

8-2016

# Advances in Raman hyperspectral compressive detection instrumentation for fast label free classification, quantitation and imaging

Bharat R. Makani  
*Purdue University*

Follow this and additional works at: [https://docs.lib.purdue.edu/open\\_access\\_dissertations](https://docs.lib.purdue.edu/open_access_dissertations)



Part of the [Analytical Chemistry Commons](#), and the [Physical Chemistry Commons](#)

---

## Recommended Citation

Makani, Bharat R., "Advances in Raman hyperspectral compressive detection instrumentation for fast label free classification, quantitation and imaging" (2016). *Open Access Dissertations*. 807.  
[https://docs.lib.purdue.edu/open\\_access\\_dissertations/807](https://docs.lib.purdue.edu/open_access_dissertations/807)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

**PURDUE UNIVERSITY**  
**GRADUATE SCHOOL**  
**Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Bharat R. Mankani

Entitled ADVANCES IN RAMAN HYPERSPECTRAL COMPRESSIVE DETECTION  
INSTRUMENTATION FOR FAST LABEL FREE CLASSIFICATION,  
QUANTITATION AND IMAGING

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Dor Ben-Amotz

Garth J. Simpson

Mary J. Wirth

Lynne Taylor

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Dor Ben-Amotz

Approved by Major Professor(s): \_\_\_\_\_

Approved by: Tim Zwier

06/30/2016

Head of the Department Graduate Program

Date



ADVANCES IN RAMAN HYPERSPECTRAL COMPRESSIVE DETECTION  
INSTRUMENTATION FOR FAST LABEL FREE CLASSIFICATION,  
QUANTITATION AND IMAGING

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Bharat R. Mankani

In Partial Fulfillment of the

Requirements of the Degree

of

Doctor of Philosophy

August 2016

Purdue University

West Lafayette, Indiana

*Dedicated to my ancestors. I would not be here without your sacrifices.*

## ACKNOWLEDGMENTS

First off I would like to thank my family. My father, Late Ramkrishan Mankani, and mother Saroj. R. Mankani, without your support and sacrifice I would never have come this far. My brother, Aditya Mankani, and sister-in-law Ria Mankani, thank you for taking over the family responsibilities so that I could pursuing my dreams. Lastly, Dennis and Vinnie Rupani, you have been my family in the U.S. and I really appreciate everything you and your family has done for me.

I would like to acknowledge all my teachers in life. In particular, Dr. Sajid Bashir, my advisor at Texas A&M University- Kingsville, who got me motivated in research and under whom I learned mass spectrometry. I would like to thank my current advisor, Dr. Dor Ben-Amotz, who took me under his wing and gave me the opportunity to learn optics, spectroscopy and Raman instrumentation. I am also grateful to all the professors at Purdue University who have helped get this far. Dr. Bradley Lucier, Dr. Gregory T. Buzzard, Dr. Mary Wirth, Dr. Garth Simpson, and Dr. Mike Ladisch thank you for making me understand the importance of mathematics, optics, chemometrics and business. I would also like to thank the Amy facility staff, Dr. Mike Everly, Tim Selby and Greg Eakins. My Ph.D. work was funded by Office of Naval Research and I am grateful for the opportunity you have provided me.

I feel that graduate school never goes easy without friends. The Ben-Amotz group members have been a huge support for me and I have had some wonderful intellectual conversations with all of you. I have had lots of support from friends outside of my lab as well. In particular, Manish Nepal, Damodar Koirala, Shane Sullivan, and Aditya Ranadive.

Lastly, I am grateful to Purdue University for having such a beautiful campus and all the facilities for the community. I have used the squash courts from my first day at Purdue. Professor Leonard Lipshitz, the advisor for the Purdue Squash Club, thank you for the wonderful organization. I have met some wonderful people while playing squash. I would miss our excellent squash group a lot.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
LIST OF ABBREVIATIONS.....	xviii
ABSTRACT.....	xx
CHAPTER 1 IMPROVEMENTS IN THE OPTIMAL BINARY COMPRESSIVE DETECTION SPECTROMETER DESIGN AND SELECT CLASSIFICATION, QUANTITATION AND IMAGING EXAMPLES.....	1
1.1 Abstract.....	1
1.2 Introduction.....	3
1.3 Methods.....	7
1.3.1 Instrument .....	7
1.3.1.1 Raman Microscope Stage .....	7
1.3.1.2 OB-CD Spectrometer.....	10
1.3.1.3 Galvo-mirrors Optics .....	12
1.3.1.4 Grating .....	13
1.3.1.4.1 Grating Formula and Groove Density Calculation .....	14
1.3.1.5 Digital Micromirror Array .....	16
1.3.1.5.1 Generating Patterns to Project on the DMD .....	18
1.3.1.5.2 Notch Scanning.....	19
1.3.1.5.3 Hadamard Scanning.....	21
1.3.2 Overview of OB-CD Filters.....	23
1.3.3 Bernstein Polynomials to Remove Fluorescence from Pure Chemical Rates.....	25
1.4 Results.....	26
1.4.1 Comparison of 514.5nm and 785nm Excitation Systems .....	26
1.4.2 Fast Chemical Classification.....	27
1.4.3 Chemical Quantitation .....	30
1.4.3.1 Quantitation of Hexane and 1-Hexene.....	31
1.4.3.2 Quantitation of Ethanol in Tequila .....	35



	Page
1.4.4 Imaging using OB-CD Spectrometer.....	38
1.4.4.1 All-On Imaging.....	39
1.4.4.1.1 Resolution of the Image Using 1951 U.S. Airforce Resolution Chart .....	39
1.4.4.2 Hyperspectral Chemical Imaging of Acetaminophen and Benzoic Acid Using Pure Spectra for OB Filter Training. ....	41
1.4.4.3 Hyperspectral Chemical Imaging of Benzoic Acid and Glucose Using MCR Derived Spectra and Removal of Fluorescence by Training on Bernstein Polynomials .....	43
1.4.4.3.1 Hadamard Spectroscopy of Multiple Points on a Line on the Sample.....	43
1.5 Conclusion .....	45
 CHAPTER 2 CLASSIFICATION ERROR ANALYSIS OF FULL SPECTRAL AND COMPRESSIVE DETECTION MEASUREMENTS.....	 47
2.1 Abstract.....	47
2.2 Introduction.....	48
2.3 Methods.....	50
2.3.1 Instrument .....	50
2.3.1.1 Raman Microscope Stage .....	50
2.3.1.2 OBCD Spectrometer .....	52
2.3.1.3 The Czerny Turner Spectrometer: .....	54
2.3.2 Aligning the CCD .....	56
2.3.3 High Speed Spectral Acquisition.....	58
2.3.4 Analysis of full spectral CCD measurements using classical least squares.....	60
2.3.5 Noise Associated with Spectral Measurement Using A CCD.....	62
2.3.6 Complementary optimal binary compressive detection (OBCD2) filters....	65
2.4 Results.....	67
2.4.1 Measured Read noise of the Detector .....	67
2.4.2 Nonlinearity in Signal a Function of Exposure Time .....	68
2.4.3 Acetone and Benzene Classification.....	70
2.4.3.1 Classification of Acetone and Benzene using OB-CD .....	71
2.4.3.2 Classification of Acetone and Benzene using OB-CD2 .....	74
2.4.3.3 Classification of Acetone and Benzene using CCD .....	76
2.4.3.3.1 EM Gain 1 .....	76
2.4.3.3.2 EM Gain 40.....	78
2.4.3.4 % Relative Standard Deviation (%RSD) in Classification Using OB-CD, OB-CD2 and CCD with and without EM-Gain .....	80
2.4.4 Hexane and Methylcyclohexane Classification.....	85
2.5 Conclusion .....	88

	Page
CHAPTER 3 RAMAN MICROSCOPY OF SINGLE CRYSTAL OF HAMMARHEAD RIBOZYME SUSPENDED IN A HANGING DROP .....	90
3.1 Abstract .....	90
3.2 Introduction.....	90
3.3 Methods.....	92
3.3.1 Sample Preparation .....	92
3.3.2 Brief Description of the Instrument .....	93
3.3.3 Data Acquisition Parameters.....	96
3.4 Results.....	97
3.4.1 Analysis of NMP in Water at Different pH .....	97
3.4.2 Analysis of Lysozyme Crystal .....	109
3.4.3 Analysis of HHRz Crystals .....	111
3.5 Conclusion .....	113
REFERENCES .....	115
VITA.....	119
PUBLICATION .....	120

## LIST OF TABLES

Table	Page
Table 2.1: Certificate of performance for ProEM 200B+ eXcelon camera.....	64
Table 3.1: The table shows the pH range and the type of buffer used with all the HHRz Samples. ....	93
Table 3.2: Structure and Pka of NMP's used in the study are shown.....	97

## LIST OF FIGURES

Figure	Page
Figure 1.1: Schematic of the new OB-CD instrument showing the VHG, DMD and two PMT's. The collection lenses after the DMD is not shown in this figure. ....	2
Figure 1.2: Jabonski diagram showing 2 two photon scattering. In Rayleigh scattering process, there is no change in the wavelength (energy) of the incoming and the scattered photons. However, in the Raman scattering process, the scattered photons gain or lose energy compared to the energy of the incoming photons. ....	4
Figure 1.3: Schematic of a generalized Raman instrument showing an incoming light source (laser) interacting with the sample. The Rayleigh scattered photons are filtered and the Raman photons are linearly dispersed into its component wavelengths and the spectrum is measured using a linear array detector. ....	5
Figure 1.4: Raman microscope optics mounted on an Olympus BX51 microscope. ....	9
Figure 1.5: Schematic of the Raman microscope stage. The dichroic mirror allows longer Stokes wavelength to pass through and reflects the laser back towards the source. ....	9
Figure 1.6: Schematic of the OB-CD spectrometer. The inset shows the spectrum of water acquired by notch scanning. ....	11
Figure 1.7: OB-CD Spectrometer optical path showing the incoming Raman beam (orange). One of the separated wavelengths shown in yellow passes through the dispersive element (VHG) and is focused on the DMD. The DMD can direct these photons in either +12° towards one of the PMT or -12° (detector not shown) ....	12
Figure 1.8: 1-D galvomirror system used to scan over the sample by moving the laser beam instead of moving the sample. ....	13
Figure 1.9: Schematic of the VHG. a) Polychromatic light is passed through the grating undiffracted and this is called the 0th order of diffraction. b) The grating is tilted to a certain angle to maximize the diffraction in one of the orders. ....	15
Figure 1.10: Texas Instruments digital micromirror device (DMD) consists of 608X684 mirrors that can be tilted $\pm 12^\circ$ to the normal. ....	16

Figure	Page
Figure 1.11: Schematic of the DMD chipset 3000 and the dimensions of the mirrors and the direction of the tilt is shown.....	17
Figure 1.12: Schematic portrays of the 2-D diffraction pattern of the 633nm light from a DMD. The diffraction pattern is dependent on the angle of incidence and the wavelength of light .....	18
Figure 1.13: 0.3” diagonal Lightcrafter DMD 3000 chipset containing 608 rows and 684 columns of micromirrors with the spectrum spanning from green(515nm) to red (640nm).....	18
Figure 1.14: All off (black) and all on (white) images modulate light towards $-12^\circ$ .....	19
Figure 1.15: 10 mirror notches set to “1” at 0,100,200,300,400 position. The region in the white show DMD mirrors set to $+12^\circ$ and the region in black are $-12^\circ$ .....	20
Figure 1.16: 1280 positive and negative Hadamard masks are applied to recover a 640 wavelength spectrum. Each row of this figure corresponds to a different mask. The 1st mask is all white corresponding to the value 1, and the next is all black with a value of 0 and so forth. ....	22
Figure 1.17: Intensity of multiplexed wavelengths of benzene through 1280 Hadamard masks. Notice the values are symmetrical because the positive and negative masks are complementary. The last 2 masks correspond to all off (background) and all on (sum of all the intensities). ....	23
Figure 1.18: Third order Bernstein polynomial curves using $x=0:341$ are shown. The basis curves are normalized to unit intensity and have positive coefficients. ....	26
Figure 1.19: A comparison of the Hadamard spectrum of benzene and the spectral window using 2 different OB-CD spectrometers.....	27
Figure 1.20: Normalized spectrum of acetone(red) and benzene(blue) are shown. OB filter(grey) generated for each chemical is also shown overlaid with the associated spectrum.....	28
Figure 1.21: Classification of acetone(red) and benzene(blue) using OB filters shown above. Each point is obtained by measuring counts for each chemical through 2 OB filters and recovering the rates. The mean rate for acetone is about $2.5 \cdot 10^6$ photons/s and the mean rate for benzene is $5.5 \cdot 10^6$ photons/s. The variance in the cloud increases with decrease in classification time but the mean rate stays the same. The ellipses represent 95% confidence interval for each classification cloud. Each classification cloud consists of 1000 independent measurements .....	29

Figure	Page
Figure 1.22: Notch spectrum of acetone(red) and benzene(blue) acquired at 10 $\mu$ s per notch filter.....	30
Figure 1.23: 100ms Hadamard spectrum of pure hexane (red), 1-hexene (blue), and 3 arbitrary mixtures are shown. ....	32
Figure 1.24: Normalized spectrum of pure hexane (red) and 1-hexene (blue) and the associated OB filters (gray) .....	33
Figure 1.25: % volume fraction of hexane, 1-hexene and mixtures of the two. Each cloud corresponds to 100 independent measurements made in 100ms total time (time in which both the OB filters were applied).....	35
Figure 1.26: 1s spectrum of pure ethanol (red) and pure water (blue) and known mixtures of ethanol and water is shown. The spectrum of tequila (unknown, dark green) is also shown.....	36
Figure 1.27: Spectrum of pure ethanol (red) and pure water (blue) and the OB filters (gray) associated with each spectrum is shown.....	37
Figure 1.28: Percent volume fraction of ethanol and water is shown in the figure. Volume fractions of 100 independent measurements of pure water (blue) and ethanol (red) and 3 known solutions of ethanol were used to make a calibration curve. Tequila has 36 $\pm$ 2% ethanol according to out measurements .....	38
Figure 1.29: Clockwise from top left: White light microscope image showing the full bar target. Top right image: shows the zoomed in all on image of the bar target. Bottom image: intensity profile of group six and element six analyzed by ImageJ.....	40
Figure 1.30: Independently measured all on images of the USAF bar target shown in false colored in red, green, and blue (RGB). The image in the lower right-hand panel (white) is a composite obtained by superposing the three RGB images.....	41
Figure 1.31: The spectrum of benzoic acid (cyan) and acetaminophen (yellow) and the OB filter for each chemical (gray) is shown.....	42
Figure 1.32: 300X300 point images of acetaminophen (yellow) and benzoic acid (cyan) acquired in 90s (left) and 0.9s (right) by collecting photons for 1ms/point and 0.01ms/point.....	42
Figure 1.33: The figure shows white light image of a 2 component sample. Hadamard spectra of 50 points were acquired along the red line. The spectra in red and green are visually chosen to show the distinct features of each. ....	43

Figure	Page
Figure 1.34: The figure shows Hadamard spectra of pure benzoic acid and glucose Solid and also entropy minimized spectra (dashed) obtained from multipoint Hadamard spectra of a mixture sample. ....	44
Figure 1.35: The figure shows OB filters generated using pure spectra and Bernstein polynomials on the left along with the image constructed using these filters. On the right OB filters were generated using entropy minimized spectra and the image constructed using these filters.....	45
Figure 2.1: The error in classification (%RSD) of acetone using CCD measurements (red) and OBCD easurements (purple) vs total signal (photon counts) of 1000 independent measurements. Analysis of CCD full spectral measurements was done using least squares, and hyperspectral CD measurements were made using OB filters...	48
Figure 2.2: Schematic of the Raman microscope stage is shown. The Raman scattered photons are detected either using the OB-CD spectrometer or the Czerny Turner Spectrometer by flipping a mirror.....	52
Figure 2.3: Schematic of the OB-CD Spectrometer. Collimated beam of light is focused on a slit at the entrance of the spectrometer. The beam is recollimated and dispersed using a VHG. The dispersed beam is focused onto the DMD. The DMD can send parts of all of the light towards either detector.....	54
Figure 2.4: Schematic of the Czerny Turner Spectrometer equipped with an EM-CCD camera. The collimated beam of Raman photons is focused onto the entrance slit of the spectrometer. A recollimating mirror reflects collimated light towards a reflective grating. The dispersed photons are focused onto the CCD detectors using a focusing mirror .....	56
Figure 2.5: A screen shot of Lightfield software used to control the ProEM. The top graph of the figure shows the Spectrum associated with a single row of pixels. Below the graph is the image acquired by exposing the CCD for 0.5s at a read rate of 1MHz in the electron multiplication (EM) mode with EM Gain 1. The most intense spots correspond to the pixels having the most signal. The horizontal line shows the spectrum associated with that row of pixels. The graph on the left of the Image shows the information associated with just that one pixel.....	57
Figure 2.6: Neon spectrum acquired on the ProEM (Red), and on the PMT with the spectrometer slit position 1 and 2 (blue and green). The FWHM of most peaks in the same region of neon spectrum acquired on a CCD and a PMT are comparable. Closing the entrance slit of the spectrometer improves resolution and a slit width at position less than 2 on the OBCD spectrometer was found to be optimal .....	58

Figure	Page
Figure 2.7: Benzene spectrum acquired on the ProEM (Red), and on the PMT with the spectrometer slit position 2 (green). Again the resolution is comparable. The sensitivity of the PMT diminishes at higher wavelength. The two benzene spectra were acquired at the same exposure time of 0.4s. The CCD had a gain setting of medium (gain 2 on a scale of 3). This may be the reason in the higher counts .....	58
Figure 2.8: Spectrum of acetone was acquired at 10ms exposure time and 1000 independent acquisitions show a continual increase in the baseline. The spectrum in black is the 1st experiment and the light red spectrum is the last experiment.....	59
Figure 2.9: Electronic baseline shift between experiments is corrected by subtracting the minimum of value of the spectrum from the whole spectrum and each experiment ..	60
Figure 2.10: Normalized pure spectra used in used in the classification of acetone and benzene .....	61
Figure 2.11: Pseudo inverse of pure spectra used in used in the classification of acetone and benzene.....	62
Figure 2.12: The figure shows one pair of OB-CD2 filters for acetone and benzene. Notice that they are complementary to each other and only one of these is needed to be applied to the DMD to make count measurements on both the detectors. ....	66
Figure 2.13: Read noise vs read rate is shown at different analog gains. The read noise increases as a factor of readout rate .....	68
Figure 2.14: Spectrum of acetone was measured at different exposure time and constant laser power and 8MHz readout speed. The figure shows a non-linearity in signal with decrease in exposure time .....	69
Figure 2.15: Log of the intensity of acetone peak at 550 pixels ( $\sim 2900\text{cm}^{-1}$ ) is plotted against the log of exposure time (red). The figure shows a non-linearity in signal with decrease at exposure times faster than 5ms .....	70
Figure 2.16: The area normalized spectrum of pure acetone(red) and pure benzene (blue) is acquired on a CCD.....	71
Figure 2.17: The spectrum of pure acetone(red) and the OB filters(gray) associated with it is shown. The training spectrum was acquired by recovering a 1s Hadamard spectrum.....	72
Figure 2.18: The spectrum of pure benzene(blue) and the OB filters(gray) associated with it is shown. The training spectrum was acquired by recovering a 1s Hadamard spectrum.....	72



Figure	Page
Figure 2.19: The figure shows the recovered rates for acetone and benzene using OB-CD to classify the two chemicals. Each cloud consists of 1000 independent measurement acquired at laser power of 30mW. There are three clouds for each chemical showing for each of the measurement times (100ms, 10ms and 1ms).....	73
Figure 2.20: The figure shows OB-CD2 filters for shaded red and blue. The spectrum of acetone (red) and benzene (blue) are also overlaid .....	74
Figure 2.21: Recovered rates for acetone and benzene using OB-CD2 to classify the two chemicals. One filter (and its complement) was used to classify the two chemicals using counts measured on 2 PMT's. Each cloud consists of 1000 independent measurement acquired at laser power of 30mW. There are three clouds for each chemical showing for each of the measurement times (100ms, 10ms and 1ms).....	75
Figure 2.22: S Matrix consists of area normalized spectrum of acetone(red), benzene(blue), and constant(black) is used as the input spectra for least squares classification of acetone and benzene .....	77
Figure 2.23: Classification clouds of acetone (shades of red) and benzene (shades of blue) are shown. Each cloud corresponds to 1000 independent measurements and the lighter shade of the color signifies faster exposure time.....	78
Figure 2.24: Spectrum of benzene measured using a CCD with (blue) and without EM gain (black) applied at laser power of 1mW and an exposure time of 1ms. The photons that get buried in the noise floor due to the amplitude of the read noise can be amplified using the EM gain.....	79
Figure 2.25: Spectrum of benzene recovered from 1ms Hadamard spectrum collected on a PMT.....	80
Figure 2.26: Comparison of %RSD in classification of acetone using least squares on full spectral measurements of a CCD (red) and compressive detection using optimal binary filtering of chemical information (purple) as a factor of total photon counts photon counts is shown. Also shown is the spectrum of acetone measured on a CCD at 61, 610 and 6000 total measured all on counts.....	82
Figure 2.27: Comparison of %RSD in classification of benzene using least squares on full spectral measurements of a CCD (red) and compressive detection using optimal binary filtering of chemical information (purple) as a factor of total photon counts photon counts is shown. Also shown is the spectrum of acetone measured on a CCD at 90, 900 and 9000 total measured all on counts.....	82

Figure	Page
Figure 2.28: Classification ellipses of acetone and benzene at 1ms and laser power of 1mW is shown. On the left is the classification of acetone and benzene using least squares to analyze full spectral measurement. On the right is the OB-CD classification. The total counts of acetone and benzene are 61 and 90 respectively. ....	83
Figure 2.29: Summary of %Relative standard deviation (RSD) in the classification of acetone using OB-CD, OB-CD2, and least squares for CCD and EMCCD data. ....	84
Figure 2.30: Summary of %RSD in the classification of benzene using OB-CD, OB-CD2, and least squares for CCD and EMCCD data. ....	84
Figure 2.31: Spectra of hexane (red) and methylcyclohexane (blue) are shown. a) OB-CD filters associated with hexane and methylcyclohexane are shown in shaded red and blue. b) OB-CD2 filters shown in shaded red and blue are shown. OB-CD filter measure the counts under the spectrum and are sequentially applied and the counts are measured on one of the two PMT's. However, OB-CD2 filters are complements of each other and the counts through each filter is measured simultaneously on two PMT's in the +12° and -12° direction. ....	86
Figure 2.32: Summary of %RSD using OB-CD, OB-CD2, and least squares for CCD and EMCCD data. a) Comparison of % RSD in the classification of hexane, and b) Comparison of % RSD in the classification of methylcyclohexane. ....	87
Figure 3.1: Secondary structure of Hammarhead ribozyme (HHRz) .....	92
Figure 3.2: Argon ion excitation laser passes through a laser line cleanup filter and a Soleil Babinet Compensator. The compensator is an optical element that is setup to allow a certain polarization of light to pass through. The laser is guided through a mirror towards the sample stage .....	94
Figure 3.3: The polarized laser beam is guided through mirrors and a dichroic beam splitter towards the 20X horizontally mounted objective. The sample is in a cuvette held on a temperature controlled sample stage. The Raman scattered photons are collected back through the same object and pass through the beam splitter. Any residual Rayleigh scattered photons are rejected by a long pass filter. The Raman beam is focused on a 100um fiber bundle and guided towards the spectrometer (not shown here). The objective can be mounted vertically by removing the 20X objective and guiding the laser vertically onto a vertically mounter objective.....	95
Figure 3.4: A schematic of vertically positioned 50X objective used to focus the laser onto HHRz crystal suspended in a hanging drop of the buffer. The sample tray is set on a XYZ stage in order to focus the laser on the crystal. A zoomed in schematic on the right shows the laser beam focused on the crystal and the Raman scattered photons being collected by the same objective .....	96

Figure	Page
Figure 3.5: Spectra of CMP at pH 2 (solid red), 4 (solid yellow), 6 (solid green) and 8 (solid red) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500 $\text{cm}^{-1}$ .....	98
Figure 3.6: The spectral range of 1500-1575 $\text{cm}^{-1}$ is shown for spectra of CMP acquired in duplicate at different pH. The spectral feature at 1547 $\text{cm}^{-1}$ seen at pH 2 shifts to 1531 $\text{cm}^{-1}$ due to deprotonation of base at the N1 position .....	99
Figure 3.7: Spectra of AMP at pH 1.5 (solid red), 3.5 (solid yellow), 5.5 (solid green) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500 $\text{cm}^{-1}$ .....	100
Figure 3.8: Background subtracted SC spectra of AMP at different pH shown in the spectral range 550-1850 $\text{cm}^{-1}$ .....	101
Figure 3.9: The spectral range of 1530-1630 $\text{cm}^{-1}$ is shown for spectra of AMP at different pH. The feature at 1562 $\text{cm}^{-1}$ seen in AMP spectra at pH 1.5 shifts to 1582 $\text{cm}^{-1}$ as seen in AMP spectra at pH 5.5. At pH 3.5 the spectrum shows a mixture of the two features. These feature arise from the deprotonation of the nitrogen in position 1 of the base .....	102
Figure 3.10: The pH profile of adenine shown by plotting the normalized intensity ratio of features at 1561 $\text{cm}^{-1}$ and 1581 $\text{cm}^{-1}$ as a function of pH (black triangles) and fit to a sigmoidal curve (solid red). The midpoint of the curve has a pH value of 3.8 which is the literature value for the pKa of nitrogen in position 1 of adenine .....	103
Figure 3.11: Spectra of GMP at pH 7.2 (solid red), 9.2 (solid yellow), 11.2 (solid green) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500 $\text{cm}^{-1}$ . Peaks that show shifts due to pH are identified by their wavenumbers. ....	104
Figure 3.12: Background subtracted SC spectra of GMP at different pH shown in the spectral range 550-1850 $\text{cm}^{-1}$ . Spectrum of water is shown in the dashed blue line. ....	105
Figure 3.13: The spectral range of 1540-1600 $\text{cm}^{-1}$ is shown for spectra of GMP at different pH. The feature at 1575 $\text{cm}^{-1}$ seen in GMP spectra at low pH shift to 1595 $\text{cm}^{-1}$ at higher pH.....	106
Figure 3.14: Spectra of IMP at pH 6.7 (solid red), 8.7 (solid yellow), 10.7 (solid green) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500 $\text{cm}^{-1}$ . Peaks that show shifts due to pH are identified by their wavenumbers. ....	107
Figure 3.15: Background subtracted SC spectra of IMP at different pH shown in the spectral range 550-1850 $\text{cm}^{-1}$ . Spectrum of water is shown in the dashed blue line. ....	108

Figure	Page
Figure 3.16: The spectral range of 1535-1630 $\text{cm}^{-1}$ is shown for spectra of IMP at different pH. The feature at 1550 $\text{cm}^{-1}$ seen in IMP spectra at lower pH shifts to 1565 $\text{cm}^{-1}$ at higher pH. ....	109
Figure 3.17: The Spectrum of Lysozyme crystal suspended in a hanging droplet. ....	110
Figure 3.18: Background subtracted SC spectrum of Lysozyme crystal suspended in a hanging droplet. ....	110
Figure 3.19: White light microscope images of HHRz crystals suspended in a hanging droplet at pH 3.8, 7.4 and 10. ....	111
Figure 3.20: Five-minute average spectrum of HHRz crystal suspended in a hanging drop and the associated buffer at pH 3.8, 7.4, and 10. Due to the range of pH's probed different buffers were used at different pH's. ....	112
Figure 3.21: Background subtracted normalized SC spectra of HHRz collected at different pH. ....	113

## LIST OF ABBREVIATIONS

AMP: Adenosine Monophosphate

APD: Avalanche Photodiode

CCD: Charge Coupled Device

CD: Compressive Detection

CMP: Cytidine Monophosphate

DAC: Data Acquisition

DMD: Digital Micromirror Device

EM: Electron Multiplying

GMP: Guanosine Monophosphate

HHRz: Hammerhead Ribozyme

IMP: Inosine Monophosphate

NI: National Instruments

OB: Optimized Binary

PI: Princeton Instruments

PMT: Photomultiplier Tube

SMCR: Self-Modeling Curve Resolution

TE: Thermoelectric

TI: Texas Instruments

VHG: Volume Holographic Grating

VI: Virtual Instrument

## ABSTRACT

Mankani, Bharat R. PhD., Purdue University, August 2016. Advances in Raman Hyperspectral Instrumentation for Fast Label Free Classification, Quantitation, and Imaging. Major Professor: Dor Ben-Amotz.

Multiple prototypes of hyperspectral compressive detection (CD) Raman spectrometers have previously been constructed in the Ben-Amotz lab and have proven to be useful for fast, label-free chemical identification, quantitation and imaging. The CD spectrometer consists of a volume holographic grating (VHG) that linearly disperses the Raman photons into its component wavelengths and all wavelengths are focused onto a digital micromirror device (DMD). The DMD is an optical modulator that consists of an array of programmable  $10\mu\text{m}$  mirrors that can reflect photons in either  $+12^\circ$  or  $-12^\circ$  to the incoming light. The DMD is tilted such that the  $+12^\circ$  photons go back through the focusing lens and the VHG and is focused onto a single  $150\mu\text{m}$  photon counting avalanche photodiode detector (APD).

In chapter 1 of the thesis I describe the construction of a new CD Raman spectrometer capable of fast hyperspectral imaging that has better photon collection efficiency and fewer photon losses compared to its predecessors. The new spectrometer consists of a VHG and a DMD, however, the DMD is not tilted but is perpendicular to the

incoming Raman photons. All the Raman photons modulated by the DMD are symmetrically detected in the  $+12^\circ$  and  $-12^\circ$  by two photon counting photomultiplier tube(PMT) detector modules. The new spectrometer avoids a double pass through the optics and hence has fewer losses associated due to reflection transmission of the optics. Full spectral measurements are made by consecutively scanning through columns of the DMD mirrors and measuring the intensity of photons associated with each wavelength. CD measurements are made by multiplexing wavelengths channels onto the detectors and can be done by applying optimal binary(OB) or Hadamard filters. The new optical design has a spectral window from  $150\text{cm}^{-1}$  to  $4000\text{cm}^{-1}$  and the improvement in the photon collection efficiency allows classification and imaging speeds of  $10\mu\text{s}$  per point with  $13\text{mW}$  of laser power on the sample, and is significantly faster than measurements made with the previous prototype.

In chapter 2 of the thesis I describe the construction of a new instrument which is equipped with both a hyperspectral CD spectrometer as well as a traditional Czerny Turner spectrometer. A flip mirror after the Raman microscope directs the Raman scattered beam either towards the CD spectrometer (with the mirror down) or towards the Czerny Turner spectrometer. This instrument allows us to perform head to head comparisons of the two spectrometers using the same Raman scattered photons emitted by the sample. The CD spectrometer uses hardware optical filters to perform compressed chemometric measurements to classify chemicals. The traditional spectrometer uses the CCD to measure full spectral data and chemometric analysis is performed to extract lower dimensional chemical information post measurement. Chemical classification results obtained using two sets of chemicals with differing degrees of spectral overlap show that CD classification



is comparable to full spectral classification in the high signal regime. However, for signals consisting of less than 1000 total photon counts, CD classification outperforms full spectral classification.

In chapter 3 of the thesis, Raman spectroscopy is used to probe changes in vibrational spectra of nucleotide solutions and hanging droplets containing RNA crystals at different pH. Self-modeling curve resolution (SMCR) applied to full Raman is used to extract solute correlated (SC) Raman spectral components that contain solute spectra with minimal interference from the surrounding solvent. The goal of these studies is to show that Raman spectroscopy can be used to study biological molecules in aqueous environments, with minimal sample preparation and without the need of labels.

## CHAPTER 1 IMPROVEMENTS IN THE OPTIMAL BINARY COMPRESSIVE DETECTION SPECTROMETER DESIGN AND SELECTED CLASSIFICATION, QUANTITATION AND IMAGING EXAMPLES

### 1.1 Abstract

Multiple prototypes of hyperspectral compressive detection (CD) Raman spectrometers have previously been constructed in the Ben-Amotz lab and have proven to be useful for fast, label-free chemical identification, quantitation and imaging. The CD spectrometer consists of a volume holographic grating (VHG) that linearly disperses the Raman photons into its component wavelengths and all wavelengths are focused onto a digital micromirror device (DMD). The DMD is an optical modulator that consists of an array of programmable  $10\mu\text{m}$  mirrors that can reflect photons in either  $+12^\circ$  or  $-12^\circ$  to the incoming light. The DMD is tilted such that the  $+12^\circ$  photons go back through the focusing lens and the VHG and is focused onto a single  $150\mu\text{m}$  photon counting avalanche photodiode detector (APD). In chapter 1 of the thesis I describe the construction of a new CD Raman spectrometer capable of fast hyperspectral imaging that has better photon collection efficiency and fewer photon losses compared to its predecessors. The new spectrometer consists of a VHG and a DMD, however, the DMD is not tilted but is perpendicular to the incoming Raman photons. All the Raman photons modulated by the DMD are symmetrically detected in the  $+12^\circ$  and  $-12^\circ$  by two photon counting

photomultiplier tube(PMT) detector modules. The new spectrometer avoids a double pass through the optics and hence has fewer losses associated due to reflection transmission of the optics. Full spectral measurements are made by consecutively scanning through columns of the DMD mirrors and measuring the intensity of photons associated with each wavelength. CD measurements are made by multiplexing wavelengths channels onto the detectors and can be done by applying optimal binary(OB) or Hadamard filters. The new optical design has a spectral window from  $150\text{cm}^{-1}$  to  $4000\text{ cm}^{-1}$  and the improvement in the photon collection efficiency allows classification and imaging speeds of  $10\mu\text{s}$  per point with  $13\text{mW}$  of laser power on the sample, and is significantly faster than measurements made with the previous prototype.

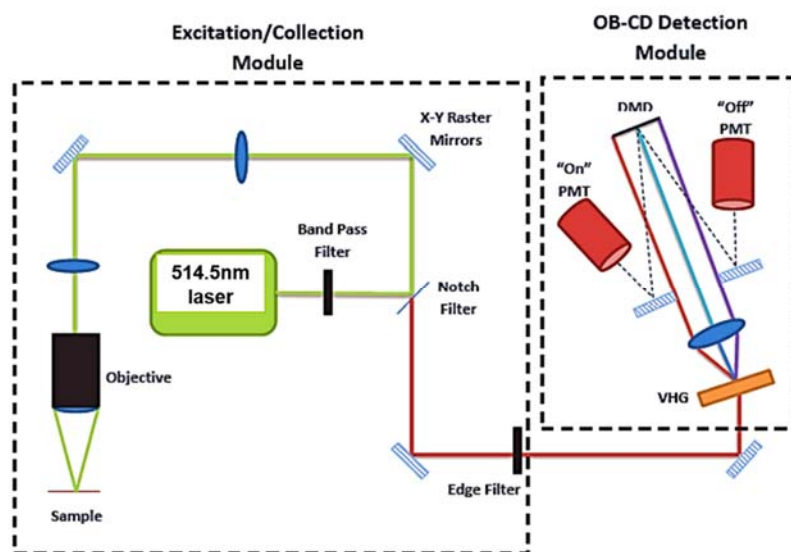


Figure 1.1: Schematic of the new OB-CD instrument showing the VHG, DMD and two PMT's. The collection lenses after the DMD is not shown in this figure.

## 1.2 Introduction

Raman spectroscopy is the study of interaction of electromagnetic radiation with matter based on the Raman effect that was theorized by Adolph Smekal in 1923, and later observed experimentally by Sir C.V. Raman and K.S. Krishnan in 1928<sup>1</sup>. When light interacts with matter, the photons undergo many different processes as a result of the interaction. One process is scattering, a two photon event that involves the incoming photon to be absorbed by the molecule and reemitted. The scattering process depends on the change in the polarizability of the molecule and increases as the fourth power of the frequency of the incident light. The emitted photon can have the same energy as the incoming photon, or may gain or lose energy as a nature of the scattering process. The most common scattering process is Rayleigh scattering where the scattered photon has the same energy as the incident photon. This is considered to be elastic scattering since there is no change in the energy of the incident photon and the scattered photon. About 1 in  $10^7$  incident photons are scattered such that the energy of the scattered photon is different from that of the incident photon. This loss (Stokes) or gain (anti-Stokes) of energy is associated with vibrational energy state of an electron in the molecule and is considered as an inelastic scattering process. This process is called Raman scattering or Raman effect and forms the basis of Raman spectroscopy<sup>2</sup>. The emitted Raman photons can have lower or higher energy compared to the incident photon. Figure 1.2 shows the energy diagram for Rayleigh and Raman (Stokes and anti-Stokes) scattering.

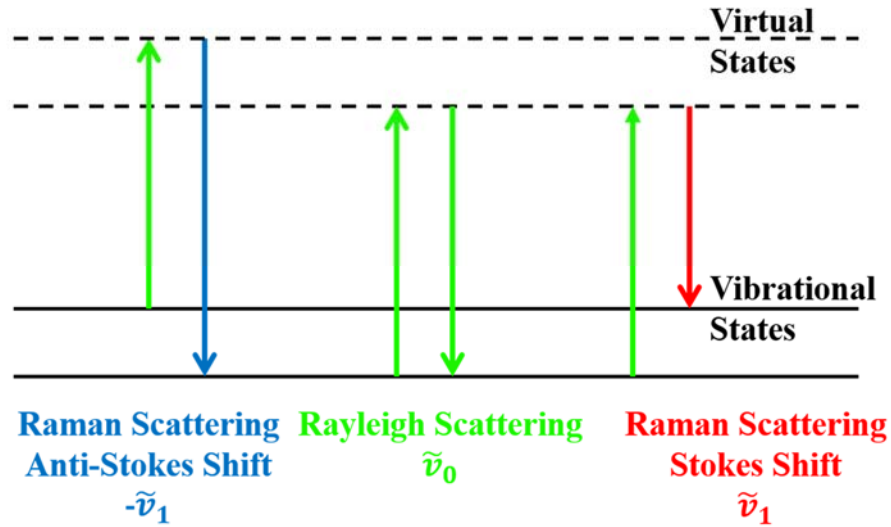


Figure 1.2: Jabonski diagram showing 2 two photon scattering. In Rayleigh scattering process, there is no change in the wavelength (energy) of the incoming and the scattered photons. However, in the Raman scattering process, the scattered photons gain or lose energy compared to the energy of the incoming photons.

Inherently Raman scattering is much weaker than Rayleigh scattering which must be filtered out in order to detect the Raman scattered photons. A Raman spectrum is a 2 dimensional plot or graph of the measured Raman scattered photons (intensity) vs pixel. A spectral pixel is the smallest independently measurable unit that contains a certain wavelength of Raman scattered photons. A pixel to wavelength calibration is performed to determine the wavelengths associated with that pixel. The abscissa of a Raman spectrum is usually represented in inverse wavelength shift with units of  $\text{cm}^{-1}$  which is proportional to energy difference of the scattered photon. Equation 1.1 describes the relation of wavenumber and wavelength and equation 1.2 is used to determine the wavenumber shift from the incident wavelength.

$$\text{Wavenumber} \equiv \tilde{\nu} = \frac{10^7}{\lambda} \quad (1.1)$$

Where  $\tilde{\nu}$  is wavenumber with units  $\text{cm}^{-1}$  and  $\lambda$  is the wavelength in nm. The wavenumber shift is given as

$$\text{Wavenumber shift} \equiv \Delta\tilde{\nu} = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right) * 10^7 \quad (1.2)$$

Where  $\Delta\tilde{\nu}$  is the wavenumber shift,  $\lambda_0$  is the wavelength of the excitation source (laser) and  $\lambda_1$  is the wavelength of the Raman scattered photons. Figure 1.3 shows a schematic of a Raman instrument. The sample is irradiated by a laser and the Rayleigh scattered photons are filtered. The Raman scattered photons are dispersed onto a detector either simultaneously or by moving a mask such that one wavelength channel is measured at any given point. The readout from the detector is plotted vs pixel (or calibrated wavenumber) to display a Raman spectrum.

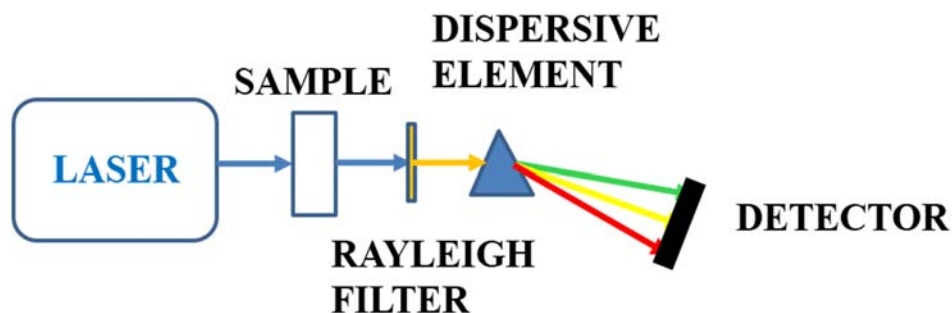


Figure 1.3: Schematic of a generalized Raman instrument showing an incoming light source (laser) interacting with the sample. The Rayleigh scattered photons are filtered and the Raman photons are linearly dispersed into its component wavelengths and the spectrum is measured using a linear array detector.

A Raman spectrum is a unique fingerprint that is be used to identify chemicals. Raman scattering is not selective to the incident wavelength, non-destructive and does not

require sample preparation. Raman spectroscopy has proved to be a useful tool many different fields of science<sup>3,4</sup>. Raman spectroscopy has been used to study biomolecules such as proteins<sup>5,6</sup>, RNA<sup>7</sup> and nucleotides<sup>8,9</sup> and has also been used in biomedical applications<sup>10,11</sup>. However, the Raman scattering yields to intrinsically low signals and it takes a hundreds of milliseconds (or longer) to collect a spectrum with good signal to noise in order to be analyzed. In applications such as Raman microscopy and imaging, Raman spectrum of thousands to millions of different spatial points are measured, and it could take a few hours to days to acquire the data<sup>12</sup>. Conventional Raman spectrum is measured concurrently on a multichannel array detector like a CCD camera that is limited by the measurement time and the inherent read noise of the detector electronics. High speed hyperspectral Raman imaging can be achieved by incorporating chemometric techniques directly on the hardware and multiplexing photons from different wavelengths that contain Raman information onto a single detector with very low noise. This has previously been demonstrated where programmable optical filters are directly applied onto the hardware using liquid crystal (LC)<sup>13-15</sup> or digital micromirror device (DMD)<sup>16-21</sup> and chemometric measurements are made.

