of interest, such as the 488nm laser line of an Ar^+ laser. Either way, new

experiments should be done on living and undamaged calli in oil palm medium to avoid stress. If contrast can still be observed, then integration into a device to sort and separate calli is the final step to bioprocess integration. Separation and sorting of pollen from other particles using flow cytometry with fluorescence as a differentiation mechanism has already been demonstrated [67]. This technique has also been applied to mouse embryoids [68] and human islets [69], among other biological samples. It is therefore clear that understanding the effect that embryogenesis has on fluorescence in oil palm calli can lead to high throughput automation with available commercial equipment.

Chapter 5

Conclusions and Future Work

5.1 Summary and Conclusions

The work presented in this thesis has focused on three areas of relevant study for integration of optical techniques into bioprocesses. The first area of study, discussed mostly in Chapter 2, is the identification of sources of systematic error in concentration estimation of chemical constituents in a bench scale bioreactor using Raman spectroscopy. This research leads to a deeper investigation into elastic scattering, which is a subject of particular depth and interest, making it the subject of Chapter 3. The third and last topic of this thesis is the integration of optical techniques into plant cell culture bioprocesses. The ultimate goal in all cases is to present new approaches to monitoring and analysis that increase throughput and decrease costs in both research and commercial bioprocess applications.

5.1.1 Online Concentration Estimates

While Raman spectroscopy has already been investigated by multiple authors for online concentration estimation, the work has largely been proof-of-concept experiments. The work presented here is a body of experiments and theoretical analysis that outlines the practical engineering issues that need to be addressed before commercial development can take place. Solutions to some of these issues are proposed here. Universal repeatable calibration spectra can be taken reliably

with a Raman spectroscopy system without the need for lengthy and cumbersome calibration procedures due to the proposed data acquisition solutions. First, a design for a rigid calibration setup using a cartridge of wells allows for high throughput, low maintenance, high repeatability, and ease of use. Second, an algorithm has been developed that adjusts for sub-pixel fluctuations in laser wavelength and power via observation of the Rayleigh scattered light, which is included in every acquisition. To facilitate the use of Raman spectroscopy, an error model that, via simulation, will predict the limits of detection before attempting a bioreaction has been developed. This error model uses error propagation matrices to propagate noise from the calibration spectra and the measured spectra. This technique will help operators of bioprocesses decide if Raman spectroscopy is the right monitoring solution before committing the resources to integrate such a system. Finally, to increase the repeatability of *in situ* and online measurements, fused silica windows have been proposed as an alternative to sapphire windows, which have been shown to exhibit a change in relative Raman peak height after autoclaving.

Other presented experiments served the purpose of identifying stability issues. Many of these experiments showed stability over a wide range of conditions. First, the repeatability of Raman spectra over successive days was established. Second, the Raman cross-sections of pure components were found not to change with temperature in the range of interest (22° - 37°). Finally, the transfer function of the fiber Raman probe was found not to change over the same range of temperatures as well. Some instabilities were discovered and investigated briefly so as to lay the ground work for a more thorough study. It was found that tightly winding or unwinding the fiber optic cable could lead to distortions in Raman spectra proportional to the relative intensity of the peak heights. This was observed in both fused silica and polycarbonate. This change could be a substantial source of distortions between the online and calibration spectra. Distortions arising from elastic scattering were confirmed at multiple impeller speeds with two separate bioreactor mediums, water and 1M glucose solution. The distortion was found to be dependent on the speed of the impellers, indicating a dependence on the size distribution of the bubbles, which is consistent with expectations. It also appeared that, like the errors caused by the fiber optic cable, distortions in the spectra were dependent on the relative intensity of the

peak heights of the Raman spectra.

5.1.2 Wavelength Dependent Scattering

The topic of scattering was developed in further depth due to its complexity. A theoretical model based on rigorous (Mie) elastic scattering theory was developed. The wavelength dependence of the scattering was predicted and observed in the forward scattering case for an aqueous suspension of polystyrene microspheres. This experiment showed that Beer's law is effective at relating the scattering cross-section to the resultant error at high turbidity, up to $\tau \approx 10$ (OD ≈ 4). This experiment was also successfully demonstrated using the fiber optic probe, and showed an agreement of better than 0.67% of the total attenuation. To observe the distortions in Raman spectra caused by elastic scatterers, an experiment where the probe excitation light was collimated to simplify the geometry was conducted. This collimated beam was used to observe the attenuation of the Raman spectrum of water by polystyrene spheres. This experiment required the development of additional theory, which was adequate at predicting the overall attenuation of the light, but failed to capture the proper wavelength dependence. The wavelength dependence was again found to be dependent on the relative peak height of the spectra, much like the air bubble scattering experiments and the fiber optic cable stability tests. Finally, generalized Lorenz-Mie theory was used to accommodate for the difference in the geometry caused by a focused Gaussian beam. An experiment using the probe's focused beam geometry showed that the theory again failed to anticipate the wavelength dependence. It is therefore of critical importance to determine how this intensity dependent distortion could be occurring, as it is present in all of the unexplained phenomena listed in this thesis.

5.1.3 Plant Cell Solutions

A number of optical solutions were implemented for various tasks in plant cell tissue culture bioprocesses. Offline concentration estimation of sugars and nitrates in samples from an oil palm (*Elaeis guineensis*) cell culture medium were acquired. While it is possible to separate the sugars, it is not possible to do so with the nitrates. Nevertheless, the offline sugar estimates

were in good agreement with HPLC, showing a comparable error performance to concentration estimates of microbial bioreactions. Another application in these bioprocesses that could benefit from a spectroscopic solution is callus identification and differentiation. An experiment using fluorescence spectroscopy was shown to exhibit contrast between calli of two separate phenotypes: hard and yellow versus friable and white. Since not all calli are embryogenic, and the embryogenic phenotype is correlated to color and friability phenotypes, this experiment shows promise as a fast and non-invasive way to quickly identify the embryogenic status of a callus. This identification can now be combined with a high performance commercial flow cytometry solution for near total automation of a callus differentiation solution. While more phenotypic information and experiments are needed to confirm this, the current data certainly indicates that fluorescence spectroscopy is a good choice for this application.

5.2 Future Work

5.2.1 Online Raman Spectroscopy

While a number of improvements have been made, systematic errors still dominate the errors in the measurements of online and offline concentration estimates using Raman spectroscopy. The most likely reasons for this are an incomplete biological model and problems with the HPLC. A more complete understanding of the analytes produced by the $E.\ coli$ is necessary to create the appropriate calibration set. For instance, Raman has shown failure in analyzing lactate, which combines with NAD+ to make NADH and pyruvate [34]. Thus, the Raman spectra of Pyruvate, NADH, and NAD+ should be acquired and inserted into the calibration matrix to see if they exist in any appreciable quantities. Another issue is the erroneous estimates in the HPLC measurements. This can be seen with glucose. The starting HPLC estimated concentration of glucose significantly overstates the nominal starting concentration, indicating that the HPLC estimates may be a significant source of systematic error. One possible solution to this problem is to generate the HPLC calibration curves with known concentrations of glucose and other analytes in the growth medium instead of in pure water. If there is interference from

the medium components, then this calibration method should account for this problem. In general, the biological component of this system has been the least tested and should probably be examined in greater detail next for improvements in the offline data results.

In addition to the biological system, optical problema worth studying are the distortions caused by moving the fiber probe and the scattering from cells and bubbles. Experiments have shown that all three exhibit the same unexplained behavior. In all cases, a transfer function that resembles the Raman signal being attenuated is extracted from the measurements. Experiments to find the root cause of this should be conducted. The first step would be to verify that the distortion seen from scattering is related in some way to the fiber probe. An experiment using a free-space setup, i.e. a setup that does not involve an optical fiber, to do the same scattering experiments should be conducted. If the problem has something to do with the way light is coupled into the fiber, then the results of these experiments should not exhibit the intensity dependant characteristics currently observed. If the data does turn out to be free of said distortions, then a thorough investigation into the specifics of the fiber optic cable and Raman probe should be conducted. Experiments to analyze the polarization dependence, for instance, could be examined. If, doing these experiments, the problems can be isolated, then an extension of the current theorem to biologic cells is all that is needed. Instead of a monodisperse solution of spheres, *E. coli* cell suspensions are polydisperse suspensions of spheroids, requiring a change in the theory. Some authors have solved this problem in the literature in such a way as to integrate it into the Mie formulism, indicating that such a change in the theory is possible [70, 71]. In addition, the cells have a refractive index very close to the refractive index of water, allowing for a simplification of the theory using the Rayleigh-Debye-Gans or Born approximation, which sets the field inside of the sphere equal to the field outside, resulting in simper expressions [37, 36]. This is probably the last major theoretical change needed to complete the scattering formalism given here.

5.2.2 Plant Cell Cultures

The next logical step for Raman spectroscopy bioprocess monitoring for plant cell cultures is to retrofit the bioreactors to allow for online measurements. The slow time scale of plant cell cultures allows for some engineering flexibility that does not exist for microbial bioreactions. Namely, it is possible to insert the probe into a separate chamber interfaced to the rest of the bioreactor through a dialysis membrane or small mesh filter, thereby eliminating the problem of scattering from biomass and bubbles. An additional challenge of integration will be the relative size of the bioreactors used for clonal propagation. In addition, any further techniques found in reducing systematic errors in bacterial bioreactions should be implemented here as well.

The only fluorescence experiments conducted on the oil palm calli for the purpose of differentiation were destructive to the calli. The next step is to get an explicit and unambiguous description of which phenotypes are expressed by the embryogenic calli. With this knowledge, the next step for these measurements would be to replicate them on whole, wet, live calli to ensure that reliable contrast can still be achieved. Finally, these measurements should be integrated into a flow cytometry system.

Another possible application to plant cell suspension is worth an investigation. Once embryogenic calli have been identified, they must be examined for possible methylation. Plant cells under stress will add methyl groups to DNA strands in an effort to stop bacterial and viral attacks. This process, known as methylation, can cause genetically clonal calli to grow into plants that express different floral organ structure [72], which lowers throughput of clonal plant propagation. As a consequence, a noninvasive test for methylation is desired, as any invasive procedures will either destroy the callus or encourage methylation. For this end, sample calli should be collected and analyzed for an exploratory investigation into the feasibility of developing a solution for these issues. This is an application that could, if possible, markedly decrease the costs and boundaries to the genetic engineering of plants.

Appendix A

Poisson Statistics for Spectroscopy

This is a brief overview of Poisson statistics as applied to spectroscopy. To determine the probability, $P(n)$, that n electrons are counted in a pixel with an integration time of t, divide t into N time bins such that no more than one electron occupies a given time bin. There will then be n filled bins and $(N - n)$ empty bins. The probability that there will be n filled bins is, from elementary statistics, the binomial distribution,

$$P(n) = \frac{N!}{n!(N - n)!}p^n q^{N-n} \tag{A.1}$$

where p is the probability that a given bin will be occupied and:

$$q = 1 - p \sim 1 \tag{A.2}$$

To get p it can be seen by inspection that:

$$p = \frac{\mu_n}{N} \ll 1 \tag{A.3}$$

where μ_n is the mean number of occupied bins. Furthermore, for the binomial distribution, the mean of the square of the deviation from the mean (i.e., the variance) is given by [33]:

$$\mathrm{var}(n) = \overline{(\mu_n - n)^2} = Npq \tag{A.4}$$

which means that the standard deviation from the mean is given by:

$$\sigma_n = \sqrt{\text{var}(n)} = \sqrt{Npq} \tag{A.5}$$

which, given that $p \ll 1$ and $q = 1 - p$, is:

$$\sigma = \sqrt{Np(1 - p)} \approx \sqrt{Np} \approx \sqrt{N(\mu_n/N)} \tag{A.6}$$

which allows us to relate the standard deviation solely in terms of the mean:

$$\sigma = \sqrt{\mu_n} \tag{A.7}$$

In order to simulate the noise, it is desirable to use Gaussian noise, as Gaussian noise is perfectly described by its first and second moment. Using the fact that $p \ll 1$ and $n \ll N$, we can see that,

$$(1 - p)^{N-n} \approx e^{-Np} \tag{A.8}$$

and,

$$\frac{N!}{(N - n)!} \approx N^n \tag{A.9}$$

Therefore, the binomial distribution in Equation A.1, when combined with Equations A.8 and A.9, is:

$$P(n) \approx \frac{(Np)^n}{n!} e^{-Np} \tag{A.10}$$

which can be further simplified using Equation A.3 to be:

$$P(n) = \frac{(\mu_n)^n}{n!} e^{-\mu_n} \tag{A.11}$$

which is the Poisson distribution. For large n, the Poisson distribution and/or the binomial distribution can be approximated by the Gaussian distribution. The proof of this is given in [33]:

$$P(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(n-\mu_n)^2/2\sigma^2} \tag{A.12}$$

which can therefore be described using only the mean of the value, μ_n:

$$P(n) = \frac{1}{\sqrt{2\pi\mu_n}} e^{-(n-\mu_n)^2/2\mu_n} \tag{A.13}$$

Thus, when n is large, the noise is completely described by the average value, μ_n.

.

Appendix B

Error Propagation Matrix Formulation

What follows is a condensed version of the formulation given by Wikstrom and Wedin in [32, 31]. Where possible, parts have been removed for clarity. For the sake of consistency, an attempt was made to use the original author's notation.

B.1 Error Propagation Matrices

The difference between the perturbed solution $\tilde{x}$ and the unperturbed x is:

$$\Delta x = \tilde{x} - x = \tilde{A}^+ \tilde{b} - A^+ b = \underbrace{(\tilde{A}^+ - A^+)\tilde{b}}_{\Delta x_A} + \underbrace{A^+ \Delta b}_{\Delta x_b} \tag{B.1}$$

where $\tilde{x} = x + \Delta x$, $\tilde{b} = b + \Delta b$ and $\tilde{A}^+ = (A + \Delta A)^+$. Below is a general theorem that describes how Δx is affected by perturbations ΔA and Δb.

Let $A \in \mathbf{R}^{m \times n}$ be a matrix with rank p and matrix $\tilde{A} = A + \Delta A$ a perturbed matrix with the same dimension and rank. Then the perturbation can be expressed as:

$$\Delta x = \underbrace{-A^+ \Delta A \tilde{x}}_{\Delta x_{A,1}} + \underbrace{(A^T A)^+ (\Delta A)^T \tilde{r}}_{\Delta x_{A,2}} + \underbrace{P_{\mathcal{N}(A)}(\Delta A)^T \tilde{\lambda}}_{\Delta x_{A,3}} + \underbrace{A^+ \Delta b}_{\Delta x_b} \tag{B.2}$$

where the first three terms correspond to perturbations in A, the last term corresponds to

123

perturbations in b, $\tilde{x}$ is the perturbed x-value, $\tilde{r}$ is the perturbed residual, $\tilde{b} - \bar{A}\tilde{x}$, and $\tilde{\lambda}$ is the perturbed vector of Lagrange multipliers, $(\tilde{A}^+)^T\tilde{x}$. If there are perturbations in both A and b, then the following terms are involved in the perturbation analysis:

- Well-determined($m = n$) $\Delta x_{A,1}, \Delta x_b$
- Overdetermined($m > n$) $\Delta x_{A,1}, \Delta x_{A,2}, \Delta x_b$
- Underdetermined($m < n$) $\Delta x_{A,1}, \Delta x_{A,3}, \Delta x_b$

and when rank(A) $= p < min(m,n)$:

- Rank-deficient $\Delta x_{A,1}, \Delta x_{A,2}, \Delta x_{A,3}, \Delta x_b$

To see the proof to get to Equation B.2, refer to Section B.2 of this appendix. If the perturbations are small enough, a first order analysis results in the following expression:

$$dx = \underbrace{-A^+dAx}_{dx_{A,1}} + \underbrace{(A^TA)^+(dA)^Tr}_{dx_{A,2}} + \underbrace{P_{\mathcal{N}(A)}(dA)^T\lambda}_{dx_{A,3}} + \underbrace{A^+db}_{dx_b} \qquad (B.3)$$

where the d, instead of Δ, denotes small perturbations and x, r, and λ are unperturbed. Technically speaking dx, dA and db are differentials. As can be seen, the perturbation dx depends in a rather complicated way on dA. One purpose of the error propagation matrices is to rewrite this expression to gain more insight. According to Equation B.2, the perturbation dx can be divided into two parts, one corresponding to perturbations in the matrix A and another to perturbations in the vector b according to:

$$
\begin{aligned}
dx &= & dx(dA, db) & \\
&= & dx_A(dA) &+& dx_b(db) & \qquad (B.4)\\
&= & dx_A(dA) &+& A^+db &
\end{aligned}
$$

It would be desirable to be able to express $dx_A(dA)$ in the same way as $dx_b(db)$, namely as a matrix times the perturbation. This is possible by use of E^+-matrices. The E_A^+-matrix can

be divided into three parts.

$$
\begin{aligned}
dx &= dx_{A,1} + dx_{A,2} + dx_{A,3} + dx_b \\
&= E^+_{A,1}dA(:) + E^+_{A,2}dA(:) + E^+_{A,3}dA(:) + E^+_b db \\
&= E^+_A dA(:) + E^+_b db
\end{aligned}
\tag{B.5}
$$

where

$$
\begin{cases}
E^+_{A,i} & \text{error propagation matrix with dimension } n \times mn \\
dA(:) & \text{big vector with dimension } mn \times 1 \text{ consisting of } dA\text{'s columns} \\
E^+_b & \text{error propagation matrix with dimension } n \times m \\
db & \text{vector with dimension } mx1 \text{ consisting of perturbations in } b \\
E^+_A & \text{sum of error propagation matrices } E^+_{A,1} + E^+_{A,2} + E^+_{A,3}
\end{cases}
$$

By splitting the compound matrix E^+ into two separate parts, E^+_A and E^+_b, it is possible to analyze the effects caused by dA and db separately. A very practical thing is that these special matrices and their SVDs can be computed by use of the SVD of A.

B.1.1   Well Determined Systems

In the case of a well determined system,

$$
\begin{aligned}
dx &= E^+_A dA(:) + E^+_b db \\
&= E^+_{A,1}dA(:) + E^+_b db \\
&= dx_{A,1} + dx_b
\end{aligned}
\tag{B.6}
$$

where $E^+_{A,1}$ can be defined in terms of its SVD:

$$
E^+_{A,1} = V[\overset{-1}{\sum}\|x\|]W^T
\tag{B.7}
$$

where W has the columns:

$$
w_i = -(u_i \frac{x^T}{\|x\|})(:), i = 1, \ldots, p
\tag{B.8}
$$

and in the full rank case, $p = n$. To prove this, in order for the expression above to be the SVD, the following two identities have to be satisfied:

$$E_{A,1}^{+} dA(:) \equiv V[\overset{-1}{\sum} \|x\|]W^T dA(:) = -A^{+} dAx \equiv dx_{A,1} \tag{B.9}$$

$$W^T W = I \tag{B.10}$$

Start to prove that Equation B.9 is true by analyzing $W^T dA(:)$ according to

$$W^T dA(:) = \begin{bmatrix} w_1^T dA(:) \\ \vdots \\ w_p^T dA(:) \end{bmatrix} \tag{B.11}$$

Moreover we have

$$w_i^T dA(:) = trace((-u_i \frac{x^T}{\|x\|})^T dA) \tag{B.12}$$

$$= -trace(\frac{x}{\|x\|} u_i^T dA) \tag{B.13}$$

$$= -trace \underbrace{(u_i^T dA \frac{x}{\|x\|})}_{\text{scalar}} \tag{B.14}$$

$$= -u_i^T dA \frac{x}{\|x\|} \tag{B.15}$$

Hence,

$$W^T dA(:) = -U^T dA \frac{x}{\|x\|} \tag{B.16}$$

and

$$V[\Sigma^{-1}\|x\|]W^T dA(:) = -V\Sigma^{-1}U^T dAx \tag{B.17}$$

$$= -A^{+} dAx \tag{B.18}$$

The next thing to prove is the orthogonality. Take two arbitrary columns w_i and w_j and study if $w_i^T w_j = \delta_{ij}$ because orthogonality means $W^T W = I$:

$$w_i^T w_j = trace((-u_i \frac{x^T}{\|x\|})^T (-u_j \frac{x^T}{\|x\|})) \tag{B.19}$$

$$= -trace(x \underbrace{\frac{u_i^T u_j}{\|x\|^2}}_{\text{scalar}} x^T) \tag{B.20}$$

$$= \frac{u_i^T u_j}{\|x\|^2} trace(xx^T) \tag{B.21}$$

$$= \delta_{ij} \tag{B.22}$$

and the proof is complete. We can now say that $E_{A,1}^+$ can be defined in terms of its SVD, which is defined in Equations B.7 and B.8.

B.1.2 Over Determined Systems

In the case of an over determined system,

$$
\begin{aligned}
dx &= & E_A^+ dA(:) & & + & E_b^+ db \\
&= E_{A,1}^+ dA(:) & + & E_{A,2}^+ dA(:) & + & E_b^+ db \\
&= dx_{A,1} & + & dx_{A,2} & + & dx_b
\end{aligned}
\tag{B.23}
$$

There is no contribution from the $dx_{A,3}$ term. The SVD of $E_{A,1}^+$ is described in Section B.1.1 and $E_{A,2}^+$ can be defined in terms of its SVD:

$$E_{A,2}^+ = V[\Sigma^{-2} \|r\|] Y^T \tag{B.24}$$

where Y has the columns

$$y_i = (\frac{r}{\|r\|} v_i^T)(:), i = 1, \ldots, p \tag{B.25}$$

and in the full rank case, $p = n$. To prove this, for this to be the wanted SVD, the following two identities have to be satisfied:

$$E_{A,2}^+ dA(:) \equiv V[\overset{-2}{\sum} \|x\|] W Y^T dA(:) = (A^T A)^+ (dA)^T r \equiv dx_{A,2} \tag{B.26}$$

$$Y^T Y = I \tag{B.27}$$

Start to prove that Equation B.26 is true by analyzing $Y^T dA(:)$ according to:

$$Y^T dA(:) = \begin{bmatrix} y_1^T dA(:) \\ \vdots \\ y_p^T dA(:) \end{bmatrix} \tag{B.28}$$

and moreover we have

$$y_i^T dA(:) = trace((\frac{r}{\|r\|} v_i^T)^T dA) \tag{B.29}$$

$$= trace(v_i \frac{r^T}{\|r\|} dA) \tag{B.30}$$

$$= trace(\underbrace{\frac{r^T}{\|r\|} dA v_i}_{scalar}) \tag{B.31}$$

$$= \frac{r^T}{\|r\|} dA v_i \tag{B.32}$$

$$= v_i^T dA^T \frac{r}{\|r\|} \tag{B.33}$$

Hence,

$$Y^T dA(:) = V^T (dA)^T \frac{r}{\|r\|} \tag{B.34}$$

and

$$V[\Sigma^- 2\|r\|] Y^T dA(:) = V\Sigma^- 2 V^T (dA)^T r \tag{B.35}$$

$$= (A^T A)^+ (dA)^T r \tag{B.36}$$

Next thing to prove is the orthogonality. Take two arbitrary columns y_i and y_j and study if $y_i^T y_j = \delta_{ij}$ because orthogonality means $Y^T Y = I$:

$$y_i^T y_j = trace((\frac{r}{\|r\|}v_i^T)^T(\frac{r}{\|r\|}v_j^T)) \tag{B.37}$$

$$= trace(v_i \underbrace{\frac{r^T r}{\|r\|^2}}_{\text{scalar}} v_j^T) \tag{B.38}$$

$$= \frac{r^T r}{\|r\|^2} trace(v_i v_j^T) \tag{B.39}$$

$$= \delta_{ij} \tag{B.40}$$

and the proof is complete. We can now say that $E_{A,2}^+$ can be defined in terms of its SVD, which is defined in Equations B.24 and B.25.

B.1.3 Under Determined Systems

In the case of an under determined system,

$$dx = E_A^+ dA(:) + E_b^+ db$$
$$= E_{A,1}^+ dA(:) + E_{A,3}^+ dA(:) + E_b^+ db \tag{B.41}$$
$$= dx_{A,1} + dx_{A,3} + dx_b$$

There is no contribution from the $dx_{A,2}$ term. The SVD of $E_{A,1}^+$ is described in Section B.1.1, and $E_{A,3}^+$ can be defined in terms of its SVD:

$$E_{A,3}^+ = V_\perp [I_{n-p} \| \lambda \|] Z^T \tag{B.42}$$

where Z has the columns

$$z_i = (\frac{\lambda}{\|\lambda\|} v_{\perp,i}^T)(:), p = 1, \ldots, n \tag{B.43}$$

and in the full rank case $p = m$. To prove this, for this to be the wanted SVD the following two identities have to be satisfied:

$$E_{A,3}^{+}dA(:) \equiv V\perp[I_{n-p}\|\lambda\|]Z^{T}dA(:)P_{\mathcal{N}(A)}(dA)^{T}\lambda \equiv dx_{A,3} \tag{B.44}$$

$$Z^{T}Z = I \tag{B.45}$$

Start to prove that Equation B.44 is true by analyzing $Z^{T}dA(:)$ according to:

$$Z^{T}dA(:) = \begin{bmatrix} z_1^T dA(:) \\ \vdots \\ z_p^T dA(:) \end{bmatrix} \tag{B.46}$$

Moreover we have

$$\begin{aligned}
z_i^T dA(:) &= trace\left(\left(\frac{\lambda}{\|\lambda\|}(v_{\perp,i}^T)^T\right)^T dA\right) \tag{B.47} \\
&= trace\left(v_{\perp,i}\frac{\lambda^T}{\|\lambda\|}dA\right) \tag{B.48} \\
&= trace\underbrace{\left(\frac{\lambda^T}{\|\lambda\|}dAv_{\perp,i}\right)}_{\text{scalar}} \tag{B.49} \\
&= \frac{\lambda^T}{\|\lambda\|}dAv_i^{\perp} \tag{B.50} \\
&= (v_{\perp,i})^T dA^T \frac{\lambda}{\|\lambda\|} \tag{B.51}
\end{aligned}$$

Hence,

$$Z^{T}dA(:) = V_{\perp}^{T}(dA)^{T}\frac{\lambda}{\|\lambda\|} \tag{B.52}$$

and

$$V_{\perp}[I_{n-p}\|\lambda\|]Z^{T}dA(:) = V_{\perp}V_{\perp}^{T}(dA)^{T}\lambda \tag{B.53}$$

$$= P_{\mathcal{N}(A)}(dA)^T\lambda \qquad (B.54)$$

By definition

$$V_\perp^T V_\perp = I_{n-p} \qquad \text{and} \qquad V_\perp V = 0 \qquad (B.55)$$

Finally it will be proven that Z also is orthogonal. Take two arbitrary columns z_i and z_j and study if $z_i^T z_j = \delta_{ij}$ because orthogonality means $Z^T Z = I$:

$$z_i^T z_j = trace((\frac{\lambda}{\|\lambda\|}(v_{\perp,i})^T)^T(\frac{\lambda}{\|\lambda\|}(v_{\perp,j})^T)) \qquad (B.56)$$

$$= trace(v_{\perp,i}\underbrace{\frac{\lambda^T\lambda}{\|\lambda\|^2}}_{\text{scalar}}(v_{\top,j})^T) \qquad (B.57)$$

$$= \frac{\lambda^T\lambda}{\|\lambda\|^2}trace(v_{\perp,i}(v_{\perp,j})^T) \qquad (B.58)$$

$$= \frac{\lambda^T\lambda}{\|\lambda\|^2}trace((v_{\perp,i})^T v_{\perp,j}) \qquad (B.59)$$

$$= \delta_{ij} \qquad (B.60)$$

and the proof is complete. We can now say that $E_{A,3}^+$ can be defined in terms of its SVD, which is defined in Equations B.42 and B.43.

B.2 Proof of Equation B.2

By definition we have

$$\Delta x = \tilde{x} - x = \tilde{A}^+\tilde{b} - A^+b = (\tilde{A}^+ - A^+)\tilde{b} + \underbrace{A^+\Delta b}_{\Delta x_b} \qquad (B.61)$$

so the term $(\tilde{A}^+ - A^+)$ is of great importance and has been extensively analyzed. Below is an extended version of the proof.

The main idea of the proof is based on the fact that the matrix A and $\tilde{A}$ are linear transformations from $\mathbf{R}^n$ into $\mathbf{R}^m$. The vector space $\mathbf{R}^n$ is spanned by $\mathcal{R}(A^T)$ and $\mathcal{N}(A)$ and $\mathbf{R}^m$ by $\mathcal{R}(A)$ and $\mathcal{N}(A^T)$. In this case we choose to span $\mathbf{R}^n$ by I_n and $\mathbf{R}^m$ by I_m. Another very crucial step in the proof is the choice of $I_n = P_{\mathcal{R}(A^T)} + P_{\mathcal{N}(A)}$ and $I_m = P_{\mathcal{R}(\tilde{A})} + P_{\mathcal{N}(\tilde{A}^T)}$ instead of $I_n = P_{\mathcal{R}(\tilde{A}^T)} + P_{\mathcal{N}(\tilde{A})}$ and $I_m = P_{\mathcal{R}(A)} + P_{\mathcal{N}(A^T)}$. Since our aim is to be able to reuse the SVD for A the former choice is better because otherwise there is a need for the SVD for $\tilde{A}$ instead.

$$
\begin{aligned}
\tilde{A}^+ - A^+ &= I_n \qquad\qquad (\tilde{A}^+ - A^+) \qquad\qquad I_m \\
&= (P_{\mathcal{R}(A^T)} + P_{\mathcal{N}(A)}) \quad (\tilde{A}^+ - A^+) \quad (P_{\mathcal{R}(\tilde{A})} + P_{\mathcal{N}(\tilde{A}^T)}) \\
&= P_{\mathcal{R}(A^T)}(\tilde{A}^+ - A^+)P_{\mathcal{R}(\tilde{A})} \quad + \quad P_{\mathcal{R}(A^T)}(\tilde{A}^+ - A^+)P_{\mathcal{N}(\tilde{A}^T)} \\
&\quad\; P_{\mathcal{N}(A)}(\tilde{A}^+ - A^+)P_{\mathcal{R}(\tilde{A})} \quad + \quad P_{\mathcal{N}(A)}(\tilde{A}^+ - A^+)P_{\mathcal{N}(\tilde{A}^T)}
\end{aligned}
$$

To simplify the expressions, the properties of the pseudoinverse is used and the fact that

$$
\mathcal{R}(A^+) = \mathcal{R}(A^T) \quad \mathcal{R}(A^T) = \mathcal{N}(A)^\perp \quad \mathcal{R}(A) = \mathcal{N}(A^T)^\perp
$$

resulting in $A^+ P_{\mathcal{N}(A^T)} = 0$ and $P_{\mathcal{N}(A)} A^+ = 0$.

$$
P_{\mathcal{R}(A^T)}(\tilde{A}^+ - A^+)P_{\mathcal{R}(\tilde{A})} = P_{\mathcal{R}(A^T)}\underbrace{\tilde{A}^+ P_{\mathcal{R}(\tilde{A})}}_{\tilde{A}^+} - \underbrace{P_{\mathcal{R}(A^T)} A^+}_{A^+} P_{\mathcal{R}(\tilde{A})} \tag{B.62}
$$

$$
P_{\mathcal{R}(A^T)}(\tilde{A}^+ - A^+)P_{\mathcal{N}(\tilde{A}^T)} = P_{\mathcal{R}(A^T)}\underbrace{\tilde{A}^+ P_{\mathcal{N}(\tilde{A}^T)}}_{0} - \underbrace{P_{\mathcal{R}(A^T)} A^+}_{A^+} P_{\mathcal{N}(\tilde{A}^T)} \tag{B.63}
$$

$$
P_{\mathcal{N}(A)}(\tilde{A}^+ - A^+)P_{\mathcal{R}(\tilde{A})} = P_{\mathcal{N}(A)}\underbrace{\tilde{A}^+ P_{\mathcal{R}(\tilde{A})}}_{tildeA^+} - \underbrace{P_{\mathcal{N}(A)} A^+}_{0} P_{\mathcal{R}(\tilde{A})} \tag{B.64}
$$

$$
P_{\mathcal{N}(A)}(\tilde{A}^+ - A^+)P_{\mathcal{N}(\tilde{A}^T)} = P_{\mathcal{N}(A)}\underbrace{\tilde{A}^+ P_{\mathcal{N}(\tilde{A}^T)}}_{0} - \underbrace{P_{\mathcal{N}(A)} A^+}_{0} P_{\mathcal{N}(\tilde{A}^T)} \tag{B.65}
$$

Summing up we then have

$$\tilde{A}^+ - A^+ = \underbrace{P_{\mathcal{R}(A^T)}\tilde{A}^+ - A^+ P_{\mathcal{R}(\tilde{A})}}_{T_1} + \underbrace{-A^+ P_{\mathcal{N}(\tilde{A}^T)}}_{T_2} + \underbrace{P_{\mathcal{N}(A)}\tilde{A}^+}_{T_3} \tag{B.66}$$

Continue by analyzing $\Delta x_{A,i} = T_i \tilde{b}$ for $i = 1, 2, 3$ in more detail

$$
\begin{aligned}
\Delta x_{A,1} &= P_{\mathcal{R}(A^T)}\underbrace{\tilde{A}^+\tilde{b}}_{\tilde{x}} - A^+ \underbrace{P_{\mathcal{R}(\tilde{A})}\tilde{b}}_{\tilde{A}\tilde{x}} = (A^+ A - A^+\tilde{A})\tilde{x} = -A^+ \Delta A \tilde{x} & \text{(B.67)}\\
\Delta x_{A,2} &= -A^+ \underbrace{P_{\mathcal{N}(\tilde{A}^T)}\tilde{b}}_{\tilde{r}} = -A^+ A A^+ \tilde{r} = -A^+(AA^+)^T \tilde{r}\\
&= -A^+(A^+)^T A^T \tilde{r} = -(A^T A)^+ A^T \tilde{r} = \{A^T = \tilde{A}^T - (\Delta A)^T\}\\
&= -(A^T A)^+ \underbrace{\tilde{A}^T \tilde{r}}_{0} + (A^T A)^+ (\Delta A)^T \tilde{r} = (A^T A)^+ (\Delta A)^T \tilde{r} & \text{(B.68)}\\
\Delta x_{A,3} &= P_{\mathcal{N}(A)}\tilde{A}^+ \tilde{b} = P_{\mathcal{N}(A)} \underbrace{\tilde{A}^+ \tilde{A}}_{(\tilde{A}^+\tilde{A})^T} \tilde{A}^+ \tilde{b} = P_{\mathcal{N}(A)}\tilde{A}^T \underbrace{(\tilde{A}^+)^T \tilde{A}^+ \tilde{b}}_{\tilde{\lambda}}\\
&= P_{\mathcal{N}(A)}\tilde{A}^T \tilde{\lambda} = \{A^T = \tilde{A}^T - (\Delta A)^T\}\\
&= \underbrace{P_{\mathcal{N}(A)}A^T \tilde{\lambda}}_{0} + P_{\mathcal{N}(A)}(\Delta A)^T \tilde{\lambda} = P_{\mathcal{N}(A)}(\Delta A)^T \tilde{\lambda} & \text{(B.69)}
\end{aligned}
$$

Observe that λ is a vector of Lagrange multipliers because the pseudoinverse solution to the underdetermined linear equation system $Ax = b$ can be formulated as

$$
\begin{aligned}
&min_x \quad \|x\|_2^2\\
&st \quad\quad Ax = b
\end{aligned}
$$

with the corresponding Lagrangian function

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^T x + \lambda^T(b - Ax) \tag{B.70}$$

and the first-order necessary conditions are

$$\frac{\partial \mathcal{L}}{\partial x} = x - A^T \lambda = 0 \rightarrow \lambda = (A^T)^+ x = (A^+)^T x$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = b - Ax = 0 \rightarrow x = A^+ b$$

Then it is proved that Δx equals

$$\underbrace{-A^+ \Delta A \tilde{x}}_{\Delta x_{A,1}} + \underbrace{(A^T A)^+ (\Delta A)^T \tilde{r}}_{\Delta x_{A,2}} + \underbrace{P_{\mathcal{N}(A)} (\Delta A)^T \tilde{\lambda}}_{\Delta x_{A,3}} + \underbrace{A^+ \Delta b}_{\Delta x_b} \qquad (B.71)$$

Appendix C

Rigorous (Mie) Scattering Derivations and Methods

C.1 Spherical Wave Equation Solution

What follows is a compressed version of the derivation of the Mie solution as synthesized from the literature [37, 36]. The format of the derivation is most similar to the treatment in Tsang et. al. and that book is recommended for more information [37].