OB-CD strategy applied to Raman spectroscopy was previously used towards fast classification, quantitation and imaging of chemicals<sup>16,17,20,22</sup>. Previous instruments used in OB-CD had photon losses associated with the optical design of the spectrometer. Improving photon collection efficiency will help with faster measurements with similar signal to noise. In this chapter a new, more efficient prototype of the OB-CD spectrometer is described. The Raman scattered photons entering the spectrometer are linearly dispersed and focused on the DMD by the volume holographic grating (VHG) and an achromatic

lens. The Raman photons are then collected onto one or 2 photomultiplier tube (PMT) detectors by reimaging the DMD onto the PMT. Collection efficiency is increased by detecting all the diffraction spots emanating from the DMD and by passing through fewer optical elements unlike the previous spectrometer design. The detection efficiency increase allows us to classify and image chemicals faster than previously described with laser power at the sample (13mW of 514nm at sample equivalent to 65mw of 785nm)

### 1.3 Methods

#### 1.3.1 Instrument

##### 1.3.1.1 Raman Microscope Stage

The Raman microscope is built by modifying an Olympus white light microscope (BX 51) and mounting a board containing the associated optics on top of the existing microscope. A 514.5nm single mode ( $TEM_{00}$ ) Argon ion laser (Modulaser Stellar-Pro-L 514/100) is used as the excitation source and is mounted on the optics table. The laser has a maximum output power of 100mW, and a 250:1 v-polarization ratio. The laser is guided through a laser line cleanup filter (Semrock Part Number: LL01-514-12.5) towards a pair of galvanometer mirrors using a 514.5nm RaserEdge dichroic beamsplitter (Semrock Part Number: LPD01-514RU-25). The laser is then guided on the back of the objective using a pair of galvanometer x/y scanning mirrors (Thor labs) and a pair of 150 mm two inch achromatic lenses arranged in a 4f configuration. This configuration is used commonly in



beam scanning techniques where the sample is stationary and the laser is moved over the sample in order to image the sample. The first lens is exactly one focal length away from the galvanometer mirrors and is two focal lengths away from the second lens, and the second lens is one focal length away from the back of the objective hence 4f configuration. The laser never moves from the back of the objective but comes in at different angles through the objective enabling to sample a different spatial spot on the sample. The microscope is equipped with an Olympus 20X objective (LMPLFL20X) with an N.A. of 0.40, and a working distance of 12.0mm. The microscope is also equipped with a programmable motorized sample stage (Prior Proscan II) that can be moved by a distance of 0.1 $\mu$ m. Rayleigh and Raman scattered light is collected through the same objective in a backscattering collection geometry. The scattered light is separated based on the wavelength at the dichroic mirror. Longer wavelength Stokes shifted Raman photons go through the dichroic beam splitter. Additional Rayleigh scattered photons are filtered using an edge filter (Semrock Part Number: LP02-514RU-25) and the Raman photons are directed towards the home built hyperspectral CD spectrometer.

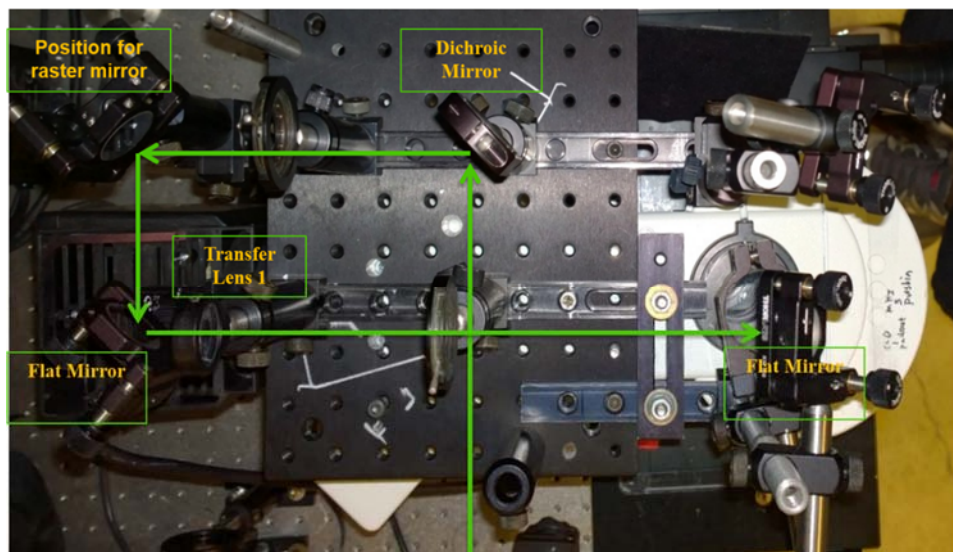


Figure 1.4: Raman microscope optics mounted on an Olympus BX51 microscope.

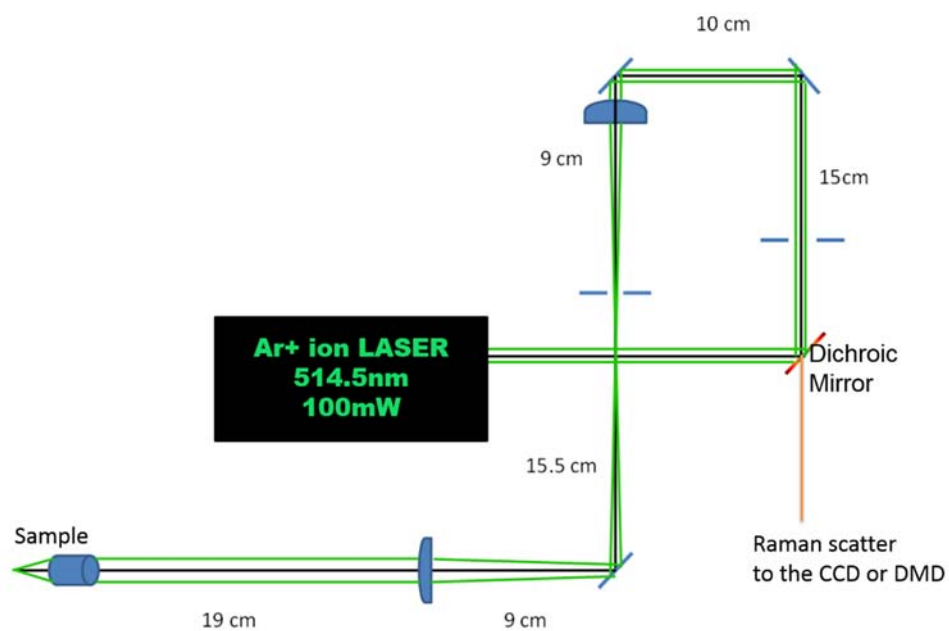


Figure 1.5: Schematic of the Raman microscope stage. The dichroic mirror allows longer Stokes wavelength to pass through and reflects the laser back towards the source.

### 1.3.1.2 OB-CD Spectrometer

The collimated Raman scattered photons are focused onto the entrance slit of the OB-CD spectrometer using a 150mm achromatic lens. The light is re-collimated using another 75mm achromatic lens and the collimated light is dispersed using a transmission volume holographic grating (VHG, 1200line/mm, cwl:514nm). A focusing lens is used to focus the dispersed wavelengths onto a digital micromirror device (DMD) (DLPLIGHTCRAFTER-DLP® LIGHTCRAFTER™ Evaluation Module). The DMD is a rectangular array of 608X684, 10 $\mu$ m pitch mirrors that can be programmed to tilt either +12° or -12° to the flat position. All the light (light spread out in different diffraction orders) is collected in one or the either direction by reimaging the DMD onto one or two photon counting photomultiplier tube modules (Hamamatsu H11890-01) using a 50mm/50mm achromatic doublet pair of lenses. The PMT counting modules output a TTL pulse when a photon is detected and National Instruments data acquisition card is used to synchronize the voltage on the galvanometer mirrors and the duration of data collection. This spectrometer can be used to obtain a Raman spectrum by wavelength scanning or by Hadamard scanning. In wavelength scanning columns of mirrors are sequentially set to direct light to one detector for a set duration of time. Hadamard scanning is achieved by applying a Hadamard mask onto the DMD and measuring the counts of photons associated with that mask. The spectrum is reconstructed by applying inverse Hadamard matrix to the measured counts. Optical filters can also be built and applied directly to the DMD as shown later in OB-CD and OB-CD2 filtering.

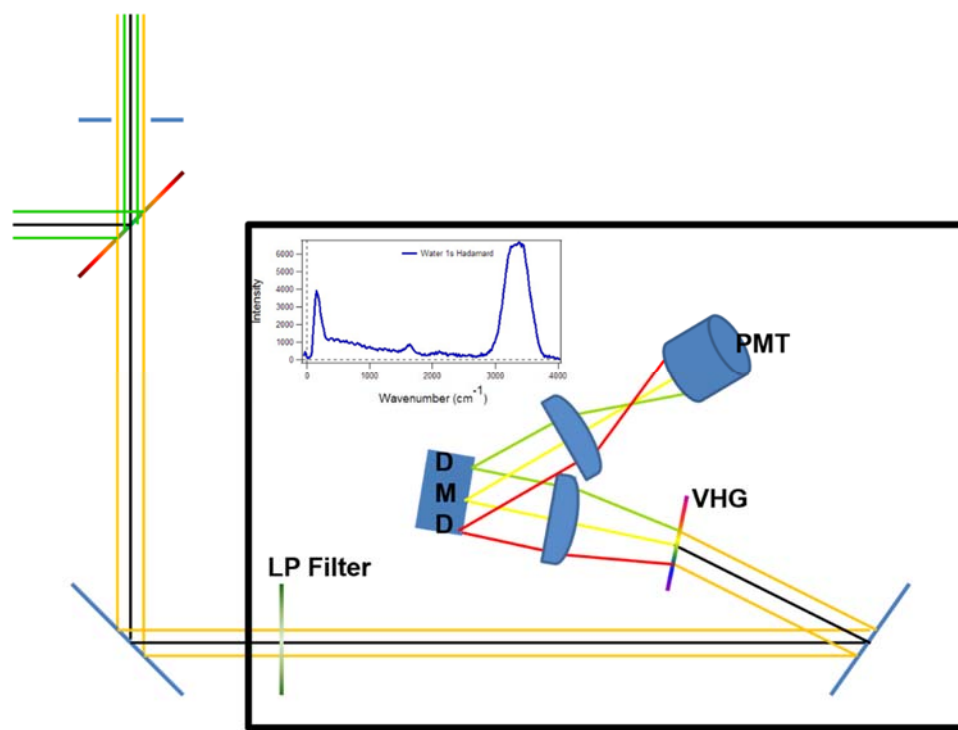


Figure 1.6: Schematic of the OB-CD spectrometer. The inset shows the spectrum of water acquired by notch scanning

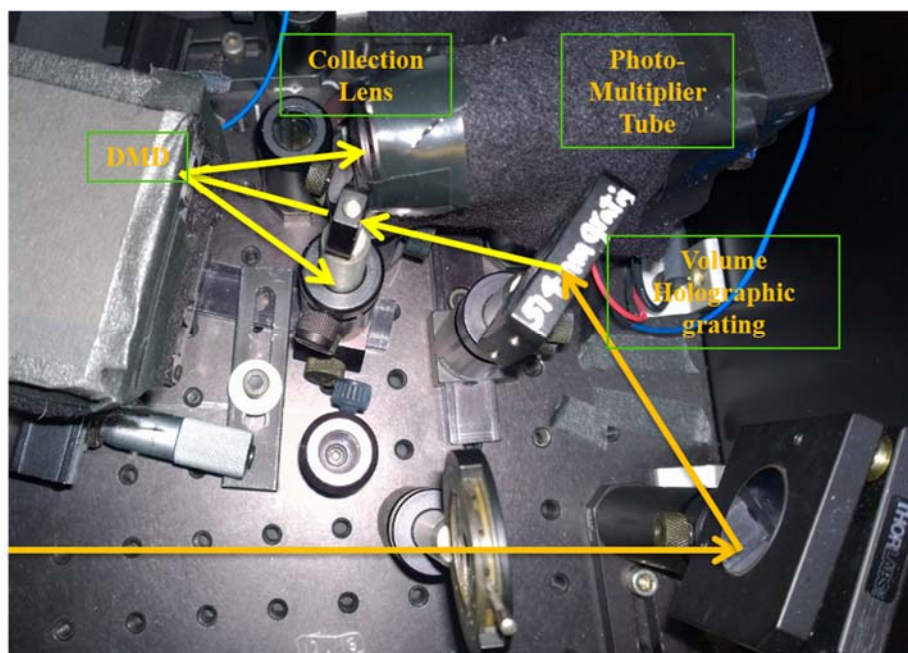


Figure 1.7: OB-CD Spectrometer optical path showing the incoming Raman beam (orange). One of the separated wavelengths shown in yellow passes through the dispersive element (VHG) and is focused on the DMD. The DMD can direct these photons in either  $+12^\circ$  towards one of the PMT or  $-12^\circ$  (detector not shown)

### 1.3.1.3 Galvo-mirrors Optics

Imaging a sample requires collecting Raman spectrum over multiple points of the sample. One way to achieve this is by moving the sample under the focal point by a known amount. Although this technique has been adopted in imaging systems, the speed of moving the sample stage is slow. Beam scanning using a galvomirror system is the faster alternative for imaging a sample. Imaging the sample under the objective can be achieved by using a 2-D galvomirror system. The schematic of a 1-D galvomirror system is shown in figure 1.8. In order to scan a sample point by point, the collimated beam of light needs to enter the back of the objective at different angles. The angle of the incoming beam

defines the focal position on the sample. Relay lenses that are set in the  $4f$  geometry are used to image the laser spot from the galvomirrors to the back of the objective. This configuration when aligned correctly, ensures that collimated laser beam can be moved using the galvomirrors and enters at the back of the objective at different angles. For small angles, the galvomirrors can scan at rates of  $300\mu\text{s}$ . Imaging any faster will create an artifact in the image as seen in the imaging section of the results.

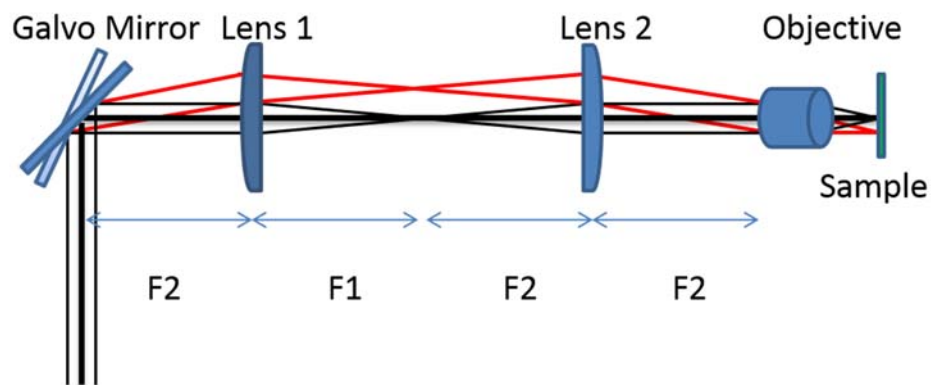


Figure 1.8: 1-D galvomirror system used to scan over the sample by moving the laser beam instead of moving the sample

#### 1.3.1.4 Grating

A diffraction grating is a periodic arrangement of reflection or transmission elements arranged on a surface and are separated by a distance that is comparable to the wavelength of light. This periodicity causes the spatial modulation in the refractive index and causes constructive and destructive interference of light by modulating the amplitude or

phase or both. A diffractive grating is used as a wavelength dispersing element and can resolve polychromatic light.

In the OB-CD spectrometers we use a volume phase holographic grating as the wavelength dispersion element. This is a type of diffraction grating that works by modulating the refractive index of the medium. VHGs are commonly used as transmission gratings and allow the light to pass through the optical element as opposed to reflective gratings that reflect the diffracted orders. The VHG has lower losses due to polarization of light and hence is a more efficient dispersion element compared to reflective gratings.

#### 1.3.1.4.1 Grating Formula and Groove Density Calculation

A VHG can diffract light into discrete directions and can be generalized by using the grating formula shown in equation 1.3

$$m\lambda f = \sin(i) + \sin(r) \quad (1.3)$$

where  $m$  is the order of the grating,  $\lambda$  is the wavelength,  $f$  is the frequency of the grooves given as lines/mm (ln/mm),  $\theta_i$  is the angle of incidence,  $\theta_r$  is the angle of diffraction.

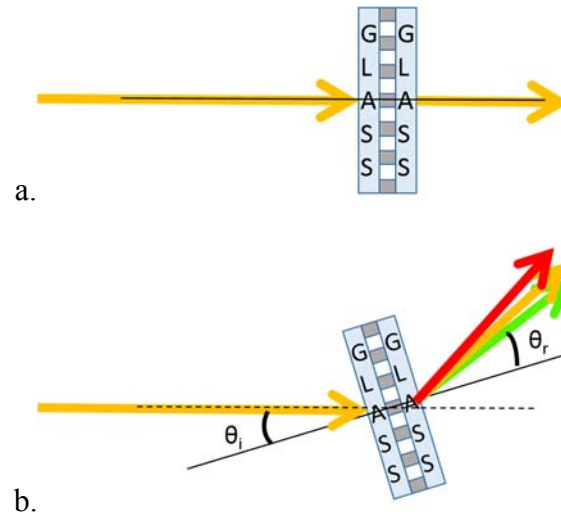


Figure 1.9: Schematic of the VHG. a) Polychromatic light is passed through the grating undiffracted and this is called the 0th order of diffraction. b) The grating is tilted to a certain angle to maximize the diffraction in one of the orders.

To calculate the groove density of the grating the angle of incidence and angle of diffraction must be measured. The diffraction grating was aligned perpendicular to the incoming light. This configuration allows all the light to pass through un-diffracted and is called the 0th order of diffraction. This is when the angle of incidence is opposite in sign to the angle of diffraction. Next the VHG was tilted such that the diffraction spot of the green wavelength was maximized at a discrete angle. The 1st order of diffraction was chosen and the VHG was tilted  $17^\circ$  relative to the incoming light. The angle of diffraction of the green light was also  $17^\circ$  to the normal of the VHG. However, the total angle shift was measured  $\sim 35^\circ$  (tilt of the VHG and the angle of diffraction of the green light)

The angle of incidence and diffraction were calculated to be approximately  $17^\circ$  for  $\lambda=514\text{nm}$ .

$$f = \frac{\sin(i) + \sin(r)}{m \cdot \lambda} \quad (1.4)$$



The groove density was calculated to be approximately 1136 ln/mm for 1st order diffraction and the wavelength of 514.5nm.

#### 1.3.1.5 Digital Micromirror Array

We used the TI DLP Lightcrafter 3000 module that consists of a 0.3” diagonal DMD chip. The DMD is shown in figure 1.10 and shows the magnified image of the diamond shaped mirrors. The mirrors have a pitch of  $10.8\mu\text{m}$  and  $7.636\mu\text{m}$  sides as shown in figure 1.11. The DMD has a tilt of  $\pm 12^\circ$  to the normal. When the DMD is powered off, all the mirrors of the DMD lie flat to the surface normal. In this state the DMD behaves as a flat mirror and only specular reflection is observed. However, when the DMD is powered on, the micromirrors can exist in one of the two states. They are aligned either  $+12^\circ$  to the surface normal (this state is also called the on state) or  $-12^\circ$  to the surface normal (also known as the off state).

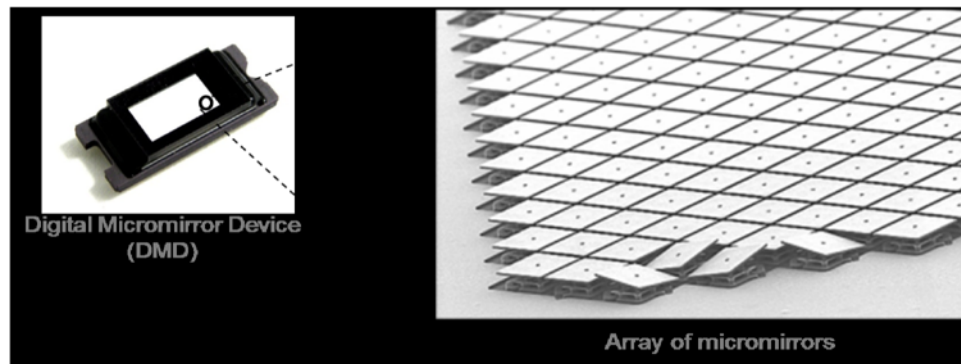


Figure 1.10: Texas Instruments digital micromirror device (DMD) consists of 608X684 mirrors that can be tilted  $\pm 12^\circ$  to the normal

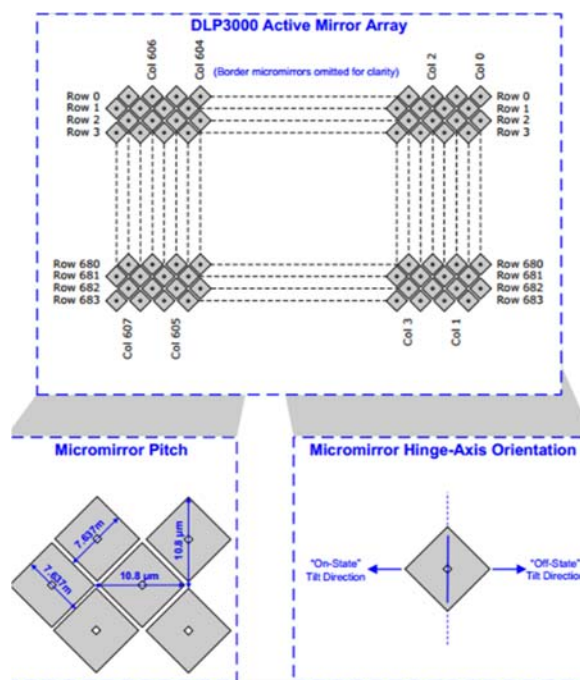


Figure 1.11: Schematic of the DMD chipset 3000 and the dimensions of the mirrors and the direction of the tilt is shown.

Since the DMD is 2-D array of periodically spaced mirrors and due to the size of the mirrors, the DMD when powered on behaves like a reflective diffraction grating. When we focus the Raman spectrum onto the DMD, the Raman photons are diffracted. The diffraction of light from the DMD follows a 2-D pattern and the schematic of the pattern can be seen in figure 1.12. The diffraction pattern dependent on the angle of incidence and the wavelength of the light. Each bright spot is constructive interference of the light and is on the order of  $\lambda/d$  apart ( $\lambda$  is the wavelength and  $d$  is the pitch between pixels). The intensity of the Raman photons being collected is distributed in this diffraction pattern and it is crucial to collect all the diffracted spots to improve the efficiency of the instrument.

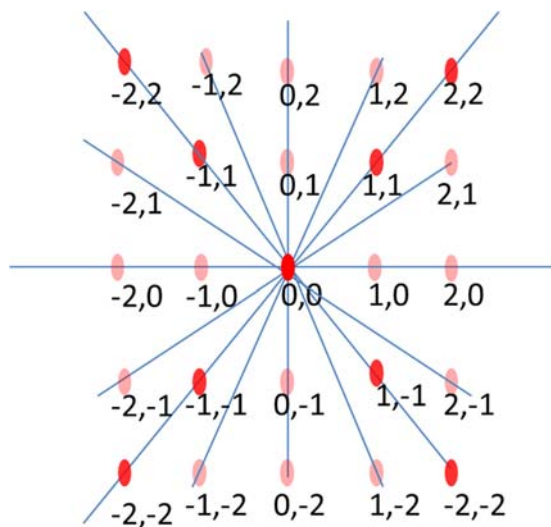


Figure 1.12: Schematic portrays of the 2-D diffraction pattern of the 633nm light from a DMD. The diffraction pattern is dependent on the angle of incidence and the wavelength of light

#### 1.3.1.5.1 Generating Patterns to Project on the DMD

The DMD is a 2-dimensional array of mirrors consisting of 608 rows by 684 columns. Raman scattered photons are linearly dispersed and focused onto the DMD.

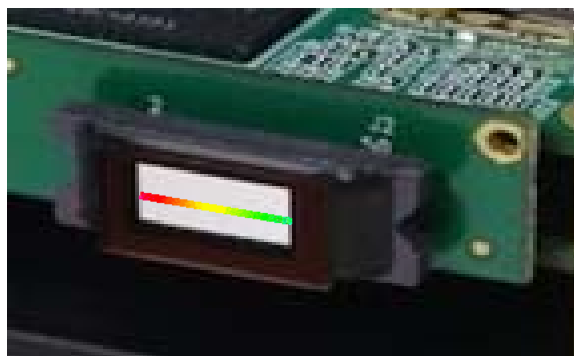


Figure 1.13: 0.3" diagonal Lightcrafter DMD 3000 chipset containing 608 rows and 684 columns of micromirrors with the spectrum spanning from green(515nm) to red (640nm)

Binary patterns are generated in Matlab and applied onto the DMD as bitmap images in order to project the spectrally separated Raman photons either in  $+12^\circ$  or  $-12^\circ$  compared to the normal. Collection lenses collect all the diffracted light from the DMD and image the DMD onto photon counting PMT modules. Unlike the scientific grade DMD (used in the previous instrument) that allows the user send to binary pattern information onto the DMD as strings the Lightcrafter accepts the patterns as bitmap images that have the same dimensions as the DMD. Binary images that send all the light in the  $+12^\circ$  (all on) or in the  $-12^\circ$  (all off) direction can be generated by setting all the mirrors to “1” or “0”. Figure 1.14 shows the pattern generated for the  $-12^\circ$  direction(black) and  $+12^\circ$  direction (white).

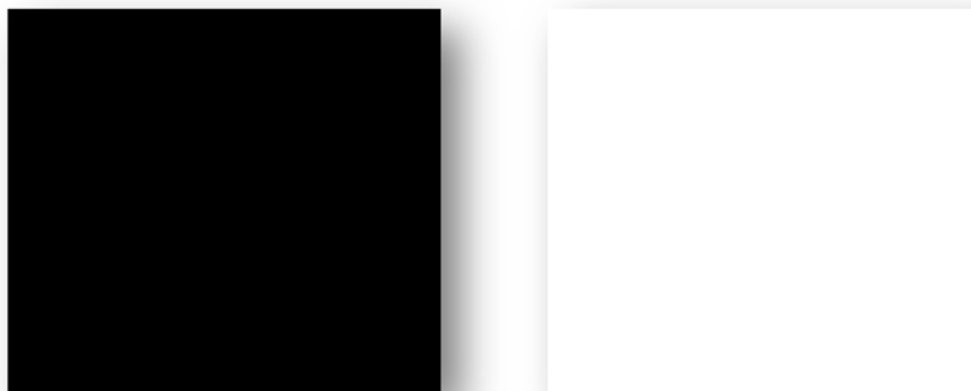


Figure 1.14: All off (black) and all on (white) images modulate light towards  $-12^\circ$

#### 1.3.1.5.2 Notch Scanning

One or a few columns of mirrors can be set from onto the detector in the  $+12^\circ$  position to project the Raman photons towards the  $+12^\circ$  direction. Sequentially switching these mirror notches through the length of the DMD and measuring the counts of each

notch onto the detector in the  $+12^\circ$  for the same given time direction produces a Raman spectrum. Figure 1.15 shows five binary bitmap images of dimensions 608X684 with 10 mirrors set to 1 and all the rest of the pixels set to 0. When these images are loaded to the DMD photons from the channels set to 1 are sent to the  $+12^\circ$  direction. The remaining photons are sent to the  $-12^\circ$  direction.



Figure 1.15: 10 mirror notches set to “1” at 0,100,200,300,400 position. The region in the white show DMD mirrors set to  $+12^\circ$  and the region in black are  $-12^\circ$

Bitmap images for 1 mirror notches are created in Matlab using the program below.

This can be modified to generate bitmaps having any size notches.

```
for i=1:684
im=zeros(608,684);
out(:,i)=1;
imwrite(logical(out),strcat('onemirror',num2str(i),'.bmp'));
end
```

#### 1.3.1.5.3 Hadamard Scanning

Hadamard transform spectroscopy takes advantage of measuring intensity by multiplexing linear combinations of multiple wavelength channels of spectrum at a time rather than measuring intensity of one wavelength channel at a time<sup>23-26</sup>. In the hyperspectral CD spectrometer, Hadamard masks can be applied to the DMD and measure photon intensity by directing photons multiplexed wavelength channels onto one almost noise free detector. Fellgett has shown that there is an improvement in the signal to noise (S/N) ratio by multiplexing wavelengths in the case of Fourier spectroscopy and a similar advantage can be seen in Hadamard spectroscopy if the primary source of noise is shot noise due to random fluctuations in signal. In traditional notch scanning, intensities of N pixels are measured with each pixel associated with a different wavelength range. The measurement accuracy is limited by the detector and the signal noise. Hadamard spectroscopy improves the S/N by making N consecutive measurements of the spectrum using N linear combination of wavelength channels. Recovery of the spectrum is done by solving for N unknowns by using a linear transform.

Hadamard matrix contains values of -1 or 1 (photons sent towards the detector). We set the values of -1 to 0 (photons sent away from the detector) and measure all the positive channels. All negative channels are measured by setting +1 to 0 and multiplying the matrix by -1. The positive intensities are subtracted from the negative intensities and the spectrum is transformed using Eq 1.5

$$H * H^T = nI_n \quad (1.5)$$

where H is the Hadamard matrix of the order n,  $I_n$  is an identity matrix of the order n. All rows and columns of the Hadamard matrix are orthogonal to each other and the size of the matrix must be a power of 2, a power of 2 multiplied by 12, or a power of 2 multiplied by 20.

Figure 1.16 shows 1280 positive and negative Hadamard masks used to generate a Hadamard spectrum of benzene shown in Figure 1.17. Each row of the figure is a different mask and is converted into the DMD dimensions (608X684) before being applied on the DMD as a bitmap image. Only 640 columns of the DMD contains the mask information. The remaining 44 columns are set to the value of 0.

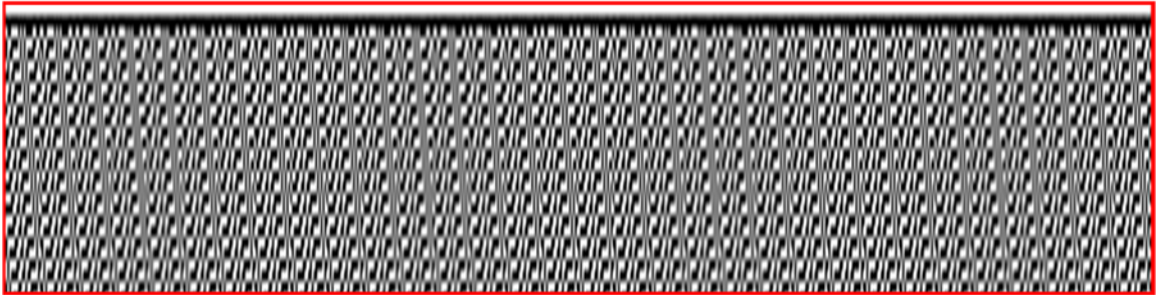


Figure 1.16: 1280 positive and negative Hadamard masks are applied to recover a 640 wavelength spectrum. Each row of this figure corresponds to a different mask. The 1<sup>st</sup> mask is all white corresponding to the value 1, and the next is all black with a value of 0 and so forth.

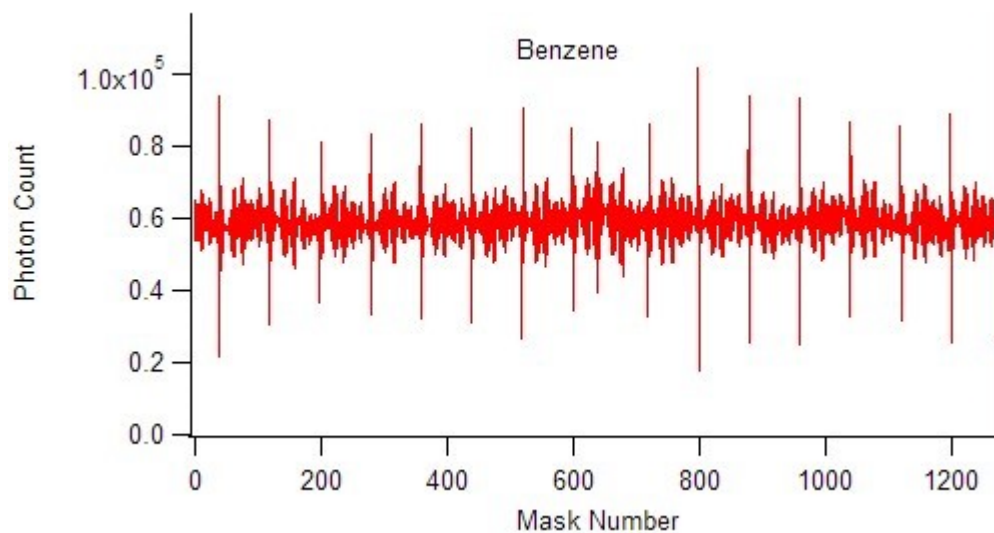


Figure 1.17: Intensity of multiplexed wavelengths of benzene through 1280 Hadamard masks. Notice the values are symmetrical because the positive and negative masks are complementary. The last 2 masks correspond to all off (background) and all on (sum of all the intensities).

### 1.3.2 Overview of OB-CD Filters

Filters are masks that are applied onto the DMD as bitmap images and they selectively turn some set DMD pixels towards the  $+12^\circ$  direction while all remaining pixels are turned towards the  $-12^\circ$  direction. This bidirectional property of the DMD makes all filters binary. In the hyperspectral compressive detection spectrometer, a spectrum is linearly dispersed onto the DMD and each pixel of the DMD is associated with a range of wavelength. OB filters are optimized filters generated for specific set of chemicals and when applied sequentially to the DMD send a combined part of the dispersed spectrum (multiplex photons associated with the DMD pixels) towards one of the detectors (either  $+12^\circ$  or  $-12^\circ$ ). Using a set of OB filters compressed parts of Raman spectra are measured



as photon counts and are converted into chemical rates (photons scattered by a particular chemical per second) of each chemical. These rates are analogous to concentration of each chemical in the mixture. The filters are optimized to minimize the error in the chemical rates.

The core of optimal binary compressive detection strategy is defined here without diving into the mathematics behind it. The mathematical theory and proofs of solving for the optimized binary filters can be found in previous work<sup>16,22</sup>. OBCD filters are generated and applied on the DMD and are specific to the sample in question. Some prior knowledge about the samples make up is essential in order to build optimized filters. The Raman spectra or MCR derived spectra can be used as training spectra to generate filters used in compressive detection. Photon counts measured through each filter is converted to chemical photon scattering rate given by equation 1.6.

$$\hat{\Lambda} = B\tau^{-1}T^{-1}\hat{\chi} \quad (1.6)$$

$A = (F^T P)$  such that  $BA = I$  and  $B = (F^T P)^{-1}$ . where  $\mathbf{A}$  and  $\mathbf{B}$  are square linear transform matrices with dimensions equal to the number of chemicals in the mixture and are used to convert rates to counts or vice versa,  $\hat{\Lambda}$  is a matrix that contains the recovered photon rates for each chemical,  $\mathbf{P}$  is the matrix consisting of the area normalized chemical spectra,  $\mathbf{F}^T$  is the transposed matrix consisting OB filters,  $T^{-1}$  is the inverse of the matrix whose diagonal values consists a fraction of total time applied per filter and the sum of the diagonal is equal to the total integration time  $\tau$  and  $\hat{\chi}$  is a matrix that contains the measured photon counts through each filter.

The OB-CD algorithm calculates OB filters to minimize the error in the square of the estimated photon rates and is given by equation 1.7

$$E\|\hat{\Lambda} - B\tau^{-1}T^{-1}\hat{\chi}\|^2 \quad (1.7)$$

### 1.3.3 Bernstein Polynomials to Remove Fluorescence from Pure Chemical Rates

Raman spectrum can be overshadowed by a broad featureless background of fluorescence and is usually removed post data collection by subtracting a polynomial background<sup>27,28</sup>. Raman spectra can be recovered if the level of fluorescence is not overwhelming. We have demonstrated that fluorescence can be removed from OB-CD measurements by training on Bernstein polynomials<sup>17</sup>. The recovered chemical rates stay fixed with varying amounts of fluorescence and this is done by building filters on polynomials that model fluorescence. We have described the use of third order Bernstein polynomials as the basis set as a model of fluorescence. Bernstein polynomials have non-negative coefficients<sup>29</sup> and is given by eq1.8

$$b_{v,n}(x) = \binom{n}{v}x^v(1-x)^{n-v}, \quad v = 0, 1, \dots, r \quad (1.8)$$

where  $\binom{n}{v}$  is binomial coefficient. The linear combination of the polynomial is given by equation 1.9, where  $\beta_v$  is the Bernstein coefficient. We use 3<sup>rd</sup> order polynomial, n=3 and is visually shown in figure 1.18.

$$B_n = \sum_{v=0}^n \beta_v b_{v,n}(x) \quad (1.9)$$

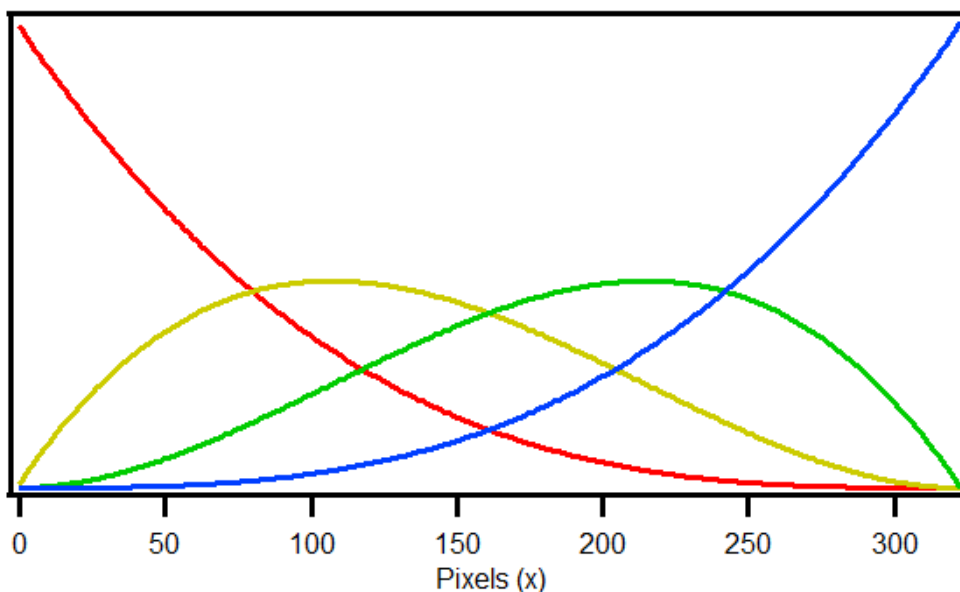


Figure 1.18: Third order Bernstein polynomial curves using  $x=0:341$  are shown. The basis curves are normalized to unit intensity and have positive coefficients.

## 1.4 Results

### 1.4.1 Comparison of 514.5nm and 785nm Excitation Systems

The older prototype of the OB-CD instrument used a 785nm near IR excitation source. The spectral window for this instrument spans approximately from 200 $\text{cm}^{-1}$ -2000 $\text{cm}^{-1}$  (wavelength range of 797-931nm). The new OB-CD prototype uses 514.5nm Argon ion excitation source and the spectral window spans from 150 $\text{cm}^{-1}$ - 4000 $\text{cm}^{-1}$  (518.5nm-650nm). The smaller spectral window for the previous prototype is due to the drop off in the sensitivity of the detector at higher wavelengths.

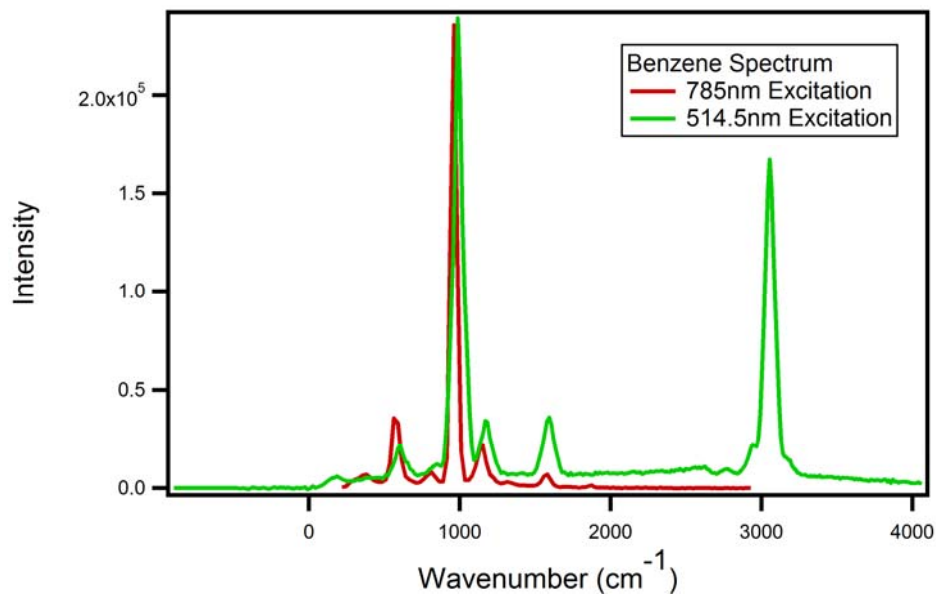


Figure 1.19: A comparison of the Hadamard spectrum of benzene and the spectral window using 2 different OB-CD spectrometers

#### 1.4.2 Fast Chemical Classification

Raman spectroscopy has previously been used towards classification of chemicals. The chemical spectrum is a unique chemical fingerprint which can be used to identify a material, and distinguish it from others. Here we demonstrate the classification of acetone and benzene using OB-CD strategy applied to Raman spectroscopy. The normalized correlation coefficient (normalized dot product) of the two spectra is 0.3 where a correlation coefficient of 1 signifies total overlap of spectra and 0 signifies no spectral similarity.