In spherical coordinates, the scalar wave equation is

$$\left[\frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \frac{\partial}{\partial r}) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \frac{\partial}{\partial \theta}) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} + k^2 \right] \psi(\bar{r}) = 0 \tag{C.1}$$

It has the outgoing wave solution

$$\psi_{mn}(kr, \theta, \phi) = h_n^{(1)}(kr) P_n^m(\cos \theta) e^{im\phi} \tag{C.2}$$

with $n = 0, 1, 2, \ldots, m = 0, \pm 1, \ldots, \pm n$. In C.2, $h_n^{(1)}$ is a spherical Hankel function of the first kind and P_n^m is the associated Legendre function. We also define regular wave function by

$$R_{g\psi_{mn}}(kr, \theta, \phi) = j_n(kr) P_n^m(\cos \theta) e^{im\phi} \tag{C.3}$$

where R_g stands for regular and j_n is the spherical Bessel function. The prefix *regular* is used to denote the fact that $h_n^{(1)}$ is replaced by j_n which is finite at $r = 0$. The associated Legendre function is defined as

$$P_n^m(x) = \frac{(-1)^m}{2^n n!}(1 - x^2)^{m/2}\frac{d^{n+m}}{dx^{n+m}}(x^2 - 1)^n \tag{C.4}$$

for $m = 0, \pm 1, \ldots, \pm n$; that is expression C.4 holds for both positive and negative values of m. The relation between positive and negative values of m is

$$P_n^{-m}(x) = (-1)^m\frac{(n - m)!}{(n + m)!}P_n^m(x) \tag{C.5}$$

The spherical harmonic is defined as

$$Y_n^m(\theta, \phi) = P_n^m(\cos\theta)e^{im\phi} \tag{C.6}$$

with orthogonality relation

$$\int_0^\pi d\theta \sin\theta \int_0^{2\pi} d\phi Y_n^m(\theta, \phi)Y_{n'}^{-m'}(\theta, \phi) = (-1)^m\frac{4\pi}{2n + 1}\delta_{mm'}\delta_{nn'} \tag{C.7}$$

The completeness relation is

$$\sum_{l=0}^\infty \sum_{m=-l}^l (-1)^m\frac{2l + 1}{4\pi}Y_l^{-m}(\theta', \phi')Y_l^m(\theta, \phi) = \delta(\phi - \phi')\delta(\cos\theta - \cos\theta') \tag{C.8}$$

The expansion of plane waves in spherical waves is given by

$$e^{i\bar{k}\cdot\bar{r}} = \sum_{l=0}^\infty \sum_{m=-l}^l (-1)^m i^l (2l + 1)j_l(kr)Y_l^m(\theta, \phi)Y_l^{-m}(\theta_k, \phi_k) \tag{C.9}$$

where (θ_k, ϕ_k) denote the angular variables describing the direction of $\bar{k}$. The vector spherical harmonics are the three vector functions $\bar{V}_{mn}^{(\alpha)}(\theta, \phi)$, $\alpha = 1, 2$, and 3. They are defined by:

$$\bar{V}_{mn}^{(1)}(\theta,\phi) = \bar{P}_{mn}(\theta,\phi) = \hat{r}Y_n^m(\theta,\phi) \qquad (n=0,1,2,\ldots) \qquad \text{(C.10)}$$

$$\bar{V}_{mn}^{(2)}(\theta,\phi) = \bar{B}_{mn}(\theta,\phi) = r\nabla[Y_n^m(\theta,\phi)] \qquad (n=1,2,3,\ldots)$$

$$= \left[\hat{\theta}\frac{d}{d\theta}P_n^m(\cos\theta) = \hat{\phi}\frac{im}{\sin\theta}P_n^m(\cos\theta)\right]e^{im\phi}$$

$$= \hat{r}\times\bar{C}_{mn}(\theta,\phi) \qquad \text{(C.11)}$$

$$\bar{V}_{mn}^{(3)}(\theta,\phi) = \bar{C}_{mn}(\theta,\phi) = \nabla\times[\bar{r}Y_n^m(\theta,\phi)] \qquad (n=1,2,3,\ldots)$$

$$= \left[\hat{\theta}\frac{im}{\sin\theta}P_n^m(\cos\theta) - \hat{\phi}\frac{d}{d\theta}P_n^m(\cos\theta)\right]e^{im\phi} \qquad \text{(C.12)}$$

Note that for $\bar{P}_{mn}$, n starts with 0; while for $\bar{B}_{mn}$ and $\bar{C}_{mn}$, n starts at 1. The orthogonality relation for vector spherical harmonics is

$$\int_0^\pi d\theta\sin\theta\int_0^{2\pi}d\phi\bar{V}_{mn}^{(\alpha)}(\theta,\phi)\cdot\bar{V}_{-m'n'}^{(\beta)}(\theta,\phi) = \delta_{\alpha\beta}\delta_{mm'}\delta_{nn'}z_{\alpha mn} \qquad \text{(C.13)}$$

$$z_{1mn} = (-1)^m\frac{4\pi}{2n+1} \qquad \text{(C.14)}$$

$$z_{2mn} = z_{3mn} = (-1)^m\frac{4\pi n(n+1)}{2n+1} \qquad \text{(C.15)}$$

The three regular vector spherical waves $R_g\bar{L}_{mn}$, $R_g\bar{M}_{mn}$, and $R_g\bar{N}_{mn}$ are defined as

$$R_g\bar{L}_{mn}(kr,\theta,\phi) = \frac{\gamma'_{mn}}{k}\nabla[R_{g\psi_{mn}}(kr,\theta,\phi)]$$

$$= \gamma'_{mn}\left[j_n'(kr)\bar{P}_{mn}(\theta,\phi) + \frac{j_n(kr)}{kr}\bar{B}_{mn}(\theta,\phi)\right] \qquad \text{(C.16)}$$

$$R_g\bar{M}_{mn}(kr,\theta,\phi) = \gamma_{mn}\nabla\times[\bar{r}R_{g\psi_{mn}}(kr,\theta,\phi)]$$

$$= \gamma_{mn}j_n(kr)\bar{C}_{mn}(\theta,\phi) \qquad \text{(C.17)}$$

$$R_g\bar{N}_{mn}(kr,\theta,\phi) = \frac{1}{k}\nabla\times R_g\bar{M}_{mn}(kr,\theta,\phi)$$

$$= \gamma_{mn} \left\{ \frac{n(n+1)j_n(kr)}{kr} \bar{P}_{mn}(\theta, \phi) + \frac{[krj_n(kr)]'}{kr} \bar{B}_{mn}(\theta, \phi) \right\} \quad \text{(C.18)}$$

In Equations C.16 through C.18:

$$\gamma'_{mn} = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} \quad \text{(C.19)}$$

$$\gamma_{mn} = \sqrt{\frac{(2n+1)(n-m)!}{4\pi n(n+1)(n+m)!}} \quad \text{(C.20)}$$

The vector spherical waves without the prefix R_g are the expressions in Equations C.16 through C.18 with j_n replace by h_n. We note that for $\bar{L}_{mn}, n = 0, 1, 2, 3, \ldots$ whereas for $\bar{M}_{mn}$ and $\bar{N}_{mn}, n = 1, 2, 3, \ldots$. The $R_g\bar{L}$ functions satisfy the equation:

$$\nabla(\nabla \cdot R_g\bar{L}_{mn}) + k^2\bar{L}_{mn} = 0 \quad \text{(C.21)}$$

whereas the $R_g\bar{M}$ and $R_g\bar{N}$ functions satisfy the vector wave equation

$$\nabla \times \nabla \times R_g\bar{M}_{mn} - k^2\bar{M}_{mn} = 0 \quad \text{(C.22)}$$

$$\nabla \times \nabla \times R_g\bar{N}_{mn} - k^2\bar{N}_{mn} = 0 \quad \text{(C.23)}$$

The vector spherical wave functions can be expressed as integral representations of vector spherical harmonics

$$R_g\bar{L}_{mn}(kr, \theta, \phi) = \frac{(-i)^{n-1}}{4\pi}\gamma'_{mn} \int_{4\pi} d\Omega' e^{ikr\hat{r}\cdot\hat{r}'} \bar{P}_{mn}(\theta', \phi') \quad \text{(C.24)}$$

$$R_g\bar{M}_{mn}(kr, \theta, \phi) = \frac{(-i)^n}{4\pi}\gamma_{mn} \int_{4\pi} d\Omega' e^{ikr\hat{r}\cdot\hat{r}'} \bar{C}_{mn}(\theta', \phi') \quad \text{(C.25)}$$

$$R_g\bar{N}_{mn}(kr, \theta, \phi) = \frac{(-i)^{n-1}}{4\pi}\gamma'_{mn} \int_{4\pi} d\Omega' e^{ikr\hat{r}\cdot\hat{r}'} \bar{B}_{mn}(\theta', \phi') \quad \text{(C.26)}$$

With the aid of C.24 through C.26, it follows that

$$\bar{I}e^{i\bar{p}\cdot\bar{r}} = \sum_{n,m}(-1)^m\frac{(2n+1)}{n(n+1)}i^n\left[-in(n-1)\frac{\bar{P}_{-mn}(\theta_p,\phi_p)}{\gamma'_{mn}}R_g\bar{L}_{mn}(pr,\theta,\phi)\right.$$

$$\left.+\frac{\bar{C}_{-mn}(\theta_p,\phi_p)}{\gamma_{mn}}R_g\bar{L}_{mn}(pr,\theta,\phi)-\frac{i\bar{B}_{-mn}(\theta_p,\phi_p)}{\gamma_{mn}}R_g\bar{L}_{mn}(pr,\theta,\phi)\right] \qquad \text{(C.27)}$$

where (θ_p,ϕ_p) are the angular variables that indicate the direction $\hat{p}$. A plane electromagnetic wave can be expressed in terms of spherical waves by taking the dot product of Equation C.27 with the incident electric field. Let the propagation direction of the incident wave be $\hat{k}_i = \sin\theta_i\cos\phi_i\hat{x} + \sin\theta_i\sin\phi_i\hat{y} + \cos\theta_i\hat{x}$, and $\hat{v}_i = \hat{\theta}_i$, $\hat{h}_i = \hat{\phi}_i$.

$$\bar{E}_i = \left(E_{vi}\hat{v}_i + E_{hi}\hat{h}_i\right)e^{i\bar{k}_i\cdot\bar{r}} \qquad \text{(C.28)}$$

$$= \sum_{n,m}(-1)^m\frac{1}{\gamma_{mn}}\frac{(2n+1)}{n(n+1)}i^n\left\{\left[E_{vi}[\hat{\theta}_i\cdot\bar{C}_{-mn}(\theta_i,\phi_i)]\right.\right.$$

$$\left.+E_{hi}[\hat{\phi}_i\cdot\bar{C}_{-mn}(\theta_i,\phi_i)]/\right]R_g\bar{M}_{mn}(kr,\theta,\phi)$$

$$+\left[E_{vi}[\hat{\theta}_i\cdot(-i\bar{B}_{-mn}(\theta_i,\phi_i))]\right.$$

$$\left.\left.+E_{hi}[\hat{\phi}_i\cdot(-i\bar{B}_{-mn}(\theta_i,\phi_i))]\right]R_g\bar{N}_{mn}(kr,\theta,\phi)\right\} \qquad \text{(C.29)}$$

To obtain a plane electromagnetic wave propagating in the $\hat{z}$ direction, we set $\theta_p = \phi_p = 0$ in C.27 and take a dot product with the polarization vector $\hat{e}_i$.

We note that

$$\bar{B}_{-mn}(\theta=0,\phi) = \bar{C}_{-mn}(\theta=0,\phi) = 0$$

unless $m = \pm 1$. Hence, from C.27

$$\hat{e}_ie^{ikz} = -\sum_{n=1}^{\infty}\sum_{m=-1.1}\frac{(2n+1)}{n(n+1)}i^n\left[\frac{\hat{e}_i\cdot\bar{C}_{-mn}(0,0)}{\gamma_{mn}}R_g\bar{M}_{mn}(kr,\theta,\phi)\right.$$

$$\left.-i\frac{\hat{e}_i\cdot\bar{B}_{-mn}(0,0)}{\gamma_{mn}}R_g\bar{N}_{mn}(kr,\theta,\phi)\right] \qquad \text{(C.30)}$$

The far-field solutions of the outgoing vector spherical waves are

$$\bar{M}_{mn}(kr, \theta, \phi) \simeq \gamma_{mn} i^{-n-1} \frac{e^{ikr}}{kr} \bar{C}_{mn}(\theta, \phi) \tag{C.31}$$

$$\bar{N}_{mn}(kr, \theta, \phi) \simeq \gamma_{mn} i^{-n} \frac{e^{ikr}}{kr} \bar{B}_{mn}(\theta, \phi) \tag{C.32}$$

In Mie scattering, the results are often expressed in functions denoted by $\pi_n(\cos\theta)$ and $\tau_n(\cos\theta)$. They are related to the associated Legendre functions by

$$\pi_n(\cos\theta) = -\frac{P_n^1(\cos\theta)}{\sin\theta} \tag{C.33}$$

$$\tau_n(\cos\theta) = -\frac{dP_n^1(\cos\theta)}{d\theta} \tag{C.34}$$

The special values when $\cos\theta = \pm 1$ are

$$\pi_n(1) = \tau_n(1) = \frac{n(n+1)}{2} \tag{C.35}$$

$$\pi_n(-1) = -\tau_n(-1) = (-1)^{n+1}\frac{n(n+1)}{2} \tag{C.36}$$

Then we can express the vector spherical harmonics for $m + 1$ as

$$\bar{B}_{ln}(\theta, \phi) = -\left[\hat{\theta}\tau_n(\cos\theta) + \hat{\phi}i\pi_n(\cos\theta)\right]e^{i\phi} \tag{C.37}$$

$$\bar{C}_{ln}(\theta, \phi) = -\left[\hat{\theta}i\pi_n(\cos\theta) - \hat{\phi}\tau_n(\cos\theta)\right]e^{i\phi} \tag{C.38}$$

and for $m = -1$

$$\bar{B}_{-ln}(\theta, \phi) = \frac{1}{n(n+1)}\left[\hat{\theta}\tau_n(\cos\theta) - \hat{\phi}i\pi_n(\cos\theta)\right]e^{-i\phi} \tag{C.39}$$

$$\bar{C}_{-ln}(\theta, \phi) = -\frac{1}{n(n+1)} \left[\hat{\theta} i \pi_n(\cos\theta) + \hat{\phi} \tau_n(\cos\theta) \right] e^{-i\phi} \tag{C.40}$$

For $\theta = 0$, we can set $\phi = 0$ without loss of generality to obtain

$$\bar{B}_{ln}(0,0) = -\frac{n(n+1)}{2} \left(\hat{\theta} + i\hat{\phi} \right) e^{i\phi} = -\frac{n(n+1)}{2}(\hat{x} + i\hat{y}) \tag{C.41}$$

$$\bar{C}_{ln}(0,0) = -\frac{n(n+1)}{2} \left(i\hat{\theta} - \hat{\phi} \right) e^{i\phi} = -i\frac{n(n+1)}{2}(\hat{x} + i\hat{y}) \tag{C.42}$$

$$\bar{B}_{-ln}(0,0) = \frac{1}{2}(\hat{x} - i\hat{y}) \tag{C.43}$$

$$\bar{B}_{-ln}(0,0) = \frac{-i}{2}(\hat{x} - i\hat{y}) \tag{C.44}$$

$$\tag{C.45}$$

Mie scattering is scattering of electromagnetic waves by a sphere of radius a and permittivity ϵ_p. Let $k_p = \omega\sqrt{\mu\epsilon_p}$ and the sphere be centered at the origin. Because of the symmetry of a sphere, it is convenient to use the scattering plane orthonormal system to express the scattering amplitudes. To get

$$\begin{bmatrix} f_{11}(\Theta) & f_{12}(\Theta) \\ f_{21}(\Theta) & f_{22}(\Theta) \end{bmatrix}$$

we let the incident wave be along the $\hat{z}$ direction, so that $\hat{k}_i = \hat{z}$. Also let the observation direction $\hat{k}_s$ be in the $y - z$ plane, with $\phi = 90°$; then

$$\Theta = \theta \tag{C.46}$$

$$\hat{l}_i = \hat{l}_s = \hat{x} = -\hat{\phi} \tag{C.47}$$

$$\hat{2}_i = \hat{y} \tag{C.48}$$

$$\hat{2}_s = \hat{k}_s \times \hat{l}_s = \hat{\theta} \tag{C.49}$$

The incident field is given by, from C.30,

$$\bar{E}_i = \hat{e}_i e^{ikz} \quad = \quad -\sum_{n+1}^{\infty} \sum_{m=-1,1} i^n \frac{(2n+1)}{n(n+1)} \left[\frac{\hat{e}_i \cdot \bar{C}_{-mn}(0,0)}{\gamma_{mn}} R_g \bar{M}_{mn}(kr, \theta, \phi) \right.$$
$$\left. -i \frac{\hat{e}_i \cdot \bar{B}_{-mn}(0,0)}{\gamma_{mn}} R_g \bar{N}_{mn}(kr, \theta, \phi) \right] \tag{C.50}$$

To solve the boundary value problem, we let the scattered field be

$$\bar{E}_s \quad = \quad -\sum_{n+1}^{\infty} \sum_{m=-1,1} i^n \frac{(2n+1)}{n(n+1)} \left[-\frac{\hat{e}_i \cdot \bar{C}_{-mn}(0,0)}{\gamma_{mn}} b_n R_g \bar{M}_{mn}(kr, \theta, \phi) \right.$$
$$\left. +i \frac{\hat{e}_i \cdot \bar{B}_{-mn}(0,0)}{\gamma_{mn}} a_n R_g \bar{N}_{mn}(kr, \theta, \phi) \right] \tag{C.51}$$

and let the internal field be

$$\bar{E}_{int} \quad = \quad -\sum_{n+1}^{\infty} \sum_{m=-1,1} i^n \frac{(2n+1)}{n(n+1)} \left[\frac{\hat{e}_i \cdot \bar{C}_{-mn}(0,0)}{\gamma_{mn}} d_n R_g \bar{M}_{mn}(k_p r, \theta, \phi) \right.$$
$$\left. -i \frac{\hat{e}_i \cdot \bar{B}_{-mn}(0,0)}{\gamma_{mn}} c_n R_g \bar{N}_{mn}(k_p r, \theta, \phi) \right] \tag{C.52}$$

Note that the internal field satisfies the vector wave equation with wavenumber k_p. The boundary conditions are the continuity of $\hat{n} \times \bar{E}$ and $\hat{n} \times \bar{H}$ which is proportional to $\hat{n} \times \nabla \times \bar{E}$. At $r = a$

$$\hat{r} \times (\bar{E}_i + \bar{E}_s) = \hat{r} \times \bar{E}_{int} \tag{C.53}$$

Balancing the $\hat{r} \times \bar{M}$ and $\hat{r} \times \bar{N}$ components gives, respectively,

$$j_n(ka) - b_n h_n(ka) = d_n j_n(k_p a) \tag{C.54}$$

$$\frac{[kaj_n(ka)]'}{ka} - a_n \frac{[kah_n(ka)]'}{ka} = c_n \frac{[k_p a j_n(k_p a)]'}{k_p a} \tag{C.55}$$

Next we match at $r = a$

$$\hat{r} \times (\nabla \times \bar{E}_i + \nabla \times \bar{E}_s) = \hat{r} \times \nabla \bar{E}_{int} \tag{C.56}$$

We have $\nabla \times \bar{M} = k\bar{N}$ and $\nabla \times \bar{N} = k\bar{M}$ with k replaced by k_p for the internal field. Thus we have, by applying Equations C.50-C.52 to Equation C.56:

$$[kaj_n(ka)]' - b_n[kah_n(ka)]' = d_n[k_paj_n(k_pa)]' \tag{C.57}$$

$$kj_n(ka) - a_nkh_n(ka) = c_nk_pj_n(k_pa) \tag{C.58}$$

Solving equations C.54 and C.57 gives b_n and d_n. Solving C.55 and C.58 gives a_n and c_n. We find:

$$a_n = -T_n^{(N)} = \frac{k_p^2a^2j_n(k_pa)[kaj_n(ka)]' - k^2A^2j_n(ka)[k_paj_n(k_pa)]'}{k_p^2a^2j_n(k_pa)[kah_n(ka)]' - k^2A^2h_n(ka)[k_paj_n(k_pa)]'} \tag{C.59}$$

$$b_n = -T_n^{(M)} = \frac{j_n(k_pa)[kaj_n(ka)]' - j_n(ka)[k_paj_n(k_pa)]'}{j_n(k_pa)[kah_n(ka)]' - h_n(ka)[k_paj_n(k_pa)]'} \tag{C.60}$$

From C.17 - C.18 and from C.51, in the far field

$$
\begin{aligned}
\bar{E}_s &= -\frac{ie^{ikr}}{kr} \sum_{n=1}^{\infty} \sum_{m=-1,1} \frac{(2n+1)}{n(n+1)} \left\{ -b_n[\hat{e}_i \cdot \hat{C}_{-mn}(0,0)]\hat{C}_{mn}(\theta,\phi) \right. \\
&\quad \left. +a_n[\hat{e}_i \cdot \hat{B}_{-mn}(0,0)]\hat{B}_{mn}(\theta,\phi) \right\}
\end{aligned}
\tag{C.61}
$$

Substituting C.37 - C.43 in C.61, we have, summing over $m = 1, -1$ in C.61,

$$
\begin{aligned}
\bar{E}_s &= -\frac{ie^{ikr}}{2kr} \sum_{n=1}^{\infty} \frac{(2n+1)}{n(n+1)} \left\{ -a_n[\hat{e}_i \cdot (\hat{x} - i\hat{y})] \left[\hat{\theta}\tau_n(\cos\theta) + \hat{\phi}i\pi_n(\cos\theta) \right] e^{i\phi} \right. \\
&\quad -a_n[\hat{e}_i \cdot (\hat{x} + i\hat{y})] \left[\hat{\theta}\tau_n(\cos\theta) - \hat{\phi}i\pi_n(\cos\theta) \right] e^{-i\phi} \\
&\quad +ib_n[\hat{e}_i \cdot (\hat{x} - i\hat{y})] \left[\hat{\theta}i\pi_n(\cos\theta) + \hat{\phi}\tau_n(\cos\theta) \right] e^{i\phi} \\
&\quad \left. +ib_n[\hat{e}_i \cdot (\hat{x} + i\hat{y})] \left[\hat{\theta}i\pi_n(\cos\theta) - \hat{\phi}\tau_n(\cos\theta) \right] e^{-i\phi} \right\}
\end{aligned}
\tag{C.62}
$$

To get $f_{11}(\Theta)$ and $f_{21}(\Theta)$, we follow equations from C.46 - C.49 and set $\hat{e}_i = \hat{x}$, and $\phi = 90°$, so

that $\hat{l}_i = \hat{l}_s = \hat{x} = -\hat{\phi}$, and $\hat{2}_s = \hat{\theta}$. Substituting these in C.62, we find

$$\bar{E}_s = -\frac{ie^{ikr}}{kr}\hat{\phi}\sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}[a_n\pi_n(\cos\theta) + b_n\tau_n(\cos\theta)] \tag{C.63}$$

Thus, since $\theta = \Theta$, we have

$$f_{11}(\Theta) = \frac{i}{k}S_1(\Theta) \tag{C.64}$$

$$f_{21}(\Theta) = 0 \tag{C.65}$$

where

$$S_1(\Theta) = \sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}[a_n\pi_n(\cos\Theta) + b_n\tau_n(\cos\Theta)] \tag{C.66}$$

To get $f_{12}(\Theta)$ and $f_{22}(\Theta)$, we set $\hat{e}_i = \hat{y}$, and $\phi = 90°$, so that $\hat{2}_i = \hat{y}$, $\hat{l}_s = -\hat{\phi}$, and $\hat{2}_s = \hat{\theta}$. This time C.62 yields

$$\bar{E}_s = \frac{ie^{ikr}}{kr}\hat{\theta}\sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}[a_n\tau_n(\cos\theta) + b_n\pi_n(\cos\theta)] \tag{C.67}$$

Thus

$$f_{12}(\Theta) = 0 \tag{C.68}$$

$$f_{22}(\Theta) = \frac{i}{k}S_2(\Theta) \tag{C.69}$$

where

$$S_2(\Theta) = \sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}[a_n\tau_n(\cos\Theta) + b_n\pi_n(\cos\Theta)] \tag{C.70}$$

The scattering amplitude matrix is then

$$\begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} = \begin{bmatrix} \frac{i}{k}S_1(\Theta) & 0 \\ 0 & \frac{i}{k}S_2(\Theta) \end{bmatrix}$$

C.2 Optical Theorem

What follows is a simple derivation for the optical theorem as synthesized from the literature [40, 37, 35]. It is most similar to the treatment in Borghese et. al. and that book is recommended for more information [40]. In this derivation it is assumed that the scattering particle is in a medium with a real refractive index. We assume, with no loss of generality, that the scattering particle lies at the origin of the frame of reference and that the incident wave is linearly polarized and propagating along the z axis, the same geometry as the forward scattering formulation in Chapter 3.

$$E_1 = E_0 \hat{e}_i e^{i(kz-\omega t)} \tag{C.71}$$

For forward scattering, $\hat{k}_s = \hat{k}_i$, the scattering amplitude will be denoted as $f(0)$ and the scattered field is:

$$E_s = E_0 \frac{e^{i(kr-\omega t)}}{r} f(0) \tag{C.72}$$

Since this is in the forward direction, any plane through the common direction of $\hat{k}_1$ and $\hat{k}_s$ can be chosen as the plane of scattering and the unit vectors $\hat{u}_{l_n}$, can be taken to coincide with the vectors $\hat{u}_{s_n}$. Then, by dot multiplying C.71 and C.72 by $\hat{e}_I = \hat{e}_S$ we get the equations

$$E_I \cdot \hat{e}_I = E_0 e^{i(kz-\omega t)} \tag{C.73}$$

$$E_S \cdot \hat{e}_I = E_0 \frac{e^{i(kr-\omega t)}}{r} f(0) \cdot \hat{e}_I \tag{C.74}$$

that can be combined into

$$E_S \cdot \hat{e}_I = E_0 \frac{e^{ik(r-z)}}{r} f(0) \cdot \hat{e}_I (E_I \cdot \hat{e}_I) \tag{C.75}$$

We now remark that, in the forward direction, at a distance z from the scattering particle, a detector captures the intensity of the incident and the scattered wave. The plane of the receiving

lens, whose diameter is d, has the equation $z = \cos t$ so that, for $z \gg d$, we can write:

$$r = (z^2 + x^2 + y^2)^{1/2} \approx z + \frac{x^2 + y^2}{2z} \tag{C.76}$$

and the total field on the lens is

$$E_S \cdot \hat{e}_I + E_I \cdot \hat{e}_I = \left[1 + \frac{e^{ik(x^2+y^2)/2z}}{z} f(0) \cdot \hat{e}_I\right] (E_I \cdot \hat{e}_I) \tag{C.77}$$

The corresponding intensity is

$$|E_S \cdot \hat{e}_I + E_I \cdot \hat{e}_I|^2 = \left\{1 + \frac{2}{z}\text{Re}[e^{ik(x^2+y^2)/2z} f(0) \cdot \hat{e}_I]\right\} |E_I \cdot \hat{e}_I|^2 \tag{C.78}$$

which has been obtained by neglecting the term $|f(0) \cdot \hat{e}_I|^2/z^2$ that is actually negligible for large z. The total energy recorded by the instrument is given by integrating the preceding expression on the surface of the front lens, i.e., letting $x, y \leq d/2$. Close examination of the integrand shows, however, that when $\sqrt{z\lambda} \ll d \ll z$, as in the present case, the integration limits can be extended from $-\infty$ to ∞ without appreciably affecting the result. Now,

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} e^{ik(x^2+y^2)/2z} dy = \frac{2\pi i z}{k} \tag{C.79}$$

so that the energy recorded by the instrument is

$$W = \left[A - \frac{4\pi}{k}Im(f(0) \cdot \hat{e}_I)\right] I_0 \tag{C.80}$$

where $I_0 = |E_1 \cdot \hat{e}_1|^2$. This equation shows that, due to the interferences of the incident and the scattered wave, the front lens of the receiving instrument was seemingly an area smaller than A and is therefore able to collect a smaller quantity of the incident wave, the term AI_0 is the energy that the instrument would collect in the absence of the particle. The second term

$$I_{ext} = \frac{4\pi}{k}\text{Im}(f(0) \cdot \hat{e}_I) I_0 \tag{C.81}$$

measures the energy that is lacking because of the presence of the particle. The lacking energy should be equal to the energy that the particle removes from the incident beam by absorption and scattering, so that the coefficient of I_0 must coincide with extinction cross-section. We are thus led to the relation,

$$\sigma_{ext} = \frac{4\pi}{k}\text{Im}|f(0) \cdot \hat{e}_I| \tag{C.82}$$

which is the expression of the optical theorem. Put in words, the optical theorem states that the total cross-section of extinction is the imaginary part of the scattering amplitude function in the forward direction.

Appendix D

MATLAB code

All of the relevant MATLAB functions and scripts used in this document are represented here. Section D.1 contains files related to data acquisition and concentration estimation. Section D.2 contains files that generate the error propagation matrices and use them to simulate the noise floor. Section D.3 contains files that numerically solve for the scattering and extinction cross-sections for elastic scatterers in the Mie regime. Section D.4 contains files for a graphical user interface for real-time monitoring of online concentration estimates. Section D.5 contains miscellaneous files used by other functions and scripts. Copies of all files can be attained by contacting Prof. Rajeev Ram.

D.1   Concentration Estimation and Data Acquisition

calibration_read5.m   reads all of the calibration files in a given directory and generates the pure component matrix.

calibration_read5.m

```
function [ps,devs,shotdevs,wdevs,wsdevs,compnames,waterdata,waterdata2,bline] ...
     = calibration_read5(dirname)
%  [ps,devs,shotdevs,wdevs,wsdevs,compnames,waterdata,waterdata2,bline]
%             =calibration_read5(dirname)
%  This is a new master version of calibration read that automatically generate
%  the correct calibration matrix and appropriate meta-data for the appropriate
```

149

% concentrations/data plots to be made.

```
if dirname(end)=='\' | dirname(end)=='/'; dirname = dirname(1:end-1); end
                                                                                    10

D = dir(dirname);
D2 = D(3:end);
c = 0;
for I = 1:length(D2);
  if lower(D2(I).name(end-3:end)) =='.txt'
    c = c+1;
    D3(c) = D2(I);
  end
end
                                                                                    20


% Look for a list of component names
J = 1; % component index
K = 1; % water index

compnames{1} = 'Water';

fprintf('reading files \n');
D = D3;
for I = 1:length(D)                                                                 30
  fprintf('.');
  D(I).name
  if isempty(findstr(lower(D(I).name),'water')) & ...
        isempty(findstr(lower(D(I).name),'dark')) & ...
        ~isempty(findstr(lower(D(I).name),'_'))
    indus = findstr(D(I).name,'_');
    compnames{J+1} = D(I).name(1:indus(1)-1);
    compconc(J) = str2num(D(I).name(indus(1)+1:indus(2)-1));
    compwat(J) = str2num(D(I).name(indus(2)+1:indus(3)-1));
    compind(J) = I;                                                                 40
    x = load([dirname,'\',D(I).name]);
    x2 = x(:,3:end)'./(ones(1340,1)*mean(x(:,3:end)',1)).*mean(mean(x(:,3:end)));
    compdata(:,J) = mean(cosmic_clean_gg(x2,300,2.5),2);
    cdevs(:,J) = std(cosmic_clean_gg(x2,300,2.5),[],2);
    J = J+1;
  end

  if findstr(lower(D(I).name),'water')
    waterind(K) = I;
    x = load([dirname,'\',D(I).name]);                                             50
    x2 = x(:,3:end)'./(ones(1340,1)*mean(x(:,3:end)',1)).*mean(mean(x(:,3:end)));
```

```
      waternames{K} = D(I).name;
      waterdata(:,K) = mean(cosmic_clean_gg(x2,300,2.5),2);
      wdevs(:,K) = std(cosmic_clean_gg(x2,300,2.5),[],2);
      K = K+1;
    end

  end

  fprintf('\n processing data \n');                                      60

  compnames = compnames.';
  compind = compind(:);
  compconc = compconc(:);
  compwat = compwat(:);
  waternames = waternames.';
  waterind = waterind(:);

  % subtract dark background all data
                                                                        70
  x = load([dirname,'\','dark_background.txt']);
  darkback = mean(cosmic_clean_gg(x(:,3:end).',300,2.5),2);

  compdata = compdata-darkback*ones(size(compdata(1,:)));
  waterdata = waterdata-darkback*ones(size(waterdata(1,:)));

  [waterdata2, shw, bline] = rayleigh_shift(waterdata);
  [compdata2, shc, bline] = rayleigh_shift(compdata,bline);
  shw
  shc                                                                   80
  devs = ray_shift_known(cdevs,shc,bline);
  for I = 1:length(compconc)
    if compwat(I)
      ps(:,I) = (compdata2(:,I)-waterdata2(:,compwat(I)))./compconc(I);
    else
      ps(:,I) = compdata2(:,I)./compconc(I);
    end
  end

  % add water to the first column of ps                                 90
  ps = [mean(waterdata2,2)./1000,ps];
  %ps = shiftcorrect
  ps = ps(254:end,:);
  devs = devs(254:end,:)./(ones(size(ps,1),1)*compconc');

  for I=2:size(ps,2)
```

```
      [c1,j1,j2,j3,j4,j5]=conc_extract(compdata2(254:end,I-1),ps,[1 I],4);
      c2=c1(1:2)./c1(1).*1000;
            corrconc(I-1)=c2(2);
end                                                                                    100


correction=[1 compconc'./corrconc(:)'];


ps = ps./(ones(size(ps,1),1)*correction);


electrons = compdata2(254:end,:).*5.19;
elecstd  = electrons.^0.5;
countstd = elecstd./5.19;
shotdevs = countstd./(ones(size(ps,1),1)*compconc');
                                                                                       110

electrons = waterdata2(254:end,:).*5.19;
elecstd  = electrons.^0.5;
countstd = elecstd./5.19;
wsdevs = countstd;


wdevs=wdevs(254:end,:);
```

`data_read_offline.m`   reads all of the offline data files in a given directory and generates the

measured spectrum matrix.

data_read_offline.m

```
function [tind,xo,devs,shotdevs,Dkeepsam] = data_read_offline(dirname,bline)
% [tind,xo,devs,shotdevs,Dkeepsam] = data_read_offline(dirname,bline)


if dirname(end)=='\' | dirname(end)=='/'; dirname = dirname(1:end-1); end


D = dir(dirname);
ind = 0;
for I = 1:length(D)
  if ~isempty(findstr('.SPE',D(I).name)) & ...
         ~isempty(findstr('Sample',D(I).name)) & ...                                   10
         length(D(I).name)>4,
    ind = ind+1;
    Dkeep(ind) = D(I);
    indus = findstr(D(I).name,'_');
    samplenum(ind) = str2num(D(I).name(indus(1)+1:indus(2)-1));
  end
end


[dns,inds]=sort(samplenum);
```

```
Dkeepsam = Dkeep(inds);                                              20

tind = dns;

fprintf('\nreading data files \n');

for I = 1:length(Dkeepsam)
    fprintf('%d',length(Dkeepsam)-I);
    fprintf('.');
    y = load([dirname,'\',Dkeepsam(I).name(1:end-4),'.txt']);
    y2 = y(:,3:end)'./(ones(1340,1)*mean(y(:,3:end)',1)).*mean(mean(y(:,3:end)));  30
    x(:,I) = mean(cosmic_clean_gg(y2,300,2.5),2);
         ddevs(:,I) = std(cosmic_clean_gg(y2,300,2.5),[],2);
end

fprintf('\n');

xo0 = x(:,1:end);
[xo, shd, bline] = rayleigh_shift(xo0,bline);
shd
devs = ray_shift_known(ddevs,shd,bline);                             40
xo = xo(254:end,:);
electrons = xo.*5.19;
elecstd  = electrons.^0.5;
countstd = elecstd./5.19;
shotdevs = countstd;

devs = devs(254:end,:);
```

data_read_cum.m reads all of the online data files in a given directory and generates the measured spectrum matrix. If a new file is added, it will only read the new file, increasing throughput.

data_read_cum.m

```
function [t,x,Dkeeps] = data_read(dirname,tfc,data)

    D = dir(dirname);
    ind = 0;
    for I = 1:length(D)
      if ~isempty(findstr('.SPE',D(I).name)) & length(D(I).name)>4,
        ind = ind+1;
        Dkeep(ind) = D(I);
        dn(ind) = datenum(D(I).date);
      end                                                           10
    end
```

```
[dns,inds]=sort(dn);
Dkeeps = Dkeep(inds);

t = [0,cumsum(diff(dns))*24];

for I = 1:length(Dkeeps)
  if strcmp('BR1_02_09_30s_10f_t00p00',Dkeeps(I).name(1:end-4))
    ind_time_1 = I;                                                        20
  end
end


for I = 1:(length(Dkeeps)-size(data,2))
  y = load([dirname,'\',Dkeeps(I+size(data,2)).name(1:end-4),'.txt']);
  x(:,I) = mean(cosmic_clean(y(:,3:end).'),2).*tfc;
end

t = t-t(ind_time_1);                                                       30

x = [data,x(85:end,:)];
```

`data_read_opalm.m` reads all of the offline data files for an oil palm fermentation in a given directory and generates the measured spectrum matrix.

data_read_oilpalm.m

```
function [tind,xo,Dkeepsam, shotdevs] = data_read_oilpalm(dirname,bline)
%function [tind,xo,Dkeepsam] = data_read_oilpalm(dirname,tfc,bline)

if dirname(end)=='\' | dirname(end)=='/'; dirname = dirname(1:end-1); end

D = dir(dirname);
ind = 0;
for I = 1:length(D)
  if ~isempty(findstr('.SPE',D(I).name)) & length(D(I).name)>4,
    ind = ind+1;                                                          10
    Dkeep(ind) = D(I);
    indus = findstr(D(I).name,'_');
        samplenum(ind) = str2num(D(I).name(indus(1)-1))*10 ...
        +str2num(D(I).name(indus(2)-1));
  end
end

[dns,inds]=sort(samplenum);
```

```
    Dkeepsam = Dkeep(inds);
```
```
    tind = dns;

    for I = 1:length(Dkeepsam)
      fprintf('%d',length(Dkeepsam)−I);
      fprintf('.');
      y = load([dirname,'\',Dkeepsam(I).name(1:end−4),'.txt']);
      y2 = y(:,3:end)'./(ones(1340,1)*mean(y(:,3:end)',1)).*mean(mean(y(:,3:end)));
      x(:,I) = mean(cosmic_clean_gg(y2,300,2.5),2);
```
```
%      if strcmp('BR1_02_09_30s_10f_t00p00',Dkeeps(I).name(1:end-4))
%         ind_time_1 = I;
%      end


    end
    fprintf('\n');

%mx = max(x)
%xn = x./(ones(1340,1)*mx);
%x = xn.*mean(mx);
```
```
%x = x(85:end,:);
% clugy removal of -1 time
tind = tind(1:end);
xo0 = x(:,1:end);
[xo, shd, bline] = rayleigh_shift(xo0,bline);
shd

xo = xo(254:end,:);
electrons = xo.*5.19;
```
```
elecstd  = electrons.^0.5;
countstd = elecstd./5.19;
shotdevs = countstd;
```

conc_extract.m estimates the concentration estimates according to the algorithm described in Chapter 2 using the pure component matrix from calibration_read5.m and the measured spectrum matrix from any one of the data read files shown above.