High signal to noise spectrum of acetone and benzene collected by Hadamard scanning and were used as training spectrum for the generation of OB filters. OB filters were constructed in Matlab using the OB filter generation algorithm. The spectral window

was masked such that the initial 50 pixels were neglected and the filters for that region was set to a value of “0”. OB filters for acetone and benzene are shown in the figure 1.20 below

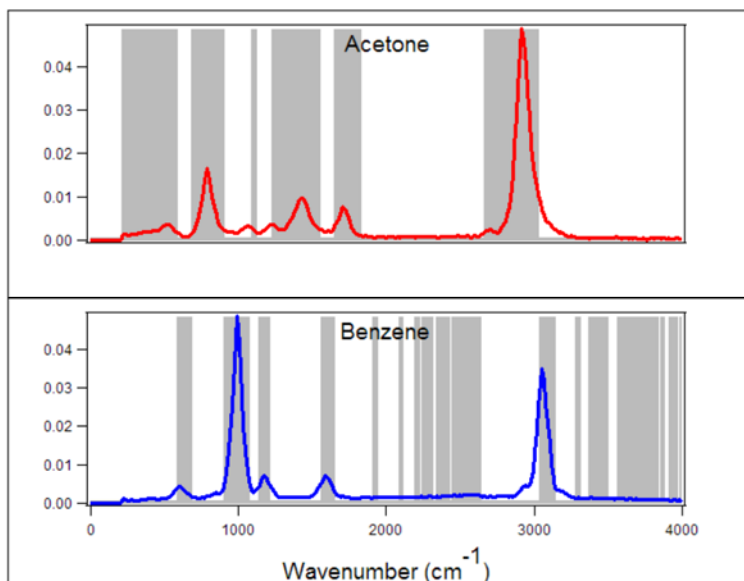


Figure 1.20: Normalized spectrum of acetone(red) and benzene(blue) are shown. OB filter(grey) generated for each chemical is also shown overlaid with the associated spectrum

Each OB filter is applied for its associated time (63% of the total time is given to the acetone filter and 37% of the total time given to the benzene filter) sequentially onto the DMD as a bitmap image. Photon counts are measured through each filter and are converted to chemical rates for each chemical. Classification map of acetone and benzene is shown in the figure 1.21 and is generated by plotting the chemical rates of acetone vs the chemical rate of benzene for each chemical. Each cloud consists of 1000 independent measurements. The classification error increases with decrease in classification time, however, the means of the clouds remain in about the same position. The figure 1.21 below

shows classification of acetone and benzene at different classification times. The ellipse around the clouds are calculated by eigenvalue decomposition of the covariance of the sample and represent 3 times the standard deviation or the 95% confidence interval. Good classification of acetone and benzene is possible even at  $10\mu\text{s}$  classification time which corresponds to 1.5 mean photon counts through acetone filter and 4.6 mean photon counts through benzene filter.

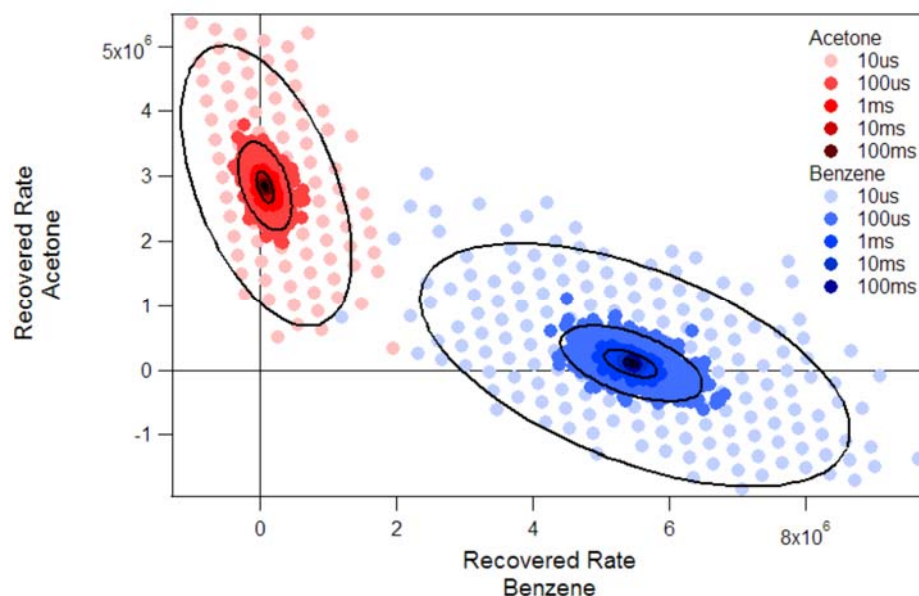


Figure 1.21: Classification of acetone(red) and benzene(blue) using OB filters shown above. Each point is obtained by measuring counts for each chemical through 2 OB filters and recovering the rates. The mean rate for acetone is about  $2.5 \times 10^6$  photons/s and the mean rate for benzene is  $5.5 \times 10^6$  photons/s. The variance in the cloud increases with decrease in classification time but the mean rate stays the same. The ellipses represent 95% confidence interval for each classification cloud. Each classification cloud consists of 1000 independent measurements

Recovered spectrum of acetone and benzene collected by applying Hadamard masks for  $10\mu\text{s}$  each is shown in figure1.22. The maximum counts for acetone is 1 count

and the maximum counts for benzene is 2 counts. Again average counts of acetone and benzene through OB filters at 10 $\mu$ s classification is 1.5 and 4.6 photon counts.

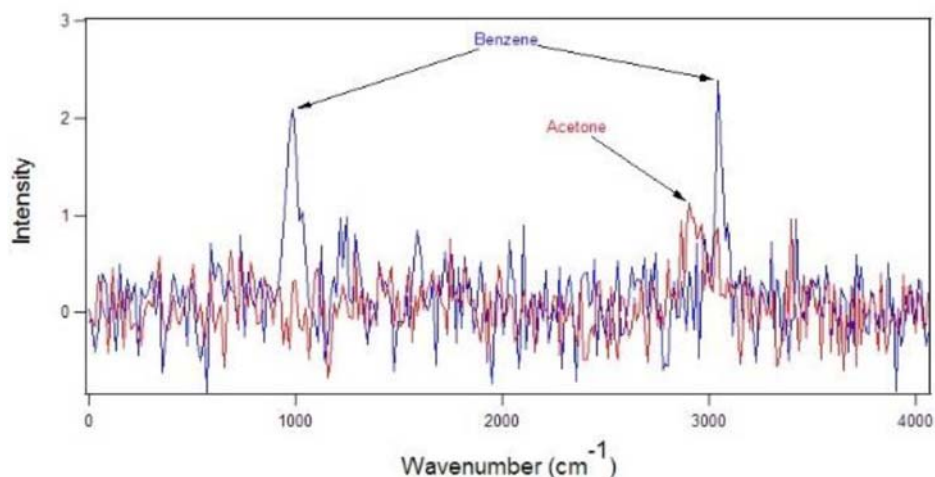


Figure 1.22: Notch spectrum of acetone(red) and benzene(blue) acquired at 10 $\mu$ s per notch filter

### 1.4.3 Chemical Quantitation

Raman spectroscopy has been used towards quantitation of chemicals in many applications. It has also been demonstrated that OB-CD strategy can be used to quantify chemical mole fractions. In chemical systems that are independent and do not react with each other, chemical quantitation is done by creating a calibration of intensity of one peak or the ratio of peaks vs concentration. Since OB filters multiplex a set of these peaks to recover the rates of the pure chemicals, the rates can be used to develop a calibration curve vs concentration. The recovered Raman scattering rate of a chemical is directly proportional to concentration of the chemical – and an unknown mixture concentration can

be recovered by extracting the concentration from the calibration curve. Here we present the quantitation of 2 chemical systems.

#### 1.4.3.1 Quantitation of Hexane and 1-Hexene

The spectra of pure hexanes and pure 1-hexene and 3 mixtures of the two are shown in the figure 1.23. The angle of overlap between hexane and 1-hexene is  $23^\circ$  and the dot product of the spectra is 0.92. (a dot product of 1 is for identical spectra with an angle of 0) These chemicals were chosen because they are moderately difficult to distinguish due to the degree of spectral similarity. Full spectral measurements were made by reconstructing 100ms Hadamard spectra. Notice that the intensity of the peak near  $1650\text{cm}^{-1}$  is the highest in pure 1-hexene and is the highest for the pure chemical. One can get relative intensity information and plot it vs concentration to get quantitative information from full spectral measurement.



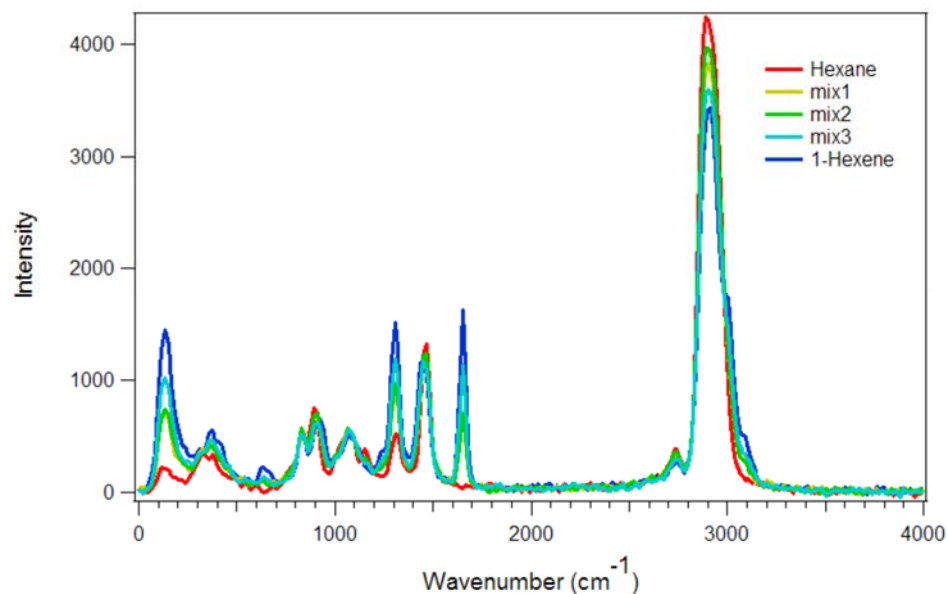


Figure 1.23: 100ms Hadamard spectrum of pure hexane (red), 1-hexene (blue), and 3 arbitrary mixtures are shown.

OB filters were constructed for pure hexane and pure 1-hexene and each filter was applied for a total time of 100ms (0.3% of total time applied to filter for hexane and 0.7% of the total time applied to filter for 1-hexene). Figure 1.24 shows the training spectrum of for pure hexane and 1-hexene. The shaded gray regions correspond to the OB filters created for each chemical. The photon counts through each OB filter were converted into the recovered rates for each chemical.

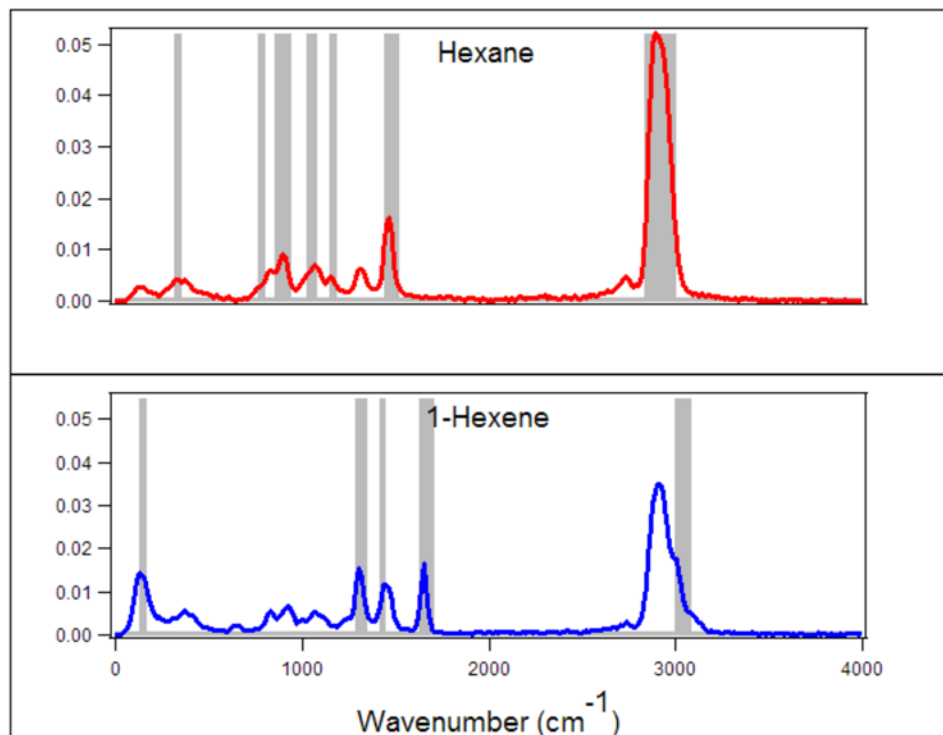


Figure 1.24: Normalized spectrum of pure hexane (red) and 1-hexene (blue) and the associated OB filters (gray)

Using the recovered Raman rates for each component, we can calculate apparent volume fractions of hexane and 1-hexene from the equations below:

$$\chi_i = \frac{w_i \hat{\lambda}_i}{\sum_i w_i \hat{\lambda}_i} \quad (1.10)$$

Where  $\chi_i$  is the mole fraction of the chemical  $i$  in the mixture,  $\hat{\lambda}_i$  is the mean recovered rate of chemical  $i$ , and  $w_i$  is the weighting factor as and is shown in equation 1.10

$$w_i = \frac{M_i}{\Lambda^{max}_i} \quad (1.11)$$

$M_i$  is the molarity (moles/L) of the pure chemical  $i$  and  $\Lambda^{max}_i$  is the recovered rate of the pure chemical (maximum recovered Raman scattering rate). Mean max scattering rate for pure hexane and pure 1-hexene is 0.8 and 1 million photon counts/s respectively.

The mole fraction can be converted into % volume fraction using equation (1.12)

$$\% \text{ vol fraction} = \frac{M_i \chi_i}{\sum_i M_i \chi_i} * 100 \quad (1.12)$$

The figure 1.25 shows the % volume fraction of 100 independent measurements of pure hexane and 1-hexene and different mixtures of hexane and 1-hexene each.

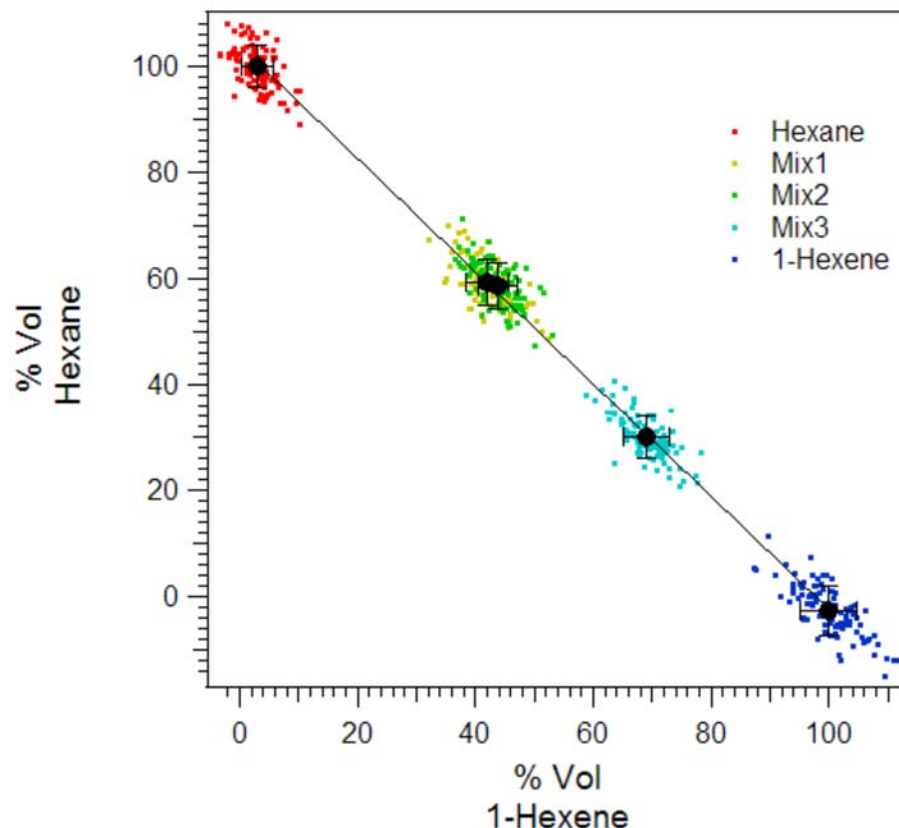


Figure 1.25: % volume fraction of hexane, 1-hexene and mixtures of the two. Each cloud corresponds to 100 independent measurements made in 100ms total time (time in which both the OB filters were applied)

#### 1.4.3.2 Quantitation of Ethanol in Tequila

The development of the new OB-CD instrument allows us to view a larger spectral window. This is essential because the CH stretch ( $2900\text{cm}^{-1}$ ) and the OH stretch ( $3200\text{--}3500\text{cm}^{-1}$ ) now can be used towards quantitation of alcohol solutions. Water has a weak OH bend ( $1600\text{cm}^{-1}$ ) and can be used in quantitation only when the signal to noise of the spectrum is high. Here we show quantitation of ethanol in a sample of silver tequila using ethanol and water as the only two components in the training spectra. The correlation coefficient of the spectrum of water and ethanol is 0.52. Is spectrum of pure ethanol, pure

water and different mixtures of ethanol and water is shown in figure 1.26. The spectrum of silver tequila is also shown and has features that resemble ethanol and water. The slightly elevated background is due to fluorescence.

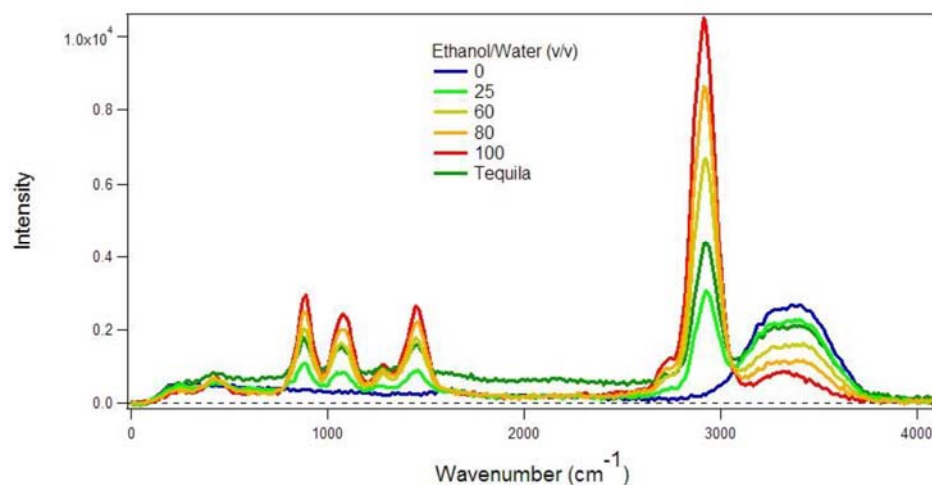


Figure 1.26: 1s spectrum of pure ethanol (red) and pure water (blue) and known mixtures of ethanol and water is shown. The spectrum of tequila (unknown, dark green) is also shown.

OB filters were generated for using pure ethanol and pure water as training spectrum for each. figure 1.27 shows the spectrum of pure ethanol and pure water recovered from 1s Hadamard spectrum. The filters for each component is shown as the shaded area overlaid on the spectrum.

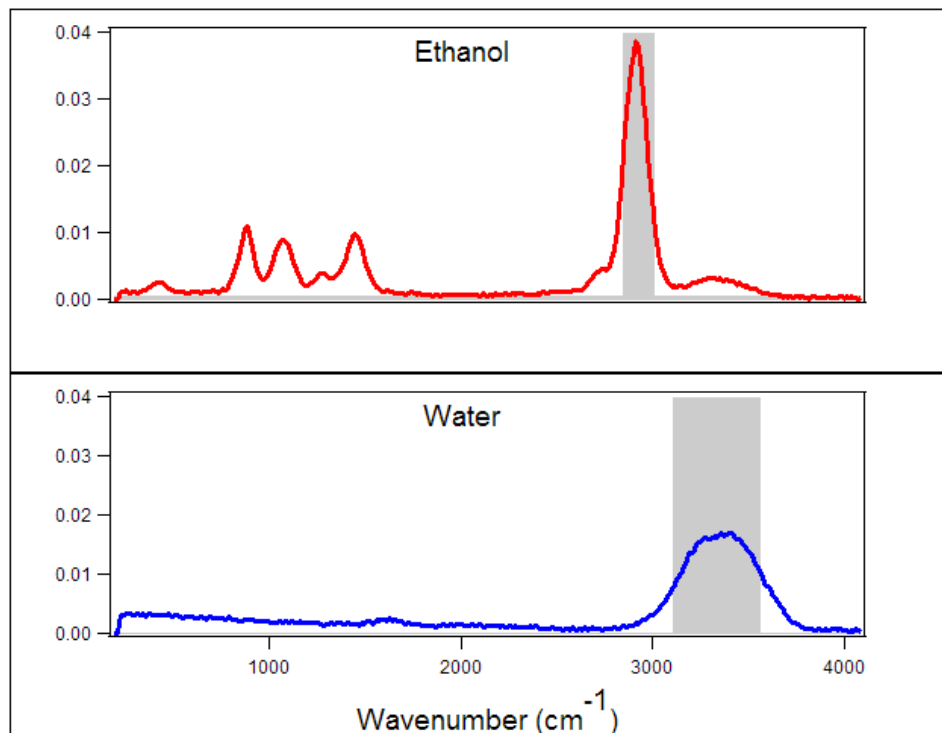


Figure 1.27: Spectrum of pure ethanol (red) and pure water (blue) and the OB filters (gray) associated with each spectrum is shown.

Recovered rates of pure ethanol, pure water and 3 known mixtures were converted into percent volume fraction using equation (1-7). This is used as a calibration curve. Recovered photon rates from tequila were converted to percent volume fraction and plotted on the calibration curve shown in figure 1.28. Our measurements show that tequila has  $36 \pm 2\%$  ethanol. The bottle of silver tequila shows 40% alcohol.

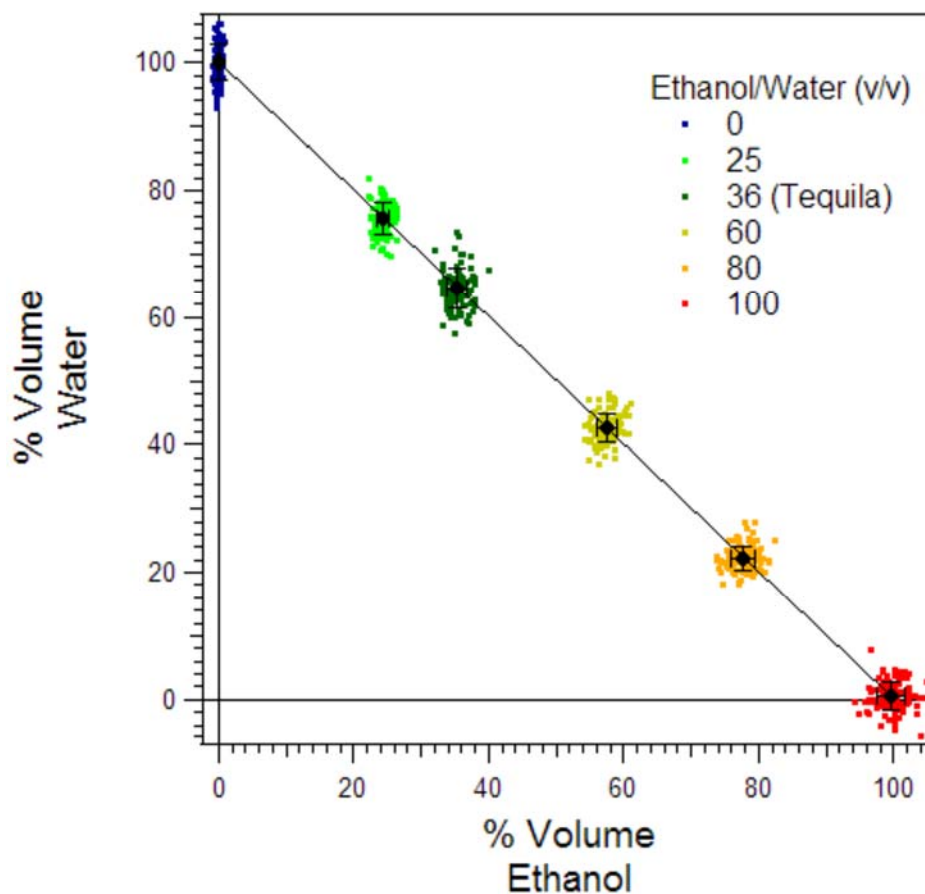


Figure 1.28: Percent volume fraction of ethanol and water is shown in the figure. Volume fractions of 100 independent measurements of pure water (blue) and ethanol (red) and 3 known solutions of ethanol were used to make a calibration curve. Tequila has  $36 \pm 2\%$  ethanol according to our measurements

#### 1.4.4 Imaging using OB-CD Spectrometer

An image is constructed by point scanning the sample using the xy galvanometer mirror scanner. The galvo mirrors are provided with a voltage and the value to voltage received by the mirror in the x and the y dimension define the point being measured. A conventional image shown here consists of a total of 90,000 points square image (300 points in the x and y dimension).

#### 1.4.4.1 All-On Imaging

A sample can be imaged by scanning the sample point by point and measuring the intensity of all the photons generated at that pixel. This is done by sending all the light from the DMD onto one detector and is analogous to a white light image seen through a white light microscope.

##### 1.4.4.1.1 Resolution of the Image Using 1951 U.S. Airforce Resolution Chart

We determined the lateral spatial resolution using a 1951 U.S. Airforce resolution chart. It is used as a widely accepted test for determining the resolution of optical instruments such as microscopes. The resolution of the image is defined by the focused laser spot size (numerical aperture of the objective) in the lateral direction and the wavelength of the laser. The diffracted limited of resolution is given by the equation 1.13:

$$R = \frac{0.61*\lambda}{NA} \quad (1.13)$$

$\lambda$  is the wavelength of the excitation source and NA is the numerical aperture of the given objective. The theoretical resolution is 784 $\mu$ m for wavelength of 0.5145nm and an NA of 0.4.

If the laser does not fill the back of the objective the NA is smaller than the given value and the resolution is larger than the theoretical resolution. Figure1.29 shows the image of U.S. bar target obtained using a white light microscope. It image shows a zoomed image of group 6 and 7 using all on image obtained by the OBCD spectrometer. Analysis



of intensity of group six period six shows that our resolution with a 20X objective and an N.A of 0.4 is  $2\mu\text{m}$  and a field of view of  $500\mu\text{m}$ .

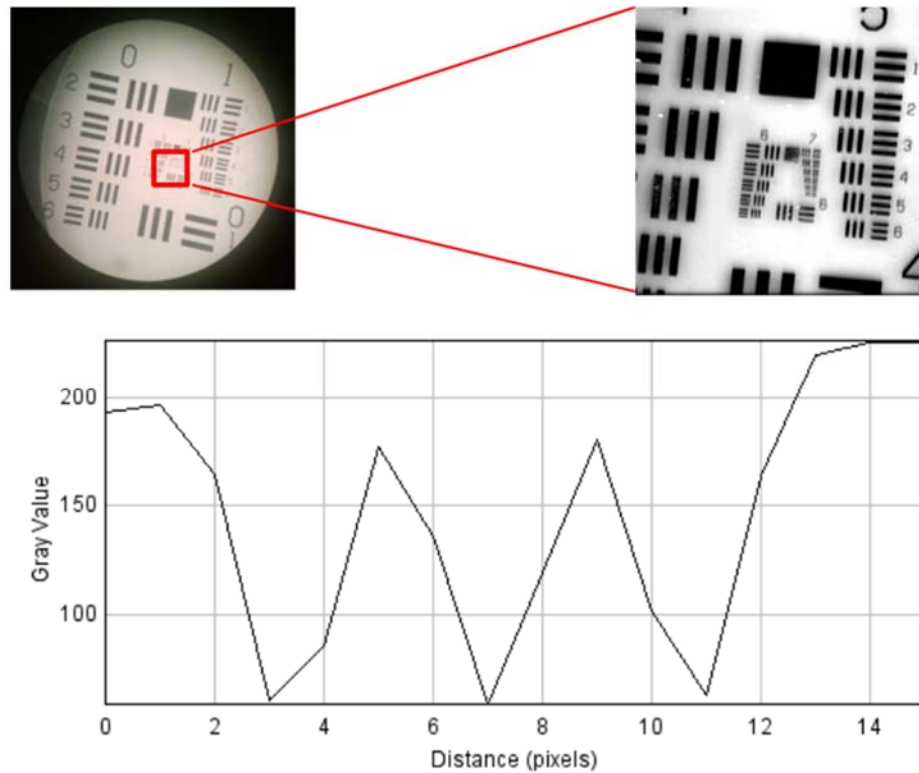


Figure 1.29: Clockwise from top left: White light microscope image showing the full bar target. Top right image: shows the zoomed in all on image of the bar target. Bottom image: intensity profile of group six and element six analyzed by ImageJ.

Three independently measured images of the USAF are shown in figure 1.30. They are false colored in red, blue green. The composite of the 3 produce the bottom right image in white and shows no distortion or shifted pixels. These images were collected with 13mW of laser power, using a 20X 0.4NA objective.

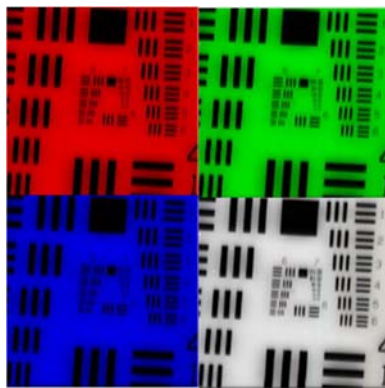


Figure 1.30: Independently measured all on images of the USAF bar target shown in false colored in red, green, and blue (RGB). The image in the lower right-hand panel (white) is a composite obtained by superposing the three RGB images

#### 1.4.4.2 Hyperspectral Chemical Imaging of Acetaminophen and Benzoic Acid Using Pure Spectra for OB Filter Training

Here we demonstrate the application of our OB compressive detection strategy to the imaging of mixtures of acetaminophen and benzoic acid. These chemicals were chosen because of their relevance in the pharmaceutical industry. The spectrum of acetaminophen and benzoic acid is shown in figure 1.31. The correlation coefficient (normalized dot product) of the two spectra is 0.88. OB filters are generated for benzoic acid and acetaminophen using the OB filter algorithm and are shown in figure 1.32. The benzoic acid filter and the acetaminophen filter is applied for 55% and 45% of the total time respectively.

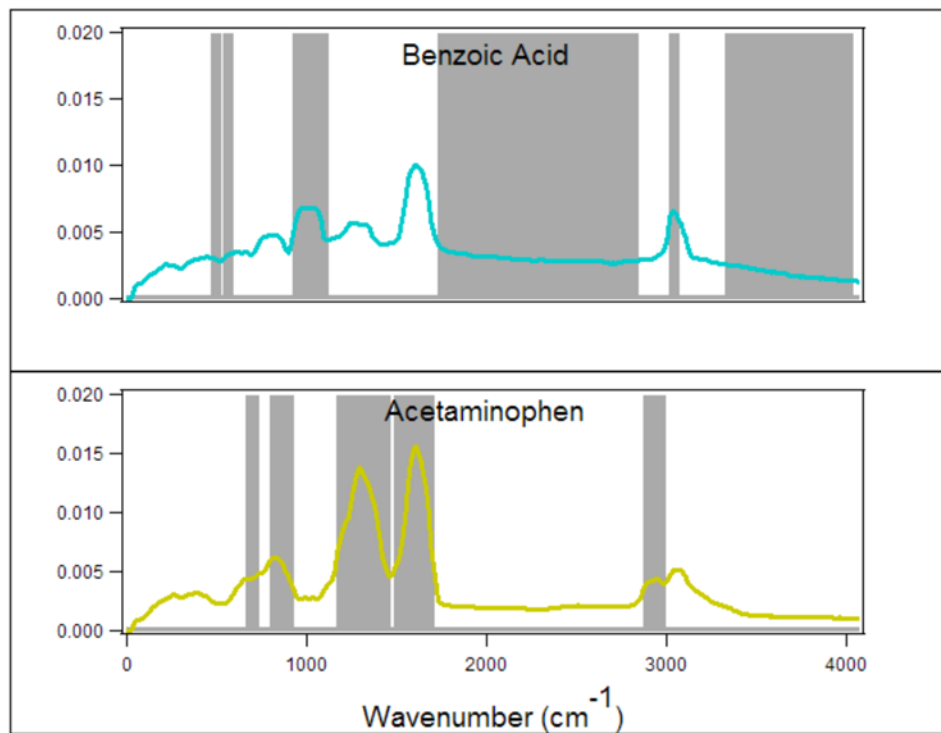


Figure 1.31: The spectrum of benzoic acid (cyan) and acetaminophen (yellow) and the OB filter for each chemical (gray) is shown.

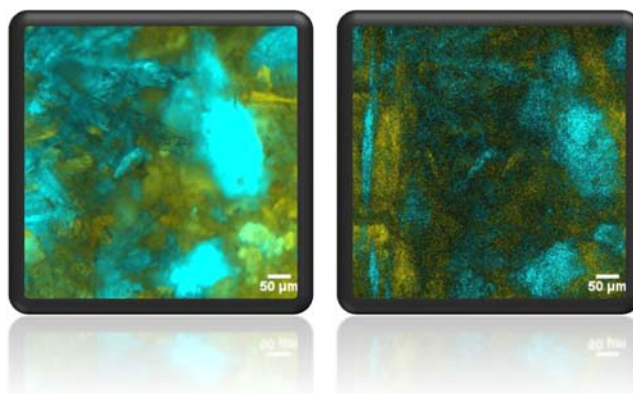


Figure 1.32: 300X300 point images of acetaminophen (yellow) and benzoic acid (cyan) acquired in 90s (left) and 0.9s (right) by collecting photons for 1ms/point and 0.01ms/point.

### 1.4.4.3 Hyperspectral Chemical Imaging of Benzoic Acid and Glucose Using MCR Derived Spectra and Removal of Fluorescence by Training on Bernstein Polynomials

#### 1.4.4.3.1 Hadamard Spectroscopy of Multiple Points on a Line on the Sample

A solid mixture of benzoic acid and glucose was prepared and Raman spectra of multiple points along the line (red) was acquired shown in figure 1.33. The two highlighted spectra are chosen to show the spectral difference at different points on the sample.

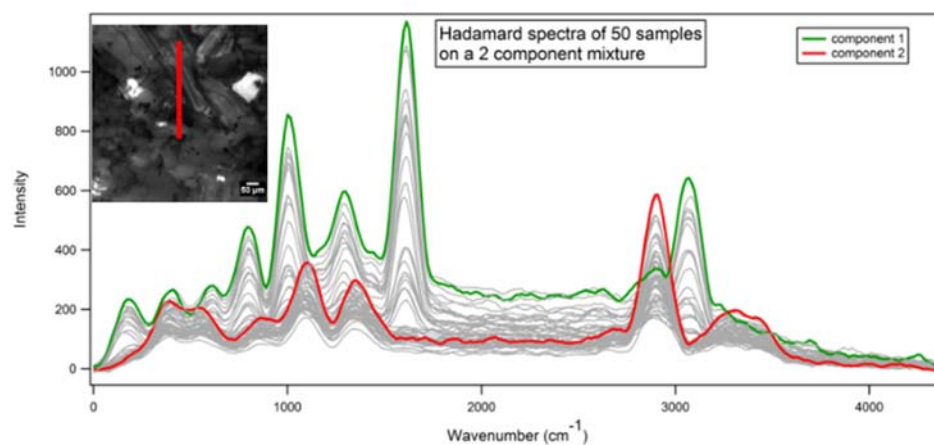


Figure 1.33: The figure shows white light image of a 2 component sample. Hadamard spectra of 50 points were acquired along the red line. The spectra in red and green are visually chosen to show the distinct features of each.

The matrix consisting of the 50 Hadamard spectra was used as input for spectral minimization algorithm. Entropy minimization is spectral minimization technique that is used to recover pure component spectra from multiple mixture spectra without assuming the spectral functional form. It relies on the fact that pure spectral entropy is lower than a

mixture. Figure 1.34 shows the Hadamard spectra of pure components (solid line) and recovered entropy minimized spectra of the two components (dashed line).

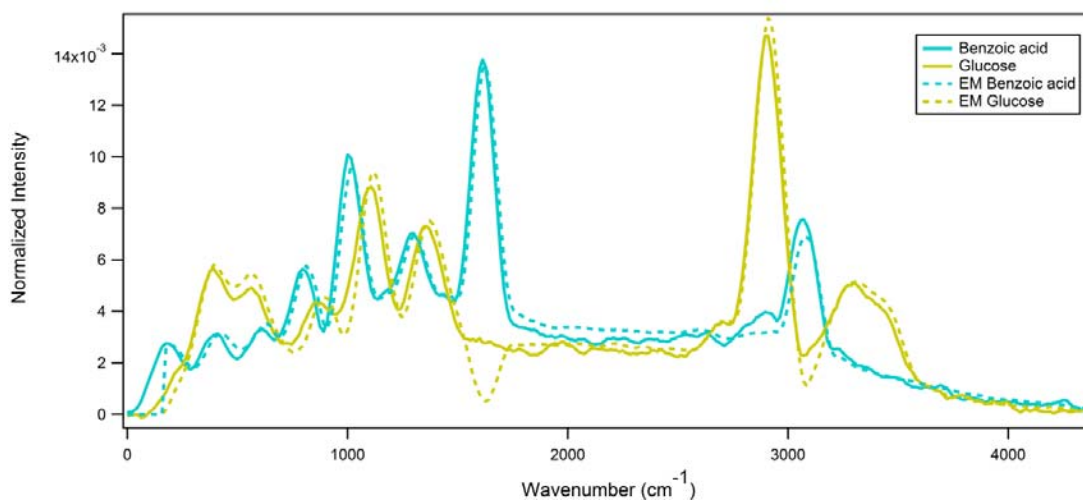


Figure 1.34: The figure shows Hadamard spectra of pure benzoic acid and glucose Solid and also entropy minimized spectra (dashed) obtained from multipoint Hadamard spectra of a mixture sample.

OB filters are generated using two pure component training spectrum and four Bernstein polynomials. OB filters built using Bernstein polynomial removes any fluctuations due to fluorescence in the sample. The sample is scanned using these set of filters and the image is reconstructed using false colors (blue for benzoic acid and yellow for glucose). Another set of OB filters are generated using entropy minimized spectra of benzoic acid and glucose and the four Bernstein polynomials. The sample is rescanned with the new filters and the images compared. The recovered rates for the two chemicals in the mixture using filters generated either way are comparable to each other as seen in figure 1.35.

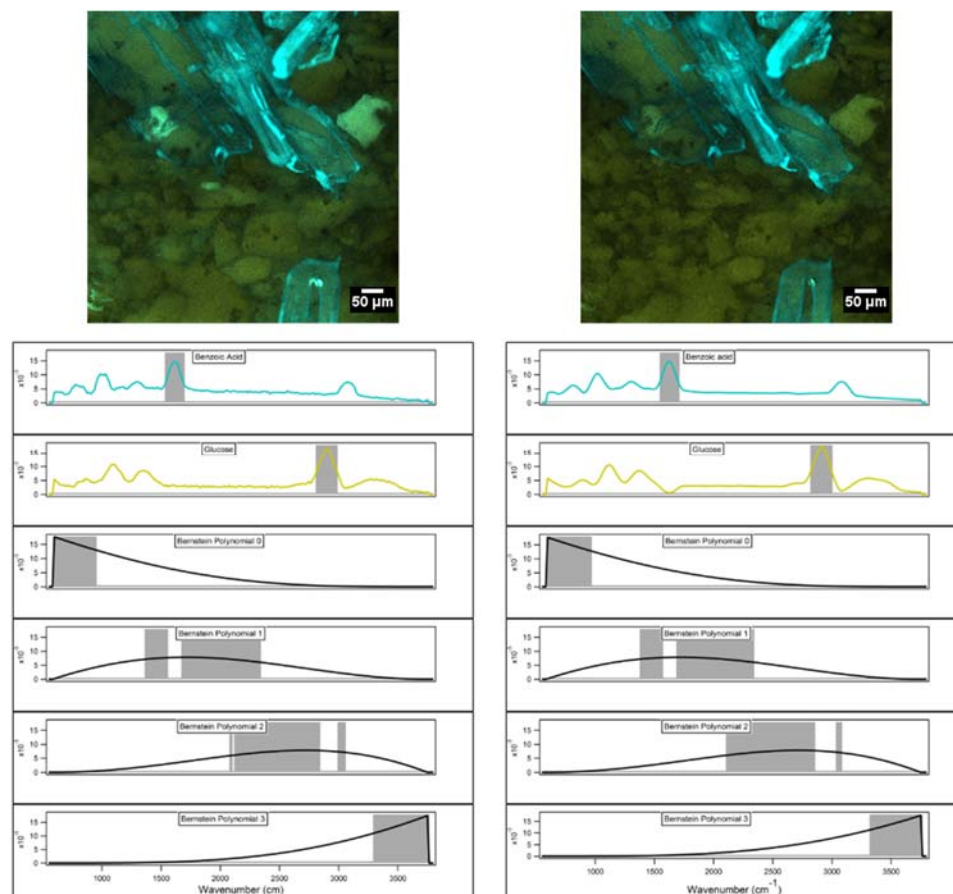


Figure 1.35: The figure shows OB filters generated using pure spectra and Bernstein polynomials on the left along with the image constructed using these filters. On the right OB filters were generated using entropy minimized spectra and the image constructed using these filters

### 1.5 Conclusion

I have described the setup of a new hyperspectral CD spectrometer used for fast label free chemical classification, quantitation, and imaging. I have also described the methods that were used to generate and analyze results obtained from this spectrometer. The new spectrometer collects light more efficiently and this allows us to make measurements with similar signal to noise reported previously faster. The classification of minimally overlapped spectra of benzene and acetone was performed in 10 $\mu$ s which is

three times better than previously reported. The spatial resolution of this instrument is also better than the its predecessor.

## CHAPTER 2 CLASSIFICATION ERROR ANALYSIS OF FULL SPECTRAL AND COMPRESSIVE DETECTION MEASUREMENTS

### 2.1 Abstract

The limiting factor for fast chemical classification using traditional Raman spectroscopy is the signal to noise of the full Raman spectral measurement. Traditionally a Raman spectrum is measured by using a spectrometer that linearly disperses Raman scattered photons and focuses them on a charge coupled device (CCD). In the low signal (fast time) region the read noise of the CCD limits the detection and classification of chemicals. In the recently developed optimized binary compressive detection (OB-CD) strategy, optimized hyperspectral optical filters are generated to identify and quantify chemical species by minimizing the error in recovered chemical concentration. These filters send chemical information (Raman photons from multiple wavelengths) to a single photon counting detector with no read noise. The total chemical concentration is recovered from the counts detected through each filter. Here we compare the results of chemical classification error using full spectral measurement and the OB-CD strategy. Our results show that in the high signal region, OB-CD classification is similar to full spectral classification. However, in the low signal regime (less than 1000 total photon counts), OB-CD Raman spectroscopy outperforms traditional full spectral CCD measurements.