conc_extract.m

```
function [conc,xm,residual,A,B,concz] = conc_extract(x,ps,indselect,pord)
% conc,xm,residual,A,B,concz] = conc_extract(x,ps,indselect,pord)
```

```matlab
if nargin<3, indselect = 1:size(ps,2);pord=4;end;
if nargin<4, pord = 4;end

ps = ps(:,indselect);

wna = linspace(0,1,size(ps,1)).' ;
%    if nopoly                                                                    10
%       A = [ps];
%    else
% % A = [ps,ones(size(ps,1),1),10.*[1-wna,1-wna.^2,1-wna.^3,1-wna.^4]];
% pord = 3;
%       A = [ps,ones(size(ps,1),1),10.*[1-wna,1-wna.^2,1-wna.^3]];
% %    A = [ps,ones(size(ps,1),1),10.*[1-wna]];
%    end
%
  A = ps;
  if pord>=0                                                                      20
    A = [ps,ones(size(ps,1),1)];
        for I = 1:pord
          A = [A,50.*(1-wna.^I)];
        end
  end


[U,S,V]=svd(A);
S = S(1:size(V,1),:);
U = U(:,1:size(V,2));
                                                                                 30
B = U*diag(1./diag(S))*V.' ;

conc_svd = B.'*x;

% for I = 1:size(x,2)
%    conc_temp = lsqlin(A,x,ones(1,size(A,2)),inf,[],[],...
%                       [zeros(size(ps,2),1);-inf*ones(5,1)],inf);
%    conc_lsq(:,I) = conc_temp;
% end
                                                                                 40




concz = conc_svd;
concc = concz(1:size(ps,2),:);
concc(concc<0) = 0;
if pord < 0
  conct = concc;
```

```
else
   conct = [concc;concz(end-pord:end,:)];                          50
end
conc = conct;

xm = A*conct;
residual = x-xm;
```

D.2   Error Propagation and Simulations

`error_prop_vec.m`   generates the error propagation matrices using the algorithm described in Chapter 2.

error_prop_vec.m

```
function [dconc, dconcb, dconcps, dconcps1,dconcps2] = ...
    error_prop_vec(A,ctdevs,ddevs,xo,residual)
% [dconc, dconcb, dconcps, dconcps1,dconcps2] = error_prop(A,ctdevs,ddevs,xo,residual)
%tic;

l=size(A,2);
%inds=1:l;
%A2=A(:,inds);
[U, S, V] = svd(A);
eS = S(1:l,:);                                                     10
eU=U(:,1:l);
eV = V;
Ebinv = eV*(eS^-1)*eU';
ps2inv = eV*(eS^-1)*eU';
q = size(ctdevs,1) * ( size(A,2) - size(ctdevs,2) );
dA=[ctdevs(:); zeros(q,1)];
eSn1=eS^-1;
eSn2=eS^-2;

b = xo;                                                           20
db = ddevs;
s = ps2inv*b;
r = residual;
mags = realsqrt(sum(s.*s));
magr = realsqrt(sum(r.*r));
ns = s/diag(mags);
nr = r/diag(magr);
```

```
eW = zeros(prod(size(eU)),size(eU,2));
eY = zeros(prod(size(eU)),size(eU,2));                                     30

dconcps1 = zeros(size(eU,2),size(b,2));
dconcps2 = zeros(size(eU,2),size(b,2));
dconcps = zeros(size(eU,2),size(b,2));
dconcb = zeros(size(eU,2),size(b,2));
dconc = zeros(size(eU,2),size(b,2));


tic;
dA_box = repmat(dA,[1,1,size(b,2)]);                                       40
EA1inv = zeros(size(V,1),size(dA,1),size(b,2));
EA2inv = zeros(size(V,1),size(dA,1),size(b,2));
for J = 1:size(b,2)
    for k = 1:size(eU,2)
        w = -(eU(:,k)*ns(:,J)') ;
        eW(:,k)=w(:);
        y = -nr(:,J)*eV(:,k)' ;
        eY(:,k)=y(:);
    end
    EA1inv(:,:,J) = V*(eSn1.*mags(J))*eW' ;                                50
    EA2inv(:,:,J) = V*(eSn2.*magr(J))*eY' ;
end
toc;
dconcps1_box = ctimes(EA1inv,dA_box);
dconcps2_box = ctimes(EA2inv,dA_box);
dconcps1 = reshape(dconcps1_box,prod(size(dconcps1_box)),1,1);
dconcps2 = reshape(dconcps2_box,prod(size(dconcps1_box)),1,1);
dconcps = dconcps1+dconcps2;
dconcb = (Ebinv*db);
dconc = dconcps + dconcb;                                                  60
```

`papersim.m` runs a simulation that uses the error propagation matrices to estimate the error in the online concentration estimates. The output is saved to a set of files.

papersim.m

```
function papersim

load online;

points = size(fvalues,2);
%points = 1;
tic;
```

```
    for k=1:points
        fprintf('\nRunning Data point %d of %d. \n',k,points);
        fname = ['output_files\online_' num2str(k)];                    10
        [conc,conc_norm,pre_conc,pre_conc_norm,dc,dcps,dcx] = ...
            papersim_loop(fps,fA,fdevs,fvalues(:,k));
        save(fname,'conc','conc_norm','pre_conc','pre_conc_norm','dc','dcps','dcx');
    end
    fprintf('\nTotal simulation run time: ');
    toc;
```

papersim_offline.m runs a simulation that uses the error propagation matrices to estimate
the error in the offline concentration estimates. The output is saved to a set of files.

papersim_offline.m

```
    function papersim_offline

    load offline2;

    points = size(ovalues,2);
    %points = 1;
    tic;
    for k=1:points
        fprintf('\nRunning Data point %d of %d. \n',k,points);
        fname = ['output_files\offline_' num2str(k)];                   10
        [conc,conc_norm,pre_conc,pre_conc_norm,dc,dcps,dcx] = ...
            papersim_loop(ops,oA,odevs,ovalues(:,k));
        save(fname,'conc','conc_norm','pre_conc','pre_conc_norm','dc','dcps','dcx');
    end
    fprintf('\nTotal simulation run time: ');
    toc;
```

papersim_opalm.m runs a simulation that uses the error propagation matrices to estimate the
error in the oil palm concentration estimates. The output is saved to a set of files.

papersim_opalm.m

```
    function papersim_opalm

    load opalm2;

    points = size(ovalues,2);
    %points = 1;
    tic;
    for k=1:points
        fprintf('\nRunning Data point %d of %d. \n',k,points);
```

```
        fname = ['output_files\opalm_' num2str(k)];                          10
        [conc,conc_norm,pre_conc,pre_conc_norm,dc,dcps,dcx] = ...
            papersim_loop(ops,oA,odevs,ovalues(:,k));
        save(fname,'conc','conc_norm','pre_conc','pre_conc_norm','dc','dcps','dcx');
    end
    fprintf('\nTotal simulation run time: ');
    toc;
```

papersim_loop.m   is the main simulation loop used by all off the above simulation functions.

papersim_loop.m

```
    function [beconcmat,beconcmat_norm,beconczmat,beconczmat_norm, ...
        dcmat, dcpsmat, dcxmat]=papersim_loop(ps,A,devs,values)


    %c1=clock;

    psdim1=size(ps,1);
    psdim2=size(ps,2);
    gain=5.19;
    bg = 84;
                                                                             10

    %basewater=ps(:,1).*1000;
    %Lac27=ps(:,6).*16.70;
    %Ace27=ps(:,2).*45.89;
    %Phe27=ps(:,9).*16.23;
    %K27=ps(:,5).*63.00;
    %NH27=ps(:,7).*15.00;
    %Na27=ps(:,8).*25.00;
    %point27=basewater+Ace27+Lac27+Phe27++K27+NH27+Na27+bg;
    %Phe=ps(:,9).*99;
    datapoint=A*values;                                                      20

    datamat=datapoint*(ones(1,1000));

    [conc,xm,residual,A,B,concz] = conc_extract(datapoint,ps);
    conc_norm=conc./(ones(size(conc,1),1)*conc(1,:)).*1000;


    %***************** Soon to be main loop



                                                                             30
    %****** 10000 combinations (100 X 100) for errors in both
    datamat_new = datapoint*(ones(1,100));
    derr_new = randn(psdim1,100).*realsqrt(datamat_new./gain);
    data_w_err_new = datamat_new + derr_new;
```

```
error_cube  =  randn(psdim1,psdim2,100);
ps_err_cube  =  (error_cube.*repmat(devs,[1,1,100]));

beconc_cube=zeros(psdim2+5,100,100);
beconcz_cube=zeros(psdim2+5,100,100);
beconc_cube_norm=zeros(psdim2+5,100,100);                                        40
beconcz_cube_norm=zeros(psdim2+5,100,100);
dcbox=zeros(psdim2+5,100,100);
dcpsbox=zeros(psdim2+5,100,100);
dcxbox=zeros(psdim2+5,100,100);
fprintf('\n Long "for" loop. . .\n\n');
for k=1:100
    %error_mat  =  randn(psdim1,psdim2);
    %fprintf('%d ',k);
    fprintf('.');
    if mod(k,20)==0                                                             50
        fprintf('\n');
    end
    %ps_err  =  (error_cube(:,:,k).*devs);
    [beconc,bexm,beresidual,beA,beB,beconcz]  =  ...
        conc_extract(data_w_err_new,ps+ps_err_cube(:,:,k));
    %beconc(1,1)
    beconc_norm=beconc./(ones(size(beconc,1),1)*beconc(1,:)).*1000;
    beconcz_norm=beconcz./(ones(size(beconcz,1),1)*beconcz(1,:)).*1000;
    beconc_cube(:,:,k)  =  beconc;
    beconcz_cube(:,:,k)  =  beconcz;                                            60
    beconc_cube_norm(:,:,k)  =  beconc_norm;
    beconcz_cube_norm(:,:,k)  =  beconcz_norm;
    [dc, dcb, dcps, dcps1,dcps2]  =  ...
        error_prop_vec_old(beA,ps_err_cube(:,:,k), ...
        derr_new,datamat_new,beresidual);
%   [dc, dcb, dcps, dcps1,dcps2]  =  ...
%       error_prop_vec_t(beA,devs,realsqrt(datamat_new./gain ), ...
%       data_w_err_new,beresidual);
    dcbox(:,:,k)=dc;
    dcpsbox(:,:,k)=dcps;                                                        70
    dcxbox(:,:,k)=dcb;
end

beconcmat  =  reshape(beconc_cube,psdim2+5,10000);
beconczmat  =  reshape(beconcz_cube,psdim2+5,10000);
beconcmat_norm  =  reshape(beconc_cube_norm,psdim2+5,10000);
beconczmat_norm  =  reshape(beconcz_cube_norm,psdim2+5,10000);
dcmat  =  reshape(dcbox,psdim2+5,10000);
dcpsmat  =  reshape(dcpsbox,psdim2+5,10000);
```

```
dcxmat = reshape(dcxbox,psdim2+5,10000);                               80
%******
```

```
%  ***********************
```

D.3   Elastic Scattering and Mie Theory

GGRB1.m   numerically calculates a Riccati-Bessel function of the 1st kind for a given size factor
and order.

GGRB1.m

```
function psi = GGRB1(nmax, x, acc)
```

```
% GGRB1  Riccati-Bessel function of the first kind
%    psi = GGRB1(nmax, x)
%    Input: highest order of psi, nmax
%        argument to the function x,
%        level of accuracy (optional, defualt is 1e-6)
%    Output: The matrix, psi of size [nmax length(x)].
%        Each row contains a R-B function of the 1st kind of order m
                                                                       10
%    Gustavo Gil 2004
```

```
if nargin < 3
  acc = 0.000001;  % desired accuracy
end
```

```
% x into a row vector
x = x(:).';
n = [1:nmax]';                                                         20
nu = n + 0.5;
```

```
%Bessel Functions of the the 1st kind.
%each row is a Bessel function of order n with respect to x
% MATLAB takes the conjugate when it transposes, so I correct for that
J = conj(besselj(nu,x));
if size(J,2) == 1
    J = J';
end
                                                                       30
```

%coefficient to convert from Bessel to Riccati-Bessel
C = **sqrt(pi.*x./2)**;
Cmat = repmat(C,nmax,1)';

%Riccati-Bessel function of the first kind
psi = Cmat.*J;

GGRB3.m numerically calculates a Riccati-Bessel function of the 3rd kind for a given size factor and order.

GGRB3.m

function zeta = GGRB3(nmax, x, acc)

```
% GGRB1 Riccati-Bessel function of the third kind
%    zeta = GGRB3(nmax, x)
%    Input: highest order of zeta, nmax
%        argument to the function x,
%        level of accuracy (optional, defualt is 1e-6)
%    Output: The matrix, zeta of size [nmax length(x)].
%        Each row contains a R-B function of the 3rd kind of order m
```
                                                                                                    10
```
% Gustavo Gil 2004
```


if nargin < 3
    acc = 0.000001; *% desired accuracy*
end

% x into a row vector
x = x(:).';
% n and nu are column vectors   20
n = [1:nmax]';
nu = n + 0.5;

% Bessel Functions of the the third kind.
% also called the first Hankel function.
% each row is a Bessel function of order n with respect to x
% MATLAB takes the conjugate when it transposes, so I correct for that
H1 = **conj**(besselh(nu,1,x));
if size(H1,2) == 1
    H1 = H1';   30
end

%coefficient to convert from Bessel to Riccati-Bessel

```
C = sqrt(pi.*x./2);
Cmat = repmat(C,nmax,1)';
```

%Riccati-Bessel function of the third kind
```
zeta = Cmat.*H1;
```

GGMieAB.m numerically calculates the Mie coefficients a_n and b_n.

GGMieAB.m

function [a,b] = GGMieAB(nmax, x, m)

```
% GGMieAB Mie Scattering coefficients a and b
%    [a, b] = MieAB(nmax, x, m)
%    Input: Highest required order of a,b nmax,
%        Size parameter x = k_m * a,
%        and refractive index of particle relative to medium m = k_p / k_m
%    Output: matrices a and b, size(a) = [nmax, length(x)];
%        where m and x must be the same length.
%        a(n,x) = a_n(x)                                                    10
%    Calls GGRB1.m, GGRB2.m


% open up m and x into row vectors
m = m(:).';
x = x(:).';
mx = m.*x; %combined argument


% Riccati-Bessel functions
% MATLAB dos not have a besselh                                            20
psi = GGRB1(nmax, x)';  % R-B function of the first kind
psi_m = GGRB1(nmax, mx)';  % R-B function of the first kind
zeta = GGRB3(nmax, x)';  % R-B function of the third kind


% temp vectors for calculating derivatives
% Tpsi_n = psi_n-1
i = sqrt(-1);
Tpsi = [sin(x) ; psi(1:(nmax-1),:)];    % psi_0 = sin(x)
Tpsi_m = [sin(mx) ; psi_m(1:(nmax-1),:)]; %psi_m_0 = sin(mx)              30
Tzeta = [-i.*exp(i.*x) ; zeta(1:(nmax-1),:)]; % zeta_0 = -cos(x)


% matrices to make arithmetic for derivatives easier
nmat = repmat((1:nmax).',1,length(x));
xmat = repmat(x,nmax,1);
mxmat = repmat(mx,nmax,1);
```

```
if length(m) == length(x)
   mmat = repmat(m,nmax,1);
else
   mmat = repmat(m,nmax,length(x));                              40
end


% derivatives are calculated using recursion relation
% Dpsi_n = psi_n-1 - (n/x)*psi
Dpsi = Tpsi−nmat.*psi./xmat;
Dpsi_m = Tpsi_m−nmat.*psi_m./(mxmat);  %
Dzeta = Tzeta−nmat.*zeta./xmat;


%Mie coeeficients a_n and b_n
a = (mmat.*psi.*Dpsi_m − Dpsi.*psi_m)./(mmat.*zeta.*Dpsi_m − Dzeta.*psi_m);    50
b = (psi.*Dpsi_m − mmat.*Dpsi.*psi_m)./(zeta.*Dpsi_m − mmat.*Dzeta.*psi_m);
```

`GGMieLegendre.m` numerically calculates the associated Legendre functions for a given size factor and orders m and n.

GGMieLegendre.m

```
function Pall = GGMieLegendre(n,m,x);
% Get P_n^m for all n and a single m


x = x(:);


Pall= zeros(length(x),n);
for k = 1:n
    P = legendre(k,x)';
    Pall(:,k)=P(:,m+1);
end                                                               10
```

`GGMieCS.m` numerically calculates the cross-sections of extinction and scattering for a given sphere size, wavenumber of excitation, and index of refraction.

GGMieCS.m

```
% MieCS Mie Scattering & Extinction Cross Sections
% [Cext, Csca] = MieCS(a, k_m, m [, s])
% Input: Scattering particle radius a,
%    wave vector of scattering light in surrounding medium k_m,
%    refractive index of particle relative to medium m = k_p / k_m
%    and the optional surface conductivity parameter s = sigma_s * c_m * mu_m.
% Output: Arrays Cext and Csca, size(Cext) = [length(x), length(m)]
```

```matlab
% Calls: MieAB.m
% Gustavo Gil 2004                                                        10


function [Cext, Csca] = GGMieCS(r, k, m)

%row vectors
r = r(:).';
k = k(:).';
m= m(:).';
x = k.*r;
                                                                          20

% Wiscombe approximation for # of terms in series
% only works for real x
%nmax = MieWn(x);


%fuck that,
nmax = 50;


%get coefficients
[a,b] = GGMieAB(nmax, x, m);
                                                                          30

% Are there NaN:s among the results?
while 1 %infinite loop
    test = find(any(any(isnan([a;b]), 3),1)); % indices of faulty x's
    if isempty(test)
        break; % break out of the while loop
    end
    disp('MieCS: NaN found');
    a(:, test, :) = 0;
    b(:, test, :) = 0; % remove the NaN:s
    nmax2 = MieWn(x(test)); % take a new (smaller) nmax              40
    if nmax2 > nmax
        error('This should never happen');
    end
    [A,B] = MieAB(nmax2, x(test), m); %compute new A, B for the invalid x values
    a(1:nmax2, test, :) = A;
    b(1:nmax2, test, :) = B; % splice the results into a, b
end


% Nmat is a summation component
n = (1:nmax).';                                                           50
N = 2*n+1;
Nmat = repmat(N,1,length(x));
```

```
% sumation components
ext = real(a+b);
sca = abs(a).^2 + abs(b).^2;


% summation
Sext = sum(Nmat.*ext,1);
Ssca = sum(Nmat.*sca,1);                                                 60


% coefficient to the sumation
Coeff = 2.*pi./(k.^2);


% Cross Sections!
% Cext = (2*pi/k^2)*real(sum((2n+1)(a_n+b_n)))
% Csca = (2*pi/k^2)*real(sum((2n+1)(a_n^2+b_n^2)))
Cext = Coeff.*Sext;
Csca = Coeff.*Ssca;
```

stavmie4_ir.m Numerically calculates the attenuation in the Raman spectrum due to scattering for a given particle size, particle concentration, and depth of field.

stavmie4_ir.m

```
function [L, tfL, lfit] = stavmie4_ir(rho0,diameter,z,slice)

strt = 3;
wllo = 784.04;
wlhi = 912.14;
%was playing with these for "fun"
%wllo = 1;
%wlhi = 2000;
pixels = 1340;
range = wlhi-wllo;                                                       10
step = range/(pixels - 1);
wlr = [ wllo : step : wlhi ];
%wavelength in nm
wl = wlr(strt-2:end);
%wavelength in m
lambda = wl.*1e-9;
%wavelength in microns
wlm = lambda.*1e6;
%wave number in m-1
k0 = 2.*pi./lambda;                                                      20
%wave number in cm-1
kcm = k0.*1e-2;
```