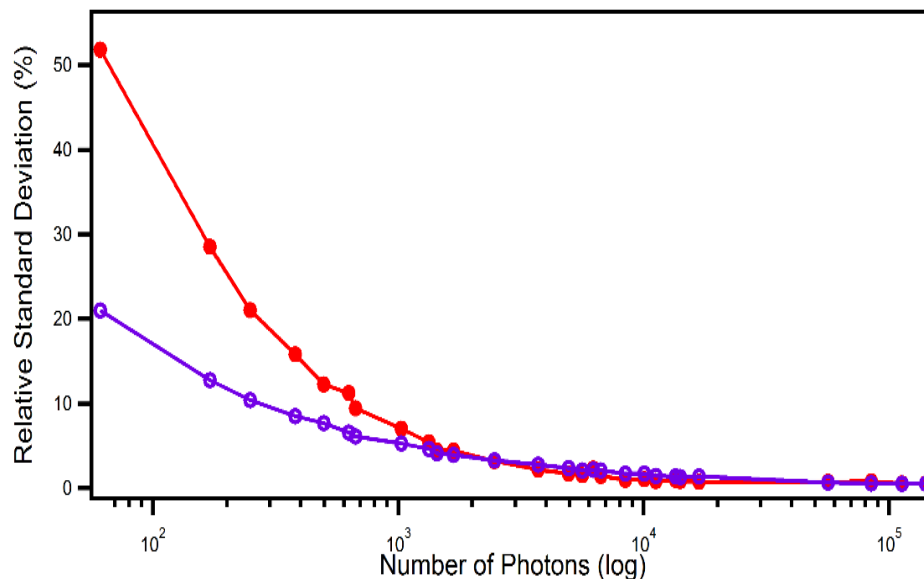


Figure 2.1: The error in classification (%RSD) of acetone using CCD measurements (red) and OBCD measurements (purple) vs total signal (photon counts) of 1000 independent measurements. Analysis of CCD full spectral measurements was done using least squares, and hyperspectral CD measurements were made using OB filters.

## 2.2 Introduction

Fast, label-free classification and quantitation of chemicals in a complex mixture is crucial in different fields of science from counterfeit food and drug detection, to dynamic biological assays and cellular imaging. The quality of rapid identification depends on the quality of measurement being made. Improving the signal to noise ratio of the measurement will help in minimizing uncertainty of estimating the chemical of interest. In the previous chapter I described new instrumentation that improves on the efficiency of Raman scattered photons detected, and also described OB-CD filtering that allows us to make multiplexed hyperspectral measurements onto one detector. In this chapter, I describe a new Raman instrument that is capable of making measurements either on the new, previously described, hyperspectral CD spectrometer or on a traditional Czerny Turner spectrometer

equaled with an electron-multiplying (EM) charge coupled device (CCD). I also compare the results of classification error associated with measurements made both the spectrometers over a range of Raman signals.

Full spectral measurements are traditionally acquired by measuring the intensity of linearly dispersed Raman scattered photons onto an array detector like a CCD. In applications where speed of the measurement is not significant, a long exposure or signal averaging multiple exposures yields good signal to noise spectrum that can be used towards chemical identification. However, for fast acquisition, the accuracy and the speed of the measurement is limited by noise associated by the detector and the amount of signal at each exposure. For example, a signal that contains a 1600 photons dispersed over a 1600 channel (assuming even distribution of the signal) detector with the read noise of 1 photon per pixel (typical read noise for 100KHz read speeds) would get buried by the noise of the detector. In contrast, the results from the previous chapter show that multiplexing wavelength channels onto a read noise free detector will have sufficient signal to noise to make successful classification even under low signal regimes.

Simulations show that at low signals the read noise dominates full spectral measurements and multiplexing photons from specified wavelengths onto one detector improves classification<sup>30</sup>. In this chapter I will compare classification results obtained experimentally by least squares analysis of full spectral measurement made on a CCD and compressive detection measurements made using OB-CD filters at low and high signal regimes.

## 2.3 Methods

### 2.3.1 Instrument

#### 2.3.1.1 Raman Microscope Stage

The Raman microscope is built by modifying an Olympus white light microscope (BX 43) and mounting a board containing the excitation source and all the associated optics on top of the existing microscope. A 532nm single mode optically pumped semiconductor laser module (Coherent Sapphire SF) is used as the excitation source. The laser has a maximum output power of 150mW, a wavelength accuracy of 0.1nm and a 100:1 v-polarization ratio. The laser is guided through a laser line cleanup filter (Semrock Part Number: LL01-532-12.5), a half wave plate (Thorlabs WPH10M-532) and a Glan polarizing cube (Thorlabs GL10-A) using half inch mirrors. The laser is directed towards a pair of galvanometer mirrors using a 532nm RaserEdge dichroic beamsplitter (Semrock Part Number: LPD02-532RU-25). The laser is then guided on the back of the objective using a pair of galvanometer x/y scanning mirrors (Thor labs) and a pair of 150 mm two inch achromatic lenses arranged in a 4f configuration. This configuration is used commonly in beam scanning techniques where the sample is stationary and the laser is moved over the sample in order to image the sample. The first lens is exactly one focal length away

from the galvanometer mirrors and is two focal lengths away from the second lens, and the second lens is one focal length away from the back of the objective hence 4f configuration. The laser never moves from the back of the objective but comes in at different angles through the objective enabling to sample a different spatial spot on the sample. The microscope is equipped with an Olympus 20X objective (LMPLFL20X) N.A. 0.40, and a working distance of 12.0mm. Rayleigh and Raman scattered light is collected through the same objective in a backscattering collection geometry. The scattered light is separated based on the wavelength at the dichroic mirror. Longer wavelength Stokes shifted Raman photons go through the dichroic beam splitter. Additional Rayleigh scattered photons are filtered using an edge filter (Semrock Part Number: LP03-532RU-25) and the Raman photons are directed either to the home built OB-CD spectrometer or towards a Czerny Turner Spectrometer (Princeton Instruments SpectraPro 2300) using a flip mirror.

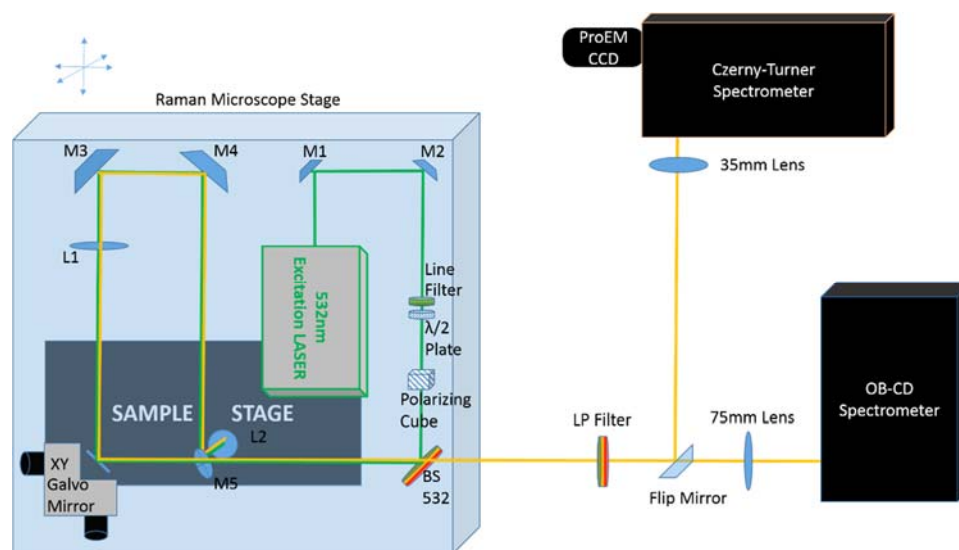


Figure 2.2: Schematic of the Raman microscope stage is shown. The Raman scattered photons are detected either using the OB-CD spectrometer or the Czerny Turner Spectrometer by flipping a mirror

### 2.3.1.2 OBCD Spectrometer

The collimated Raman scattered photons are focused onto the entrance slit of the OB-CD spectrometer using a 75mm achromatic lens. The light is re-collimated using another 75mm achromatic lens and the collimated light is dispersed using a transmission volume holographic grating (VHG, 600line/mm, cwl:600nm). A focusing lens is used to focus the dispersed wavelengths onto a digital micromirror device (DMD). The DMD is a rectangular array of 608X684, 10 $\mu$ m pitch mirrors that can be programmed to tilt either +12° or -12° to the flat position. All the light (light spread out in different diffraction orders) is collected in one or the either direction by reimaging the DMD onto 2 photomultiplier tube (PMT) counting modules using a 75mm/35mm achromatic doublet pair of lenses. The PMT counting modules output a TTL pulse when a photon is detected and National

Instruments data acquisition card is used to synchronize the voltage on the galvanometer mirrors and the duration of data collection. This spectrometer can be used to obtain a Raman spectrum by wavelength scanning or by Hadamard scanning. In wavelength scanning columns of mirrors are sequentially set to direct light to one detector for a set duration of time. Hadamard scanning is achieved by applying a Hadamard mask onto the DMD and measuring the counts of photons associated with that mask. The spectrum is reconstructed by applying inverse Hadamard matrix to the measured counts. Optical filters can also be built and applied directly to the DMD as shown later in OB-CD and OB-CD2 filtering.

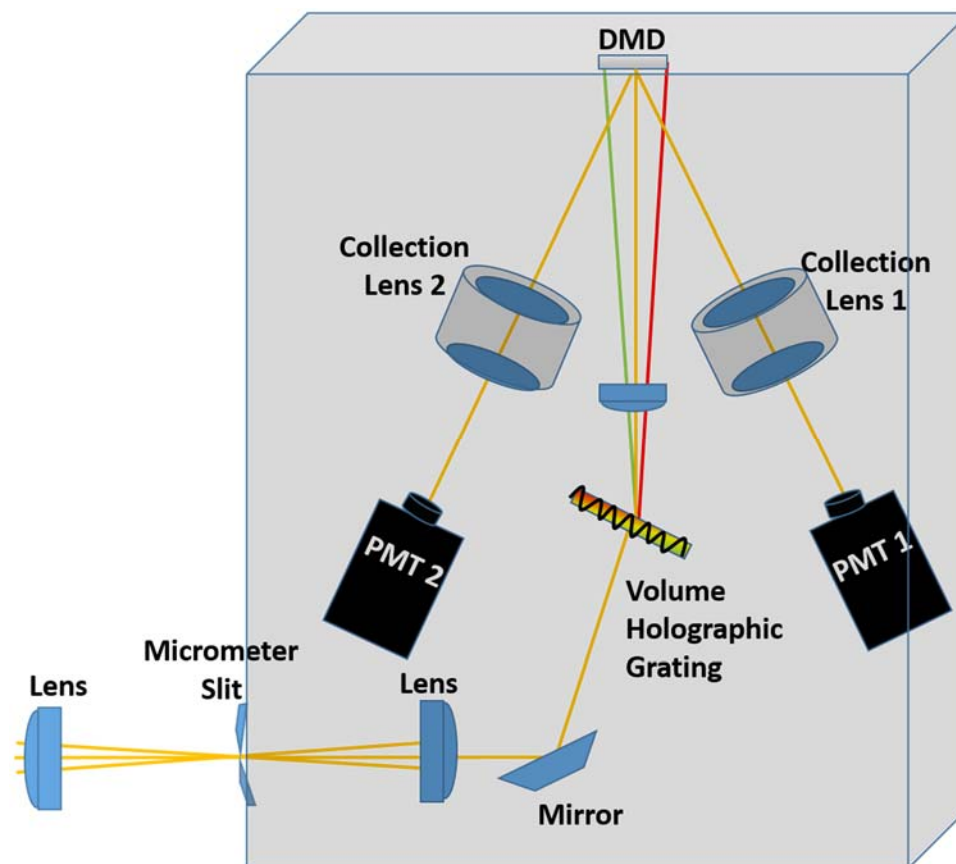


Figure 2.3: Schematic of the OB-CD Spectrometer. Collimated beam of light is focused on a slit at the entrance of the spectrometer. The beam is recollimated and dispersed using a VHG. The dispersed beam is focused onto the DMD. The DMD can send parts of all of the light towards either detector

### 2.3.1.3 The Czerny Turner Spectrometer:

The Spectrometer is a conventional spectrometer containing a collimating mirror, a reflective grating (300gr/mm) a focusing mirror. An electron multiplying charge coupled device (EM-CCD) is used as a detector. The incoming Raman beam is focused on the entrance slit of the spectrometer using a 35mm achromatic lens. The lens was chosen to match the  $F\#$  of the spectrometer such that the grating is filled with the collimated light. The grating disperses the collimated light into its component wavelengths and a focusing

mirror focuses each wavelength onto a CCD detector. The CCD detector can be used either as a low noise/high signal detector where measurements are made either in 100KHz or 1MHz read rate or as a High capacity detector where faster measurements can be made at 1MHz, 4MHz or 8MHz read rate. An analog gain of low medium or high can also be applied in any of the two modes with all read rates. In spectrum with low photon distribution, the signal can get buried under read noise especially when fast measurements are made. In this case another type of gain, the electron multiplying (EM) gain can be applied. When the EM gain is applied, the photoelectrons generated are accelerated through a series of gain registers that amplify these electrons by impact ionization before the signal is read by the read register. Therefore, the signal is amplified but not the read noise. Full spectral measurements are made using a CCD and the data is post processed using chemometric algorithms to separate out useful information.



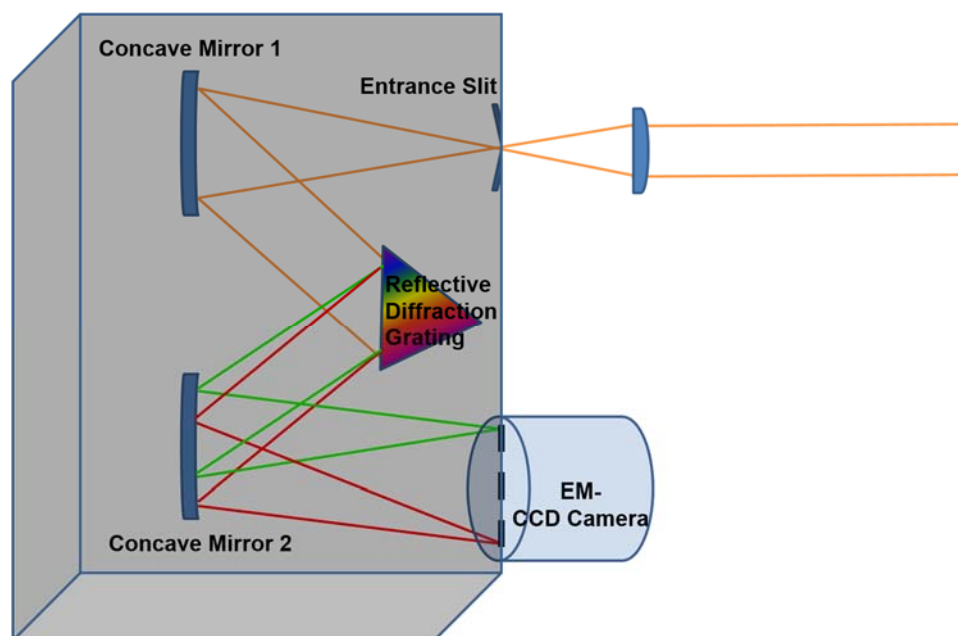


Figure 2.4: Schematic of the Czerny Turner Spectrometer equipped with an EM-CCD camera. The collimated beam of Raman photons is focused onto the entrance slit of the spectrometer. A recollimating mirror reflects collimated light towards a reflective grating. The dispersed photons are focused onto the CCD detectors using a focusing mirror

### 2.3.2 Aligning the CCD

This section describes the method for alignment for the P.I. ProEM camera. Once the camera is attached to the Acton SpectraPro 2300 spectrometer, the slit on the spectrometer is set to  $10\mu\text{m}$  (1 on the slit). A neon pen is used as the light source. The 300gr/mm grating is chosen with the central wavelength as 668nm. Alignment of the camera requires adjustment of the rotation of the CCD and the adjustment of the focus of the CCD. Rotation of the CCD ensures that spectral lines are perpendicular to the rows of the CCD array. The rotation of the CCD is done by physically rotating the CCD while monitoring the image of the CCD using the LightField software. The use of the cursor helps in the rotation and is shown in figure 2.5. Adjusting the focus of the spectrometer is

achieved by moving the camera back and forth until the spectrometer's focal plane lies on the CCD.

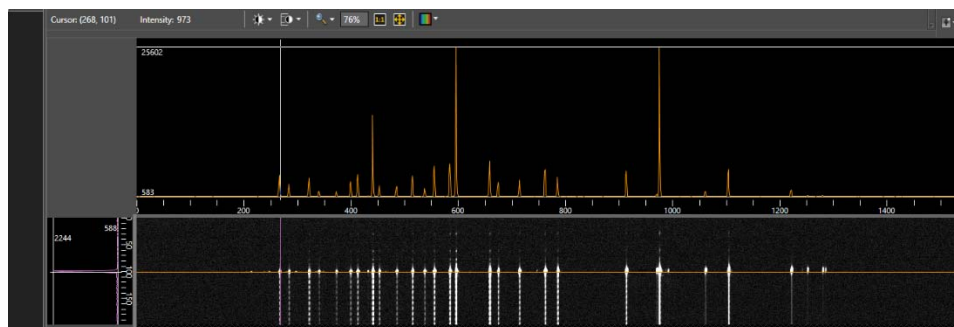


Figure 2.5: A screen shot of Lightfield software used to control the ProEM. The top graph of the figure shows the Spectrum associated with a single row of pixels. Below the graph is the image acquired by exposing the CCD for 0.5s at a read rate of 1MHz in the electron multiplication (EM) mode with EM Gain 1. The most intense spots correspond to the pixels having the most signal. The horizontal line shows the spectrum associated with that row of pixels. The graph on the left of the Image shows the information associated with just that one pixel

Figure 2.6 and 2.7 show the resolution of the neon and benzene spectrum using the Czerny turner spectrometer and CCD detector and on the hyperspectral CD spectrometer using the PMT detector. Figure 2.6 shows the spectral resolution of benzene with respect to entrance slit width of the hyperspectral CD spectrometer. The spectrum acquired using the CCD has an electronic baseline of about 600 counts. Figure 2.7 shows the spectrum of benzene acquired using the two different spectrometers. The benzene peak at around 635nm corresponds to the CH stretch ( $3055\text{cm}^{-1}$ ) is less intense in the spectrum acquired by notch scanning the DMD and measured by the PMT than that acquired on the CCD. This may be due to the fact that the sensitivity of the PMT is lower at higher wavelengths.

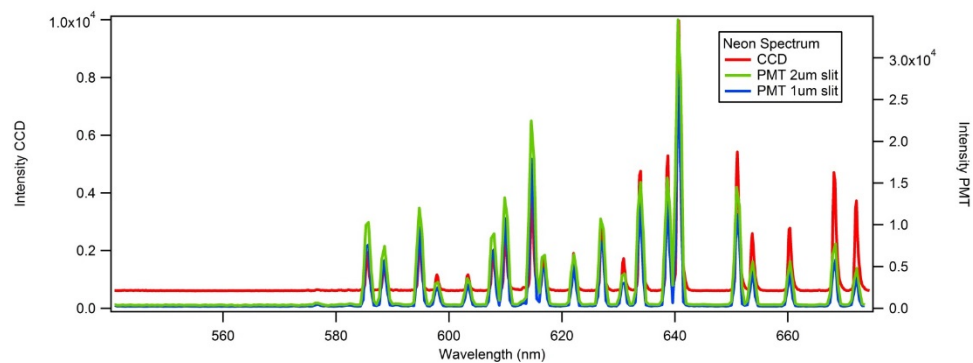


Figure 2.6: Neon spectrum acquired on the ProEM (Red), and on the PMT with the spectrometer slit position 1 and 2 (blue and green). The FWHM of most peaks in the same region of neon spectrum acquired on a CCD and a PMT are comparable. Closing the entrance slit of the spectrometer improves resolution and a slit width at position less than 2 on the OBCD spectrometer was found to be optimal

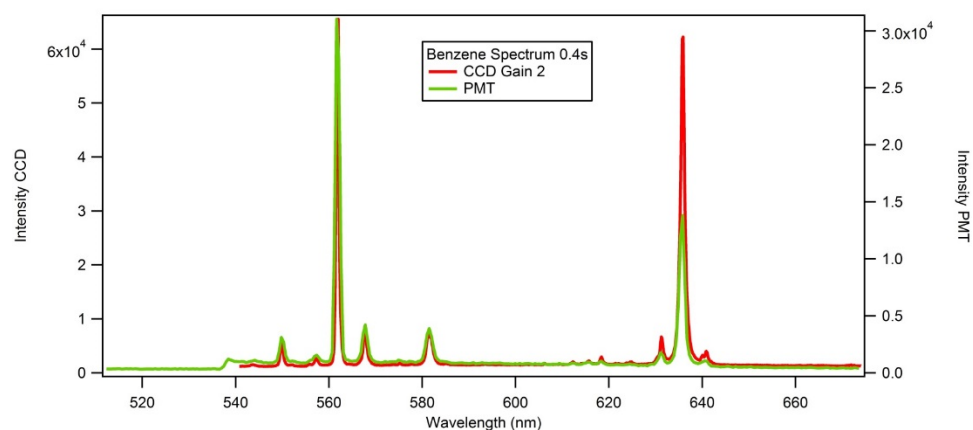


Figure 2.7: Benzene spectrum acquired on the ProEM (Red), and on the PMT with the spectrometer slit position 2 (green). Again the resolution is comparable. The sensitivity of the PMT diminishes at higher wavelength. The two benzene spectra were acquired at the same exposure time of 0.4s. The CCD had a gain setting of medium (gain 2 on a scale of 3). This may be the reason in the higher counts

### 2.3.3 High Speed Spectral Acquisition

Under the add-in tab of LightField 4.0 software select highspeed camera. This changes the experimental settings to minimize readout time and maximize the number of

frames of spectra per second. In this setting we bin 20 rows of the CCD. Figure 2.8 shows 1000 independently measured spectra of acetone acquired at 10ms exposure using 8MHz read speed. There is an electronic offset in each measurement that can be removed by subtracting the minimum of each spectrum with itself. Figure 2.9 shows the spectra of acetone collected in figure 2.8 with the minimum subtracted from each spectrum.

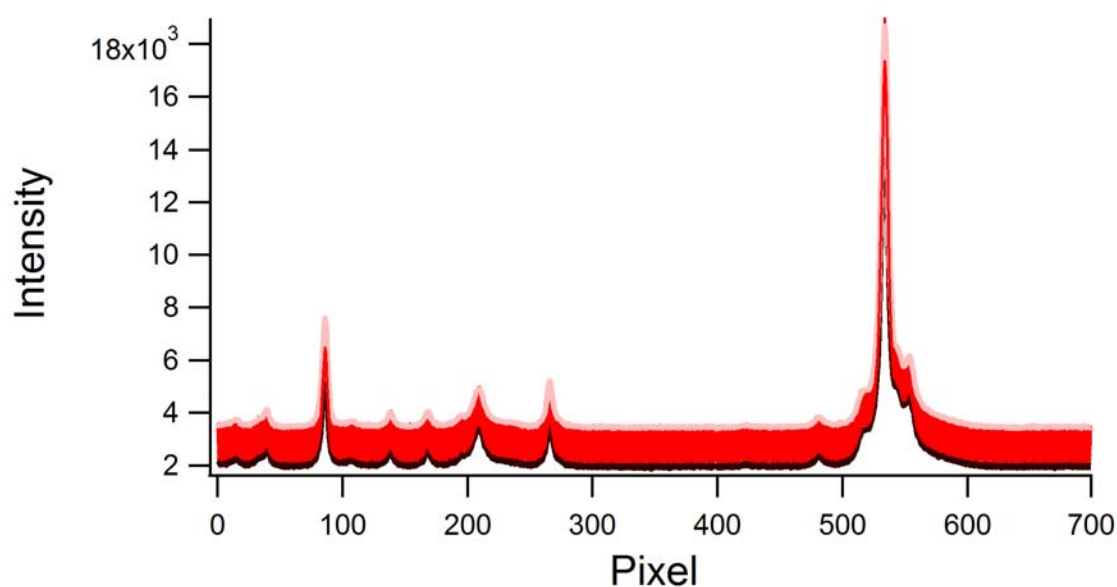


Figure 2.8: Spectrum of acetone was acquired at 10ms exposure time and 1000 independent acquisitions show a continual increase in the baseline. The spectrum in black is the 1st experiment and the light red spectrum is the last experiment

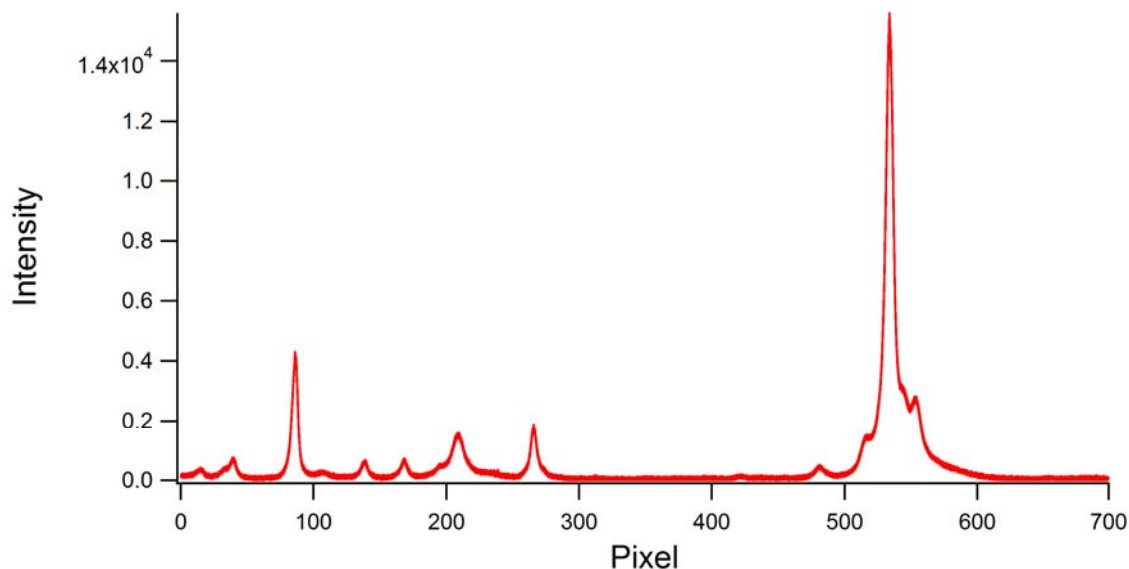


Figure 2.9: Electronic baseline shift between experiments is corrected by subtracting the minimum of value of the spectrum from the whole spectrum and each experiment

#### 2.3.4 Analysis of full spectral CCD measurements using classical least squares

Spectra of a mixture are assumed to be a linear combination of small number of pure components and the concentration of the pure components can be solved using classical least squares as shown in equation 2.1.

$$\mathbf{D} = \mathbf{cS} \quad (2.1)$$

where  $\mathbf{D}$  is the data matrix containing  $n$  independent experiments and  $m$  wavelength channels (pixels),  $\mathbf{c}$  is a column matrix containing concentration of each component in the sample in  $n$  experiments and  $k$  components, and  $\mathbf{S}$  is the area normalized spectra of the pure components in the sample with  $k$  pure spectra in rows and  $m$  wavelength channels in columns.

In the classification of 2 components such as acetone and benzene (results shown in section 2.4.3.3.1) 1s spectrum of pure acetone and benzene was acquired. Classification

was performed by solving for concentration of 1000 acquisitions of pure acetone of benzene. To remove any electronic baseline, minimum value was subtracted from each spectrum and a constant value was used as a third component of training spectra. The spectra were cropped to contain 700pixels since there was no Raman information in the remaining 800 pixels. Each pure spectrum was normalized to unit area and concentration solved by solving equation 2.2

$$\mathbf{c} = \mathbf{DS}^T(\mathbf{SS}^T)^{-1} \quad (2.2)$$

The normalized spectra used for training for classification of acetone and benzene is shown in figure 2.10

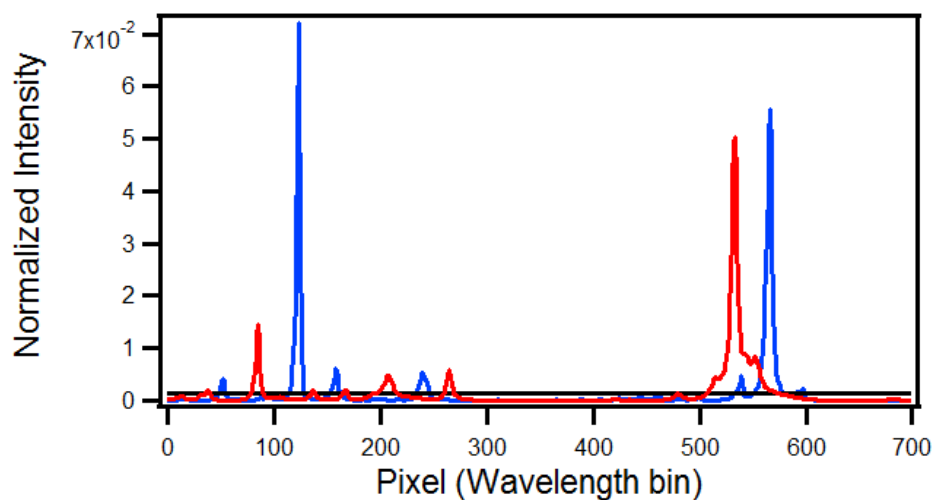


Figure 2.10: Normalized pure spectra used in used in the classification of acetone and benzene

The pseudo inverse  $\mathbf{S}^T(\mathbf{S}\mathbf{S}^T)^{-1}$  spectra is shown in figure 2.11.

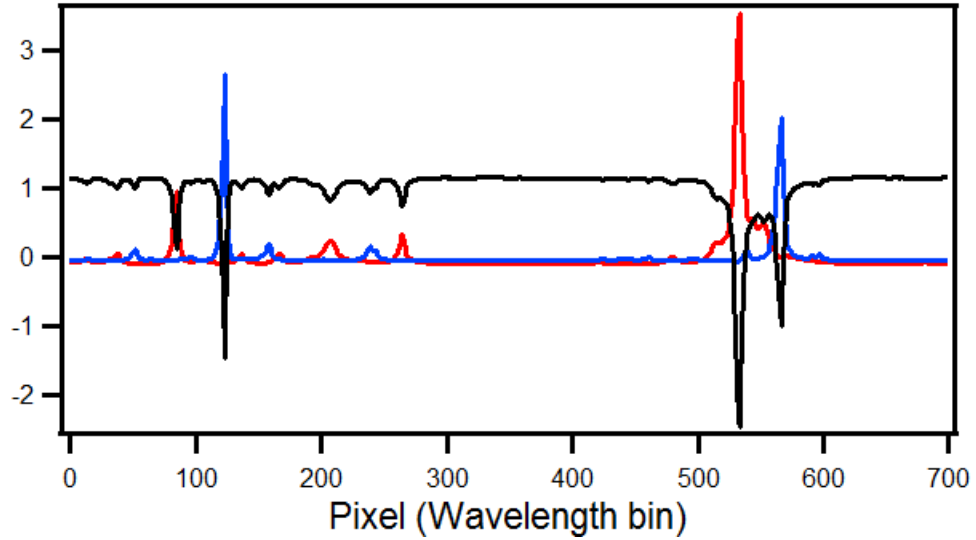


Figure 2.11: Pseudo inverse of pure spectra used in used in the classification of acetone and benzene

### 2.3.5 Noise Associated with Spectral Measurement Using A CCD

The primary sources of noise in a CCD is defined in the equation below

$$\text{Overall Noise} = \sqrt{\text{Read Noise}^2 + \text{Dark Noise}^2 + \text{Shot Noise}^2} \quad (2.3)$$

Shot noise arises from intensity variation of the signal and is modeled by a Poisson process. The noise is proportional to the square root of the intensity of the signal. In the case of high signal, this is the predominant source of noise.

$$\text{Shot Noise} = \sqrt{\text{Signal}} \quad (2.4)$$

Dark noise also known as thermal noise is generated due to randomly generated electrons that is independent of light falling on the CCD. It is due to thermal energy within

the silicon lattice of the CCD. Thermal noise is also modeled by Poisson statistics. Thermal noise is time dependent and is significantly decreased by cooling the CCD. The ProEM CCD is TE cooled to -60°C and has a dark current of 0.0036e/pixel/s.

$$\text{Dark Noise} = \sqrt{\text{Dark Current} * \text{Integration Time}} \quad (2.5)$$

Read noise is a term used for the collection of electronic noise inherent to the CCD and the electronics associated with it. It represents the error associated in the quantification (read) of electronic signal on the CCD. Read noise is modeled by Gaussian statistics. In low signal regions (especially due to fast data acquisition), the read noise is the limiting factor limit of detection.

Table 2.1 is the certificate of performance provided for the ProEM 200B+ eXcelon (1600X200) camera used to measure the spectra in all the experiments described in this chapter.



Table 2.1: Certificate of performance for ProEM 200B+ eXcelon camera

Princeton Instruments		Certification of Performance			
Read Noise (e- rms)		Full Well (Ke-)			
Multiplication Gain Amplifier		Measured			
Read Speed	Measured Read electrons	Single Pixel Output node (Multi. Gain)		130 900	
8MHz	71.74				
4MHz	28.79				
1MHz	18.98				
Traditional Gain Amplifier					
1MHz	6.63				
100KHz	3.01				
Dark Charge (e-/pixel/sec)					
Measured:		0.0063 @ -60°C			
Conversion Gain (e-/cts) and Non-Linearity					
Multiplication Gain Amplifier					
Speed	Gain#1	Gain#2	Gain#3	Non-Linearity	
8MHz	14.31	8.01	4.23	<2%	
4MHz	11.18	6.11	3.09	<1%	
1MHz	8.92	4.95	2.43	<1%	
Traditional Gain Amplifier					
1MHz	3.51	1.72	0.86	<1%	
100KHz	3.21	1.60	0.78	<1%	

### 2.3.6 Complementary optimal binary compressive detection (OBCD2) filters

OB-CD filters used in chemical classification multiplex selective wavelength channels onto one photon counting PMT while photons from all the other wavelength channels are not detected. OB-CD2 filters similar to OBCD filters classify chemicals by multiplexing wavelength channels but takes advantage of detecting all the photons from all wavelength channels onto 2 PMT's. In classification of a binary mixtures, only one filter is needed to classify both chemicals for information from both detectors are utilized.

OB-CD2 filters classification algorithm minimizes the total volume of the multidimensional ellipsoidal confidence region derived from the covariance matrix of the estimated Raman scattering rates unlike the OB-CD algorithm that minimizes the sum of the variance of the estimated Raman scattering rates. Normalized pure component spectra are as training spectra and is stored in a matrix  $\mathbf{P}$ . OB-CD2 filters have a value of "0" where the photons are sent in the  $-12^\circ$  direction and counted by the PMT or "1" where the photons are sent towards another PMT in the  $+12^\circ$  direction. The time used in the measurement is equally distributed to 50:50 of the total integration time. The counts  $\hat{\chi}$  through each filter is measured and converted into chemical scattering rates (photons emanating from a pure component per second) given by equation 2.6

$$\hat{\Lambda} = B\tau^{-1}T^{-1}\hat{\chi} \quad (2.6)$$

where  $B$  is  $(F^T P)^{-1}$ , the left inverse of the  $F^T P$  where  $B F^T P = I_n$ . This is similar to OB-CD filters except OB-CD filters are optimized to minimize the determinant of the variance-covariance matrix of the mean rates.

$$Q(B, T, F, \bar{\Lambda}) = \det(\tau^{-1} B T^{-1} \text{diag}(T F^T P \bar{\Lambda} T^{-1} B^T)) \quad 2.7$$

where T is a matrix whose diagonal entries indicate the fraction of total measurement time.

Unlike OB-CD where the number of filters equals the number of chemicals classified, OB-CD2 requires m pairs of filters where  $m=n-1$  pair of filters and n is the number of chemicals. In 2 component mixtures such as the ones classified in all the examples here, 1 pair of filter was needed to classify 2 components. The filter was applied on the DMD and the counts were measured on PMT1 and the complementary filter counts were measured on PMT2. Figure 2.12 shows an example of OB-CD2 filters where all the white region of the bitmap corresponds to the DMD pixels that send light in the  $+12^\circ$  direction while all the black region corresponds to the region of the DMD facing the  $-12^\circ$ .



Figure 2.12: The figure shows one pair of OB-CD2 filters for acetone and benzene. Notice that they are complementary to each other and only one of these is needed to be applied to the DMD to make count measurements on both the detectors.

## 2.4 Results

### 2.4.1 Measured Read noise of the Detector

Read noise is time independent and is associated with number of times the CCD is readouts. We determined the read noise of the ProEM camera by shuttering the camera and TE cooling the camera to  $-60^{\circ}\text{C}$ . One thousand independent measurements were made at exposure time of 10ms and at different analog gain and read settings. EM gain was set to 1 for read speeds of 1MHz, 4MHz and 8MHz. Figure 2.13 shows Read noise/pixel (the mean of the standard deviation the thousand independent measurements of all pixels) plotted verses read rate. The read noise increases with an increase in read rate and hence becomes the limiting factor in the detection of low signals (fast measurement times). The read noise measured experimentally matches the CCD certification provided by the manufacturer.

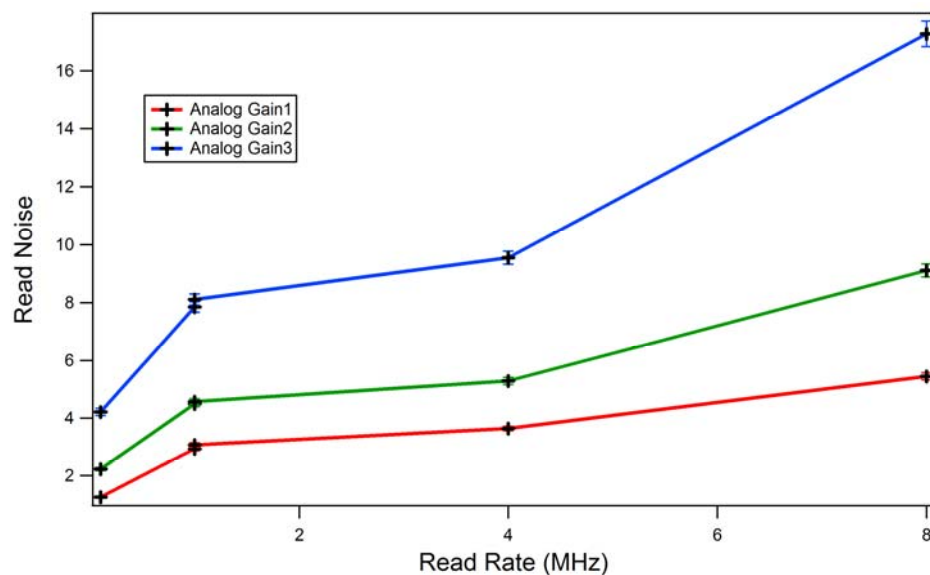


Figure 2.13: Read noise vs read rate is shown at different analog gains. The read noise increases as a factor of readout rate

#### 2.4.2 Nonlinearity in Signal a Function of Exposure Time

The Raman photon counts should scale as a function of exposure time. That is if the exposure time is decreased from 100ms to 10ms, the total Raman photons (and also the intensity of peaks) should decrease by a factor of 10. In order to test if this was true we measured the Raman signal of acetone at a constant laser power and fastest read speed (8MHz) but at different exposure times. The CCD was configured to collect at the fastest possible mode using the fast spectral measurement add-in with settings described in the methods section. Figure 2.14 shows the spectra of acetone collected at exposure times of 100ms, 10ms and 1ms. The intensity of the CH stretch at around 535pixels was used to check the linearity of the CCD as function of exposure time. Background was subtracted from each of the spectrum and the intensity of the CH stretch after background correction

was approximately 15000, 1500, and 200 counts for 100ms, 10ms and 1ms exposure time respectively.

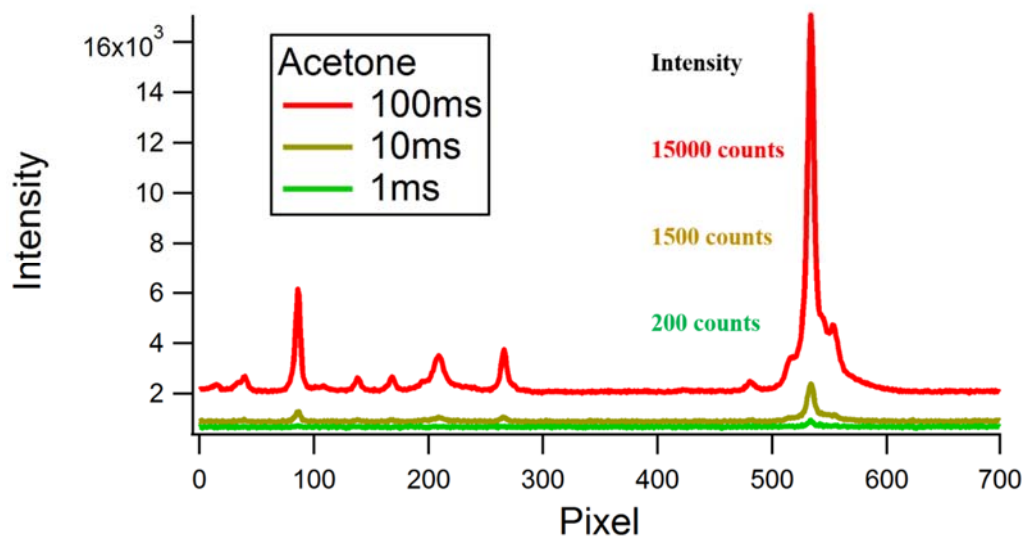


Figure 2.14: Spectrum of acetone was measured at different exposure time and constant laser power and 8MHz readout speed. The figure shows a non-linearity in signal with decrease in exposure time

The detector shows a non-linearity in measured intensity as a function of exposure time for fast exposures. Figure 2.15 shows the log of maximum intensity of CH stretch of acetone vs. the log of exposure time in  $\mu\text{s}$ . The red curve is the experimental value and the black curve is the theoretical value scaled linearly to exposure time. The experimental curve deviates from the theoretical value for times below 10ms. The experimental value can be fit using a 3rd order polynomial and the exposure time value can be derived to correct for the deviation.

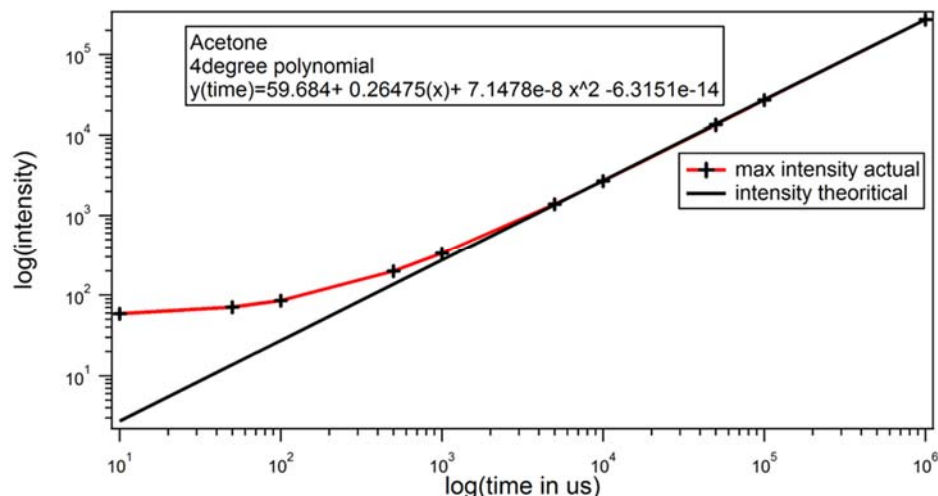


Figure 2.15: Log of the intensity of acetone peak at 550 pixels ( $\sim 2900\text{cm}^{-1}$ ) is plotted against the log of exposure time (red). The figure shows a non-linearity in signal with decrease at exposure times faster than 5ms

A correction factor is applied for the measurements that were made at readouts faster than 5ms. For example, to collect Raman signal that corresponds to 1ms, the CCD had to be read out at approximately  $800\mu\text{s}$ .

### 2.4.3 Acetone and Benzene Classification

The spectrum of acetone and benzene is shown in Figure 2.16. The correlation coefficient (normalized dot product) of the two spectra is 0.3. Four different strategies are demonstrated for the classification of these chemicals. First, we demonstrate OB-CD strategy that has also been previously demonstrated. Next we introduce the results of the new OB-CD2 strategy that uses complementary filters and 2 detectors to measure all the Raman scattered photons. We also demonstrate classification of acetone and benzene using least squares applied to full spectral measurement of the 2 chemicals using a CCD in the

traditional mode and in the EM mode where the EM gain is set to 40. All the measurements were performed at laser power of 30mW, 3mW and 1mW and time points of 100ms, 10ms and 1ms for each laser power. The total photons at each time point were measured on the OB-CD spectrometer by setting all the DMD mirrors towards +12° direction and collecting on one of the 2 PMT's. The total photons counted on the PMT only shows the amount of light detected in the OB-CD spectrometer and does not account difference in losses between the Czerny Turner spectrometer and the OB-CD spectrometer.

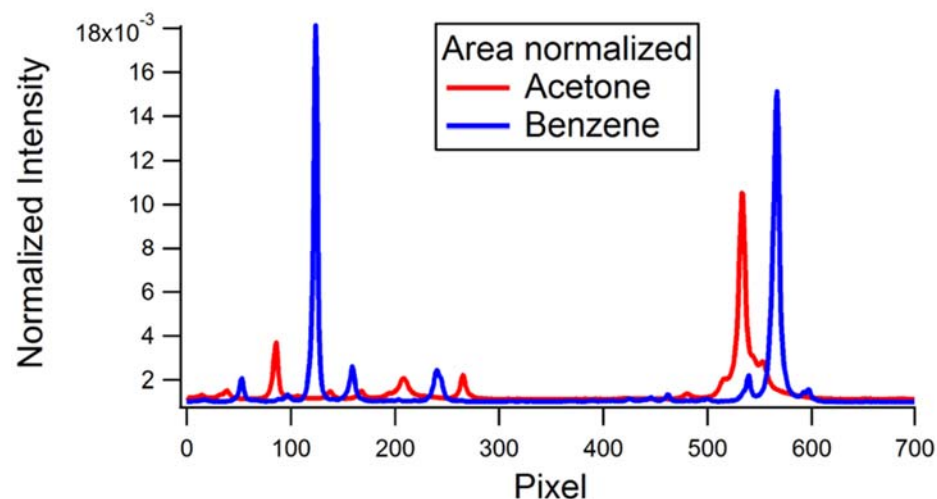


Figure 2.16: The area normalized spectrum of pure acetone (red) and pure benzene (blue) is acquired on a CCD

#### 2.4.3.1 Classification of Acetone and Benzene using OB-CD

OB-filters were generated for pure acetone and pure benzene as described in the previous chapter. Figure 2.17 and figure 2.18 shows the spectrum of acetone and benzene and the OB filters associated with each chemical. The training spectrum was acquired by recovering a 1s Hadamard spectrum of each of the chemicals with 30mW laser power on



the sample. The OB filters generated here were used for all the OB-CD for acetone and benzene.

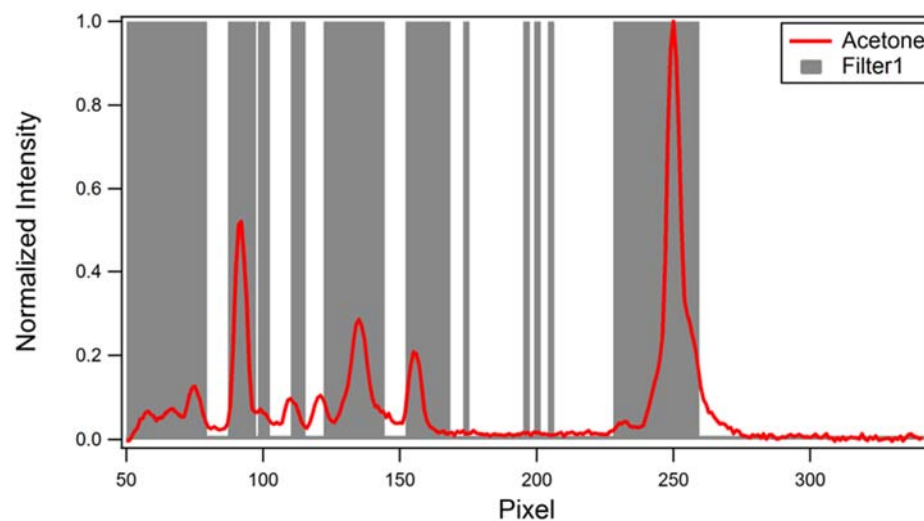


Figure 2.17: The spectrum of pure acetone(red) and the OB filters(gray) associated with it is shown. The training spectrum was acquired by recovering a 1s Hadamard spectrum

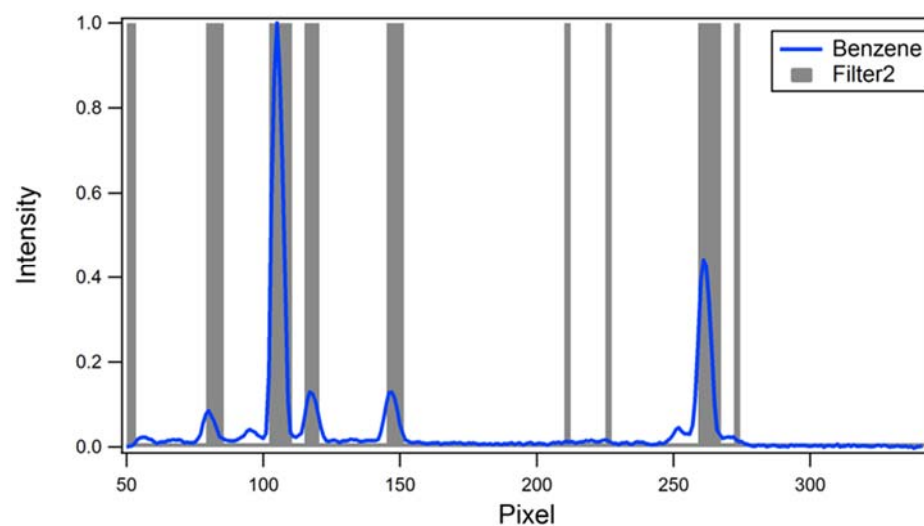


Figure 2.18: The spectrum of pure benzene(blue) and the OB filters(gray) associated with it is shown. The training spectrum was acquired by recovering a 1s Hadamard spectrum

OB filters are applied to the DMD for a given amount of time and the Raman photons associated with those pixels are binned onto a single channel PMT. The counts from the PMT are measured and the chemical rates are recovered. Figure 2.19 shows the recovered rates classification clouds for acetone and benzene. The laser power was set to about 30mW and one thousand independent measurements were made for total measurement time of 100ms, 10ms and 1ms. As the measurement time is decreased, the error associated in the classification increases as well. This is seen by the dimensions of the cloud. The mean of each of the cloud per chemical remains the same.

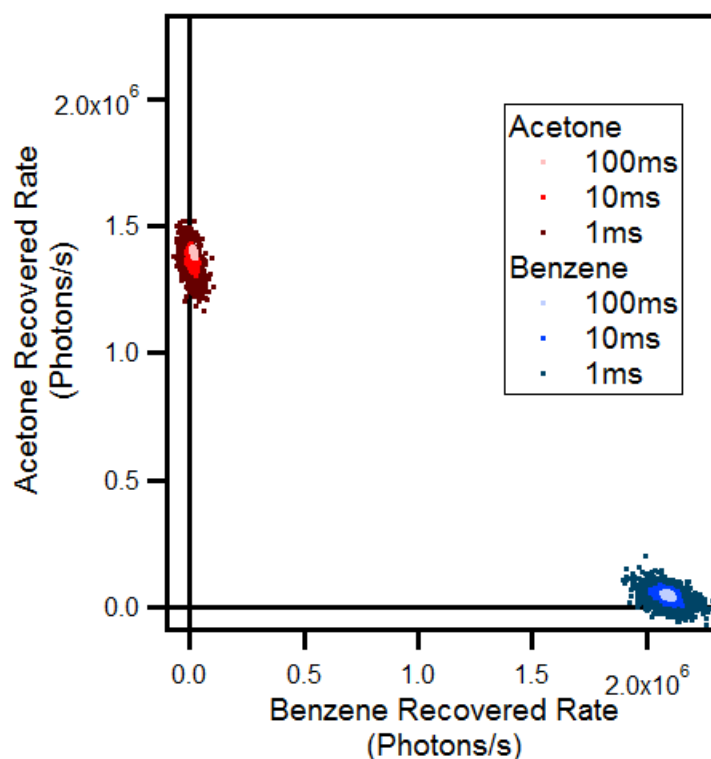


Figure 2.19: The figure shows the recovered rates for acetone and benzene using OB-CD to classify the two chemicals. Each cloud consists of 1000 independent measurement acquired at laser power of 30mW. There are three clouds for each chemical showing for each of the measurement times (100ms, 10ms and 1ms)

### 2.4.3.2 Classification of Acetone and Benzene using OB-CD2

OB2-filters were generated for pure acetone and benzene. For a pair of chemicals, one set of complementary filters are generated. Figure 2.20 shows the spectrum of acetone and benzene and the OB2 filters associated with each chemical. The training spectrum was acquired by recovering a 1s Hadamard spectrum of each of the chemicals with 30mW laser power on the sample. The filters generated are used for all OB-CD2 classification experiments of acetone and benzene at all laser powers and all time points.

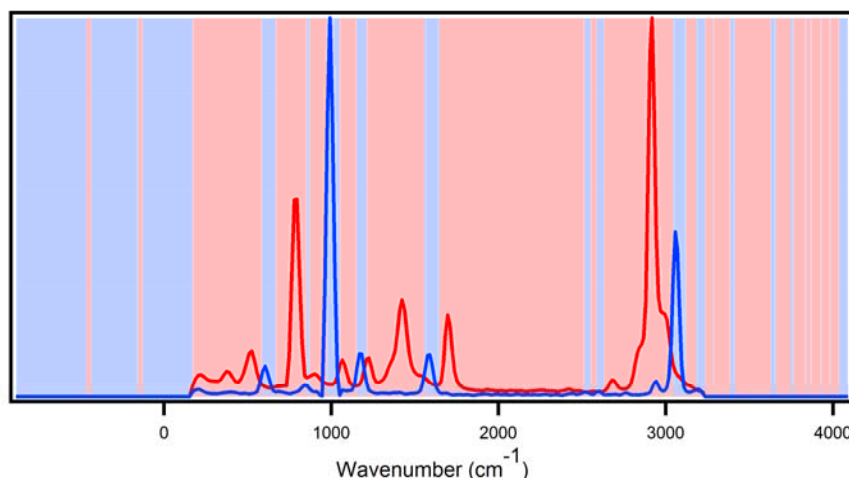


Figure 2.20: The figure shows OB-CD2 filters for shaded red and blue. The spectrum of acetone (red) and benzene (blue) are also overlaid

The DMD consists of micromirrors that can be set in one of 2 directions and applying one OB filters sends photons associated those wavelength bins towards the +12° PMT while sending all the other photons associated with the complementary filter towards the -12° PMT. This allows all the photons to be collected by applying one filter and collecting all the photons onto 2 PMT's. The time associated with each of the filter is the

same and the chemical rates are recovered from the counts measured by both the PMT's. Figure 2.21 shows the recovered rates for acetone and benzene classification using OB-CD2 complementary filtering strategy. The laser power was set to about 30mW and one thousand independent measurements were made for total measurement time of 100ms, 10ms and 1ms. As the measurement time is decreased, the error associated in the classification increases as well. This is seen by the dimensions of the cloud. The mean of each of the cloud per chemical remains the same.

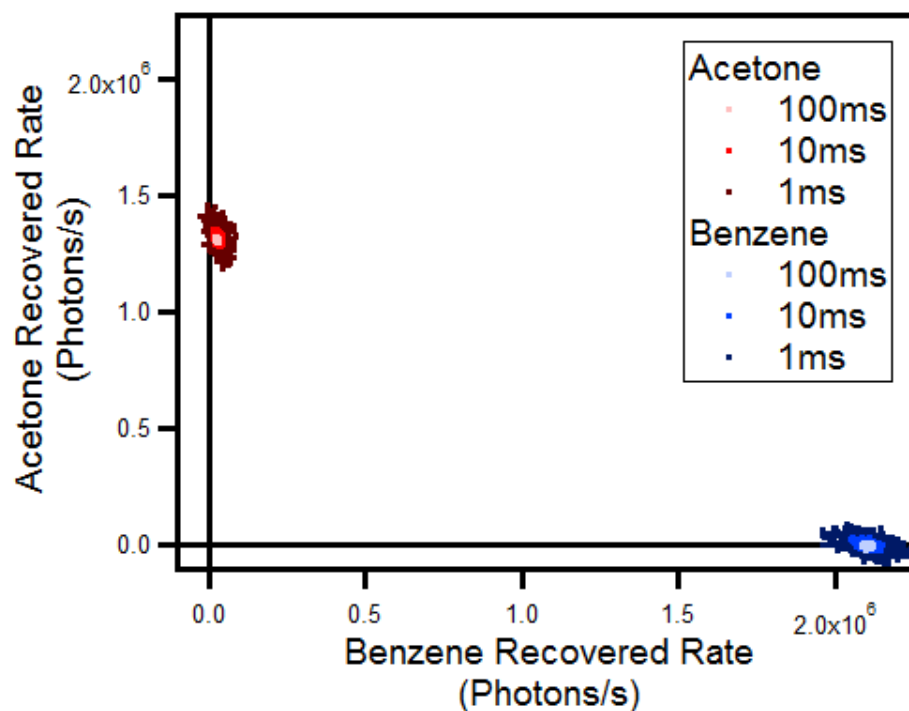


Figure 2.21: Recovered rates for acetone and benzene using OB-CD2 to classify the two chemicals. One filter (and its complement) was used to classify the two chemicals using counts measured on 2 PMT's. Each cloud consists of 1000 independent measurement acquired at laser power of 30mW. There are three clouds for each chemical showing for each of the measurement times (100ms, 10ms and 1ms)

### 2.4.3.3 Classification of Acetone and Benzene using CCD

#### 2.4.3.3.1 EM Gain 1

Analysis of the CCD data was done using least squares method as described in the methods section. One second spectra of pure acetone and pure benzene were acquired by setting the laser power to approximately 30mW. One thousand independent measurement of pure acetone and pure benzene were made at 100ms, 10ms, and 1ms exposure time. The measurements were multiplied by a factor of 10, 100 or 1000 to make them equivalent to acquiring a 1s spectrum. The spectra were also multiplied with the factor of laser power since decreasing the laser power by a factor of 10 is equivalent to acquiring spectra 10 times faster. Therefore, a factor of 1, 10,30 is multiplied to data sets collected at 30mw, 10mW, 1mW respectively The baseline was corrected by subtracting each pixel of the spectrum by its minimum value. This corrected any electronic baseline shift in each set of a thousand measurements. There is no Raman information after pixel 700 hence all pixels thereafter were cropped for all the spectra collected. The data matrix D consists of experiments arranged in rows and pixels (wavelength channels) arranged in columns. The S matrix consists of normalized spectra of the pure components and also a constant value to remove any constant baseline arranged in columns. Figure 2.22 shows the S matrix.

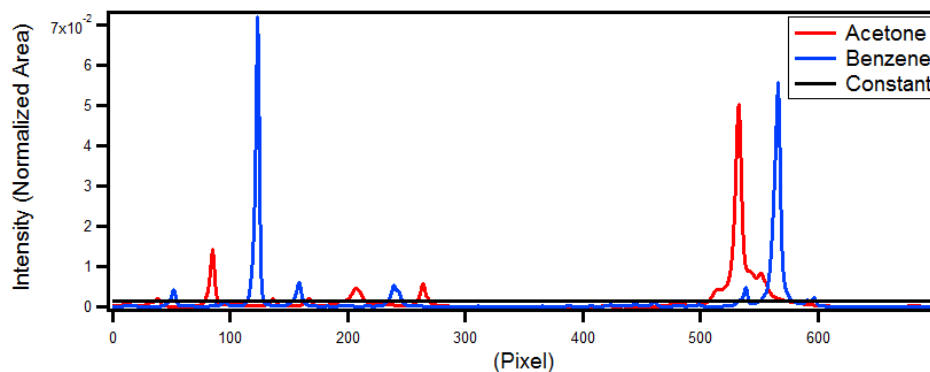


Figure 2.22: S Matrix consists of area normalized spectrum of acetone(red), benzene(blue), and constant(black) is used as the input spectra for least squares classification of acetone and benzene

The output of least squares classification of acetone and benzene is a  $1000 \times 3$  matrix C containing the weight of acetone, benzene and constant. The classification clouds of acetone and benzene at 30mW laser power collected at 100ms, 10ms and 1ms is shown. In Figure 2.23. The constant does not have any significant value in this experiment and is ignored. Each cloud corresponds to 1000 independent experiments. The lighter clouds signify faster exposure time.

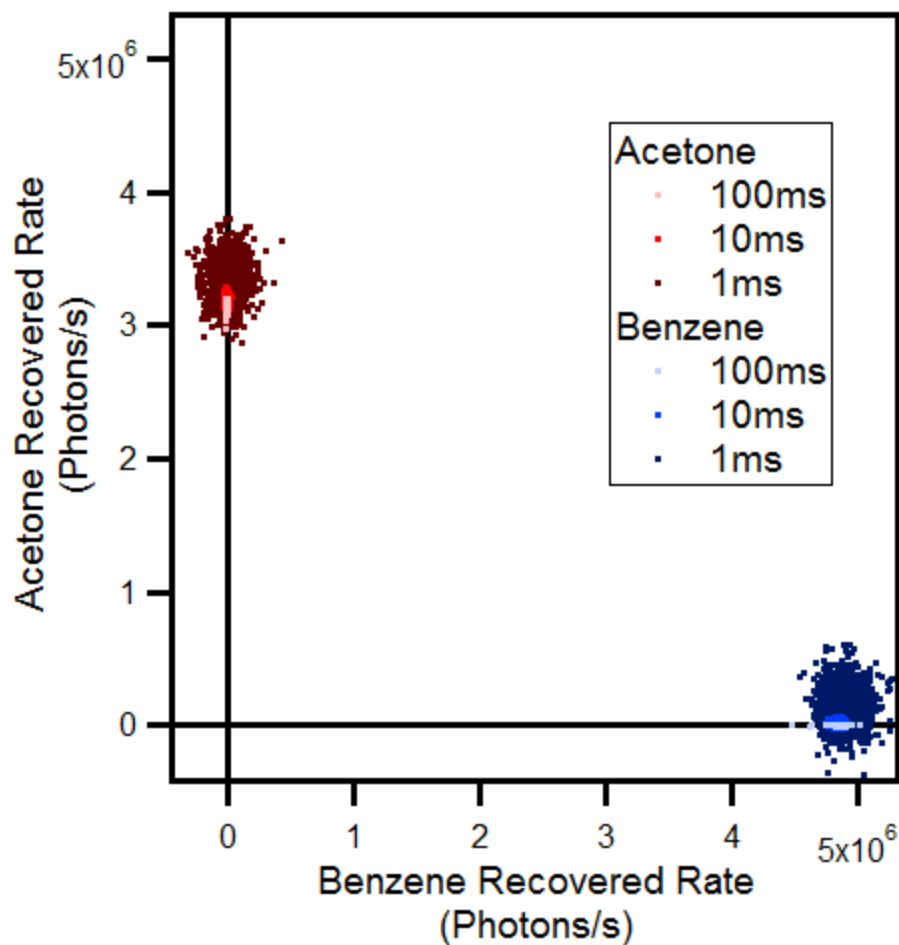


Figure 2.23: Classification clouds of acetone (shades of red) and benzene (shades of blue) are shown. Each cloud corresponds to 1000 independent measurements and the lighter shade of the color signifies faster exposure time

#### 2.4.3.3.2 EM Gain 40

Analysis of the EM- CCD data was done using least squares method as described in the methods section. EM Gain of 40 was applied for all spectra collected for exposure times of 10ms or faster for laser power 3mW and exposure time of 100ms and lower for laser power of 1mW at the sample. There is no benefit of applying EM gain to signals that have lower than 300 photon counts per bin. Figures 2.24 and 2.25 compares the spectrum

of benzene collected at exposure time 1ms at laser power of 1mW using CCD, EM-CCD and Hadamard scanning.

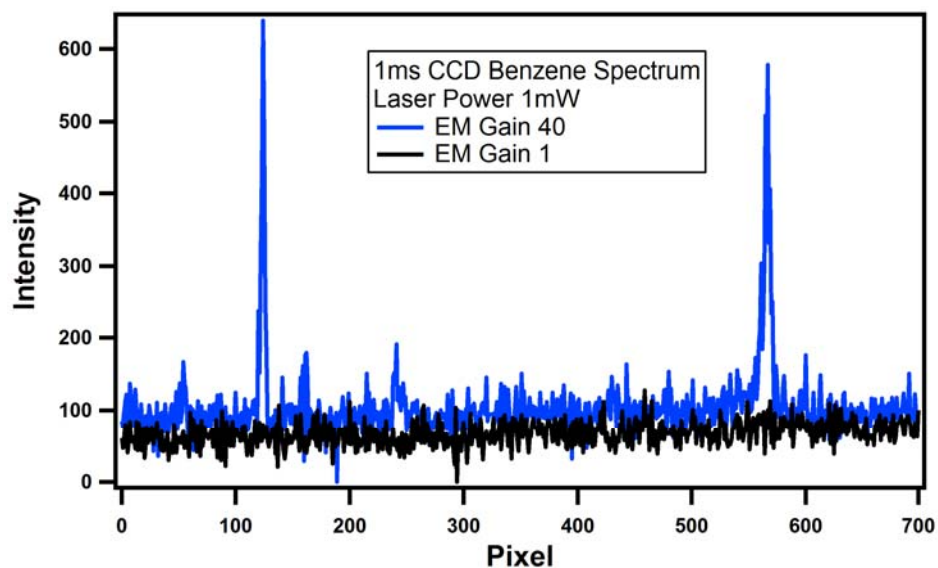


Figure 2.24: Spectrum of benzene measured using a CCD with (blue) and without EM gain (black) applied at laser power of 1mW and an exposure time of 1ms. The photons that get buried in the noise floor due to the amplitude of the read noise can be amplified using the EM gain



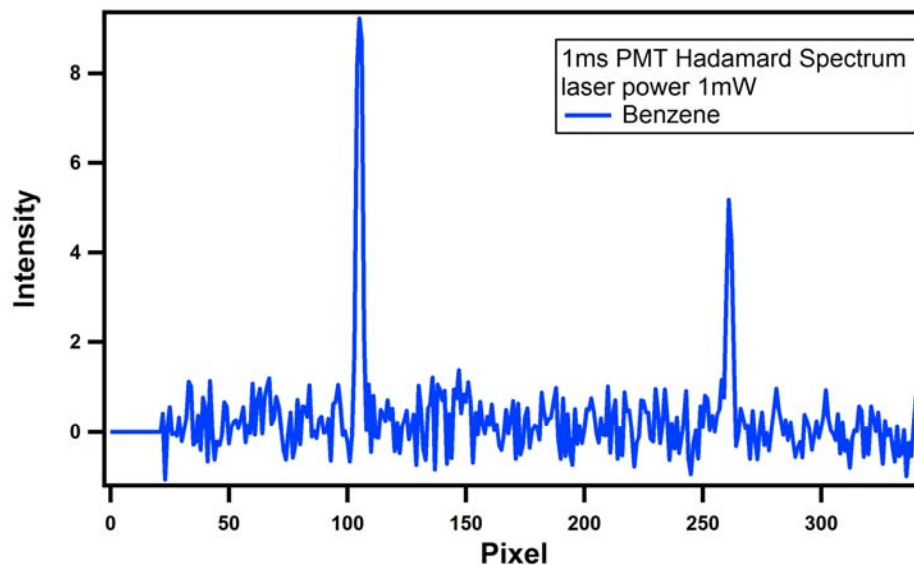


Figure 2.25: Spectrum of benzene recovered from 1ms Hadamard spectrum collected on a PMT

#### 2.4.3.4 % Relative Standard Deviation (%RSD) in Classification Using OB-CD, OB-CD2 and CCD with and without EM-Gain

The above classification clouds for acetone and benzene generated with either OB-CD strategy or from CCD data using least squares used 30mW of laser power at the sample. Since the CCD is unable to make precise measurements faster than 1ms, faster measurement times were mimicked by decreasing the laser power. The laser was lowered by approximately a factor of 10 (3mW) and 30 (1mW) and acetone and benzene spectra was collected at the same time points (data not shown). Exposure times of 10ms at laser power of 30mW is similar to exposure time of 100ms at laser power of 3mW. The % relative standard deviation for each cloud point was calculated. The %RSD is the measure of the distribution of data about the mean. Equation 2.8 shows the calculation for % relative

standard deviation. Figure 2.26 shows the %RSD of acetone for OB-CD classification and CCD classification plotted against log total counts.

$$\%RSD = \frac{\sigma * 100}{\mu} \quad (2.8)$$

The figure also shows the spectrum of acetone collected using a CCD at different signal levels. Classification of acetone for regions of high signal (>1000 total photons) full CCD measurements marginally outperforms OC-CD. However, for signals below 1000 total measured photons, OBCD has less classification error compared to full CCD measurement. At low light levels, the read noise of the CCD dominates the signal and the error in classification using least squared becomes much higher than binning high information parts of the spectrum onto a single channel detector that has comparatively low noise. This is also seen in Figure 2.27 in the classification of benzene using OB-CD and full spectral CCD measurements.

Classification ellipses of acetone and benzene at 1ms and 1mW laser power is shown in figure 2.28. On the left is the classification of acetone and benzene using least squares to analyze full spectral measurement. On the right is the OB-CD classification. The total counts of acetone and benzene are 61 and 90 respectively. It can be seen that there is a significant overlap in the ellipses of the CCD classification, however, OB-CD is able to classify acetone and benzene fairly well.

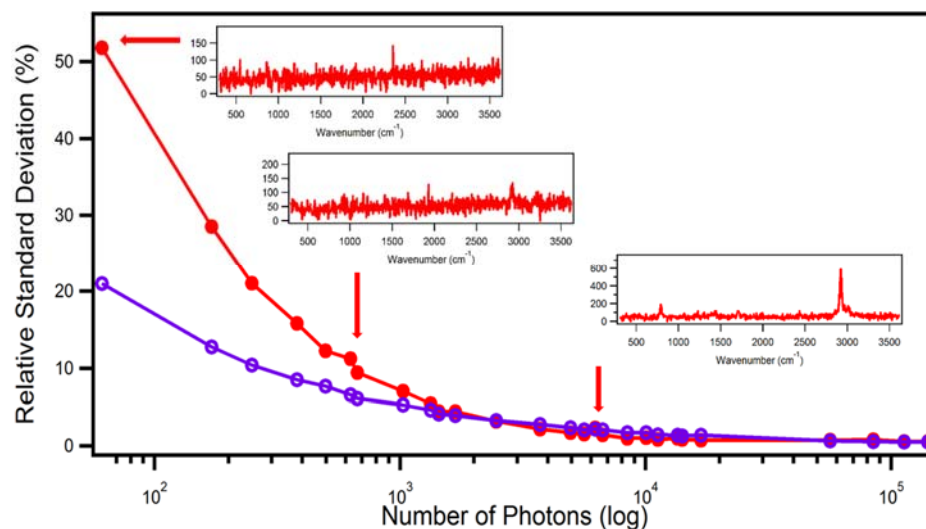


Figure 2.26: Comparison of %RSD in classification of acetone using least squares on full spectral measurements of a CCD (red) and compressive detection using optimal binary filtering of chemical information (purple) as a factor of total photon counts photon counts is shown. Also shown is the spectrum of acetone measured on a CCD at 61, 610 and 6000 total measured all on counts

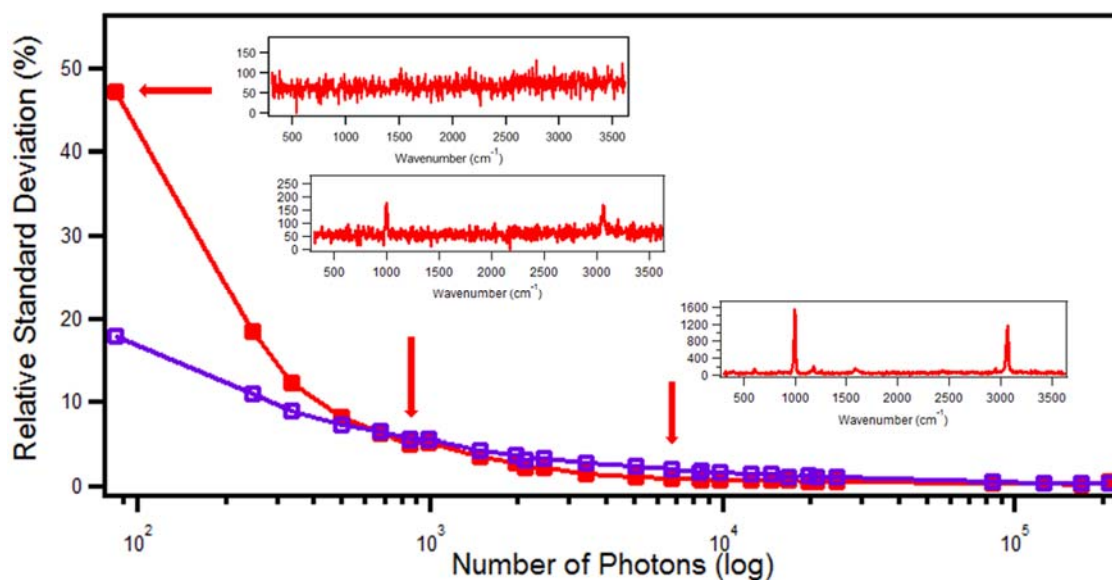


Figure 2.27: Comparison of %RSD in classification of benzene using least squares on full spectral measurements of a CCD (red) and compressive detection using optimal binary filtering of chemical information (purple) as a factor of total photon counts photon counts is shown. Also shown is the spectrum of acetone measured on a CCD at 90, 900 and 9000 total measured all on counts

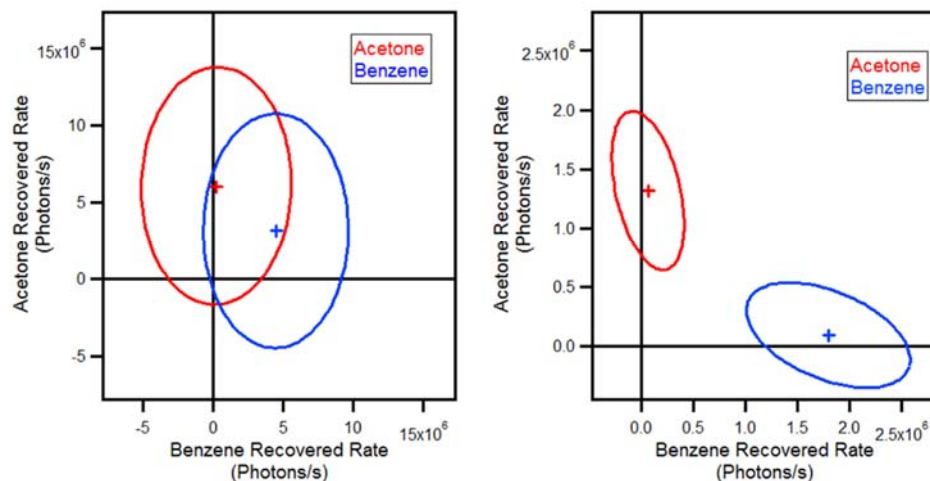


Figure 2.28: Classification ellipses of acetone and benzene at 1ms and laser power of 1mW is shown. On the left is the classification of acetone and benzene using least squares to analyze full spectral measurement. On the right is the OB-CD classification. The total counts of acetone and benzene are 61 and 90 respectively.

Figure 2.29 and 2.30 show the summary of %RSD in the classification of acetone and benzene using OB-CD, OB-CD2, and least squares analyzed CCD and EM-CCD data. EM-CCD shows the lowest %RSD in classification in low count regime. This is attributed to the fact that the CCD detects 2 times more photons than the CD spectrometer seen in the classification clouds in figure 2.19 and 2.23. The EM gain multiplies the photo electrons without multiplying the noise before the data is read off the CCD.

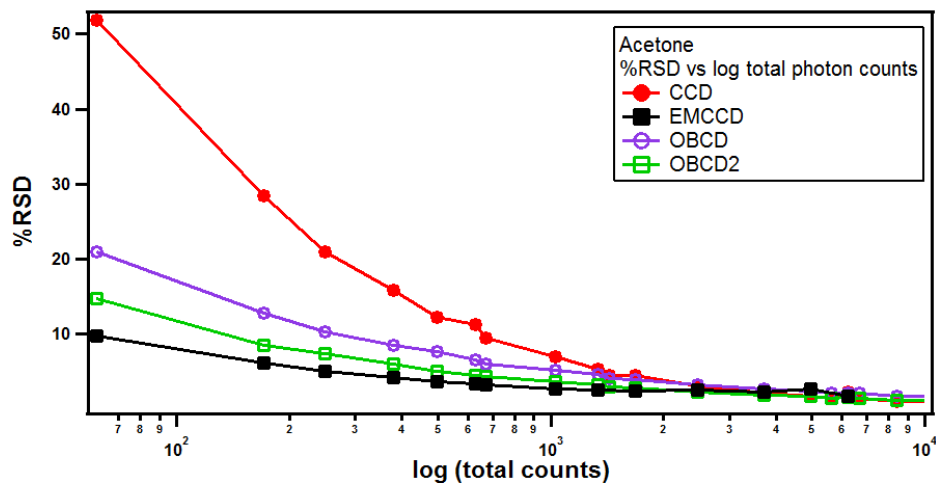


Figure 2.29: Summary of %Relative standard deviation (RSD) in the classification of acetone using OB-CD, OB-CD2, and least squares for CCD and EMCCD data.

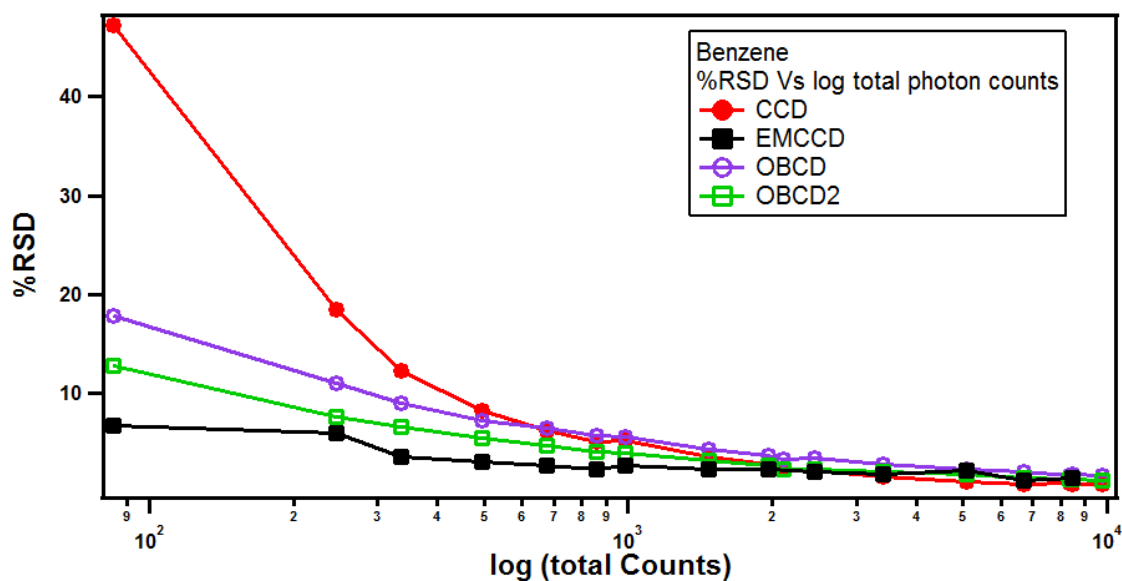


Figure 2.30: Summary of %RSD in the classification of benzene using OB-CD, OB-CD2, and least squares for CCD and EMCCD data.

#### 2.4.4 Hexane and Methylcyclohexane Classification

Hexane and methylcyclohexane were classified using OB-CD, CCD, OB-CD2 and EM-CCD with a gain of 1 and gain 40 as described in the previous section. The correlation coefficient (normalized dot product) of the two is 0.94 and is a more challenging classification problem than acetone and benzene. The OBCD and OB-CD2 filters are shown in figure 2.31a and b respectively. along with the spectra of hexane and methylcyclohexane.

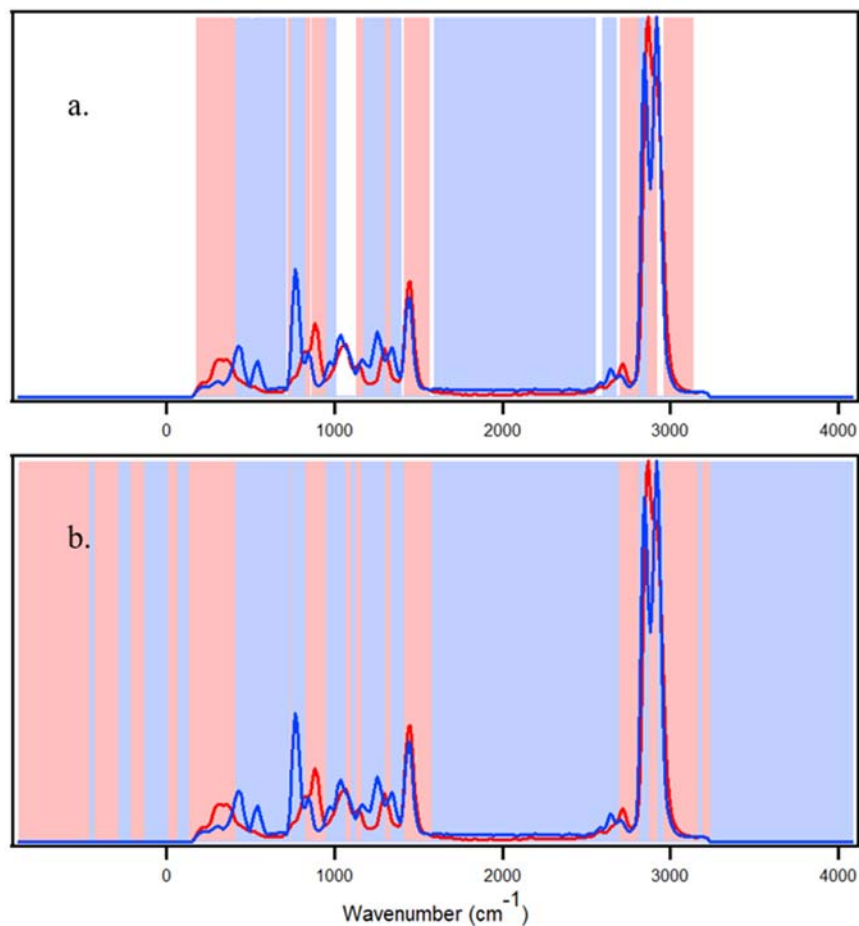


Figure 2.31: Spectra of hexane (red) and methylcyclohexane (blue) are shown. a) OB-CD filters associated with hexane and methylcyclohexane are shown in shaded red and blue. b) OB-CD2 filters shown in shaded red and blue are shown. OB-CD filter measure the counts under the spectrum and are sequentially applied and the counts are measured on one of the two PMT's. However, OB-CD2 filters are complements of each other and the counts through each filter is measured simultaneously on two PMT's in the  $+12^\circ$  and  $-12^\circ$  direction.

Figure 2.32 summarizes the %RSD in the classification clouds. The comparison of OB-CD and CCD analyzed by least squares show that OB-CD outperforms CCD classification under signals of 1000 photon counts. OB-CD 2 performs better than OB-CD filters and EM-CCD with gain 40 outperforms all others in the low noise regime.