.

```
%Particle Density, rho, in particles per cubic meter
rho = rho0/slice;


%Water is full complex refractive index
% still need imaginary part for polystyrene
%ignoring absorbion in the visible
ncH2O = real(ref_water(wlm));
%ncH2O = 1.33;
nrStyrene = 1.5663 + .00785./(wlm.^2) + .000334./(wlm.^4);
%nrStyrene = 1.59;
niStyrene = 0;
ncStyrene = nrStyrene + i*niStyrene;
% m = refractive index of the particle relative to the medium
m = ncStyrene./ncH2O;
%ignore complex part of refractive index to compute k
kH2O = k0.*real(ncH2O);


%optical path length
% is now an input
%z = ncH2O*.01;
%z = .000025;


%all in meters
sphere_diameter = diameter;
sphere_radius = sphere_diameter/2;
size_factor_x = sphere_radius.*kH2O;



Cgeom = pi*sphere_radius^2;


%cross sections
[CSext CSsca] = GGMieCS(sphere_radius,kH2O,m);


%Qext from 2 different models
Qext1 = CSext./Cgeom;


%Q2 = zeros(length(size_factor_x),6);
%for i = 1:length(size_factor_x)
%   Q2(i,:) = Mie(m,size_factor_x(i));
%end
%Qext2 = Q2(:,1);


%error between models
%errQ = Qext1-Qext2';
```

30

40

50

60

```
%optical thickness tau:
tau = Qext1.*z.*rho.*Cgeom;                                          70

R = (2.*lambda.*tau.*size_factor_x)./(3.*pi.*z.*Qext1);

lp = mean(z./tau)

lp_in_microns = lp*1e6

L = exp(-tau);
tfL = L./max(L);
pc = polyfit(wl,tfL,1);                                             80
lfit = polyval(pc,wl);
%mse = (tfL-lfit).^2;
```

`multi_angle.m` numerically calculates the attenuation in the Raman spectrum due to scattering using the probe geometry, and integrates the beam shaping coefficient, g_n.

multi_angle.m

```
tic;
% input sphere_radius
%angle_max
sphere_radius = 1.530e-6;
%angle_max = atan(3*25.4/32/5);
angle_max = pi;

wo=90*1e-6;
z = .001
rho = 2.0631e+013;                                                 10

%make vector fof thetas from 0 to theta_max
thetas = linspace(0+.000001,angle_max-.000001,1000)';

%take the cosine of these thetas
cost = cos(thetas)';

%make the wavelngth vector
wllo = 784.04;
wlhi = 912.14;                                                      20
wl=linspace(wllo,wlhi,1340)';
wlm=wl*1e-3;

% refractive index of Polystyrene
```

```
nrStyrene = 1.5663 + .00785./(wlm.^2) + .000334./(wlm.^4);
nc = [5e−4, 6e−4,6e−4,7e−4,8e−4,8e−4,9e−4];
wlc = [770, 800, 830, 860, 890, 920, 950];
niStyrene = interp1(wlc,nc,wl,'pchip');
ncStyrene = nrStyrene + i.*niStyrene;
```

<div align="right">30</div>

```
% refractive index of water
nrH2O = real(ref_water(wlm));


% relative refractive index
m = ncStyrene./nrH2O;


% k in free space
k0 = 2.*pi./(wl*1e−9);
```

<div align="right">40</div>

```
% k in water
kH2O = k0.*real(nrH2O);


% size factor, x, of the spheres to the wavelength
size_factor_x = sphere_radius.*kH2O;
nmax=50;


% Mie coefficients. Note that these are matrices of size
% nmax X number of pixels (or length(m) to be specific)
% these are independant of angle, but not of wavlength
[an,bn] = GGMieAB(nmax, size_factor_x, m);
```

<div align="right">50</div>

```
% Associated legendre function values for the selected wavelengths
% Note that this is a matrix of size length(thetas) X nmax. These
% is independant of wavelength, but not of angle
P1n = GGMieLegendre(nmax,1,cost);


% Computation of pi_n(cos(theta)).
% Note that this is a matrix of size length(thetas) X nmax. This
% is independant of wavelength, but not of angle
sinmat=sin(thetas)*ones(1,nmax);
pin = −P1n./sinmat;
pin(1,:)=(.5*[1:nmax].*[2:nmax+1]);
```

<div align="right">60</div>

```
% Computation of tau_n(cos(theta)).
% Note that this is a matrix of size length(thetas) X nmax. This
% is independant of wavelength, but not of angle
dthetas = mean(diff(thetas));
taun = −diff(P1n)./dthetas;
```

```
taun=[(.5*[1:nmax].*[2:nmax+1]); taun];                                            70

%coefficient vector for sums
n = 1:nmax;
coeff = ((2.*n+1)./(n.*(n+1)))';
cmat = coeff*ones(1,1340);

%allocate S matrices
S1 = zeros(1000,1340);
S2 = zeros(1000,1340);
                                                                                   80


[Cext, Csca] = GGMieCS(sphere_radius, kH2O, m);
[Cextr, Cscar] = GGMieCS(sphere_radius, kH2O, real(m));


s=(kH2O.*wo)'.^-1;
smat=repmat(s,50,1);
a=(n+-1)'.*(n+2)';
amat=repmat(a,1,1340);
gn=exp(-amat.*(smat.^2));                                                          90

toc;
tic;
for k = 1:1000
    if mod(k,10)==0
        fprintf('.');
    end
    if mod(k,200)==0
        fprintf('\n');
    end                                                                          100
    pin_k = (ones(1340,1)*pin(k,:))';
    taun_k = (ones(1340,1)*taun(k,:))';
    S1(k,:)= sum( cmat.*gn.*(an.*pin_k + bn.*taun_k) , 1 );
    S2(k,:)= sum( cmat.*gn.*(an.*taun_k + bn.*pin_k) , 1 );
end
toc;



integrand= ( sin(thetas)*ones(1,1340)).*(abs(S1).^2 + abs(S2).^2).*dthetas;       110
tCSsca = sum(integrand).*pi./(kH2O'.^2);
kCSsca = sum(integrand(1:131,:)).*pi./(kH2O'.^2);

% relevant cross-section
```

```
%rCSsca= zeros(1000,1340);
%for k=1:1000
%   if mod(k,10)==0
%       fprintf('.');
%   end
%   if mod(k,200)==0                                                   120
%       fprintf('\n');
%   end
%   rCSsca(k,:) = sum(integrand(k:end,:)).*pi./(kH2O'.^2);
%end


rCSsca = sum(integrand(132:end,:)).*pi./(kH2O'.^2);

wtr = load('20050719_water_500mim_30s_10f.txt');
bgd = load('20050719_bg_30s_10f.txt');                                130
PS = load('20050719_PS_taup1_500mim_30s_10f.txt');

mwtr = mean(cosmic_clean_gg(wtr(:,3:end,1)',300,2.5),2);
mbgd = mean(cosmic_clean_gg(bgd(:,3:end,1)',300,2.5),2);
mPS = mean(cosmic_clean_gg(PS(:,3:end,1)',300,2.5),2);

w = mwtr-mbgd;
P = mPS-mbgd;

b = fir1(16,.01);                                                     140
experiment = conv(b,P./w);
experiment = experiment(1:end-16);

N = 1 - exp(-rho.*z.*(tCSsca+tCSsca(28)));
D = z.*rho.*(tCSsca+tCSsca(28));
N2 = 1 - exp(-rho.*z.*(rCSsca+tCSsca(28)));
D2 = z.*rho.*(rCSsca+tCSsca(28));

theory = N./D;
theory2 = N2./D2;                                                     150
pc = polyfit(wl(254:end-8),experiment(262:end),1);
lfit = polyval(pc,wl(254:end-8));

save multi
```

D.4 Graphical User Interface for Online Results

Ferment.m   creates a GUI for viewing online concentration estimates and defines all of the objects in it.

Ferment.m

```
% set environment
clear all;
close all;

global ps compnames tfc t x concrelevent Dkeeps conc xm residual ...
     A B concz dflag lnp mm tim;

%flag indicating if data has been taken before
dflag = 0;
                                                                          10

% default options
lnp = 3;
mm = 1;
tim = 1;

%figure properties
f = figure;
set(f,'numbertitle','off');
set(f,'name','Online Raman Fermenteation Graphical User Interface');
set(f,'menubar','none');                                                  20
set(f,'units','normalized');

%create axes
a_glu = axes('position',[.05 .73 .25 .23],'box','on');
a_phe = axes('position',[.05 .39 .25 .23],'box','on');
a_od  = axes('position',[.05 .05 .25 .23],'box','on');
a_ace = axes('position',[.37 .73 .25 .23],'box','on');
a_for = axes('position',[.37 .39 .25 .23],'box','on');
a_lac = axes('position',[.37 .05 .25 .23],'box','on');
                                                                          30
% format axes
axes(a_glu);
xl_glu = xlabel('time','parent',a_glu,...
    'units','normalized','position',[.5 -.13 0]);
yl_glu = ylabel('concetration','parent',a_glu,...
    'units','normalized','position',[-.11 .5 0]);
ttl_glu = title('Glucose','parent',a_glu,...
    'units','normalized','fontsize',12,'fontweight','bold');
```

```
axes(a_phe);                                                                          40
xl_phe = xlabel('time','parent',a_phe,...
    'units','normalized','position',[.5 -.13 0]);
yl_phe = ylabel('concetration','parent',a_phe,...
    'units','normalized','position',[-.11 .5 0]);
ttl_phe = title('Phenylalinine','parent',a_phe,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_od);
xl_od  = xlabel('time','parent',a_od,...
    'units','normalized','position',[.5 -.13 0]);                                     50
yl_od  = ylabel('OD Value','parent',a_od,...
    'units','normalized','position',[-.11 .5 0]);
ttl_od = title('Optical Density','parent',a_od,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_ace);
xl_ace = xlabel('time','parent',a_ace,...
    'units','normalized','position',[.5 -.13 0]);
yl_ace = ylabel('concetration','parent',a_ace,...
    'units','normalized','position',[-.11 .5 0]);                                     60
ttl_ace = title('Acetate','parent',a_ace,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_for);
xl_for  = xlabel('time','parent',a_for,...
    'units','normalized','position',[.5 -.13 0]);
yl_for  = ylabel('concetration','parent',a_for,...
    'units','normalized','position',[-.11 .5 0]);
ttl_for = title('Formate','parent',a_for,...
    'units','normalized','fontsize',12,'fontweight','bold');                          70


axes(a_lac);
xl_lac  = xlabel('time','parent',a_lac,...
    'units','normalized','position',[.5 -.13 0]);
yl_lac  = ylabel('concetration','parent',a_lac,...
    'units','normalized','position',[-.11 .5 0]);
ttl_lac = title('Lactate','parent',a_lac,...
    'units','normalized','fontsize',12,'fontweight','bold');


                                                                                      80
% Create pushbuttons
btn_cal = uicontrol('style','pushbutton','units','normalized',...
    'fontsize',20,'FontWeight','bold','fontname','times new roman',...
```

```
                'position',[.69 .825 .25 .05],'string','Load Calibration Data',...
                'callback','ferment_cal_button');
btn_bsc = uicontrol('style','pushbutton','units','normalized',...
                'fontsize',10,'FontWeight','bold','fontname','times new roman',...
                'position',[.89 .745 .05 .03],'string','Browse...',...
                'callback','ferment_browse_cal_button');
btn_dat = uicontrol('style','pushbutton','units','normalized',...                        90
                'fontsize',20,'FontWeight','bold','fontname','times new roman',...
                'position',[.69 .49 .25 .05],'enable','off',...
                'string','Update Fermentation Data','callback','ferment_data_button');
btn_bsd = uicontrol('style','pushbutton','units','normalized',...
                'fontsize',10,'FontWeight','bold','fontname','times new roman',...
                'position',[.89 .41 .05 .03],'enable','off',...
                'string','Browse...','callback','ferment_browse_data_button');


% create text boxes
txt_cal = uicontrol('style','text','units','normalized',...                              100
                'position',[.69 .905 .25 .031],'backgroundcolor','red',...
                'fontsize',18,'FontWeight','bold','fontname','times new roman',...
                    'string','No Calibration Data Loaded',...
                'tooltipstring','Click button below to load calibration data');
txt_drc = uicontrol('style','text','units','normalized',...
                'fontsize',10,'FontWeight','normal','fontname','times new roman',...
                'position',[.69 .745 .19 .03],'HorizontalAlignment','right',...
                'string','F:\Stav\MATLAB\ferm_2_9_05\Calibration_data\test3');
txt_dat = uicontrol('style','text','units','normalized',...
                'fontsize',18,'FontWeight','bold','fontname','times new roman',...     110
                'position',[.69 .57 .25 .031],'backgroundcolor','red',...
                'string','Load calibration data first');
txt_drd = uicontrol('style','text','units','normalized',...
                'fontsize',10,'FontWeight','normal','fontname','times new roman',...
                'position',[.69 .41 .19 .03],'HorizontalAlignment','right',...
                'string','F:\Stav\MATLAB\ferm_2_9_05\Raman_data');


% create labels
lbl_drc = uicontrol('style','text','units','normalized',...
                'position',[.69 .775 .133 .024],'backgroundcolor',[0.8 0.8 0.8],...     120
                'fontsize',12,'FontWeight','bold','fontname','times new roman',...
                'string','Calibration files path: ');
lbl_drd = uicontrol('style','text','units','normalized',...
                'position',[.69 .44 .091 .024],'backgroundcolor',[0.8 0.8 0.8],...
                'fontsize',12,'FontWeight','bold','fontname','times new roman',...
                'string','Data files path: ');
lbl_tim = uicontrol('style','text','units','normalized',...
                'position',[.69 .1 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
```

```
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'horizontalalignment','left','string','time units:');                    130
lbl_mol = uicontrol('style','text','units','normalized',...
        'position',[.69 .2 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'horizontalalignment','left','string','concentration units:');
lbl_lin = uicontrol('style','text','units','normalized',...
        'position',[.69 .3 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'horizontalalignment','left','string','line style:');
```

```
% create radio option buttons                                                    140
rdb_lin = uicontrol('style','radiobutton','units','normalized',...
        'position',[.82 .3 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'string','Lines only','enable','off','value',0,...
        'callback','ferment_lines_radiobutton');
rdb_dot = uicontrol('style','radiobutton','units','normalized',...
        'position',[.82 .279 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'string','Points only','enable','off','value',0,...
        'callback','ferment_points_radiobutton');                               150
rdb_lnd = uicontrol('style','radiobutton','units','normalized',...
        'position',[.82 .258 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'string','Lines and points','enable','off','value',1,...
        'callback','ferment_lines_n_points_rbutton');

rdb_mim = uicontrol('style','radiobutton','units','normalized',...
        'position',[.82 .2 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'string','milliMolarity','enable','off','value',1,...                   160
        'callback','ferment_millimolar_rbutton');
rdb_mol = uicontrol('style','radiobutton','units','normalized',...
        'position',[.82 .179 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...
        'string','Grams/Liter','enable','off','value',0,...
        'callback','ferment_molar_rbutton');

rdb_hrs = uicontrol('style','radiobutton','units','normalized',...
        'position',[.82 .1 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
        'fontsize',12,'FontWeight','bold','fontname','times new roman',...      170
        'string','Hours','enable','off','value',1,...
        'callback','ferment_hours_rbutton');
rdb_min = uicontrol('style','radiobutton','units','normalized',...
```

```
'position',[.82 .079 .121 .024],'backgroundcolor',[0.8 0.8 0.8],...
'fontsize',12,'FontWeight','bold','fontname','times new roman',...
'string','Minutes','enable','off','value',0,...
'callback','ferment_minutes_rbutton');
```

ferment_get_OD.m uses the concentration estimate of water to compute the OD, as described in Section 3.3.

ferment_get_OD.m

function OD = ferment_get_OD(normwater)
%function OD = ferment_get_OD(normwater)

od = −**log**(normwater);

%p = [-2.9276831516201108e-005 ...
%   9.0109294370817499e-004 ...
%   -1.3136269067777562e-002 ...
%   1.4286296729834461e-001 ...
%   -2.4880157964629778e-002];

p = [−1.071255349468067e+002 ...
        1.676477062914177e+002 ...
       −6.646131693645528e+001 ...
        1.626102622239855e+001 ...
        1.019551494851869e−001];

OD = **polyval**(p,od);

ferment_hours_rbutton.m defines the "hours" radiobutton.

ferment_hours_rbutton.m

set(rdb_hrs,'value',1);
set(rdb_min,'value',0);

tim = 1;

if dflag
    ferment_replot;
end

ferment_lines_n_points_rbutton.m defines the "lines and points" radiobutton.

ferment_lines_n_points_rbutton.m

```
set(rdb_lin,'value',0);
set(rdb_dot,'value',0);
set(rdb_lnd,'value',1);

lnp = 3;

if dflag
    ferment_replot;
end
```

`ferment_lines_rbutton.m` defines the "lines" radiobutton.

ferment_lines_radiobutton.m

```
set(rdb_lin,'value',1);
set(rdb_dot,'value',0);
set(rdb_lnd,'value',0);

lnp = 1;

if dflag
    ferment_replot;
end
```

`ferment_millimolar_rbutton.m` defines the "millimolarity" radiobutton.

ferment_millimolar_rbutton.m

```
set(rdb_mol,'value',0);
set(rdb_mim,'value',1);

mm = 1;

if dflag
    ferment_replot;
end
```

`ferment_minutes_rbutton.m` defines the "minutes" radiobutton.

ferment_minutes_rbutton.m

```
set(rdb_mol,'value',1);
set(rdb_mim,'value',0);

mm = 2;

if dflag
    ferment_replot;
end
```

ferment_molar_rbutton.m defines the "molarity" radiobutton.

ferment_molar_rbutton.m

```
set(rdb_mol,'value',1);
set(rdb_mim,'value',0);

mm = 2;

if dflag
    ferment_replot;
end
```

ferment_points_rbutton.m defines the "points" radiobutton.

ferment_points_radiobutton.m

```
set(rdb_lin,'value',0);
set(rdb_dot,'value',1);
set(rdb_lnd,'value',0);

lnp = 2;

if dflag
    ferment_replot;
end
```

ferment_relevent.m extracts the relevant concentration estimates from the concentration matrix for plotting.

ferment_relevent.m

```
function concrelevent = ferment_relevent(compnames,conc);

concrelevent = zeros(5,size(conc,2));
wat_mask = strcmpi('Water',compnames);
glu_mask = strcmpi('Glucose',compnames);
ace_mask = strcmpi('Acetate',compnames);
for_mask = strcmpi('Formate',compnames);
lac_mask = strcmpi('Lactate',compnames);
phe_mask = strcmpi('Phenylalanine',compnames);

for I = 1:length(compnames)
    if wat_mask(I)
        concrelevent(1,:) = conc(I,:);
    elseif glu_mask(I)
        concrelevent(2,:) = conc(I,:);
    elseif ace_mask(I)
```

10

```
            concrelevent(3,:) = conc(I,:);
        elseif for_mask(I)
            concrelevent(4,:) = conc(I,:);
        elseif lac_mask(I)                                        20
            concrelevent(5,:) = conc(I,:);
        elseif phe_mask(I)
            concrelevent(6,:) = conc(I,:);
        end
    end
```

ferment_replot.m   is envoked anytime a change is made. This function plots all of the data using the specified options.