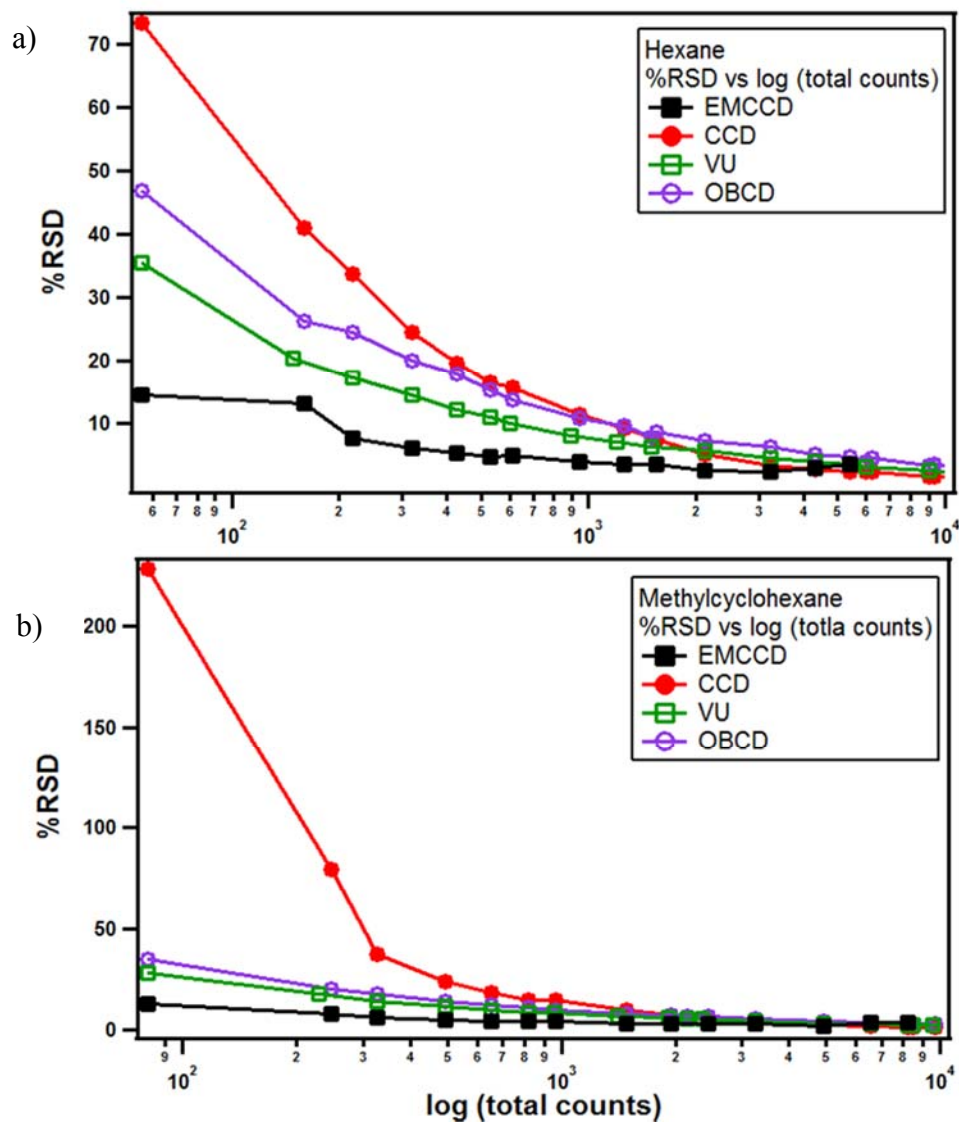


Figure 2.32: Summary of %RSD using OB-CD, OB-CD2, and least squares for CCD and EMCCD data. a) Comparison of % RSD in the classification of hexane, and b) Comparison of % RSD in the classification of methylcyclohexane.



## 2.5 Conclusion

Classification of two sets of chemicals was performed using OB-CD and least squares analyzed full spectral measurements on a CCD. Here I highlight the advantages and disadvantages of the techniques.

The fastest reliable readout on the CCD was made using the fast spectral acquisition at 1ms per experiment. This is significantly slower than measurements made on a PMT using the hyperspectral CD spectrometer. The CD spectrometer is limited by the Data acquisition card and can acquire CD measurements at  $3\mu\text{s}$  per experiment. In order to perform the experiment above, the laser power was turned down 30 fold to measure high and low range of the signal.

There was less classification error associated with high signals. OB-CD, OB-CD2 and CCD perform comparable to each other. Since it is possible to collect full spectrum in this regime, the CCD has an advantage of collecting all the data and not losing any information. In this regime, shot noise is the dominant source of noise. There is no added advantage of using the EM gain at the high signal regime. As a matter of fact, low noise mode (readout speed of 100KHz or 1MHz) gives the best signal to noise when the acquisition times are longer than 50ms.

In low signal regime (mimicking fast acquisition times) where total photon counts is comparable to the number of wavelength channels on the CCD, OB-CD outperforms CCD classification with lower %RSD in the classification clouds. The new OB-CD2 filters send all the Raman photons onto 2 PMT's performs slightly better than OB-CD filters. The use of EM gain improves the signal to noise of full spectral CCD measurements and least squares classification shows the least %RSD. However, the speed of spectral acquisition is again limited to 1ms per experiment.

## CHAPTER 3 RAMAN MICROSCOPY OF SINGLE CRYSTALS OF HAMMARHEAD RIBOZYME SUSPENDED IN A HANGING DROP

### 3.1 Abstract

In this chapter, I use Raman spectroscopy to analyze biological molecules. More specifically Raman spectroscopy is used to probe vibrational changes in nucleotide solutions and hanging droplets containing RNA crystals at different pH environments. Self-modeling curve resolution (SMCR) applied to Raman spectroscopy is used to extract solute correlated (SC) Raman spectra. The goal of this chapter is to show that Raman spectroscopy can be used to study biological molecules in aqueous environment, with minimal sample preparation and without the need of labels.

### 3.2 Introduction

Raman microscopy can be used to probe molecular interactions occurring in biomolecules without the need for labels<sup>31-33</sup>. Raman spectroscopy is a spectroscopic method based on the inelastic scattering of light due to molecular vibrational modes. Unlike IR spectroscopy, Raman spectroscopy is compatible with samples that are in aqueous environment and this makes Raman spectroscopy a useful tool to study biomolecules in its native environment<sup>34,35</sup>. In this chapter we will use Raman microscopy to study changes in

molecular vibrations signatures of a single crystal of hammerhead ribozyme (HHRz) suspended in a hanging drop with respect to pH. A set of control experiments were performed to understand the changes of vibrational fingerprint on single nucleotide level with respect to pH.

Hammerhead ribozymes (HHRz) are small (50-150 base) catalytic self-cleaving RNAs that cleave phosphodiester bond. They are found throughout all kingdoms of life but the physiological function of the ribozyme is still unclear. RNAs are comprised of four nucleotide bases and by forming secondary structures (folding in a specific orientation), are able to catalyze biochemical reactions near physiological pHs even though individual bases have pKa's far from physiological pH. The secondary structure of HHRz is shown in figure 3.1. The cleavage reaction of the ribozyme depicts the involvement of base guanine at position 8 and 12 and base adenine at position 6 of the secondary structure. It has been shown that it is possible to measure changes in the Raman spectrum of a single nucleotide in a ribozyme<sup>36</sup>. Here we try to understand the mechanism of HHRz catalysis by identifying the base involved in catalysis using Raman spectroscopy. We study the Raman spectrum of individual base monophosphate (NMP) at different pHs and will try to correlate the base involved in catalysis of HHRz based on the results of individual NTP's

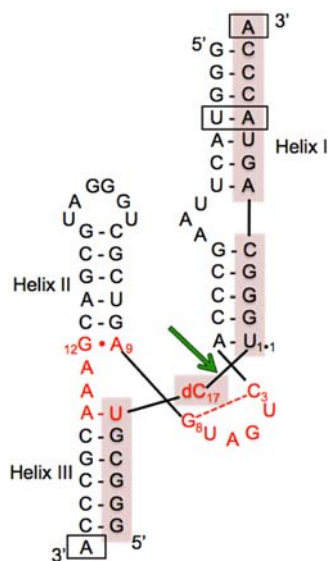


Figure 3.1: Secondary structure of Hammarhead ribozyme (HHRz)

### 3.3 Methods

#### 3.3.1 Sample Preparation

Samples were made by 50mM NMP were dissolved in Nanopure water and the pH was adjusted by adding either HCl or NaOH. HHRz crystals were prepared by graduate student Aamir Mir from the Barbara Golden lab<sup>37,38</sup>. Table 3.1 show the buffer used with HHRz crystals at different pHs. All samples were prepared in the Golden lab. The pH ranges were chosen to monitor changes on the nucleotide bases. The pKa of the primary phosphate is reported to be around 1.4 and the sugar ring is at pKa of 12<sup>39,40</sup>. The samples were transferred to the Ben-Amotz lab for Raman data acquisition and analysis. SMCR has been proved as a very useful tool to analyze solute and solvent contributions and was used to subtract contribution due to the water and buffers<sup>41,42</sup>.

Table 3.1: The table shows the pH range and the type of buffer used with all the HHRz Samples.

pH	Buffer
2.2-3.6	Glycine-HCl buffer
3.6-5.6	KAcetate
5.0-7.5	KCacodylate
7.4-8.0	Tris-HCl
8.6-10.6	Glycine-NaOH

### 3.3.2 Brief Description of the Instrument

The instrument details are described in more details elsewhere<sup>42</sup>. A brief description of the instrument is explained here. The instrument used to make the Raman measurements comprised of a 514.5nm Argon ion laser with a highly polarized beam guided through mirrors towards the sample. A 20X or 50X objective was mounted either horizontally or vertically to collect scattered photons from the sample. Raman scattered photons were separated from the Rayleigh scattered photons using a dichroic mirror and a long pass filter, and focused onto a circular fiber-optic bundle consisting of seven fibers. The other end of the bundle, the fibers are arranged linearly and is coupled to the spectrometer consisting of 300, 600, and 1200 grooves/mm gratings. The Raman photons are linearly separated and focused onto a CCD array of 400X1340 detectors. Figures 3.2 and 3.3 show the instrumental setup. In order to focus the laser onto the HHRz crystal, the microscope objective had to be setup vertically. A XYZ sample stage was also setup to help focus the

laser onto the crystal. Figure 3.4 shows the schematic of vertical setup of the objective and the hanging droplets in the sample tray. A zoomed in schematic of the crystal in the hanging droplet is also shown.

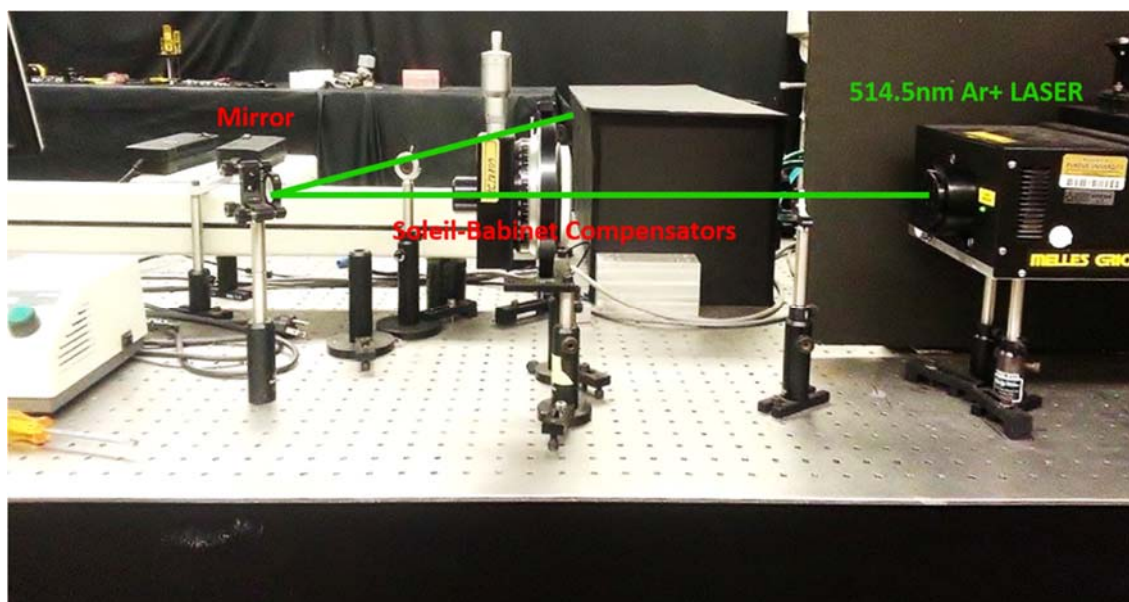


Figure 3.2: Argon ion excitation laser passes through a laser line cleanup filter and a Soleil Babinet Compensator. The compensator is an optical element that is setup to allow a certain polarization of light to pass through. The laser is guided through a mirror towards the sample stage

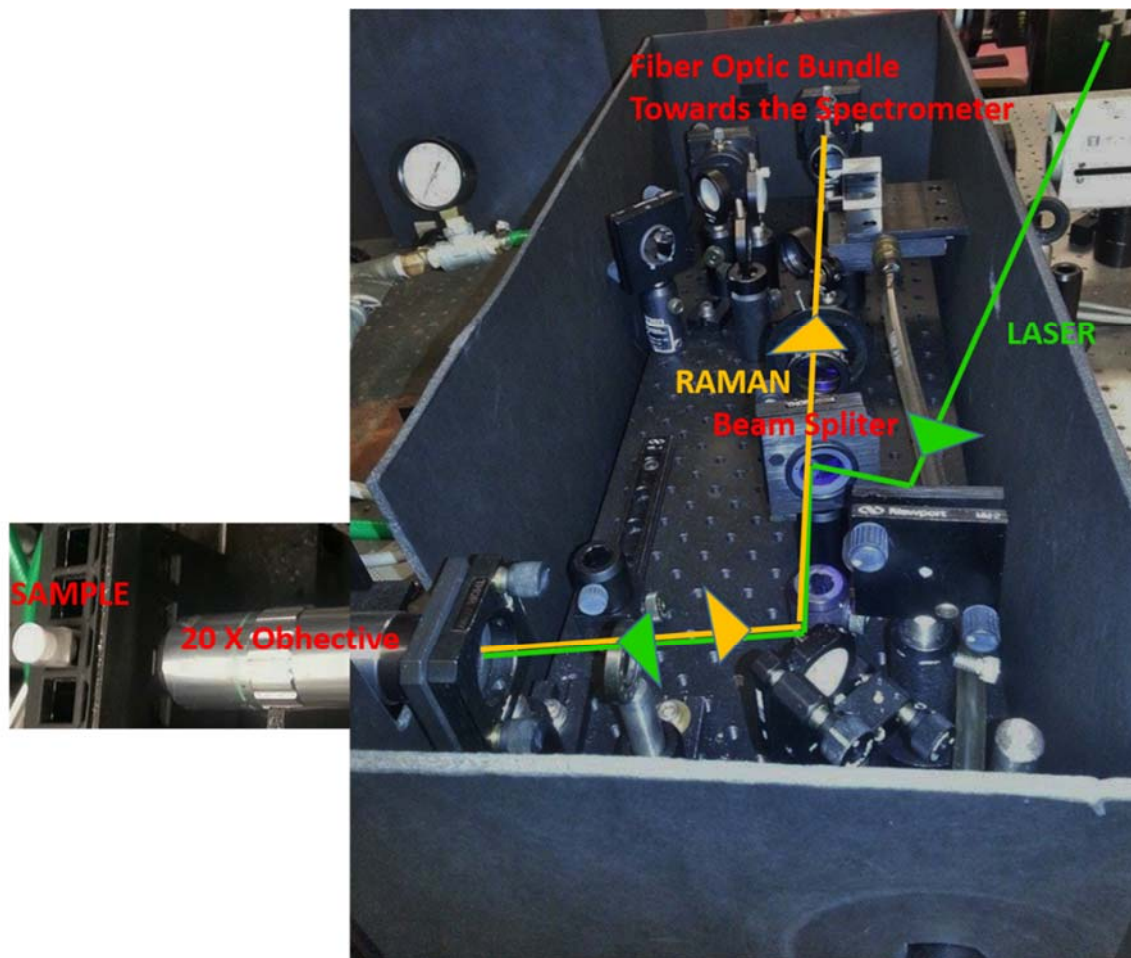


Figure 3.3: The polarized laser beam is guided through mirrors and a dichroic beam splitter towards the 20X horizontally mounted objective. The sample is in a cuvette held on a temperature controlled sample stage. The Raman scattered photons are collected back through the same object and pass through the beam splitter. Any residual Rayleigh scattered photons are rejected by a long pass filter. The Raman beam is focused on a 100um fiber bundle and guided towards the spectrometer (not shown here). The objective can be mounted vertically by removing the 20X objective and guiding the laser vertically onto a vertically mounter objective



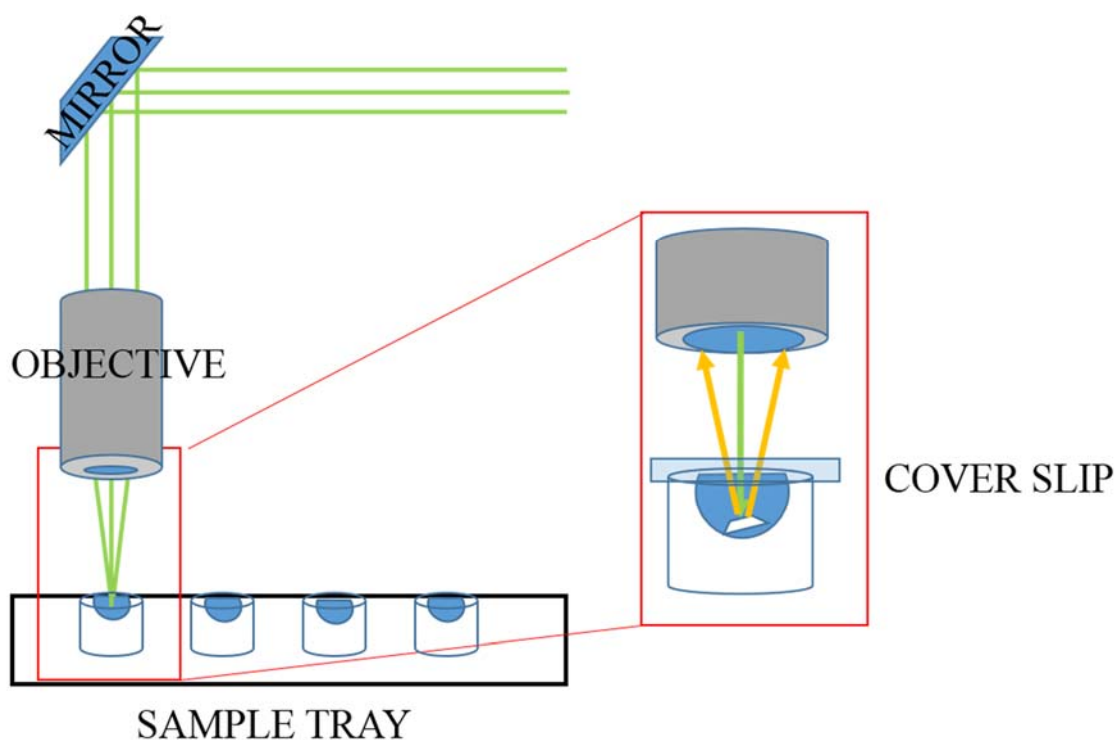


Figure 3.4: A schematic of vertically positioned 50X objective used to focus the laser onto HHRz crystal suspended in a hanging drop of the buffer. The sample tray is set on a XYZ stage in order to focus the laser on the crystal. A zoomed in schematic on the right shows the laser beam focused on the crystal and the Raman scattered photons being collected by the same objective

### 3.3.3 Data Acquisition Parameters

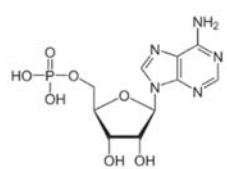
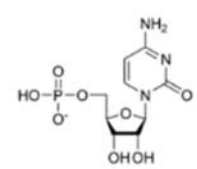
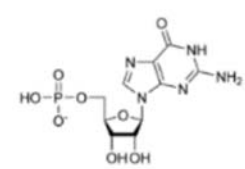
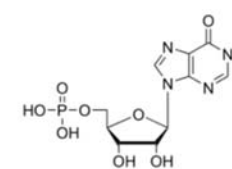
The Data was collected using PI software WinSpec. All experiments were collected in duplicate for a total time of 5 minutes by averaging 750 acquisitions at 0.4s each. The laser power at the sample was 13mW for all NTP experiments and 50mW for the crystals.

## 3.4 Results

### 3.4.1 Analysis of NMP in Water at Different pH

50mM NMP was dissolved in water and the pH adjusted to match the pH range to be analyzed. A water solution consisting of the same amount of acid or base was used as a reference. Raman spectrum of the NMP sample or the reference solution was acquired in duplicate for 5 minutes by averaging the signal of 750 acquisitions for an exposure time of 0.4s each. Table 3.2 Shows the structures of the NMP's and the associated pKa<sup>43</sup> for the bases. In this study we monitored the Raman spectrum around the N-1 pKa for all the NMP's.

Table 3.2: Structure and Pka of NMP's used in the study are shown

 <p>AMP</p>	 <p>CMP</p>	 <p>GMP</p>	 <p>IMP</p>
<p>pKa: Base-N 3.8 (N-1)</p>	<p>pKa: Base-N 4.5 (N-1)</p>	<p>pKa: Base-N 9.4 (N-1) 2.4 (N-7)</p>	<p>pKa: Base-N 9.4 (N-1) 2.4 (N-7)</p>

The spectra of CMP were collected by summing 750 spectra acquired at 0.4s each for a total time of 5 minutes at about 13mW laser power at the sample. The feature arising at 1650cm<sup>-1</sup> seen in all the spectra is the HOH bend of water. Figure 3.5 shows the raw

spectra of CMP at pH of 2, 4, 6 and 8 (solid lines) in the spectral window of approximately 500-2500  $\text{cm}^{-1}$ . The figure also shows the spectrum of water at pH 2 (dashed line).

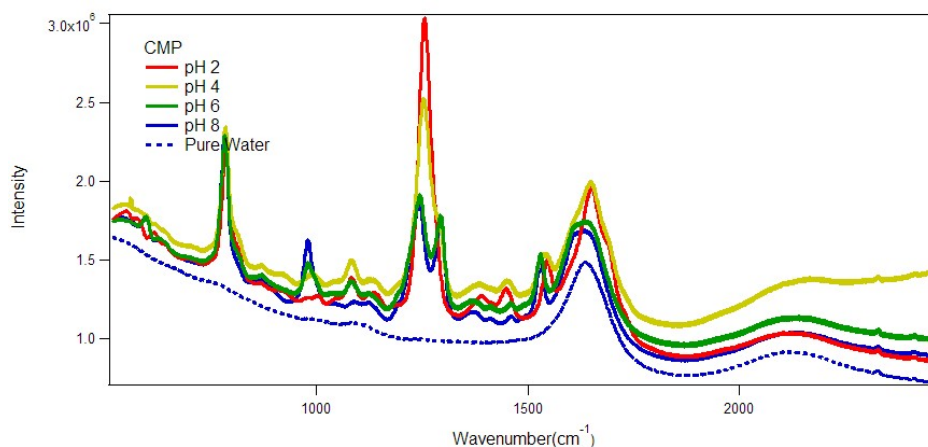


Figure 3.5: Spectra of CMP at pH 2 (solid red), 4 (solid yellow), 6 (solid green) and 8 (solid red) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500  $\text{cm}^{-1}$

A zoomed in region of CMP is shown in figure 3.6. The window covers the spectral range of 1500-1580  $\text{cm}^{-1}$ . The feature at 1547  $\text{cm}^{-1}$  seen in CMP spectra at pH 2 shifts to 1531  $\text{cm}^{-1}$  as seen in CMP spectra at pH 6 and pH 8. At pH 4 the spectrum shows a mixture of the two features. The pKa of CMP shown in Table 3.2 is 4.5. The spectra shown here is in agreement with experiments from literature<sup>36</sup>. The Raman features shown here are due to the deprotonation of the base at the N1 position. The dashed line is the spectrum of water that shows no features in this region.

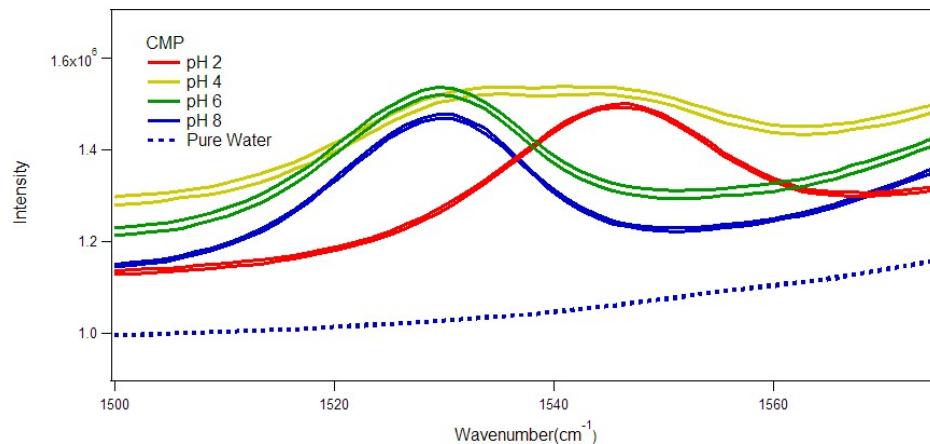


Figure 3.6: The spectral range of 1500-1575  $\text{cm}^{-1}$  is shown for spectra of CMP acquired in duplicate at different pH. The spectral feature at 1547 $\text{cm}^{-1}$  seen at pH 2 shifts to 1531 $\text{cm}^{-1}$  due to deprotonation of base at the N1 position

The spectra of AMP were collected by summing 750 spectra acquired at 0.4s each for a total time of 5 minutes at about 13mW laser power at the sample. Figure 3.7 shows the raw spectra of AMP at pH of 1.5, 3.5, and 5.5 (solid lines) in the spectral window of approximately 500-2500  $\text{cm}^{-1}$ . The figure also shows the spectrum of water at pH 2 (dashed line) and feature arising at 1650 $\text{cm}^{-1}$  is the HOH bend of water that can be seen in all raw spectra. The spectrum of sample at different pH's show different baseline an indication of slight florescence increase due to impurities in the sample.

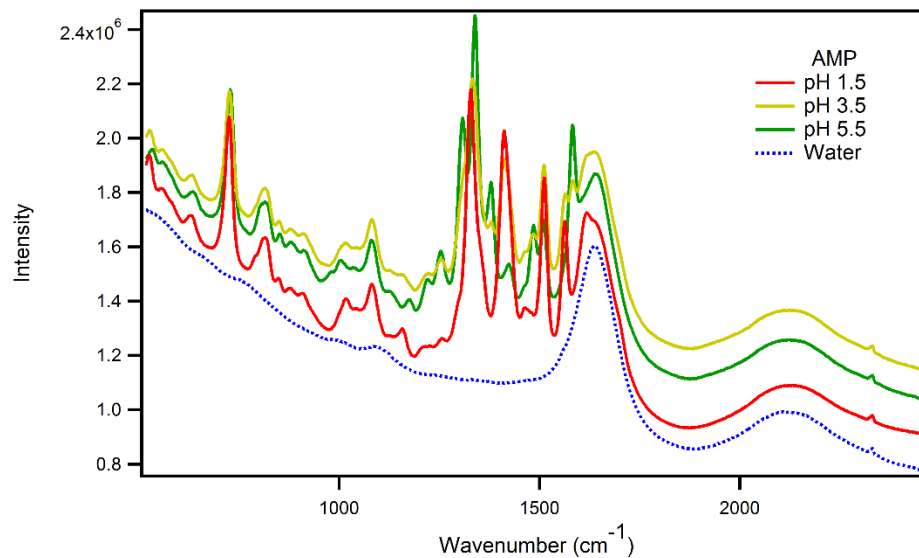


Figure 3.7: Spectra of AMP at pH 1.5 (solid red), 3.5 (solid yellow), 5.5 (solid green) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500 $\text{cm}^{-1}$

Raman MCR or SMCR was used to generate solute correlated (SC) spectra. This is similar to difference spectra in which a reference spectrum (usually a solvent) is subtracted out of the spectrum of a solution. The remaining spectrum contains information of the solute and any perturbations in the solvent caused due to the solute solvent interactions. The background of the spectrum can also be removed by subtracting a linear or non-linear polynomial function. Figure 3.8 shows the background subtracted SC spectra of ATP at different pH's.

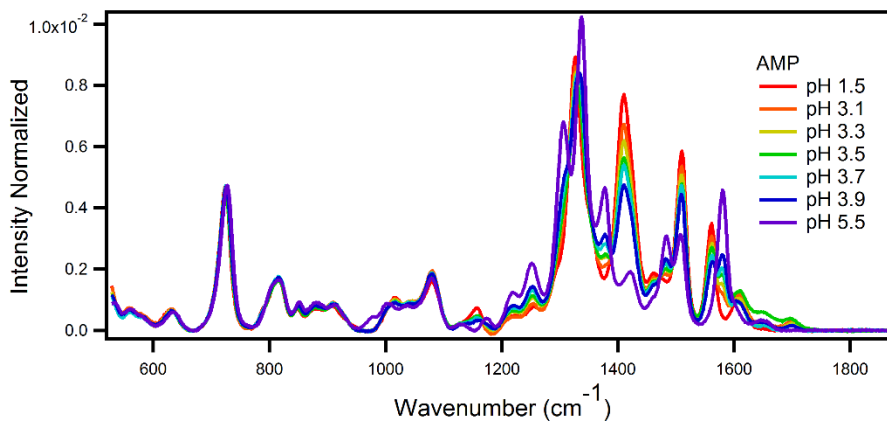


Figure 3.8: Background subtracted SC spectra of AMP at different pH shown in the spectral range 550-1850 $\text{cm}^{-1}$

A zoomed in region of AMP is shown in figure 3.9. The window covers the spectral range of 1530-1630  $\text{cm}^{-1}$ . The feature at 1561 $\text{cm}^{-1}$  seen in AMP spectra at pH 2 shifts to 1581 $\text{cm}^{-1}$  as seen in CMP spectra at pH 1.5 and pH 5.5. At pH 3.9 the spectrum shows an approximately equal mixture of the two features. The pKa of nitrogen in position 1 of the base shown in Table 3.2 is 3.8.

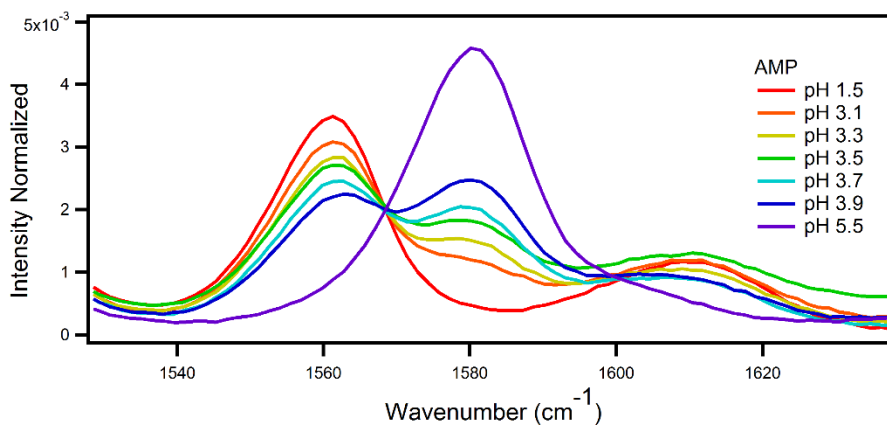


Figure 3.9: The spectral range of 1530-1630 $\text{cm}^{-1}$  is shown for spectra of AMP at different pH. The feature at 1562 $\text{cm}^{-1}$  seen in AMP spectra at pH 1.5 shifts to 1582 $\text{cm}^{-1}$  as seen in AMP spectra at pH 5.5. At pH 3.5 the spectrum shows a mixture of the two features. These feature arise from the deprotonation of the nitrogen in position 1 of the base

The ratio of peak intensities at 1561 $\text{cm}^{-1}$  and 1581 $\text{cm}^{-1}$  were normalized and plotted against pH and is shown in figure 3.10. The points were fit using a sigmoidal fit and the midpoint of the fit calculated to be 3.8. This is the value listed in table 3.2 as the pK<sub>a</sub> value for the N1 position of Adenine base.

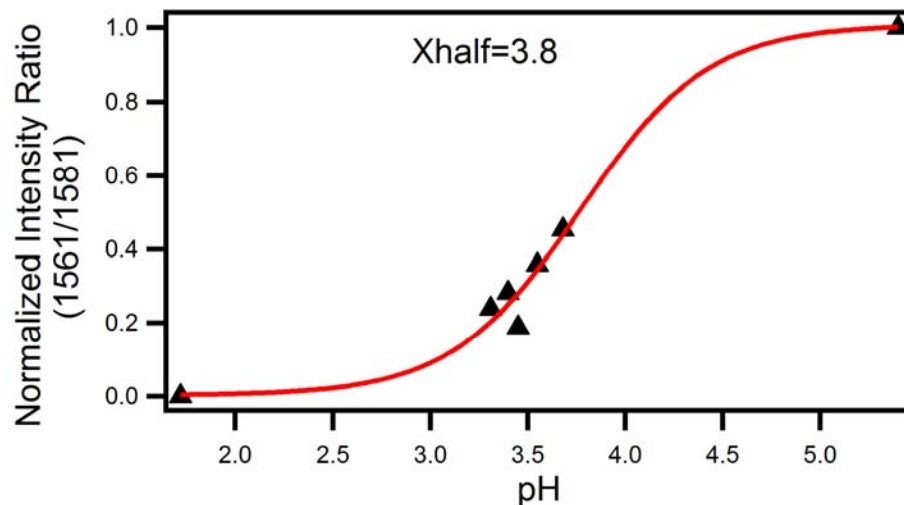


Figure 3.10: The pH profile of adenine shown by plotting the normalized intensity ratio of features at  $1561\text{cm}^{-1}$  and  $1581\text{cm}^{-1}$  as a function of pH (black triangles) and fit to a sigmoidal curve (solid red). The midpoint of the curve has a pH value of 3.8 which is the literature value for the pKa of nitrogen in position 1 of adenine

The spectra of GMP were collected by summing 750 spectra acquired at 0.4s each for a total time of 5 minutes at about 13mW laser power at the sample. Figure 3.11 shows the raw spectra of GMP at pH of 7.2, 9.2, and 11.2 (solid lines) offset vertically and a spectral window of approximately  $500\text{-}2500\text{ cm}^{-1}$ . The figure also identifies wavenumbers for peaks that show significant changes due to pH.



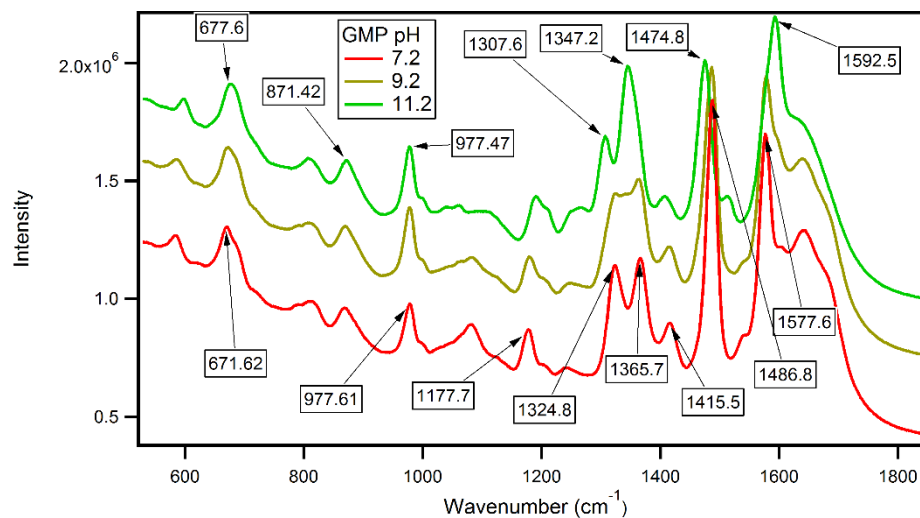


Figure 3.11: Spectra of GMP at pH 7.2 (solid red), 9.2 (solid yellow), 11.2 (solid green) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500  $\text{cm}^{-1}$ . Peaks that show shifts due to pH are identified by their wavenumbers.

SMCR was used to generate solute correlated (SC) spectra of GMP. The spectra in figure 3.12 contains information of the solute and any perturbations in the solvent caused due to the solute solvent interactions. The background of the spectrum can also be removed by subtracting a linear or non-linear polynomial function.

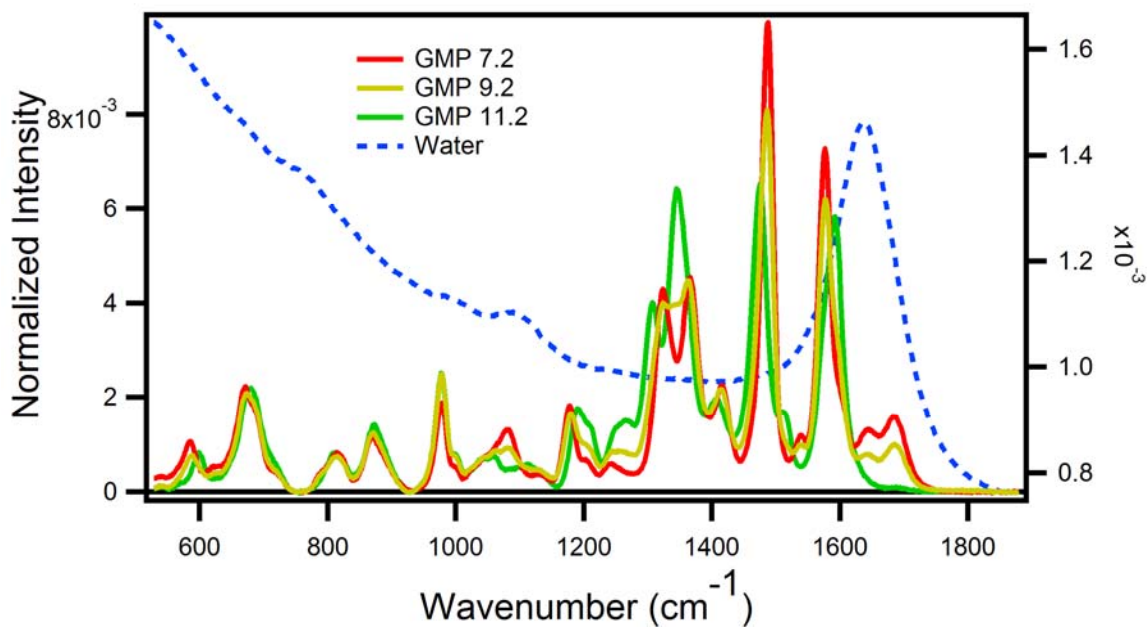


Figure 3.12: Background subtracted SC spectra of GMP at different pH shown in the spectral range 550-1850 $\text{cm}^{-1}$ . Spectrum of water is shown in the dashed blue line.

A zoomed in region of GMP is shown in figure 3.13. The window covers the spectral range of 1540-1650  $\text{cm}^{-1}$ . There is a led drastic shift in this region of GMP compared to the last two CMP and AMP. This could be due to the reason that base guanine self assembles to form tetramers known as G-quartet.

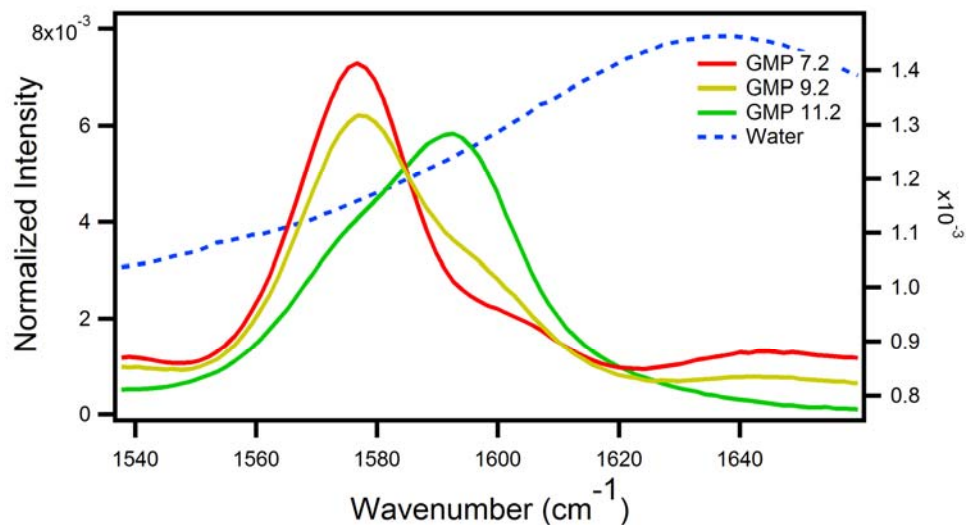


Figure 3.13: The spectral range of 1540-1600 $\text{cm}^{-1}$  is shown for spectra of GMP at different pH. The feature at 1575 $\text{cm}^{-1}$  seen in GMP spectra at low pH shift to 1595 $\text{cm}^{-1}$  at higher pH

Inosine is a base and a sugar that is deaminated version of adenosine and has similar pKa to guanine. However, IMP does not form tetramers and can be used instead of GMP to study the changes of Raman spectra at pH. The spectra of IMP were collected by summing 750 spectra acquired at 0.4s each for a total time of 5 minutes at about 13mW laser power at the sample. Figure 3.14 shows the raw spectra of IMP at pH of 6.7, 8.7, and 10.7 (solid lines) offset vertically and a spectral window of approximately 500-2500  $\text{cm}^{-1}$ . The figure also identifies wavenumbers for peaks that show significant changes due to pH.

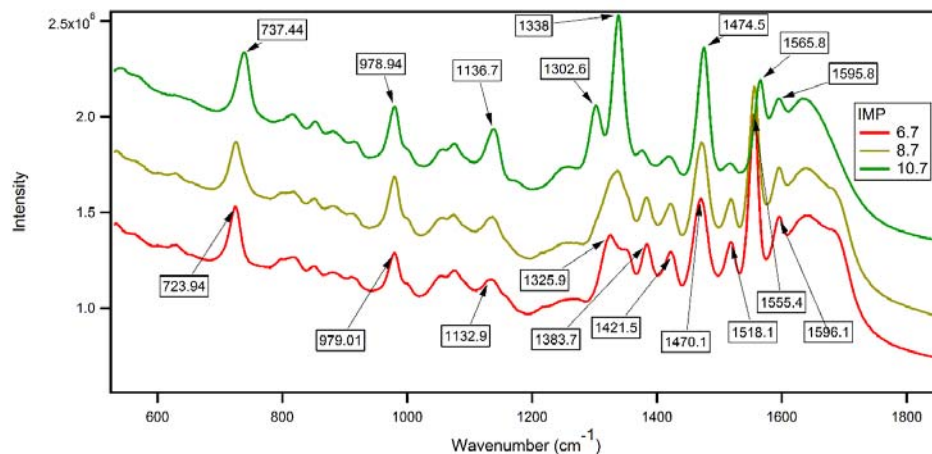


Figure 3.14: Spectra of IMP at pH 6.7 (solid red), 8.7 (solid yellow), 10.7 (solid green) and the spectrum of water (dashed blue) is shown from wavenumber range of 500 – 2500  $\text{cm}^{-1}$ . Peaks that show shifts due to pH are identified by their wavenumbers.

SMCR was used to generate solute correlated (SC) spectra of IMP. The spectra in figure 3.15 contains information of the solute and any perturbations in the solvent caused due to the solute solvent interactions. The background of the spectrum can also be removed by subtracting a linear or non-linear polynomial function.

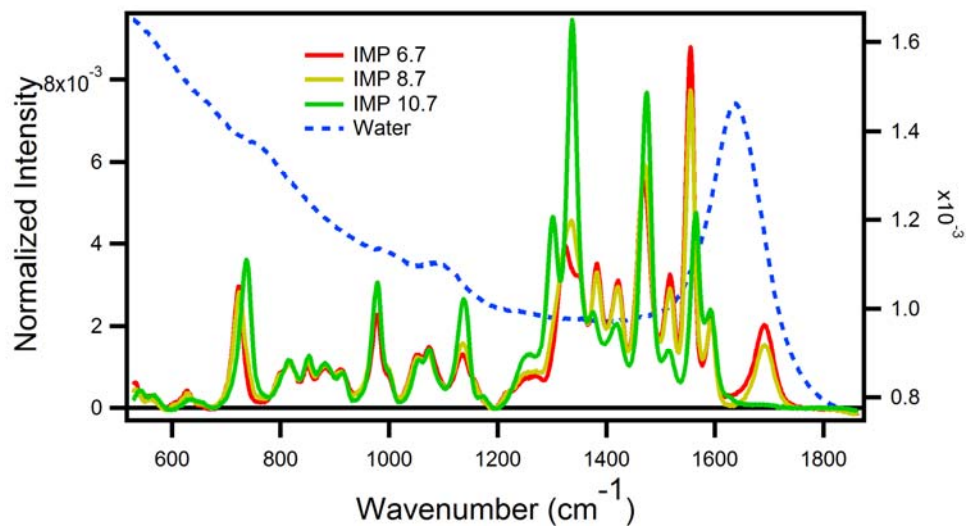


Figure 3.15: Background subtracted SC spectra of IMP at different pH shown in the spectral range 550-1850 $\text{cm}^{-1}$ . Spectrum of water is shown in the dashed blue line.

A zoomed in region of IMP is shown in figure 3.16. The window covers the spectral range of 1535-1635  $\text{cm}^{-1}$ . There is a significant shift seen in the spectrum, however, to be able to analyze the data for pKa, narrower pH range and more data points should be acquired.

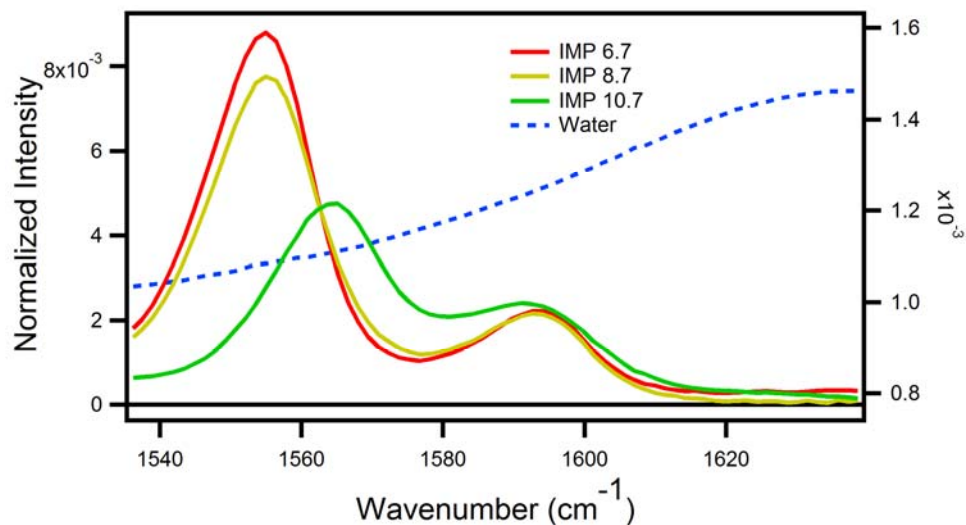


Figure 3.16: The spectral range of 1535-1630cm<sup>-1</sup> is shown for spectra of IMP at different pH. The feature at 1550cm<sup>-1</sup> seen in IMP spectra at lower pH shifts to 1565cm<sup>-1</sup> at higher pH.

#### 3.4.2 Analysis of Lysozyme Crystal

Lysozyme crystal was suspended in a hanging droplet on a hydrophobic coverslip. The 20X microscope objective was set up vertically in order to collect the Raman spectrum of lysozyme crystal. Figure 3.17 shows the raw spectra of obtained from a crystal of lysozyme in a hanging droplet suspended in a buffer (red) and the buffer (blue). The spectrum of buffer is used as the reference solution that is subtracted out using SMCR. SC spectrum arising predominantly from lysozyme crystal with very little buffer contribution is shown in figure 3.18.

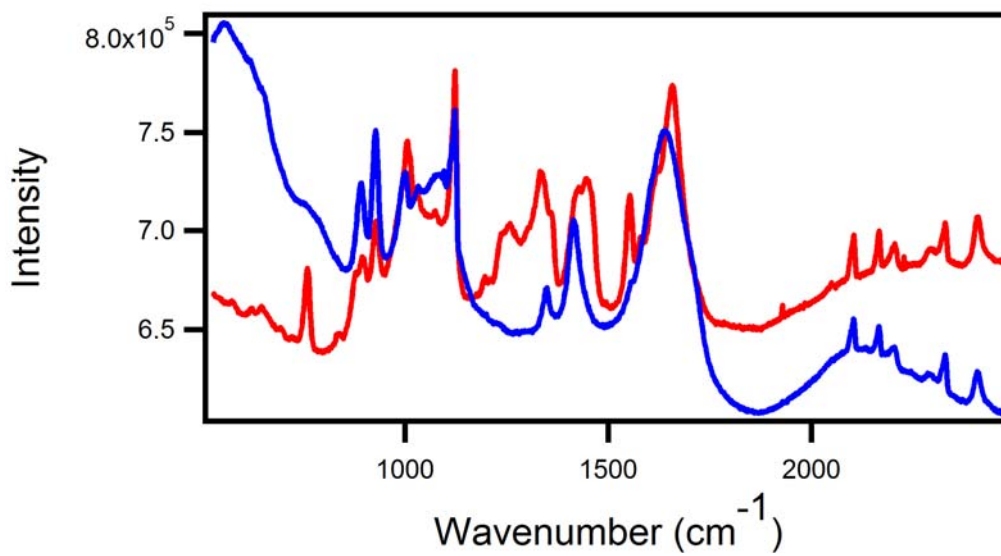


Figure 3.17: The Spectrum of Lysozyme crystal suspended in a hanging droplet.

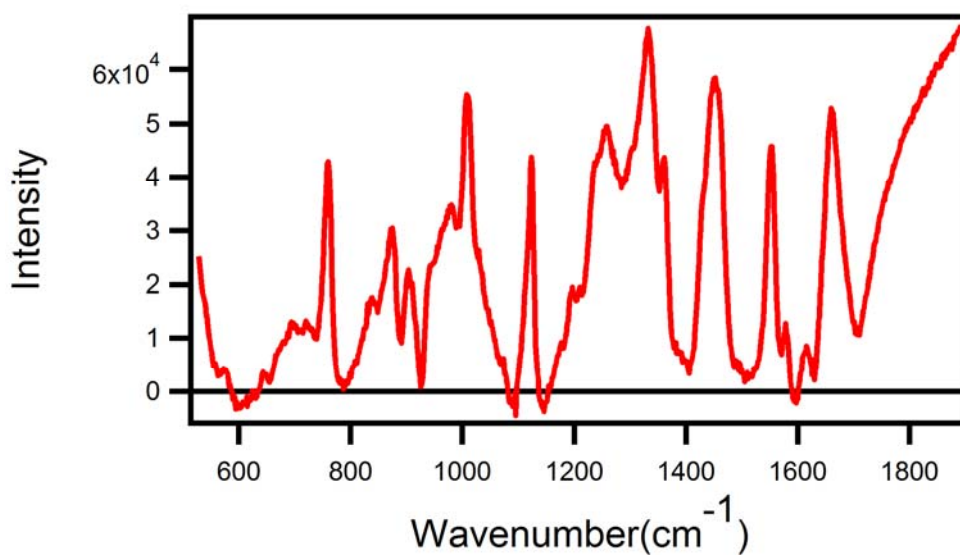


Figure 3.18: Background subtracted SC spectrum of Lysozyme crystal suspended in a hanging droplet.

### 3.4.3 Analysis of HHRz Crystals

HHRz crystals were grown in the Barbara Golden lab in biochemistry and transferred to the Ben-Amotz lab on the day of the experiment. Figure 3.19 shows images of HHRz crystal that were incubated with buffers of different pHs. The length of the crystals ranges from 90-150 $\mu\text{m}$ . These crystals were contained on a hydrophobic glass coverslip inside a drop of hanging buffer at a constant pH. The coverslip was sealed on one well of a 16 well cell culture plate using high vacuum grease. A reservoir of buffer was added to the well of cell culture plate to prevent buffer in the hanging drop from evaporating.

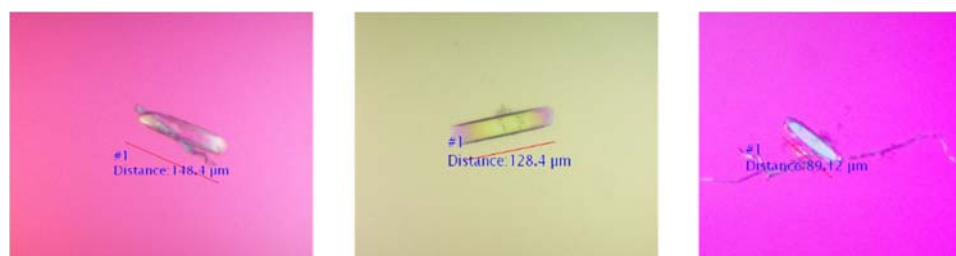


Figure 3.19: White light microscope images of HHRz crystals suspended in a hanging droplet at pH 3.8, 7.4 and 10.

Raman Spectra of HHRz and the surrounding buffer was acquired using 55mW laser power at the sample through a 50X objective. The set of 2 spectra were acquired for 5 minutes each by summing 750 acquisition at 0.4s exposure. Figure 3.2 shows the spectra of HHRz and surrounding buffer at pH of 3.8, 7.4 and 10.0. different buffers were used to match and maintain the pH range chosen. Focusing the laser on the HHRz crystal at high



pH caused the crystal to move out of the laser focal spot making good spectral acquisition difficult.

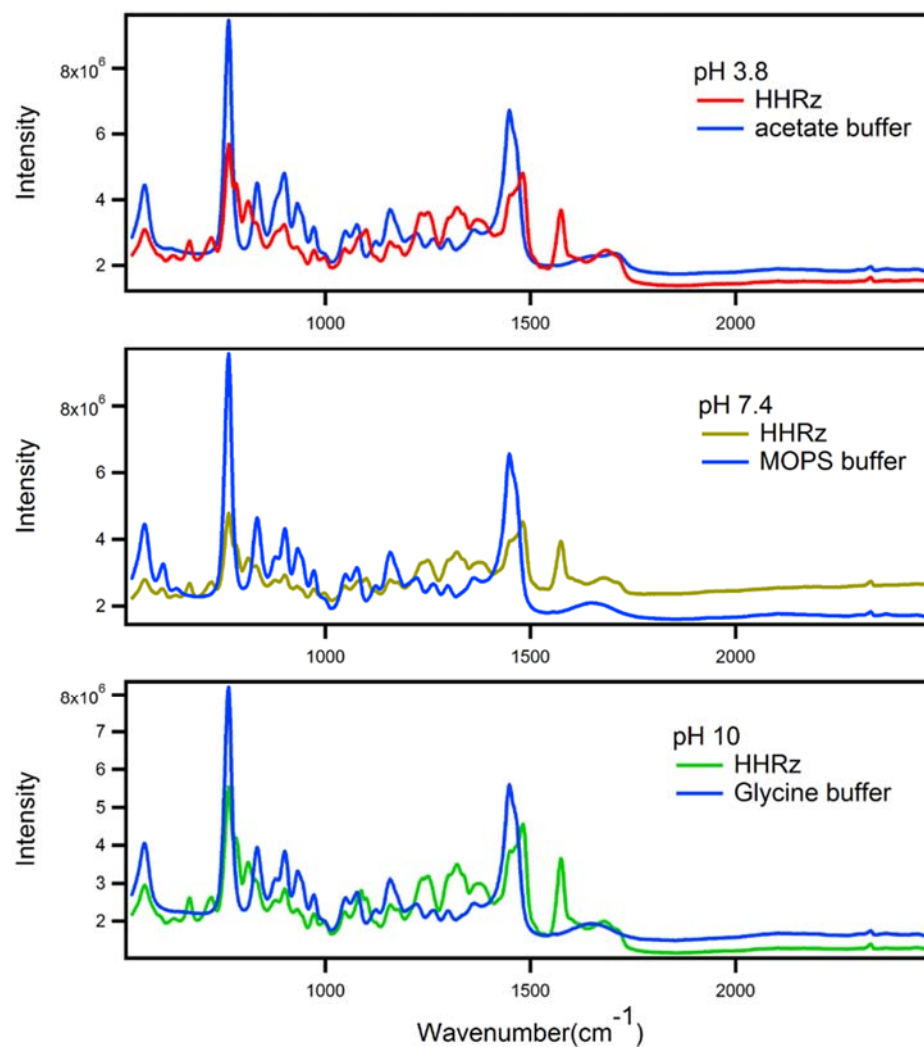


Figure 3.20: Five-minute average spectrum of HHRz crystal suspended in a hanging drop and the associated buffer at pH 3.8, 7.4, and 10. Due to the range of pH's probed different buffers were used at different pH's.

The Raman spectral contribution of the buffer was subtracted from HHRz spectra and the background removed. Figure 3.21 shows the intensity normalized spectra of HHRz at

different pH. There is no significant change seen in the Raman spectra of HHRz due to the deprotonation of 1 base. This could be due to the fact that the change in the Raman spectra is extremely small (on the order of 1%) and buried under other Raman bands.

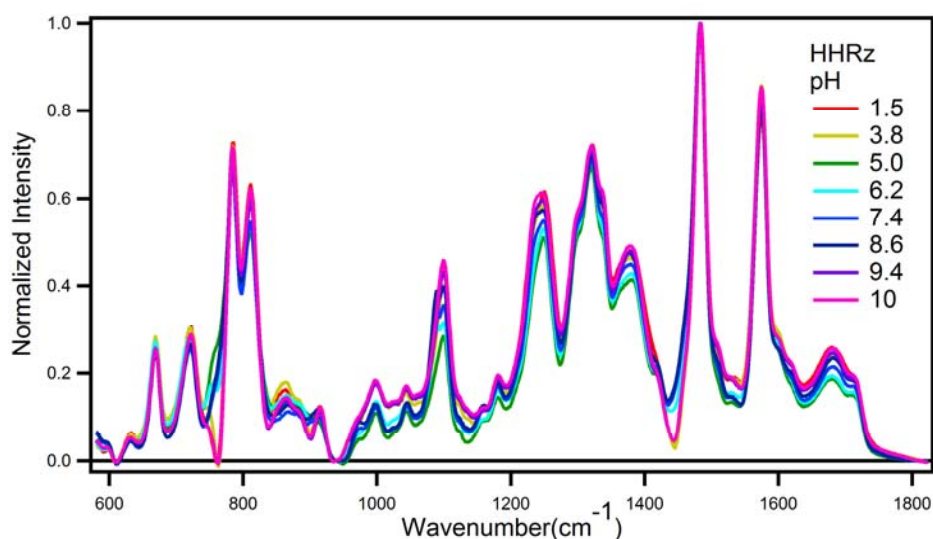


Figure 3.21: Background subtracted normalized SC spectra of HHRz collected at different pH

### 3.5 Conclusion

Raman spectroscopy can be used as a useful tool to monitor reactions in biological samples in aqueous environments. Raman spectra of NMP models at different pH's show significant changes due to changes in pH. A pH curve can be made by using Raman band intensity ratios and plot against pH. The midpoint of the sigmoidal curve yields the pKa of the reaction. Raman spectroscopy can also be used to detect changes in nucleotide bases in more complex molecules. Although we did not successfully probe pH changes of a single nucleotide in HHRz, changes in protein and RNA crystals can be studied in the aqueous

environment which is an advantage Raman spectroscopy offers over many other spectroscopic techniques.

## REFERENCES

## REFERENCES

1. Ewen. Smith. *Modern Raman spectroscopy: a practical approach*. (JWiley, 2005).
2. Ball, D. W. & Society of Photo-optical Instrumentation Engineers. *Field guide to spectroscopy*. (SPIE, 2006).
3. Kumar, S. *et al.* Raman and infra-red microspectroscopy: towards quantitative evaluation for clinical research by ratiometric analysis. *Chem Soc Rev* **45**, 1879–1900 (2016).
4. Nafie, L. A. Recent advances in linear and non-linear Raman spectroscopy. Part IX: Recent advances in linear and non-linear Raman spectroscopy. *J. Raman Spectrosc.* **46**, 1173–1190 (2015).
5. Rygula, A. *et al.* Raman spectroscopy of proteins: a review: Raman spectroscopy of proteins. *J. Raman Spectrosc.* **44**, 1061–1076 (2013).
6. Carey, P. R. RAMAN CRYSTALLOGRAPHY AND OTHER BIOCHEMICAL APPLICATIONS OF RAMAN MICROSCOPY. *Annu. Rev. Phys. Chem.* **57**, 527–554 (2006).
7. Gong, B. *et al.* Raman crystallography of RNA. *Methods* **49**, 101–111 (2009).
8. O'Connor, T., Johnson, C. & Scovell, W., M. RAMAN pH PROFILES FOR NUCLEIC ACID CONSTITUENTS I. CYTIDINE AND URIDINE RIBONUCLEOSIDES. *Biochim. Biophys. Acta* **447**, 484–494 (1976).
9. O'Connor, T., Johnson, C. & Scovell, W., M. Raman ph profiles for nucleic acid constituents. II. 5'-AMP and 5'-GMP ribonucleotides..pdf. *Biochim. Biophys. Acta* **447**, 495–508 (1976).
10. Klein, K. *et al.* Label-Free Live-Cell Imaging with Confocal Raman Microscopy. *Biophys. J.* **102**, 360–368 (2012).
11. Antonio, K. A. & Schultz, Z. D. Advances in Biomedical Raman Microscopy. *Anal. Chem.* **86**, 30–46 (2014).

12. Majzner, K. *et al.* 3D confocal Raman imaging of endothelial cells and vascular wall: perspectives in analytical spectroscopy of biomedical research. *The Analyst* **138**, 603–610 (2013).
13. Uzunbajakava, N., de Peinder, P., 't Hoof, G. W. & van Gogh, A. T. Low-cost spectroscopy with a variable multivariate optical element. *Anal. Chem.* **78**, 7302–7308 (2006).
14. Davis, B. M. *et al.* Multivariate Hyperspectral Raman Imaging Using Compressive Detection. *Anal. Chem.* **83**, 5086–5092 (2011).
15. Cebeci Maltaş, D., Kwok, K., Wang, P., Taylor, L. S. & Ben-Amotz, D. Rapid classification of pharmaceutical ingredients with Raman spectroscopy using compressive detection strategy with PLS-DA multivariate filters. *J. Pharm. Biomed. Anal.* **80**, 63–68 (2013).
16. Wilcox, D. S. *et al.* Digital compressive chemical quantitation and hyperspectral imaging. *The Analyst* **138**, 4982 (2013).
17. Rehrauer, O. G., Mankani, B. R., Buzzard, G. T., Lucier, B. J. & Ben-Amotz, D. Fluorescence modeling for optimized-binary compressive detection Raman spectroscopy. *Opt. Express* **23**, 23935 (2015).
18. Quyen, Ng. T., Jouan, M. D., Dao, N. Q., Da Silva, E. & Phuong, D. A. New Raman spectrometer using a digital micromirror device and a photomultiplier tube detector for rapid on-line industrial analysis. Part II: Choice of analytical methods. *Appl. Spectrosc.* **62**, 279–284 (2008).
19. Quyen, Ng. T., Da Silva, E., Dao, N. Q. & Jouan, M. D. New Raman Spectrometer Using a Digital Micromirror Device I.pdf. *Appl. Spectrosc.* **62**, 273–278 (2008).
20. Wilcox, D. S., Buzzard, G. T., Lucier, B. J., Wang, P. & Ben-Amotz, D. Photon level chemical classification using digital compressive detection. *Anal. Chim. Acta* **755**, 17–27 (2012).
21. Smith, Z. J., Strombom, S. & Wachsmann-Hogiu, S. Multivariate optical computing using a digital micromirror device for fluorescence and Raman spectroscopy. *Opt. Express* **19**, 16950–16962 (2011).
22. Buzzard, G. T. & Lucier, B. J. Optimal filters for high-speed compressive detection in spectroscopy. in (eds. Bouman, C. A., Pollak, I. & Wolfe, P. J.) 865707 (2013). doi:10.1117/12.2012700
23. Nelson, E. D. Hadamard Spectroscopy.pdf. *J. Opt. Soc. Am.* **60**, 1664–1669 (1970).
24. Treado, P. J. & Morris, M. D. A thousand points of light: the Hadamard transform in chemical analysis and instrumentation. *Anal. Chem.* **61**, 723A–734A (1989).

25. Marshall, A. G. & Comisarow, M. B. Fourier and Hadamard transform methods in spectroscopy. *Anal. Chem.* **47**, 491A–504a (1975).
26. Damian, C., Crisan, A., Vasile, T., Coltuc, D. & Damian, V. Noise evaluation in Hadamard spectroscopy. in *Signals, Circuits and Systems (ISSCS), 2015 International Symposium on 1–4* (IEEE, 2015).
27. Lieber, C. A. & Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Appl. Spectrosc.* **57**, 1363–1367 (2003).
28. Zhao, J., Lui, H., McLean, D. I. & Zeng, H. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. *Appl. Spectrosc.* **61**, 1225–1232 (2007).
29. Farouki, R. T. The Bernstein polynomial basis: A centennial retrospective. *Comput. Aided Geom. Des.* **29**, 379–419 (2012).
30. Buzzard, G. T. & Lucier, B. J. Optimal filters for high-speed compressive detection in spectroscopy. in (eds. Bouman, C. A., Pollak, I. & Wolfe, P. J.) 865707 (2013). doi:10.1117/12.2012700
31. Majzner, K. *et al.* 3D confocal Raman imaging of endothelial cells and vascular wall: perspectives in analytical spectroscopy of biomedical research. *The Analyst* **138**, 603–610 (2013).
32. Antonio, K. A. & Schultz, Z. D. Advances in Biomedical Raman Microscopy. *Anal. Chem.* **86**, 30–46 (2014).
33. Camp Jr, C. H. & Cicerone, M. T. Chemically sensitive bioimaging with coherent Raman scattering. *Nat. Photonics* **9**, 295–305 (2015).
34. Carey, P. R. RAMAN CRYSTALLOGRAPHY AND OTHER BIOCHEMICAL APPLICATIONS OF RAMAN MICROSCOPY. *Annu. Rev. Phys. Chem.* **57**, 527–554 (2006).
35. Gong, B. *et al.* Raman crystallography of RNA. *Methods* **49**, 101–111 (2009).
36. Gong, B. *et al.* Direct Measurement of a  $pK_a$  near Neutrality for the Catalytic Cytosine in the Genomic HDV Ribozyme Using Raman Crystallography. *J. Am. Chem. Soc.* **129**, 13335–13342 (2007).
37. Mir, A. *et al.* Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction. *Biochemistry (Mosc.)* **54**, 6369–6381 (2015).

38. Mir, A. & Golden, B. L. Two Active Site Divalent Ions in the Crystal Structure of the Hammerhead Ribozyme Bound to a Transition State Analogue. *Biochemistry (Mosc.)* **55**, 633–636 (2016).
39. O'Connor, T., Johnson, C. & Scovell, W., M. RAMAN pH PROFILES FOR NUCLEIC ACID CONSTITUENTS I. CYTIDINE AND URIDINE RIBONUCLEOSIDES. *Biochim. Biophys. Acta* **447**, 484–494 (1976).
40. O'Connor, T., Johnson, C. & Scovell, W., M. Raman ph profiles for nucleic acid constituents. II. 5'-AMP and 5'-GMP ribonucleotides..pdf. *Biochim. Biophys. Acta* **447**, 495–508 (1976).
41. Lawton, W. H. & Sylvestre, E. A. Self Modeling Curve Resolution. *Technometrics* **13**, 617–633 (1971).
42. Davis, J. G., Gierszal, K. P., Wang, P. & Ben-Amotz, D. Water structural transformation at molecular hydrophobic interfaces. *Nature* **491**, 582–585 (2012).
43. Garrett, R. H. & Grisham, C. M. *Biochemistry*. (Brooks/Cole, Center of Learning, 2013).



VITA

## VITA

Bharat R. Mankani was born on May 28, 1982 in Hyderabad, India. He attended Hyderabad Public School and completed high school at St. Mary's Jr. College, Yousufguda. In 2007, Bharat graduated from Texas A&M University- Kingsville with a Bachelor of Science in chemistry with minor in biology. In 2010 Bharat graduated with a Master of science in chemistry from Texas A&M University- Kingsville. Bharat worked at Promega Corporation, WI, where he was involved in chemistry development of forensic kits for Maxwell 16 automated system. In fall 2011 Bharat joined Purdue University to pursue Ph.D. in chemistry.

Bharat enjoys playing the piano, violin and the flute and has passed certified exams from the Trinity College of London. Bharat has played squash in many different clubs in India and the U.S and has played multiple open tournaments at the University Club of Chicago.

PUBLICATION

# Fluorescence modeling for optimized-binary compressive detection Raman spectroscopy

Owen G. Rehrauer,<sup>1</sup> Bharat R. Mankani,<sup>1</sup> Gregery T. Buzzard,<sup>2</sup>  
Bradley J. Lucier,<sup>3</sup> and Dor Ben-Amotz<sup>1,\*</sup>

<sup>1</sup>*Purdue University, Department of Chemistry, West Lafayette, IN, USA*

<sup>2</sup>*Purdue University, Department of Mathematics, West Lafayette, IN, USA*

<sup>3</sup>*Purdue University, Departments of Mathematics and Computer Science, West Lafayette, IN, USA*

[\\*bendor@purdue.edu](mailto:bendor@purdue.edu)

**Abstract:** The recently-developed optimized binary compressive detection (OB-CD) strategy has been shown to be capable of using Raman spectral signatures to rapidly classify and quantify liquid samples and to image solid samples. Here we demonstrate that OB-CD can also be used to quantitatively separate Raman and fluorescence features, and thus facilitate Raman-based chemical analyses in the presence of fluorescence background. More specifically, we describe a general strategy for fitting and suppressing fluorescence background using OB-CD filters trained on third-degree Bernstein polynomials. We present results that demonstrate the utility of this strategy by comparing classification and quantitation results obtained from liquids and powdered mixtures, both with and without fluorescence. Our results demonstrate high-speed Raman-based quantitation in the presence of moderate fluorescence. Moreover, we show that this OB-CD based method is effective in suppressing fluorescence of variable shape, as well as fluorescence that changes during the measurement process, as a result of photobleaching.

© 2015 Optical Society of America

**OCIS codes:** (300.6450) Spectroscopy, Raman; (120.6200) Spectrometers and spectroscopic instrumentation; (070.6120) Spatial light modulators.

---

## References and links

1. J. W. Chan, D. S. Taylor, T. Zwerdling, S. M. Lane, K. Ihara, and T. Huser, "Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells," *Biophys. J.* **90**, 648–656 (2006).
2. A. M. K. Enejder, T.-W. Koo, J. Oh, M. Hunter, S. Sasic, M. S. Feld, and G. L. Horowitz, "Blood analysis by Raman spectroscopy," *Opt. Lett.* **27**, 2004–2006 (2002).
3. O. O. Soyemi, F. G. Haiback, P. J. Gemperline, and M. L. Myrick, "Nonlinear optimization algorithm for multivariate optical element design," *Appl. Spectros.* **56**, 477–487 (2002).
4. M. F. Duarte, M. A. Davenport, D. Tarkhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.* **25**, 83–101 (2008).
5. W. C. Sweatt, C. A. A. Boye, S. M. Gentry, M. R. Descour, B. R. Stallard, and C. L. Grotbeck, "ISIS: an information-efficient spectral imaging system," *Proc. SPIE* **3438**, 98–106 (1998).
6. N. Uzunbajakava, P. de Peinder, G. W. t Hooft, and A. T. M. van Gogh, "Low-cost spectroscopy with a variable multivariate optical element," *Anal. Chem.* **78**, 7302–7308 (2006).
7. J. E. Vornehm, A. J. Dong, R. W. Boyd, and Z. Shi, "Multiple-output multivariate optical computing for spectrum recognition," *Opt. Express* **21**, 25005–25014 (2014).

8. D. S. Wilcox, G. T. Buzzard, B. J. Lucier, P. Wang, and D. Ben-Amotz, "Photon level chemical classification using digital compressive detection," *Anal. Chim. Acta* **755**, 17–27 (2012).
9. D. S. Wilcox, G. T. Buzzard, B. J. Lucier, O. G. Rehrauer, P. Wang, and D. Ben-Amotz, "Digital compressive chemical quantitation and hyperspectral imaging," *Analyst* **138**, 4982–4990 (2013).
10. G. T. Buzzard and B. J. Lucier, "Optimal filters for high-speed compressive detection in spectroscopy," *Proc. SPIE* **8657**, 865707 (2013).
11. J. Zhao, H. Lui, D. I. McLean, and H. Zeng, "Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy," *Appl. Spectrosc.* **61**, 1225–1232 (2007).
12. C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl. Spectrosc.* **57**, 1363–1367 (2003).
13. D. V. Martyshev, R. C. Ahuja, A. Kudriavtsev, and S. B. Mirov, "Effective suppression of fluorescence light in Raman measurements using ultrafasttime gated charge coupled device camera," *Rev. Sci. Instrum.* **75** 630–635 (2004).
14. N. Everall, R. W. Jackson, J. Howard, and K. Hutchinson, "Fluorescence rejection in Raman spectroscopy using a gated intensified diode array detector," *J. Raman Spectrosc.* **17** 415–423 (1986).
15. N. J. Everall, J. P. Partanen, J. R.M. Barr and M. J. Shaw, "Threshold measurements of stimulated Raman scattering in gases using picosecond KrF laser pulses," *Opt. Commun.* **64** 393–397 (1987).
16. D. Zhang and D. Ben-Amotz "Enhanced chemical classification of Raman images in the presence of strong fluorescence interference," *Appl. Spectrosc.* **54** 1379–1383 (2000).
17. F. Pukelsheim, *Optimal design of experiments*, (John Wiley & Sons Inc. 1993).
18. A. C. Atkinson, A. N. Donev, and R. D. Tobias, *Optimum experimental designs with SAS*, (Oxford University Press, 2007).
19. K. Golcuk, G. S. Mandair, A. F. Callender, N. Sahar and D. H. Kohn and M. D. Morris, "Is photobleaching necessary for Raman imaging of bone tissue using a green laser?" *BBA - Biomembranes* **1758**, 868–873 (2006).
20. S. Bernstein, "Demonstration du theoreme de weierstrass fondee sur le calcul des probabilites," *Comm. Soc. Math. Kharkov* **13**, 1–2 (1912).
21. R. T. Farouki, "The Bernstein polynomial basis: a centennial retrospective," *Comput. Aided Geom. Design* **29**, 379–419 (2012).
22. C. Frausto-Reyes, C. Medina-Gutiérrez, R. Sato-Berrú, and L. R. Sahagún, "Qualitative study of ethanol content in tequilas by Raman spectroscopy and principal component analysis," *Spectrochim. Acta A* **61**, 2657–2662 (2005).
23. R. Sato-Berrú, J. Medina-Valtierra, C. Medina-Gutiérrez, and C. Frausto-Reyes, "Quantitative NIR Raman analysis in liquid mixtures," *Spectrochim. Acta A* **60**, 2225–2229 (2004).

---

## 1. Introduction

Current spectroscopic chemical analysis instruments are capable of generating data sets that are so large that they require transforming the data to a lower-dimensional space, using methods such as principal component analysis (PCA) [1] or partial-least squares (PLS) [2] to facilitate subsequent automated chemical classification and quantitation. Moreover, because the collection of high-dimensional data is often the slowest step in the process, a number of compressive detection strategies [3–7] have been introduced with the goal of increasing data collection speed by making measurements only in the low-dimensional space containing the information of interest. One such method is our previously described optimized binary compressive detection (OB-CD) strategy, in which OB filters are applied to a digital mirror microarray (DMD) to redirect or collect photons of (multiple) selected colors, for detection using a single channel detector, such as a photon counting photomultiplier tube (PMT) or an avalanche photodiode (APD) [8, 9].

Previously [8–10], we demonstrated that the OB-CD strategy enabled high-speed chemical classification, quantitation, and imaging. Here we demonstrate an extension of the OB-CD method that facilitates Raman classification and quantitation in the presence of fluorescence background. Key advantages of this method, relative to conventional fluorescence subtraction strategies [11–16], include its compatibility with automated high-speed chemical analysis in the presence of variable fluorescence backgrounds. Here we show that fluorescence backgrounds may be quantified and subtracted on-the-fly by including Bernstein polynomial spectral functions in the OB-CD training set, along with Raman spectra of the components of interest. In

other words, we augment the Raman spectral training set with Bernstein polynomial spectral shapes to model fluorescence and thus obtain OB-CD filters that are used to quantify both the chemical components of interest and the fluorescence background spectra.

## 2. Theory

We begin by giving a short summary of the OB-CD mathematical framework [8]; some of the wording is taken from the cited paper and we refer the reader there for more details and discussion, including our assumption of linear additive spectra, which justifies in part the form of our model. This framework is based on the subfield of statistics known as “Optimal Design of Experiments” [17]. We then discuss the notion of *nuisance parameters* and how this notion can be exploited to reduce the variance of the estimated photon emission rates pertaining to the (non-nuisance) components of interest. Finally, we introduce a new strategy for modeling, quantifying, and suppressing fluorescence background signals in Raman spectra.

### 2.1. Review of mathematical model

We assume that our chemical sample consists of a mixture of  $n$  known chemical species  $S_j$ ,  $j = 1, \dots, n$ . In a particular sample, the species  $S_j$  emits photons at a rate  $\Lambda_j$ , so the number of photons emitted in time  $t$  is a Poisson random variable with mean  $t\Lambda_j$ ; our goal is to estimate these rates as accurately as possible, in order to quantify the composition of a sample containing such components.

The wavelength, or energy, of each photon observed in the experiments can be labeled with an integer  $i \in \{1, \dots, N\}$ . Assume that we know the shape of the spectrum associated with species  $S_j$ ; denote the probability that a photon from species  $S_j$  has label  $i$  by  $P_{ij}$ , so  $\sum_{i=1}^N P_{ij} = 1$ . In other words, the  $P_{ij}$ ,  $i = 1, \dots, N$ , form the spectrum of the  $j$ th compound, normalized so that the sum is 1. Thus the stream of labeled photons emanating from a sample is modeled by a vector Poisson process with rates  $P\Lambda$ , where  $\Lambda = (\Lambda_1, \dots, \Lambda_n)^T$ , and  $P = (P_{ij})_{N \times n}$ . (Here and later, superscript  $T$  denotes “transpose.”) If we measure the number of photons that arrive in each energy bin for time  $t$  then the number of photons with label  $i$  entering our instrument from all chemical species has a Poisson distribution with mean  $t(P\Lambda)_i = t \sum_{j=1}^n P_{ij}\Lambda_j$ . We assume that the number of wavelength channels,  $N$ , is greater than the number of chemical species  $n$ , and that the columns of  $P$  are linearly independent, i.e.,  $P$  has full rank. (In other words, we assume that no spectrum can be written as a linear combination of the other spectra.)

We consider taking  $m$  independent measurements with  $m \geq n$ . In the  $k$ th measurement, we set in our *optical filter* the transmittance of all photons with energy level  $i$  to be a number  $F_{ik}$  with  $0 \leq F_{ik} \leq 1$ ; i.e., the probability that in the  $k$ th measurement a photon with energy label  $i$  is counted is  $F_{ik}$ . Our observation in the  $k$ th measurement is the total photon count, summed over all energy levels  $i$ , from observing the photon stream for time  $T_{kk}$ , which will be a Poisson random variable with mean

$$T_{kk} \sum_{i=1}^N F_{ik} \left( \sum_{j=1}^n P_{ij}\Lambda_j \right) = T_{kk} \sum_{i=1}^N \sum_{j=1}^n F_{ik} P_{ij}\Lambda_j.$$

(We use a double subscript on  $T_{kk}$  because we will make these numbers the diagonal of a matrix  $T$ .) We refer to the columns of the matrix  $F = (F_{ik})_{N \times m}$  as *filters*, and the entries of  $F$  can be chosen as we wish, since they are parameters of our measuring device. Based on [17] and as in [8], we choose filters  $F$  and corresponding measurement times  $T$  to minimize the variance in estimated rates, subject to a total time constraint expressed as  $\sum_{k=1}^m T_{kk} = 1$ .

For a DMD, we can choose only  $F_{ik} = 0$  or  $F_{ik} = 1$ , while for an analog spatial light modulator (SLM) we can in principle choose any  $0 \leq F_{ik} \leq 1$ . (Moreover, one could use a DMD as an

analogue filter by varying time for which each individual energy bin in a given filter is turned on, although we have not done so in the present studies.)

We denote by  $\mathbf{x}$  our complete observation, a vector of  $m$  independent Poisson random variables with means and variances given by the vector  $TF^T P\Lambda$ , where  $T$  is the  $m \times m$  diagonal matrix with diagonal entries  $T_{11}, \dots, T_{mm}$  and  $F^T P$  is an  $m \times n$  matrix. We assume that  $F$  is chosen so that  $F^T P$  has rank  $n$  (which is possible since  $P$  has rank  $n$  and  $m \geq n$ ).

If we denote by  $\hat{\mathbf{x}}$  a sample from this random variable, then our estimate  $\hat{\Lambda}$  of the true rates  $\Lambda$  is given by  $\hat{\Lambda} = BT^{-1}\hat{\mathbf{x}}$ , where the  $n \times m$  matrix  $B = (b_{ik})_{n \times m}$  is a *left inverse* of  $F^T P$ , i.e.,  $B(F^T P) = I$ , the  $n \times n$  identity matrix. (If  $n = m$  then  $F^T P$  is a square matrix and  $B$  is simply  $(F^T P)^{-1}$ .) We note that the expected value of  $\hat{\Lambda}$  satisfies  $E(\hat{\Lambda}) = BT^{-1}E(\hat{\mathbf{x}}) = BT^{-1}(TF^T P\Lambda) = \Lambda$ , so  $\hat{\Lambda}$  is an unbiased estimator of  $\Lambda$ .

It was shown earlier [8] that the variance of the estimate  $\hat{\Lambda}_j$  of the  $j$ th rate  $\Lambda_j$  is given by

$$E(|(BT^{-1}\mathbf{x})_j - \Lambda_j|^2) = \sum_{k=1}^m b_{jk}^2 T_{kk}^{-1} (F^T P\Lambda)_k.$$

In the cited paper the variances were summed over *all*  $\Lambda_j$  to derive

$$E(\|BT^{-1}\mathbf{x} - \Lambda\|^2) = \sum_{j=1}^n \sum_{k=1}^m b_{jk}^2 T_{kk}^{-1} (F^T P\Lambda)_k = \sum_{k=1}^m \frac{((F^T P)\Lambda)_k}{T_{kk}} \|\mathbf{B}\mathbf{e}_k\|^2. \quad (1)$$

Here  $\mathbf{e}_k$  is a vector whose components are zero except for a 1 in the  $k$ th component and  $\|y\|^2 = \sum_{j=1}^n y_j^2$ .

We now deviate a bit from the earlier exposition [8], where it is perhaps not stated clearly that because the sum of variances (1) depends on the unknown rates  $\Lambda$ , we cannot choose  $m$ ,  $F$ ,  $T$ , and  $B$  to minimize (1) for *all*  $\Lambda$  simultaneously. Therefore we pick a single  $\bar{\Lambda} = (1, \dots, 1)^T$  and choose  $m$ ,  $F$ ,  $T$ , and  $B$  to minimize

$$E(\|BT^{-1}\mathbf{x} - \bar{\Lambda}\|^2) = \sum_{k=1}^m \frac{((F^T P)\bar{\Lambda})_k}{T_{kk}} \|\mathbf{B}\mathbf{e}_k\|^2 \quad (2)$$

under the constraints  $m \geq n$ ;  $0 \leq F_{ik} \leq 1$ ,  $1 \leq i \leq n$ ,  $1 \leq k \leq m$ ;  $0 < T_{kk}$ ,  $1 \leq k \leq m$ , and  $\sum_{k=1}^m T_{kk} = 1$ ; and  $B(F^T P) = I$ . This choice of  $\bar{\Lambda}$  has proved useful absent additional information [8]. Our experience indicates that the resulting filters and measurement times are relatively insensitive to changing individual coefficients in  $\bar{\Lambda}$  by up to a factor of 100.

## 2.2. Nuisance parameters

It often happens that the photon emission rates of some chemical species are of more interest than others. One might have a contaminant with a broad (known) spectrum; while this may be one of the chemical species  $S_j$ , we don't really care about the accuracy of estimation of that particular  $\Lambda_j$ . In the field of Optimal Design of Experiments, such variables  $\Lambda_j$  are known as *nuisance parameters*: They are a necessary part of the model, but we don't care about the accuracy of their estimates except insofar as it affects the estimates of the other variables [18].