ferment_replot.m

```
    global ps compnames tfc t x concrelevent Dkeeps conc xm residual ...
        A B concz dflag lnp mm tim;


    if lnp == 1
        cstring = '-r';
    elseif lnp == 2
        cstring = 'ro';
    elseif lnp == 3
        cstring = '-ro';                                          10
    end

    if mm == 1
        ymult = [1 1 1 1 1];
    elseif mm == 2
        ymult = 1e-3.*[180.2 136.08 68.02 112.06 165.19];
    end

    if tim == 1
        tmult = 1;                                                20
    elseif tim == 2
        tmult = 60;
    end

    atten = concrelevent(1,:)./1000;
    newconc = concrelevent./repmat(atten,6,1);

    OD = ferment_get_OD(atten.*1000./concrelevent(1,1));

    axes(a_glu)                                                   30
```

```
plot(t.*tmult,newconc(2,:).*ymult(1),cstring);
xl_glu  = xlabel('time','parent',a_glu,...
    'units','normalized','position',[.5 -.13 0]);
yl_glu  = ylabel('concetration','parent',a_glu,...
    'units','normalized','position',[-.11 .5 0]);
ttl_glu = title('Glucose','parent',a_glu,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_ace)
plot(t.*tmult,newconc(3,:).*ymult(2),cstring);                    40
xl_ace  = xlabel('time','parent',a_ace,...
    'units','normalized','position',[.5 -.13 0]);
yl_ace  = ylabel('concetration','parent',a_ace,...
    'units','normalized','position',[-.11 .5 0]);
ttl_ace = title('Acetate','parent',a_ace,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_for)
plot(t.*tmult,newconc(4,:).*ymult(3),cstring);
xl_for  = xlabel('time','parent',a_for,...                        50
    'units','normalized','position',[.5 -.13 0]);
yl_for  = ylabel('concetration','parent',a_for,...
    'units','normalized','position',[-.11 .5 0]);
ttl_for = title('Formate','parent',a_for,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_lac)
plot(t.*tmult,newconc(5,:).*ymult(4),cstring);
xl_lac  = xlabel('time','parent',a_lac,...
    'units','normalized','position',[.5 -.13 0]);                 60
yl_lac  = ylabel('concetration','parent',a_lac,...
    'units','normalized','position',[-.11 .5 0]);
ttl_lac = title('Lactate','parent',a_lac,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_phe)
plot(t.*tmult,newconc(6,:).*ymult(5),cstring);
xl_phe  = xlabel('time','parent',a_phe,...
    'units','normalized','position',[.5 -.13 0]);
yl_phe  = ylabel('concetration','parent',a_phe,...                70
    'units','normalized','position',[-.11 .5 0]);
ttl_phe = title('Phenylalinine','parent',a_phe,...
    'units','normalized','fontsize',12,'fontweight','bold');


axes(a_od)
```

```
plot(t.*tmult,OD,cstring);
xl_od  =  xlabel('od value','parent',a_od,...
    'units','normalized','position',[.5 -.13 0]);
yl_od  =  ylabel('concetration','parent',a_od,...
    'units','normalized','position',[-.11 .5 0]);
ttl_od  =  title('Optical Density','parent',a_od,...
    'units','normalized','fontsize',12,'fontweight','bold');
```

ferment_browse_cal_button.m   defines the "Browse" pushbutton under the "Load Calibration

Data" pushbutton.

ferment_browse_cal_button.m

```
def  =  char({get(txt_drc,'string')});
answer  =  uigetdir(def,'Enter Folder Name');

a  =  char(answer);

if ~isempty(a)
        set(txt_drc,'string',a);
end
```

ferment_browse_data_button.m   defines the "Browse" pushbutton under the "Update Fermen-

tation Data" pushbutton.

ferment_browse_data_button.m

```
def  =  char({get(txt_drd,'string')});
answer  =  uigetdir(def,'Enter Folder Name');

a  =  char(answer);

if ~isempty(a)
        set(txt_drd,'string',a);
end
```

ferment_cal_button.m   defines the "Load Calibration Data" pushbutton.

ferment_cal_button.m

```
global ps compnames tfc t x concrelevent Dkeeps conc xm residual A B ...
    concz dflag lnp mm tim compdata waterdata compconc;

dirname  =  get(txt_drc,'string');
```

```
[ps,compnames,tfc, compdata, waterdata, compconc] = ferment_calibration_read(dirname);

set(btn_dat,'enable','on');
set(btn_bsd,'enable','on');
set(rdb_lin,'enable','on');                                                10
set(rdb_dot,'enable','on');
set(rdb_lnd,'enable','on');

set(rdb_mim,'enable','on');
set(rdb_mol,'enable','on');

set(rdb_hrs,'enable','on');
set(rdb_min,'enable','on');

set(txt_cal,'string','Calibration data loaded','backgroundcolor','green');    20
set(txt_dat,'string','Ready for the first data point','backgroundcolor','green');
```

`ferment_calibration_read.m` reads all of the calibration files in the directory shown and generates the pure component matrix. This will be invoked whenever the "Load Calibration Data" pushbutton is pressed.

ferment_calibration_read.m

```
function [ps,compnames,tfc, compdata, waterdata, compconc] = ...
    ferment_calibration_read(dirname)

% This is a new master version of calibration read that automatically generate
% the correct calibration matrix and appropriate meta-data for the appropriate
% concentrations/data plots to be made.

if dirname(end)=='\' | dirname(end)=='/'; dirname = dirname(1:end-1); end

D = dir(dirname);                                                          10

h = waitbar(0,'loading calibration data, please wait');
set(h,'units','normalized','position',[.36 .465 .28 .07],'name','Loading...');

% Look for a list of component names
J = 1; % component index
K = 1; % water index
M = 1; % acetone index
N = 1; % empty index
P = 1; % acewater                                                          20
Q = 1; % empwater
```

```matlab
compnames{1} = 'Water';

for I = 1:length(D)


    if isempty(findstr(lower(D(I).name),'acetone')) & ...
          isempty(findstr(lower(D(I).name),'empty')) & ...
          isempty(findstr(lower(D(I).name),'water')) & ...
          isempty(findstr(lower(D(I).name),'dark')) & ...
          isempty(findstr(lower(D(I).name),'acewater')) & ...
          ~isempty(findstr(lower(D(I).name),'_'))
      indus = findstr(D(I).name,'_');
      compnames{J+1} = D(I).name(1:indus(1)-1);
      compconc(J) = str2num(D(I).name(indus(1)+1:indus(2)-1));
      compwat(J) = str2num(D(I).name(indus(2)+1:indus(3)-1));
      compind(J) = I;
      x = load([dirname,'\',D(I).name]);
      compdata(:,J) = mean(cosmic_clean(x(:,3:end).'),2);
      J = J+1;
    end


    if findstr(lower(D(I).name),'water') & ...
          isempty(findstr(lower(D(I).name),'acewater')) & ...
          isempty(findstr(lower(D(I).name),'empwater'))
      waterind(K) = I;
      x = load([dirname,'\',D(I).name]);
      waternames{K} = D(I).name;
      waterdata(:,K) = mean(cosmic_clean(x(:,3:end).'),2);
      K = K+1;
    end


    if findstr(lower(D(I).name),'acewater')
      acewaterind(P) = I;
      x = load([dirname,'\',D(I).name]);
      acewaternames{P} = D(I).name;
      acewaterdata(:,P) = mean(cosmic_clean(x(:,3:end).'),2);
      P = P+1;
    end


    if findstr(lower(D(I).name),'acetone')
      acetoneind(M) = I;
      x = load([dirname,'\',D(I).name]);
      acetonenames{M} = D(I).name;
      acetonedata(:,M) = mean(cosmic_clean(x(:,3:end).'),2);
```

30

40

50

60

```
    M = M+1;
  end

  if findstr(lower(D(I).name),'empty')                                70
    emptyind(N) = I;
    x = load([dirname,'\',D(I).name]);
    emptynames{N} = D(I).name;
    emptydata(:,N) = mean(cosmic_clean(x(:,3:end).'),2);
    N = N+1;
  end

  if findstr(lower(D(I).name),'empwater')
    empwaterind(Q) = I;
    x = load([dirname,'\',D(I).name]);                                80
    empwaternames{Q} = D(I).name;
    empwaterdata(:,Q) = mean(cosmic_clean(x(:,3:end).'),2);
    Q = Q+1;
  end

  waitbar(I/length(D),h);

end

compnames = compnames.';                                              90
compind = compind(:);
compconc = compconc(:);
compwat = compwat(:);
acetonenames = acetonenames.';
acetoneind = acetoneind(:);
waternames = waternames.';
waterind = waterind(:);
acewaternames = acewaternames.';
acewaterind = acewaterind(:);
emptynames = emptynames.';                                           100
emptyind = emptyind(:);
empwaternames = empwaternames.';
empwaterind = empwaterind(:);

% Since dir results in alphabeticallly ordered lists, they acetone and empty data
% should be aligned.

% subtract dark background all data

x = load([dirname,'\','dark_background.txt']);                       110
darkback = mean(cosmic_clean(x(:,3:end).'),2);
```

```
compdata = compdata-darkback*ones(size(compdata(1,:)));
waterdata = waterdata-darkback*ones(size(waterdata(1,:)));
acetonedata2 = acetonedata-emptydata;
acetonedata3 = acetonedata-darkback*ones(size(acetonedata(1,:)));
%emptydata = emptydata-darkback*ones(size(emptydata(1,:)));
acewaterdata2 = acewaterdata-empwaterdata;
acewaterdata3 = acewaterdata-darkback*ones(size(acewaterdata(1,:)));
%empwaterdata = empwaterdata-darkback*ones(size(empwaterdata(1,:)));          120


% Now generate the water correction vectors and calculate the pure spectra
% The idea is to use the acewater data and acetone data to normalize the water contribution
% to the correct level

for I = 1:length(compconc)
  if compwat(I)
    watnorm(I) = mean(acetonedata(85:end,I)./acewaterdata(85:end,compwat(I)));
    %watnorm2(I) = mean(acetonedata2(85:end,I)./acewaterdata2(85:end,compwat(I)));     130
    %watnorm3(I) = mean(acetonedata2(85:end,I)./acewaterdata3(85:end,compwat(I)));
    ps(:,I) = (compdata(:,I)-waterdata(:,compwat(I)).*watnorm(I))./compconc(I);
    %ps2(:,I) = (compdata(:,I)-waterdata(:,compwat(I)))./compconc(I);
    %ps2(:,I) = (compdata(:,I)-waterdata(:,compwat(I)).*watnorm2(I))./compconc(I);
    %ps3(:,I) = (compdata(:,I)-waterdata(:,compwat(I)).*watnorm3(I))./compconc(I);
  else
    ps(:,I) = compdata(:,I)./compconc(I);
    %ps2(:,I) = ps(:,I);
    %ps3(:,I) = ps(:,I);
  end                                                                          140
end


% add water to the first column of ps
ps = [mean(waterdata,2)./1000,ps];
ps = ps(85:end,:);
%ps2 = [mean(waterdata,2)./1000,ps2];
%ps2 = ps2(85:end,:);
%ps3 = [mean(waterdata,2)./1000,ps3];
%ps3 = ps3(85:end,:);

                                                                             150
% for now, don't do any white light correction (not necessary);
tfc = ones(1340,1);


%toddo = load('toddo.txt');
%tod = toddo(:,1);
%OD = toddo(:,2);
```

```
%DO  =  toddo(:,3);
%dcw  =  OD./8.8.*4.269;
```

close(h);   160

ferment_data_button.m   defines the "Update Fermentation Data" pushbutton.

ferment_data_button.m

```
global ps compnames tfc t x concrelevent Dkeeps conc xm residual ...
    A B concz dflag lnp mm tim;

dirname = get(txt_drd,'string');

if dflag == 0
  [t,x,Dkeeps] = ferment_data_read(dirname,tfc);
  dflag = 1;
else
  [t,x,Dkeeps] = ferment_data_read_cum(dirname,tfc,x);                        10
end


[conc,xm,residual,A,B,concz] = ferment_conc_extract(x,ps);

concrelevent = ferment_relevent(compnames,conc);

ferment_replot;
```

ferment_data_read.m   reads all of the data files in the directory shown and generates the measured spectrum matrix. This will be invoked whenever the "Update Fermentation Data" pushbutton is pressed for the first time.

ferment_data_read.m

```
function [t,x,Dkeeps] = ferment_data_read(dirname,tfc)

h = waitbar(0,'loading online data, please wait');
set(h,'units','normalized','position',[.36 .465 .28 .07],'name','Loading...');

D = dir(dirname);
ind = 0;
for I = 1:length(D)
  if ~isempty(findstr('.SPE',D(I).name)) & length(D(I).name)>4,
    ind = ind+1;                                                                       10
    Dkeep(ind) = D(I);
    dn(ind) = datenum(D(I).date);
```

```
      end
   end

   [dns,inds]=sort(dn);
   Dkeeps  =  Dkeep(inds);

   t  =  [0,cumsum(diff(dns))*24];
                                                                              20
   for I  =  1:length(Dkeeps)
      y  =  load([dirname,'\',Dkeeps(I).name(1:end-4),'.txt']);
      x(:,I)  =  mean(cosmic_clean(y(:,3:end).'),2).*tfc;
      waitbar(I/length(Dkeeps),h);

%      if strcmp('BR1_02_09_30s_10f_t00p00',Dkeeps(I).name(1:end-4))
%          ind_time_1  =  I;
%      end

   end                                                                        30

   ind_time_1  =  1;
   t  =  t-t(ind_time_1);

   x  =  x(85:end,:);
   close(h);
```

ferment_data_read_cun.m reads only the new data files in the directory shown and generates

the measured spectrum matrix. This will be invoked whenever the "Update Fermentation Data"

pushbutton is pressed other than the first time.

ferment_data_read_cum.m

```
   function [t,x,Dkeeps]  =  ferment_data_read_cum(dirname,tfc,data)

   h  =  waitbar(0,'loading online data, please wait');
   set(h,'units','normalized','position',[.36 .465 .28 .07],'name','Loading...');

   D  =  dir(dirname);
   ind  =  0;
   for I  =  1:length(D)
      if ~isempty(findstr('.SPE',D(I).name)) & length(D(I).name)>4,
         ind  =  ind+1;                                                        10
         Dkeep(ind)  =  D(I);
         dn(ind)  =  datenum(D(I).date);
      end
   end
```

```
[dns,inds]=sort(dn);
Dkeeps = Dkeep(inds);

t = [0,cumsum(diff(dns))*24];
                                                                            20
for I = 1:length(Dkeeps)
   if strcmp('BR1_02_09_30s_10f_t00p00',Dkeeps(I).name(1:end−4))
      ind_time_1 = I;
   end
end


if length(Dkeeps) > size(data,2)
    for I = 1:(length(Dkeeps)−size(data,2))
       y = load([dirname,'\',Dkeeps(I+size(data,2)).name(1:end−4),'.txt']);     30
       x(:,I) = mean(cosmic_clean(y(:,3:end).'),2).*tfc;
       waitbar(I/(length(Dkeeps)−size(data,2)),h);
    end
    x = [data,x(85:end,:)];
else
    x = data;
end

t = t−t(ind_time_1);
                                                                            40

close(h);
```

ferment_conc_extract.m estimates the concentration estimates according to the algorithm described in Chapter 2 using the pure component matrix from ferment_calibration_read.m and the measured spectrum matrix from any one of the data read files shown above. This will be invoked whenever the ferment_replot.m function is invoked.

ferment_conc_extract.m

```
function [conc,xm,residual,A,B,concz] = ferment_conc_extract(x,ps,indselect,pord)
% [conc,xm,residual,A,B,concz] = conc_extract(x,ps,indselect);

if nargin<3, indselect = 1:size(ps,2);pord=4;end;
if nargin<4, pord = 4;end

ps = ps(:,indselect);
```

```
wna = linspace(0,1,size(ps,1)).' ;
%   if nopoly                                                          10
%      A = [ps];
%      else
% % A = [ps,ones(size(ps,1),1),10.*[1-wna,1-wna.^2,1-wna.^3,1-wna.^4]];
% pord = 3;
%        A = [ps,ones(size(ps,1),1),10.*[1-wna,1-wna.^2,1-wna.^3]];
% %    A = [ps,ones(size(ps,1),1),10.*[1-wna]];
%   end
%
 A = ps;
 if pord>=0                                                           20
   A = [ps,ones(size(ps,1),1)];
       for I = 1:pord
       A = [A,50.*(1-wna.^I)];
         end
 end


[U,S,V]=svd(A);
S = S(1:size(V,1),:);
U = U(:,1:size(V,2));
                                                                      30

B = U*diag(1./diag(S))*V.' ;


conc_svd = B.'*x;

% for I = 1:size(x,2)
%    conc_temp = lsqlin(A,x,ones(1,size(A,2)),inf,[],[],...
%                      [zeros(size(ps,2),1);-inf*ones(5,1)],inf);
%    conc_lsq(:,I) = conc_temp;
% end
                                                                      40

concz = conc_svd;
concc = concz(1:size(ps,2),:);
concc(concc<0) = 0;
if pord < 0
  conct = concc;
else
  conct = [concc;concz(end-pord:end,:)];
end
conc = conct;                                                         50

xm = A*conct;
```

residual = x−xm;

D.5   Miscellaneous

`cosmic_clean_gg.m`   Is used by the data acquisition functions in Section D.1 to remove cosmic rays.

cosmic_clean_gg.m

```
function y = cosmic_clean_gg(x,thresh,stdthresh)
% function y = cosmic_clean_gg(x,thresh)
%
% Cosmic Ray removal for multiple frame data or, single vector
% data.  This works better with multiple frame data


% If x is a matrix where columns are spectra, look
% across rows
if nargin<2, thresh = 300;stdthresh = 2;end;                       10
if nargin < 3, stdthresh = 2;end;
%stdthresh
y = x;
if size(x,2) > 1
        for J = 1:size(x,1)
                ind = peakfind_gg(x(J,:),stdthresh*std(x(J,:)));
                y(J,ind) = (x(J,:)*ones(size(x(J,:).')))-. . .
                        x(J,ind)*ones(size(x(J,ind).'))) ./ . . .
                        (length(x(J,:))−length(ind));
                                                                   20
                %mean(x(J,setdiff(1:size(x,2),ind)));
        end;
        % interleave and repeat
        iint = [1:2:size(x,2),2:2:size(x,2)];
        x = y(:,iint);
        for J = 1:size(x,1)
                ind = peakfind_gg(x(J,:),stdthresh*std(x(J,:)));
                y(J,iint(ind)) = (x(J,:)*ones(size(x(J,:).')))-. . .
                        x(J,ind)*ones(size(x(J,ind).'))) ./ . . .
                        (length(x(J,:))−length(ind));              30
                %mean(x(J,setdiff(1:size(x,2),ind)));
        end;

else
```

```
% for single vector, linearly interpolate between points
ind = peakfind(x,thresh);
if ind(1)==1, ind = ind(2:end);end
if ind(end)== length(x), ind = ind(1:end-1);end;
y(ind) = (y(ind-1)+y(ind+1))./2;
```
                                                                                                     40
```
end
```

peakfind_gg.m is used by cosmic_clean_gg.m to find the cosmic ray peaks.