We assume that we are truly interested in the photon emission rates of only the first  $n' < n$  chemical species. For such systems, we don't care about the variances of the estimates of  $\Lambda_j$  for  $n' < j \leq n$  and so instead of minimizing (2) we minimize the sum of the variances of the estimates of only the first  $n'$  emission rates:

$$\sum_{j=1}^{n'} E(|(BT^{-1}\mathbf{x})_j - \bar{\Lambda}_j|^2) = \sum_{j=1}^{n'} \sum_{k=1}^m b_{jk}^2 T_{kk}^{-1} (F^T P\bar{\Lambda})_k.$$

In effect, we estimate  $\Lambda_j$  for  $n' < j \leq n$  only well enough to minimize the error of the sum of variances of the first  $n'$  emission rates.

### 2.3. Estimating fluorescence using Bernstein polynomials

A fluorescent spectrum is generally smooth and broad, in contrast to the narrow peaks found in Raman spectra; it may vary from one sample to another, or from time to time for the same sample.

One can reduce the amount of fluorescence in a sample by photobleaching the sample prior to analysis [19]. Another commonly-used procedure for subtracting such fluorescence backgrounds from Raman spectra is to fit the fluorescent spectrum to a polynomial [11, 12]; in effect, this amounts to using a fixed polynomial of specified degree to model a fluorescent spectrum. This may work well if in a set of samples the fluorescence doesn't vary over space or time; this is often not the case, however, in Raman (and particularly Raman imaging) applications, for which one must estimate the shape of the fluorescence dynamically.

Instead of somehow applying conventional fluorescence fitting and subtraction strategies to OB-CD measurements, we use OB-CD filters that are derived using either actual fluorescence spectra of known shape or a family of polynomials that models general fluorescence spectra.

More specifically, if the sample of interest contains a fixed fluorescent signal of known shape, we do the following. We simply add the fluorescence spectrum as an extra column of  $P$  with associated rate variable  $\Lambda_j$ , and then treat  $\Lambda_j$  as a nuisance parameter, not adding its variance to the sum of variances being minimized.

If the fluorescence has a spectrum that varies over space or time, however, we cannot apply the previous procedure and must model the fluorescence dynamically, as follows.

Since fluorescence backgrounds can often be fit reasonably well by a cubic polynomial, we would like to identify a polynomial basis for cubic polynomials such that (1) all basis elements are nonnegative on  $[0, 1]$  and (2) every nonnegative polynomial can be written as a linear combination of the basis elements with nonnegative coefficients. (Negative coefficients, which are supposed to model rates, are nonphysical and increase the variance of our estimates.) Unfortunately, for  $r > 1$ , no such basis exists.

Of interest, however, is the *Bernstein basis* [20, 21] of polynomials of degree  $r$ , given by

$$B_{v,r}(x) = \binom{r}{v} x^v (1-x)^{r-v}, \quad v = 0, 1, \dots, r,$$

are nonnegative on  $[0, 1]$  (so any linear combination of them with nonnegative coefficients is nonnegative on  $[0, 1]$ ). They have other nice properties: They form a basis for the space of polynomials of degree  $r$ ; they have optimal stability in some sense [21]; and “many” if not “most” nonnegative polynomials that come up in practice as models for fluorescent spectra have nonnegative coefficients in the Bernstein basis (or at least, any negative coefficients are not “overly” large). The Bernstein polynomials even resemble single-peak spectra.

In the specific case  $r = 3$  we have  $B_{0,3}(x) = (1-x)^3$ ,  $B_{1,3}(x) = 3x(1-x)^2$ ,  $B_{2,3}(x) = 3x^2(1-x)$ ,  $B_{3,3}(x) = x^3$ . These polynomials are shown in Fig. 1.

In this work we model fluorescent spectra with linear combinations of  $B_{v,3}(x)$ ,  $v = 0, 1, 2, 3$ , and often treat the Bernstein coefficients as nuisance parameters. Because fluorescent spectra are not precisely polynomials, this introduces *model error*, which we discuss in the Results and Discussion section. We have also found that for some purposes using quartic polynomials ( $r = 4$ ) gives more accurate results.

## 3. Methods and materials

### 3.1. OB-CD spectrometers

We have previously described an OB-CD spectrometer using a 785 nm laser for excitation with 75 mW at the sample [8]. In the present studies we have utilized that spectrometer as well as



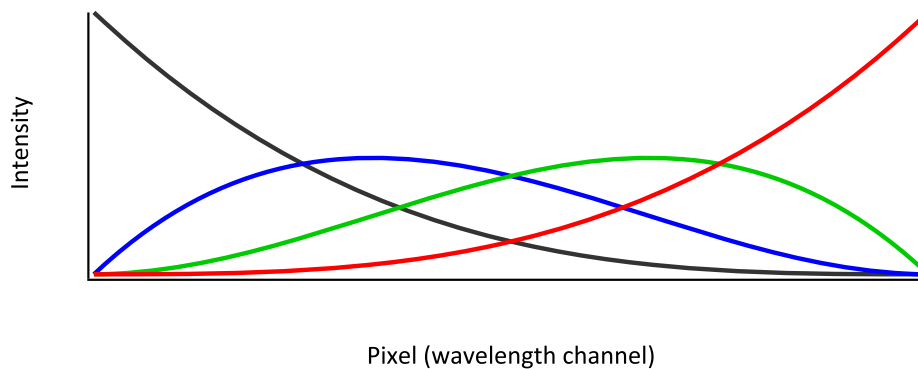


Fig. 1. Plot of the four degree-three Bernstein polynomials as a function of wavelength channel (scaled to be over the interval  $[0, 1]$ .) The colors denote the various polynomials: black,  $B_{3,0}(x)$ ; blue,  $B_{3,1}(x)$ ; green,  $B_{3,2}(x)$ ; and red,  $B_{3,3}(x)$ .

a new OB-CD spectrometer using a 514 nm laser for excitation whose schematic is shown in Fig. 2 (the 785 nm excitation laser system is similar in design, as previously described in [8]). Much like the previously described system, our microscope is configured to collect the back-scattered Raman signal with the same objective lens that is used to focus the argon ion laser (Modu Laser Stellar Pro L 100 mW) onto the sample. The laser passes through a laser-line bandpass filter (Semrock RazorEdge) before it is focused onto the sample using a microscope objective (Nikon MPlan, 20x, 0.4 NA), and unless indicated otherwise the laser power at the sample was about 12 mW for all experiments described in this paper. The backscattered light is collected and then separated from the laser Rayleigh scattering using a dichroic mirror (Semrock RazorEdge). Then, the Raman scattered light is sent to the spectrometer (right portion of Fig. 2), where it is filtered first using a long pass (edge) filter (Semrock RazorEdge), followed by passing through a volume holographic grating (Wasatch Photonics,  $\sim 1000$  lines  $\text{mm}^{-1}$ ). This light is then dispersed onto the DMD (Texas Instruments, DLP3000,  $608 \times 684$  mirror array with  $10.8 \mu\text{m}$  mirror pitch). The spectral window in this system is  $\sim 200\text{--}4100 \text{ cm}^{-1}$ . For all data collected in this paper, we binned two columns of adjacent DMD mirrors together, yielding a total of 342 “bins” with each energy bin corresponding to  $\sim 12 \text{ cm}^{-1}$ . Light from the DMD is then focused onto a photon-counting photomultiplier tube (PMT) (Hamamatsu model #H10682-01) with a dark count rate of  $\sim 500 \text{ photons s}^{-1}$ . TTL pulses from the PMT are counted using a USB data acquisition (DAQ) card (National Instruments, USB-6212BNC). The system is controlled with interface software written in Labview 2013. Binary filter generation is performed as previously described using Matlab (Matlab 7.13 R2011b) [8]. Data was in some cases further processed and manipulated using Igor Pro 6.04.

### 3.2. Chemicals used in classification/quantitation

Acetone and benzene were purchased from Macron (batch #0000070736) and OmniSolv (lot #42282), respectively. Hexanes were acquired from Baxter (lot #901141). Methylcyclohexane was acquired from Mallinckroft (lot #1906 KCBN). Aniline and toluene were acquired from Aldrich (batch #05925CB) and Mallinckroft (lot #8608 X14752), respectively. Aniline was purified via distillation by heating aniline to  $190^\circ \text{C}$  in a round bottom flask, which was connected to a chilled condenser. Ethanol was acquired from Koptec (200 Proof, lot #225411) and water was ultrapurified in our lab (Milli-Q UF Plus,  $18.2 \text{ m}\Omega \text{ cm}$ , Millipore). The overhead

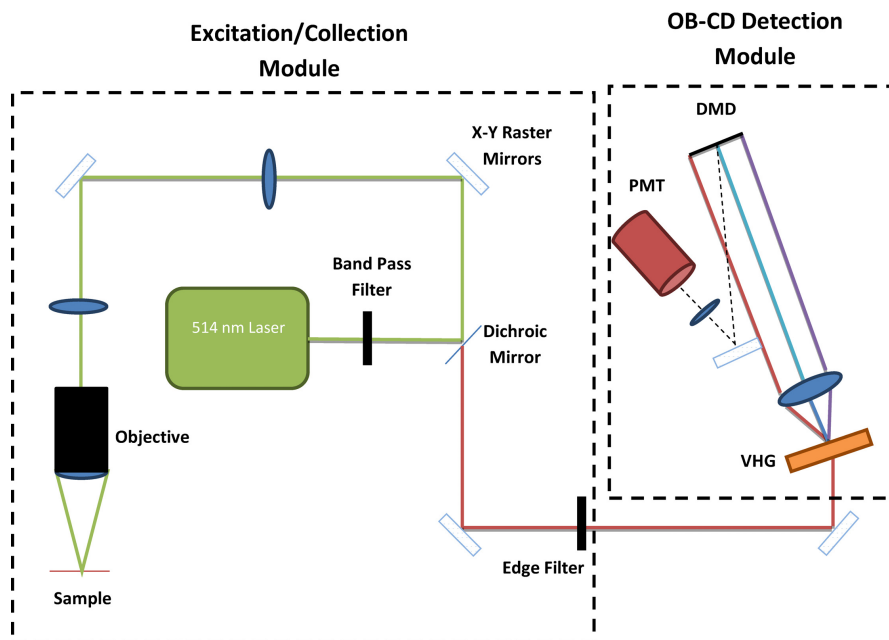


Fig. 2. Schematic of the OB-CD Raman system based upon a 514 nm laser excitation source.

transparency was 3M brand (model #PP2500).

## 4. Results and discussion

### 4.1. Treatment of nuisance parameters

Here we demonstrate how the OB-CD detection strategy is influenced by whether the photon emission rates of some of the spectra used for training OB-CD filters are treated as nuisance parameters; we designate such spectra as *nuisance spectra*. More specifically, we generated an OB-CD training set containing the spectra of hexane, methylcyclohexane, a spectral feature arising from the NIR objective in the 785 nm OB-CD system, and the four Bernstein polynomials shown in Fig. 1. Using this training set, we calculated two sets of OB-CD filters (explicitly, both the matrix of binary filters,  $F$ , and the measurement time matrix,  $T$ ): one set for which no spectral features were considered to be nuisance spectra and a second set for which the spectral feature arising from the NIR objective and the Bernstein polynomials were considered to be nuisance spectra. The optimal filters turned out to be identical in both cases (although that need not in general be the case), while the measurement time matrices,  $T$ , were quite different as shown in Fig. 3. Notice that the OB-CD filters associated with non-nuisance spectra are turned on for longer percentages of the total measurement time relative to OB-CD filters for the same spectral components when no portion of the training set is considered a nuisance spectra.

One might expect that additional time spent measuring the non-nuisance spectra would result in lower variance in the recovered Raman rates for these components. To test this, the two sets of OB-CD filters (and the associated measurement time matrices) shown in Fig. 3 were used to classify hexane and methylcyclohexane (explicitly, there was no added fluorescence in these samples despite including Bernstein polynomials in both OB-CD training sets). Each chemical was measured using each of the sets of OB-CD filters for 1,000 measurement with 10 ms total

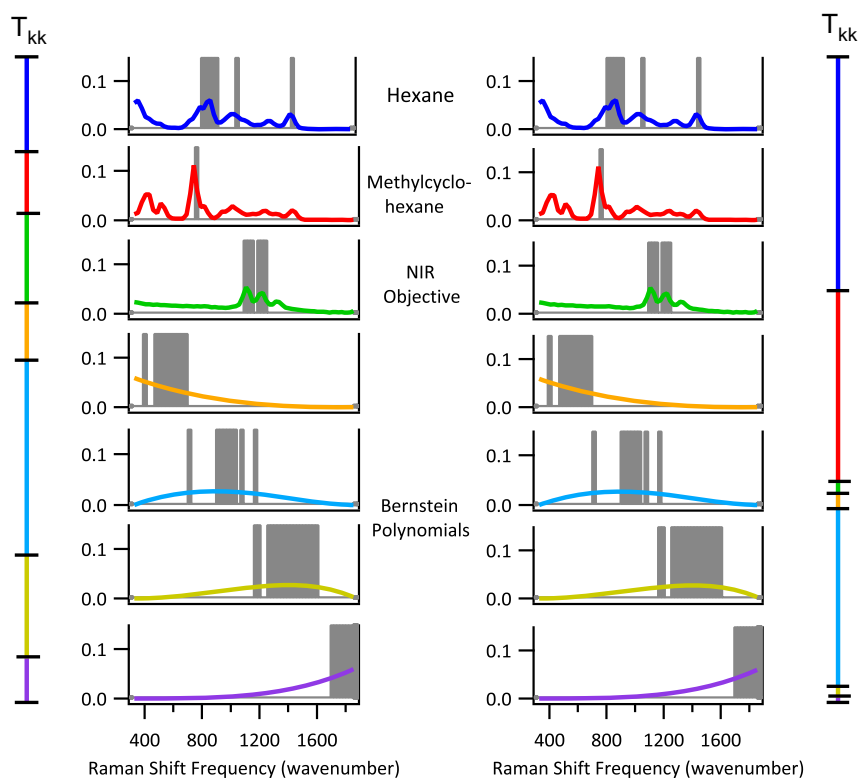


Fig. 3. The colored curves are training spectra (each normalized to unit area) and the gray bands indicate regions in which the OB-CD filters are on (i.e., direct light towards the detector). The Raman spectra were obtained with a spectral resolution of  $30 \text{ cm}^{-1}$ . The lines labeled  $T_{kk}$  correspond to the fraction of the total measurement time that data is collected using the filter associated with each spectral component (denoted by color). The OB-CD filter and  $T_{kk}$  results on the left were obtained without considering any components to be nuisance spectra, while those on the right were obtained when considering the NIR objective and Bernstein polynomials to be nuisance spectra.

integration time. The results of these measurements, shown in Fig. 4, clearly reveal the reduced variance (smaller 95% confidence bands) obtained when treating only the two components of interest as non- nuisance spectra. Note in particular that the mean recovered Raman rates for hexane and methylcyclohexane generated from both sets of OB-CD filters differed very little.

#### 4.2. Validation of Bernstein polynomials

To test and validate our OB-CD fluorescence suppression strategy we either used a white light source to simulate fluorescence or samples containing fluorescent components. The results described in this section were obtained using a white light illuminator as a convenient fluorescence mimic, as its intensity can readily be varied, and its shape resembles typical fluorescence backgrounds (and has a different shape in the 514 nm and 785 nm spectral region). In subsequent subsections, we describe results obtained using samples with fluorescent impurities, rather than white light, to validate our OB-CD strategy. Here we produced OB-CD filters by training using Bernstein polynomials as well as the Raman spectra of n-hexane and methylcyclohexane (and a

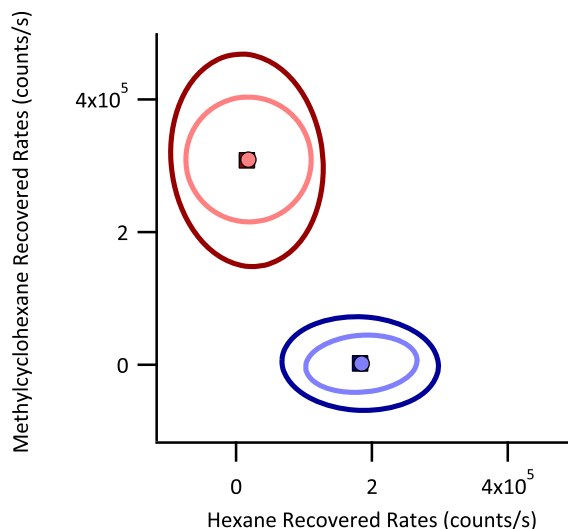


Fig. 4. Recovered Raman rates for hexane (blue) and methylcyclohexane (red) generated using OB-CD filters that considered components of the training set to be nuisance spectra (dark blue and dark red) and OB-CD filters that considered the spectral component arising from the NIR objective and the four Bernstein polynomials to be nuisance spectra (light blue and light red). In all cases, 1,000 OB-CD measurements were taken with a total integration time of 10 ms. The ellipses represent the 95% confidence interval of the recovered Rates for each sample. The large markers in the center of each ellipse represent the mean recovered Raman rates.

spectral component arising from the NIR objective for the 785 nm system). We treat the coefficients of the Bernstein polynomials and the spectral component arising from the NIR objective for the 785 nm system as nuisance parameters. The resulting filters and training spectra from the 785 nm system are virtually identical to those shown in Fig. 3 (and are provided in the Appendix, along with the filters generated for the 514 nm OB-CD system).

The following results were obtained by holding the white light at constant intensity (of about 4 million counts per second) such that the total Raman/fluorescence signal intensity never exceeds 5 million counts per second. We then varied the Raman excitation laser intensity (using neutral density filters) in order to vary the relative amount of Raman and fluorescence in the measured spectra and OB-CD signals, and thus determine how fluorescence background influenced our the recovery of apparent Raman rates using OB-CD.

At each Raman signal intensity, we recovered Raman rates using the OB-CD filters as described in Section 2.1, both with and without the white light background. This allowed us to determine the error in recovered Raman rates as a result of adding fluorescent (white light) background. Figure 5 compares the Raman rates recovered with ( $y$ -axis) and without ( $x$ -axis) added white light when using 30 ms total integration time for each measurement. The number at the top indicates the ratio of the integrated area of the fluorescence and Raman signals and the error bars represent the standard deviations of the Raman rates for each component (see the figure caption for further details).

If fluorescence is not perfectly modeled, then the recovered Raman rates can contain a systematic modeling error whose magnitude increases with fluorescent intensity. When using a constant intensity white light to model fluorescence such modeling error introduces an approximately constant offset to the recovered Raman rates. The magnitude of this offset can be de-

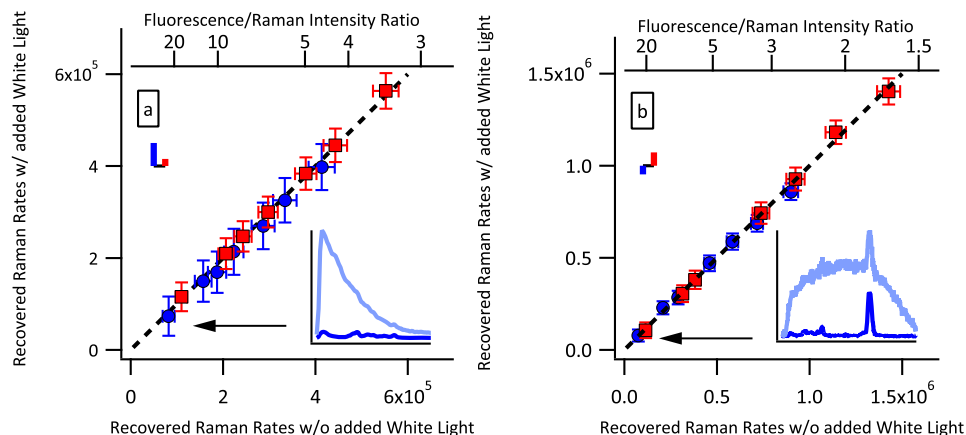


Fig. 5. Plot of hexane (blue) and methylcyclohexane (red) recovered Raman rates measured without added white light versus recovered Raman rates measured with added white light on the (a) 785 nm laser excitation system and (b) 514 nm laser excitation system. Rates have each been corrected by removing a small constant vertical offset (modeling error) whose magnitude was determined by measuring the apparent recovered Raman rates obtained in measurements performed on white light without Raman. The magnitude of this correction is represented by the colored bars in the upper left of each plot. Each point represents the means of 1,000 measurements (each obtained using a 30 ms total integration time) with error bars representing 1 standard deviation. Top axis denotes the ratio of the total (integrated) number of the white light/Raman photons. The inset spectra were obtained from hexane with and without added white light with  $\sim 1$  OD neutral density filter and correspond to measurements made at the points denoted by the arrows.

terminated by measuring the apparent recovered rates of the Raman components obtained when measuring only white light (containing no Raman photons). For the measurements shown in Fig. 5, these modeling errors were relatively small and have magnitudes indicated by the bars in the upper left of each plot. These modeling errors have been subtracted from each of the points in Fig. 5. In other words, before correcting for this modeling error, all of the recovered Raman rates were slightly offset from the dashed diagonal line (of slope 1).

If the spectrum of white light was modeled perfectly by a degree-three polynomial (or was corrected for modeling error, as described above), we would expect that the mean recovered Raman rates for samples with added white light would not significantly differ from the mean recovered Raman scattering rates without added white light. In other words, we would expect the points to lie on a line with slope one as indicated by the dashed lines in Fig. 5. Thus, the agreement between the points and dashed line in Fig. 5 clearly demonstrate that Raman components can be quantified accurately in the presence of fluorescence backgrounds whose integrated intensity is up to 20 times that of the Raman component of interest. Note that the factor of 20 is obtained from the results shown in Fig. 5, as this is when the Raman signal-to-noise approaches 1:1.

Samples with larger fluorescence/Raman intensity ratios can in principle be accurately analyzed using OB-CD with longer measurement times. However, when the integration time approaches one second it may be appropriate to use conventional full spectral measurements and fluorescence subtraction procedures as the OB-CD strategy is primarily advantageous for performing high speed (or low light level) measurements that are not compatible with CCD detection. Additionally, performing OB-CD measurements on samples in which fluorescence is more than 20 times as intense as the Raman signal of interest would require careful modeling

error correction (as the modeling error would become large relative to the Raman intensities). The results presented in Section 4.3 demonstrate the accuracy with which high speed OB-CD Raman classification and quantitative measurements may be performed without correcting for model error so long as the fluorescence background intensity does not exceed 20 times the Raman signal intensity.

### 4.3. Raman quantitation and classification of fluorescent samples

#### 4.3.1. Toluene and fluorescent aniline

The following results were obtained using liquid mixtures of toluene and an aniline sample that was partially oxidized and, as a result, developed a fluorescent impurity that could be removed by distillation. We selected these two liquids because of their significant spectral overlap (the dot product of the two normalized spectral vectors is 0.91) and thus successful classification of aniline/toluene mixtures may be used to demonstrate that our fluorescent mitigation strategy is compatible with the Raman-based quantification of such spectrally overlapped mixtures.

We trained OB-CD filters using 785 nm spectra obtained from distilled aniline, toluene, the NIR objective spectrum describe above, as well as the four Bernstein polynomials shown in Fig. 1 (all of the resulting spectra and OB-CD filter functions are given in the Appendix). For this experiment, we treated the spectral component arising from the NIR objective and the four Bernstein polynomials as nuisance spectra. We used OB-CD to recover Raman rates for toluene, distilled aniline, fluorescent aniline, and mixtures of both types of aniline and toluene. Using these recovered Raman rates, we calculated apparent volume fractions of aniline and toluene as follows:

$$\chi_i = \frac{w_i \hat{\Lambda}_i}{\sum_j w_j \hat{\Lambda}_j},$$

where  $w_i$  is equal to  $M_i / \hat{\Lambda}_i^{\text{Max}}$ ,  $M_i$  is equal to molarity of the  $i^{\text{th}}$  pure liquid, and  $\hat{\Lambda}_i^{\text{Max}}$  is equal to the mean recovered Raman scattering rate for the  $i^{\text{th}}$  pure liquid as previously reported [9]. We then estimated the apparent volume fraction ( $\Phi$ ) for aniline and toluene in each sample. We did this by dividing the apparent mole fraction (for either aniline or toluene) by the molarity of each pure liquid as follows:

$$\Phi_i = \frac{M_i \chi_i}{\sum_j M_j \chi_j}.$$

Figure 6 plots the resulting apparent volume fractions of toluene and aniline as well as mixtures of the two. The results in Fig. 6 demonstrate that the mean recovered Raman rates are insensitive to the fluorescence arising from the impure aniline sample. The variance of the measurements without fluorescence, however, is less than the variance of samples with fluorescence (as a result of additional shot noise in the latter measurements).

#### 4.3.2. Aqueous ethanol and fluorescent tequila

Previous work [22, 23] has demonstrated that Raman spectroscopy can be used to quantify the volume percentage of ethanol in tequila samples and qualitatively distinguish distilled (“silver”) and highly fluorescent, aged (or “golden”) tequilas, even in the presence of fluorescence (more common in aged, so-called “golden” tequila). Here we show that our OB-CD fluorescence mitigation strategy can be used to quantify the volume percentage of ethanol in tequila, even for fluorescent “golden” tequila samples at speeds much greater than those previously reported for this application.

We used the 514 nm laser-based OB-CD system for these studies, both because there was more fluorescence produced at this wavelength than when using the 785 nm excitation and because the C-H and O-H stretch vibrational bands are not readily detectable using the 785 nm

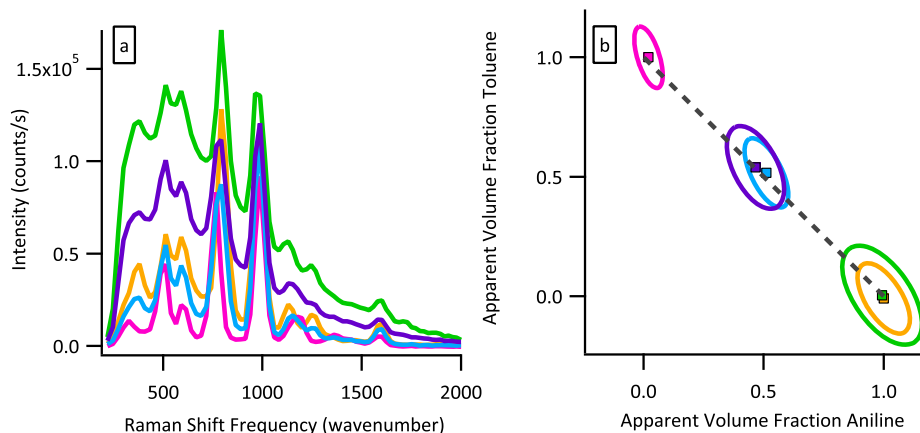


Fig. 6. (a) Spectra of distilled aniline (orange), toluene (magenta), fluorescent aniline (green), a 47:53 volume-by-volume mixture of distilled aniline and toluene (dark green), and a 52:48 mixture of fluorescent aniline and toluene (cyan) measured on the 785 nm OB-CD system. (b) Apparent volume fractions of distilled aniline (orange), toluene (magenta), fluorescent aniline (green), a 47:53 volume-by-volume mixture of distilled aniline and toluene (dark green), and a 52:48 mixture of fluorescent aniline and toluene (cyan). Each chemical was sampled 1,000 times at 20 ms per experiment. Ellipses correspond to the 95% confidence interval of the recovered rates for each sample. The large squares with black borders in the center of each ellipse represent the mean of each sample.

system. Spectra were collected and OB-CD filters were trained using ethanol, water, and Bernstein polynomials (shown in the Appendix). Note that we treated the four Bernstein polynomials as nuisance spectra for OB-CD filter generation. After this, OB-CD was used to recover Raman rates for ethanol and water in a silver tequila (“Arandas” brand) and a golden tequila (“Casamigos” brand). The corresponding apparent volume fractions were obtained from the recovered rates as shown in Fig. 7. In order to keep the fluorescence photon rates within the linear regime of the PMT detector, the laser intensity at the sample was reduced to  $\sim 2$  mW using a neutral density filter placed in front of excitation laser and the integration time per sample was increased to 100 ms.

The inset table in Fig. 7(b) shows the mean apparent ethanol volume fractions obtained in a measurement times of 100 ms and, parenthetically, the label volume percent ethanol for each sample. As can be seen, our predicted volume percentages of ethanol very nearly match the label percentage. However, the variance of the fluorescent “golden” tequila measurements is much greater than that of the “silver” tequila measurements due to the increased shot noise resulting from the fluorescence background. In spite of this, we note that by using OB-CD, we can accurately predict the volume percentage of ethanol in tequila samples even when the integrated intensity of the fluorescence of the sample is 20 times larger than the integrated Raman signal. Since the signal-to-noise of such measurements is typically limited by photon (shot) noise rather than read noise, the total time required to obtain a given precision depends on the available laser power and thus is expected to be comparable to that obtained using full spectral measurements (under otherwise identical conditions).

#### 4.3.3. Fluorescent plastic film photobleaching

Here we show an imaging application of our OB-CD fluorescence modeling technique to demonstrate that the recovery of Raman scattering rates is unaffected by fluorescence pho-

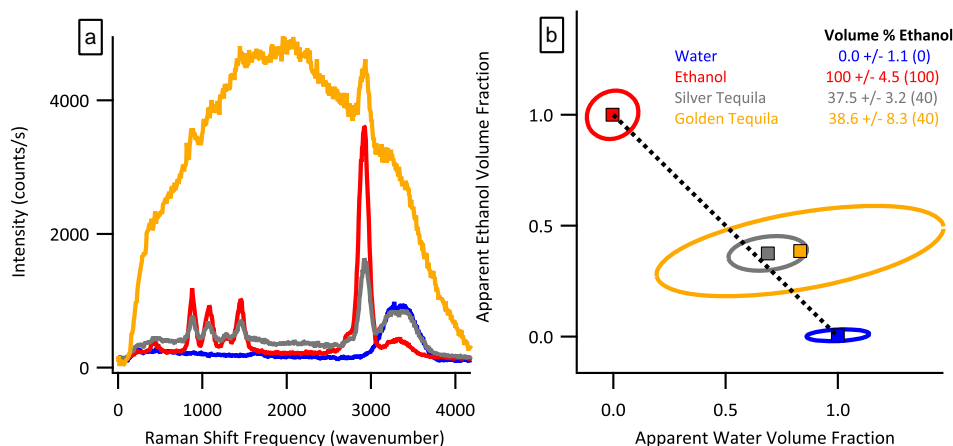


Fig. 7. (a) Spectra of water (blue), ethanol (red), Arandas brand silver tequila (gray), and Casamigos brand golden tequila (dark yellow) measured on the 514 nm OB-CD system (b) Apparent volume fractions of water (blue), ethanol (red), silver tequila (gray), and golden tequila (dark yellow) are compared with the nominal volume fractions (as obtained from the label on the tequila bottles). Each chemical was sampled 1,000 times at 100 ms per OB-CD measurement. Ellipses correspond to the 95% confidence interval of the recovered rates for each sample. Large squares with black borders represent the mean of each sample. The dashed line corresponds to line with slope  $-1$ . Inset table reports the mean apparent volume fraction of ethanol (plus/minus 1 standard deviation) for each sample and then, parenthetically, the label ethanol volume percentage for each sample.

photobleaching. For this purpose, we used a plastic film sample consisting of a clear overhead transparency of  $\sim 1.7$  mm thickness. This sample was chosen as it was found to contain both Raman and fluorescence signals when excited at 785 nm and the fluorescence could be photobleached by exposure to the excitation laser. Additionally, the film exhibited fluorescence with an integrated intensity 10 times that of the integrated intensity Raman features.

While photobleaching decreased the fluorescence background intensity of the sample by  $\sim 50\%$ , the remaining fluorescence could not readily be further photobleached. Thus, for OB-CD training purposes, we generated a Raman spectrum of the plastic by manually performing a polynomial background subtraction from a spectrum of photobleached plastic in the  $328 \text{ nm}^{-1}$  to  $2057 \text{ nm}^{-1}$  region. More specifically, the polynomial subtraction was performed using the “backcor” MATLAB algorithm (Vincent Mazet, 2010), using an Asymmetric Huber cost function, a threshold of 0.01, and four third-degree Bernstein polynomials as a basis. The resulting background subtracted Raman spectrum, as well as the spectra of the overhead transparency before and after photobleaching are shown in Fig. 8.

Next, OB-CD filters were calculated by training on the fluorescence-subtracted Raman features of the plastic film, the spectral component arising from the NIR objective, and the four third-degree Bernstein polynomials (and the resulting training spectra are provided in supplementary material). Note that unlike the filters constructed for previous examples, no components were considered nuisance spectra when calculating OB-CD filters, as we wanted to accurately estimate the intensity of the fluorescence before and after photobleaching. Using these filters, a  $200 \text{ pixel} \times 200 \text{ pixel}$  region of the plastic film (approximately  $1 \text{ mm} \times 1 \text{ mm}$ ) was imaged with an integration time of 10 ms per pixel. Once this image was collected, two lines were photobleached in the transparency to form a photobleached “+” pattern near the center of the field of view (as shown in Fig. 10). Each line, consisting of 50 pixels, was photobleached for



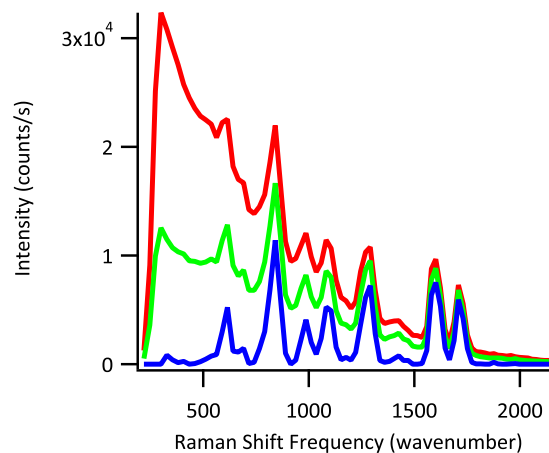


Fig. 8. The measured spectra of a cellulose acetate overhead transparency are plotted before photobleaching (red) and after 20 minutes of photobleaching (green). The output of the polynomial baseline subtraction is also plotted (blue).

10 minutes by scanning the laser repeatedly over the “+” pattern at a rate of 1 second per pixel. After photobleaching, the same field of view was reimaged and OB-CD was used to recover the Raman and fluorescence rates. Images were generated using the recovered rates using a method we have previously described [9]. The fluorescence intensity in this image, was determined from the sum of the recovered rates for all four Bernstein polynomials. Note that several spots on the plastic film were highly fluorescent (likely due to a fluorescent impurity, or dust particle, in the film), with counts well outside the linear region of the PMT and a fluorescence background intensity much greater than 20 times the average Raman signal. These pixels also had unusually high recovered apparent Raman rates, which we attributed to model error. These points were removed from the image, as indicated by black dots in the images shown in Fig. 9.

The upper two panels in Fig. 9 show the apparent recovered Raman rates before (left) and after (right) photobleaching, while the lower two panels show the corresponding apparent recovered fluorescence rates. Note that there is no evidence of a “+” pattern in the upper right panel; this indicates that photobleaching did not alter the apparent recovered Raman rates obtained from the film. There was a small ( $\sim 10\%$ ) decrease in average fluorescence intensity after photobleaching. We attributed this to the photobleaching that occurred while scanning the laser over the entire region during the OB-CD measurement. The similarity of the two upper images in Fig. 9 clearly demonstrates that we are able to obtain high-speed Raman intensity measurements in the presence of a fluorescent background with variable intensity.

## 5. Conclusions

We have demonstrated an OB-CD fluorescence mitigation strategy that can be used to accurately recover Raman rates from samples with moderate fluorescence intensity (that is, up to 20 times more intense than that of the integrated Raman signal). These results were achieved by quantifying fluorescence using OB-CD filters trained on cubic Bernstein polynomials. We have validated this strategy using both white light as a fluorescence mimic, as well as using fluorescent liquid and solid samples. Thus, the present results demonstrate the feasibility of fast (sub-second) OB-CD based Raman classification and quantitation of moderately fluorescent

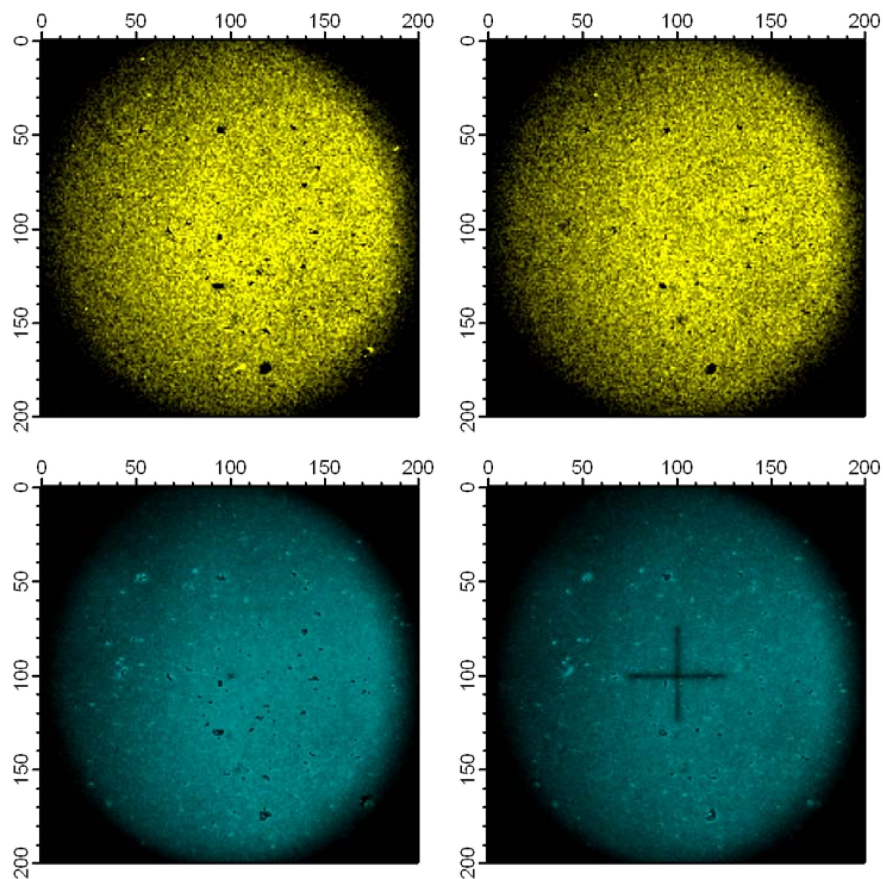


Fig. 9. Images showing the recovered Raman (yellow) and fluorescence (cyan) rates of a cellulose acetate overhead transparency before (images on the left) and after (images on the right) photobleaching a “+” pattern into the center of the imaged area. All images were collected with an integration time of 10 ms per pixel. The circular nature of these images arises from the field of view of the objective, as the images were obtained by raster-scanning the angle of the laser as it enters the back of the objective (while remaining centered in the objective).

samples. This approach can be extended to systems with a fluorescence/Raman intensity ratios greater than 20:1, but would likely require turning down the laser intensity (to avoid detector saturation) and using much longer integration times. Thus, the presented OB-CD strategy is expected to be most useful in applications requiring fast analysis of liquid and solid samples whose fluorescence does not overwhelm the underlying Raman chemical fingerprints. This is consistent with previous results [10], which indicated that the trade-off between higher read-noise and higher spectral information content of full-spectral CCD measurements relative to the OB-CD detection strategy would indicate that OB-CD is most advantageous (relative to CCD measurements) in fast (low-signal) applications that inaccessible to CCD-based measurements.

## 6. Appendix

Fig. 10 shows the training spectra and calculated OB-CD filters for several of the samples in this paper.

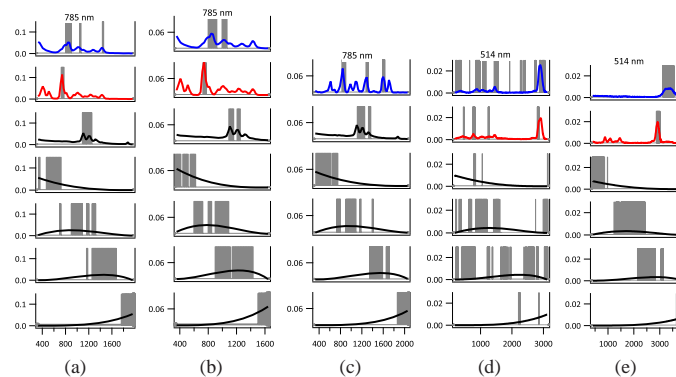


Fig. 10. Area-normalized spectra plotted against Raman shift frequency (wavenumber) shown here pertain to results obtained using a 785 nm and 514 nm excitation laser as noted above each subfigure. Note that the 785 nm spectra were measured with a resolution of  $30 \text{ cm}^{-1}$  and the 514 nm spectra were measured with a resolution of  $12 \text{ cm}^{-1}$ . Unless otherwise noted, the spectral component arising from the NIR objective and the Bernstein polynomials were always considered nuisance spectra. (a) Spectra and the resulting OB-CD filters for (in order from top down): n-hexane, methylcyclohexane, the spectral component arising from the NIR objective and the four degree-three Bernstein polynomials. The fraction of the total measurement time that each filter was collecting was 0.385, 0.219, 0.055, 0.034, 0.268, 0.031, and 0.009, respectively. (b) Spectra and the resulting OB-CD filters for (in order from top down): n-hexane, methylcyclohexane, and the four degree-three Bernstein polynomials. The fraction of the total measurement time that each filter was collecting was 0.248, 0.388, 0.010, 0.032, 0.277 and 0.045, respectively. (c) Spectra and the resulting OB-CD filters for (in order from top down): aniline, toluene, spectral component arising from the NIR objective, and the four degree-three Bernstein polynomials. The fraction of the total measurement time that each filter was collecting was 0.484, 0.265, 0.025, 0.003, 0.154, 0.059 and 0.010, respectively. (d) Spectra and the resulting OB-CD filters for (in order from top down): ethanol, water, and the four degree-three Bernstein polynomials. The fraction of the total measurement time that each filter was collecting was 0.276, 0.261, 0.027, 0.116, 0.232, and 0.088, respectively. (e) Spectra and the resulting OB-CD filters for (in order from top down): Raman features of the plastic film, the spectral component arising from the NIR objective, and the four degree-three Bernstein polynomials. The fraction of the total measurement time that each filter was collecting was 0.099, 0.210, 0.099, 0.271, 0.212, and 0.109, respectively. No components were considered nuisance spectra, as we wanted to accurately estimate the intensity of the fluorescence before and after photobleaching.

### Acknowledgments

The work was supported in part by the Office of Naval Research (Contract N0001413-0394 to DBA, GTB, and BJL) and the Simons Foundation (Award #209418 to BJL).