peakfind_gg.m

```
function ind = peakfind_gg(x,thresh);
% ind = peakfind_gg(x);
% finds the peaks
if nargin < 2, thresh = min(x);end;

ind = [];

%for I = 2:length(x)-1
% if x(I)>=x(I-1) & x(I) >= x(I+1) & x(I)>=thresh, ind = [ind,I]; end
%end
```
                                                                                                     10
```
ind = find(x(2:end-1)>=x(1:end-2) & x(2:end-1)>=x(3:end) ...
    & x(2:end-1)>=(thresh+median(x)))+1;

if  x(1) > x(2)& x(1)>=( thresh + median(x) )
   dummy = ind; ind(1) = 1;
   ind(2:length(dummy)+1) = dummy;
end

if x(end)>x(end-1) & x(end)>=( thresh + median(x) )
   ind(end+1) = length(x);
end
```
                                                                                                     20
```
%ind = find(x(2:end-1)-x(1:end-2)>thresh & x(2:end-1)-x(3:end)>thresh)+1;
%if x(1)-x(2)>thresh, dummy = ind; ind(1) = 1;ind(2:length(dummy)+1) = dummy;end
%if x(end)-x(end-1)>thresh, ind(end+1) = length(x);end
```

rayleigh_shift.m is used by some of the data acquisition functions in Section D.1 to correct

for wavelength shifts by using the Rayleigh line, as discussed in Section 2.3.2. This particular

function is used if the shift amount is not already known.

rayleigh_shift.m

```
function [newspectra, shifts, baselineout] = rayleigh_shift(oldspectra,baseline,upsam,rlen)
% [newspectra, shifts, baselineout] = rayleigh_shift(oldspectra,baseline,upsam,rlen)
% Takes a matrix of raman spectra that include the rayleigh line
% and performs a non integer shift to help match up the lines
% Note: spectra must be the columns
if nargin < 2
    rlen = 50;
    upsam = 100;
    baseline = interp(oldspectra(1:rlen,1), upsam);
elseif nargin < 3                                                          10
    rlen = 50;
    upsam = 100;
elseif nargin < 4
    rlen = 50;
end

fprintf('upsampling...\n');

for I = 1:size(oldspectra,2)
    upspectra(:,I) = interp(oldspectra(:,I), upsam);                       20
    fprintf('%d',size(oldspectra,2)-I+1);
    fprintf('.');
end
fprintf('\ncorrecting...\n');

uprline = upspectra(1:rlen*upsam,:);

for I = 1:size(uprline,2)
    fprintf('%d',size(uprline,2)-I+1);
    fprintf('.');                                                          30
    rcorr = xcorr(baseline,uprline(:,I));
    [m ind] = max(rcorr);
    shifts(I) = ind-rlen*upsam;
    shift = ind-rlen*upsam;
    if shift > 0
        shifted(:,I) = [zeros(shift,1); upspectra(1:end-shift,I)];
    elseif shift < 0
        shifted(:,I) = [upspectra(1+abs(shift):end,I); upspectra(end,I).*ones(abs(shift),1)];
    else
        shifted(:,I) = upspectra(:,I);                                     40
    end
    newspectra(:,I) = decimate(shifted(:,I),100,'FIR');
end
fprintf('\n');
```

baselineout = baseline;

`rayleigh_shift_known.m` is used by some of the data acquisition functions in Section D.1 to correct for wavelength shifts by using the Rayleigh line, as discussed in Section 2.3.2. This particular function is used if the shift amount is already known.

ray_shift_known.m

```matlab
function newspectra = ray_shift_known(oldspectra,shifts,baseline,upsam)

if nargin < 4
   upsam = 100;
end

fprintf('upsampling. . .\n');
for I = 1:size(oldspectra,2)
   upspectra(:,I) = interp(oldspectra(:,I), upsam);
   fprintf('%d',size(oldspectra,2)-I);                              10
   fprintf('.');
end
fprintf('\ncorrecting. . .\n');

for I = 1:size(oldspectra,2)
   fprintf('%d',size(oldspectra,2)-I);
   fprintf('.');
   shift = shifts(I);
   if shift > 0
      shifted(:,I) = [zeros(shift,1); upspectra(1:end-shift,I)];        20
   elseif shift < 0
      shifted(:,I) = [upspectra(1+abs(shift):end,I); upspectra(end,I).*ones(abs(shift),1)];
   else
      shifted(:,I) = upspectra(:,I);
   end
   newspectra(:,I) = decimate(shifted(:,I),100,'FIR');
end
fprintf('\n');
```

Bibliography

[1] Michel Chartrain, Peter M. Salmon, David K. Robinson, and Barry C Buckland. Metabolic engineering and directed evolution for the production of pharmaceuticals. *Current Opinion in Biotechnology*, 11(2):209–214, April 2000.

[2] Thomas Hermann. Industrial production of amino acids by *coryneform* bacteria. *Journal of Biotechnology*, 104(1-3):155–172, Sept. 2003.

[3] Sang Yup Lee, Jong il Choi, and Heng Ho Wong. Recent advances in polyhydroxyalkanoate production by bacterial fermentation: mini-review. *International Journal of Biological Macromolecules*, 25(1-3):31–36, June 1999.

[4] Peter Harms, Yordan Koston, and Govind Rao. Bioprocess monitoring. *Current Opinion in Biotechnology*, 13:124–127, 2002.

[5] E. Stärk, B. Hizmann, K. Schügerl, T. Scheper, C. Fuchs, D. Köster, and H. Märkl. In-situ-fluorescence-probes: A useful tool for non-invasive bioprocess monitoring. *Advances in Biochemical Engineering*, 74:21–38, 2002.

[6] Sakhamuri Sivakesava, Joseph Irudayaraj, and Demirci Ali. Simultaneous determination of multiple components in lactic acid fermentation using FT-MIR, NIR, and FT-Raman spectroscopic techniques. *Process Biochemistry*, 37:371–378, 2001.

[7] Bruce Chase. New generation of raman instruments. *Applied Spectroscopy*, 48(7):14A, July 1994.

195

[8] A. Y. Tikhomirov, J. Munch, T. McKay, and J. Staromlyska. Construction of high-quality narrow-band holographic filters. *Optical and Quantum Electronics*, 31(5):535–543, 1999.

[9] Adrian D. Shaw, Naheed Kaderbhai, Alun Jones, Andrew M. Woodward, Royston Goodacre, Jem J. Rowland, and Douglas B. Kell. Noninvasive, on-line monitoring of the biotransformation by yeast of glucose to ethanol using dispersive raman spectroscopy and chemometrics. *Applied Spectroscopy*, 53(11):1419–1428, November 1999.

[10] Aoife C. McGovern, David Broadhurst, Janet Taylor, Naheed Kaderbhai, MIchael Winson, David A. Small, Jem J. Rowland, Douglas B. Kell, and Royston Goodacre. Monitoring of complex industrial bioprocesses for metabolite concentrations using modern spectroscopies and machine learning: Application to gibberellic acid production. *Biotechnology and Bioengineering*, 78(5):527–538, June 2002.

[11] Harry L.T. Lee, Paolo Boccazzi, Nathalie Gorret, Rajeev J. Ram, and Anthony J. Sinskey. *In situ* bioprocess monitoring of *Escherichia coli* bioreactions using raman spectroscopy. *Vibrational Spectroscopy*, 35(1):131–137, June 2004.

[12] G. Janelt, N. Gerbsch, and R. Buchholz. A novel fiber optic probe for on-line monitoring of biomass concentrations. *Bioprocess Engineering*, 22(3):275–279, March 2000.

[13] Richard L. McCreery. *Raman spectroscopy for Chemical Analysis*, volume 157 of *Chemical Analysis*. Wiley-Interscience, New York, 2000.

[14] C. V. Raman. A new radiation. *Indian Journal of Physics*, 2:387–398, March 1928.

[15] C. V. Raman and K. S. Krishnan. New class of spectra due to secondary radiation. *Indian Journal of Physics*, 2:399–419, July 1928.

[16] Bernard J. Bulkin. *Analytical Raman Spectroscopy*, volume 114 of *Chemical Analysis*. Wiley-Interscience, New York, 1991.

[17] D. A. Long. *Raman Spectroscopy*. McGraw Hill, New York, 1977.

[18] J. R. Ferraro and K. Nakamoto. *Introductory Raman Spectroscopy*. Academic Press, Boston, 1994.

[19] M. Diem. *Introduction to Modern Vibrational Spectroscopy*. Wiley, New York, 1996.

[20] K. A. Harvey and C. J. Myatt. External-cavity diode laser uding a grazing-incidence diffraction grating. *Optics Letters*, 16(12):910–912, June 1991.

[21] A. S. Arnold, J. S. Wilson, and M. G. Boshier. A simple extended-cavity diode laser. *Review of Scientific Instruments*, 69(3):1236–1239, March 1998.

[22] G. Lidén. Understanding the bioreactor. *Bioprocess and Biosystems Engineering*, 24(5):273–279, Jan. 2002.

[23] Fabrice Guillard and Christian Trägårdh. Modeling of the performance of industrial bioreactors with a dynamic microenvironmental approach: A critical review. *Chemical Engineering and Technology*, 22(3):187–195, March 1999.

[24] Sakhamuri Sivakesava, Joseph Irudayaraj, and Demirci Ali. Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. *Journal of Industrial Microbiology and Biotechnology*, 26(4):185–190, April 2001.

[25] M. Kadleíkova, J. Breza, and M. Veselý. Raman spectra of synthetic sapphire. *Microelectronics Journal*, 32(12):955–958, Dec. 2001.

[26] Todd W. Simpson, Qingzhe Wen, Ning Yu, and David R. Clarke. Kinetics of the amorphous $\rightarrow$ γ $\rightarrow$ α transformations in aluminum oxide: Effect of crystallographic orientation. *Journal of the American Ceramic Society*, 81(1):61–66, Jan. 1998.

[27] Philippe Colomban and Mickael Havel. Raman imaging of stress-induced phase transformation in transparent ZnSe ceramic and sapphire single crystals. *Journal of Raman Spectroscopy*, 33(10):789–795, Oct. 2002.

[28] I. H. Malitson. Interspecimen comparison of the refractive index of fused silica. *Journal of the Optical Society of America*, 55(10):1205–1209, Oct. 1965.

[29] Lloyd N. Trefethen and David Bau. *Numeric Linear Algebra*. Siam, Philadelphia, 1997.

[30] Mark van den Brink, Michel Pepers, and Alex M. van Herk. Raman spectroscopy of polymer latexes. *Journal of Raman Spectroscopy*, 33(4):264–272, March 2002.

[31] Per Åke Wedin and Gunilla Wikström. Interpretation and practical use of error propagation matrices. Technical Report UMINF-02.15, Department of Computing Science, Umeå University, Sweden, 2002.

[32] Per Åke Wedin and Gunilla Wikström. First order error analysis of a linear system of equations by use of error propagation matrices connected to the pseudo inverse solution. Technical Report UMINF-02.16, Department of Computing Science, Umeå University, Sweden, 2002.

[33] Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to Probability*. Athena Scientific, 2002.

[34] Mark R. De Graef, Svetlana Alexeeva, Jacky L. Snoep, and M. Joost Teixeira De Mattosi. The steady-state internal redox state (NADH/NAD) reflects the external redox state and is correlated with catabolic adaptation in *Escherichia coli*. *Journal of Bacteriology*, 181(8):2351–2357, April 1999.

[35] H. C. van de Hulst. *Light Scattering by Small Particles*. Dover Press, New York, NY, 1957.

[36] L. P. Bayvel and A. R. Jones. *Electromagnetic Scattering and its Applications*. Applied Science, Englewood, NJ, 1981.

[37] Leung Tsang, Jin Au Kong, and Kung-Hau Ding. *Scattering of Electromagnetic Waves: Theories and Applications*. Series on Remote Sensing. Wiley, New York, 2000.

[38] Thomas Wriedt. A review of elastic light scattering theories. *Particle and Particle Systems Characterization*, 15(2):67–74, Dec. 1998.

[39] Gustav Mie. Optics of turbid media. *Annalan der Physik*, 25(3):377–445, March 1908.

[40] Ferdinando Borhese, Paolo Denti, and Roalba Saija. *Scattering from Model Nonsherical Particles: Theory and Applications to Environmental Physics*. Springer, Berlin, 2003.

[41] Gerard Gouesbat and Gerard Grehan. Sur la généralisation de la théorie de lorenz-mie. *Journal of Optics*, 13(2):97–103, 1982.

[42] Gerard Gouesbet, Gerard Grehan, and B. Maheu. Scattering of a gaussian beam by a Mie scatter center using a Bromwich formalism. *Journal of Optics*, 16(2):88–93, March 1985.

[43] Gerared Grehan, B. Maheu, and Gerard Gouesbet. Scattering of laser beams by Mie scatter centers: numerical results using a localized approximation. *Applied Optics*, 25(19):3539–3548, Oct. 1986.

[44] Gerard Gouesbet, B. Maheau, and Gerard Grehan. Light scattering from a sphere arbitrarily located in a gaussian beam, using a Bromwich formalism. *Journal of the Optical Society of America A*, 5(9):1427–1443, Sept. 1988.

[45] J. P. Barton, D. R. Alexander, and S. A. Schaub. Internal and near surface electromagnetic fields for spherical particle irradiated by a focused laser beam. *Journal of Applied Optics*, 64(Aug.):1632–1639, 4 1988.

[46] E. E. M. Khaled, S. C. Hill, and P. W. Barber. Scattered and internal intensity of a sphere illuminated with a gaussian beam. *IEEE Transactions on Antennas and Propagation*, 41(3):295–303, March 1993.

[47] A. Doicu and Thomas Wriedt. Computation of the beam-shape coefficients in the generalized Lorenz-Mie theory by using the translational addition theorem for spherical vector wave functions. *Applied Optics*, 36(13):2971–2978, May 1997.

[48] James A. Lock and Gerard Gouesbet. Rigorous justification of the localized approximation to the beam shape coefficients in generalized Mie theory. I. On-axis beams. *Journal of the Optical Society of America A*, 11(9):2503–2515, Sept. 1994.

[49] Gerard Gouesbet and James A. Lock. Rigorous justification of the localized approximation to the beam shape coefficients in generalized Mie theory. II. Off-axis beams. *Journal of the Optical Society of America A*, 11(9):2516–2525, Sept. 1994.

[50] Nancy L. Swanson, Barton D. Billard, and Theresa L. Gennaro. Limits of optical transmission measurements with application to particle sizing techniques. *Applied Optics*, 38(27):5887–5893, Sept. 1999.

[51] Nancy L. Swanson and D. B. Billard. Multiple scattering efficiency and optical extinction. *Physical Review E*, 61(4):4518–4522, April 2000.

[52] J. R. Hodkinson. *Aerosol Science*, chapter 10, pages 287–315. Academic Press, New York, NY, 1966.

[53] Xiaoyan Ma, Lu Jun Q, R. Scott Brock, Kenneth M. Jacobs, Ping Yuan, and Xin-Hua Hu. Determination of complex refractive index of polystyrene microspheres from 370 to 1610nm. *Physics in Medicine and Biology*, 48(24):4165–4172, Dec. 2003.

[54] N. E. Gregory and N. F. Thornhill. The effects of aeration and agitation on the measurement of yeast biomass using a laser turbidity probe. *Bioprocess Engineering*, 16(6):339–344, May 1997.

[55] P. M. Kieran, P. F. MacLoughlin, and D. D. Malone. Plant cell suspension cultures: some engineering considerations. *Journal of Biotechnology*, 59(1-2):39–52, Dec. 1997.

[56] Yasoumi Ibaraki and Kenji Kurata. Automation of somatic embryo production. *Plant Cell, Tissue and Organ Culture*, 65(3):179–199, Jan. 2001.

[57] X. Sun and J. C. Linden. Shear stress effects on plant cell suspension cultures in a rotating wall vessel bioreactor. *Journal of Industrial Microbiology and Biotechnology*, 22(1):44–47, Jan. 1999.

[58] Beverly Ingram and Ferda Mavituna. Effect of bioreactor configuration on the growth and maturation of *picea sitchensis* somatic embryo cultures. *Plant Cell, Tissue and Organ Culture*, 61(2):87–96, Jan. 2000.

[59] Alan C. Cassells and Rosario F. Curry. Oxidative stress and physiological, epigenetic and genetic variability in plant tissue culture: implications for micropropagators and genetic engineers. *Plant Cell, Tissue and Organ Culture*, 64(2-3):145–157, Jan. 2001.

[60] S. Certa and S. A. Weitzman. Influence of oxygen radical injury on DNA methylation. *Mutation Research*, 386(2):141–152, Apr. 1997.

[61] Nathalie Gorret, Samsul Kamal bin Rosli, Sheldon F. Oppenheim, Laura B. Willis, Phillip A. Lessard, ChoKyun Rha, and Anthony J. Sinskey. Bioreactor culture of oil palm (*Elaeis guineensis*) and effects of nitrogen source, inoculum size, and conditioned medium on biomass production. *Journal of Biotechnology*, 108(3):253–263, Mar. 2004.

[62] Yijing Xu, Joseph F. Ford, Charles K. Mann, and Thomas J. Vickers. Raman measurement of glucose in bioreactor materials. *Proceedings of the SPIE - The International Society for Optical Engineering*, 2976:10–19, Feb. 1997.

[63] James N. Spencer, George M. Bodner, and Lyman H. Richard. *Chemistry: Structure and Dynamics*. Wiley, New York, 2nd edition, 2003.

[64] Ana Belen Martin, Yolanda Cuardado, Hilario Guerra, Piedad Gallego, Oscar Hita, Luisa Martin, Ana Dorado, and Nieves Villalobos. Differences in the contents of total sugars, reducing sugars, starch and sucrose in embryogenic and non-embryogenic calli from *Medicago arborea L. Plant Science*, 154(2):143–151, May 2000.

[65] Dr. Philip Lessard. Research Scientist, Department of Biology, Massachusetts Institute of Technology. Personal Coomunication, March 2005.

[66] DR. Ahmad Parveez Ghulam Kadir. Research Scientist, Malaysia Palm Oil Board. Personal Communication, June 2005.

[67] Roger Byrne, Jungjae Park, Lynn Ingram, and Tak Hung. Cytometric sorting of *Pinaceae* pollen and its implications for radiocarbon dating and stable isotope analyses. Presented at The 20th Annual Pacific Climate Workshop on Climate Variability of the Eastern North Pacific and Western North America, April 2003.

[68] H. Bohlen, S. Schwengberg, R. Bongaarts, and R. Pulak. Flow cytometry for isolation and optical analysis of embryoid bodies from cultured mouse embryonic stem cells. Presented at 2nd Annual International Society of Stem Cell Researchers Meeting, June 2004.

[69] M. S. Hanson, B. Armann, E. W. Hatch, D. A. Hullett, J. S. Odorico, H. W. Sollinger, and L. A. Fernandez1. Rapid and sensitive assessment of human islet viability by multiparameter flow cytometry on intact and dissociated islets. Presented at International Pancreas and Islet Transplant Association (IPITA) 2005 Meeting, May 2005.

[70] Michael I. Mishchenko and Larry D. Travis. Light scattering by polydispersions of randomly oriented spheroids with sizes comparable to wavelengths of observation. *Applied Optics*, 33(30):7206–7225, Oct. 1994.

[71] Georges R. Fournier and Blair T. N. Evans. Approximation to extinction efficiency for randomly oriented spheroids. *Applied Optics*, 30(15):2042–2048, May 1991.

[72] E. Jaligot, T. Beulé, and A. Rival. Methylation-sensitive RFLPs: characterisation of two oil palm markers showing somaclonal variation-associated polymorphism. *Theoretical and Applied Genetics*, 104(8):1263–1269, April 2002.

3233- 37