

**DYNAMIC CHEMICAL IMAGING AND ANALYSIS WITHIN  
BIOLOGICALLY ACTIVE MATERIALS**

by

**Alex M. Sherman**

**A Dissertation**

*Submitted to the Faculty of Purdue University*

*In Partial Fulfillment of the Requirements for the degree of*

**Doctor of Philosophy**



Department of Chemistry

West Lafayette, Indiana

May 2021

**THE PURDUE UNIVERSITY GRADUATE SCHOOL**  
**STATEMENT OF COMMITTEE APPROVAL**

**Dr. Garth Simpson, Chair**

Department of Chemistry

**Dr. Sabre Kais**

Department of Chemistry

**Dr. Julia Laskin**

Department of Chemistry

**Dr. Chi Zhang**

Department of Chemistry

**Approved by:**

Dr. Christine Hrycyna

*Dedicated to my family who has helped me throughout my academic career*

## ACKNOWLEDGMENTS

I would like to acknowledge and thank those who had helped me during my graduate career. Garth, thank you for helping me improve everyday as an independent thinker and a better more well-rounded scientist. Thank you for creating a lab environment that fosters quality and exciting science, as well as creativity and innovation. Stanley Chan, thank you for helping to broaden my interests in the world of computer science through our collaboration together. Also, thank you to those I had collaborated with in the pharmaceutical industry, especially Paul Stroud. My time collaborating with industry has helped to improve my scientific abilities as well as motivate me into choosing a career in industry.

Thank you to all of my friends during my time at Purdue University. Hilary, thank you for being a friend that kept me grounded all of these years. Helping me through the hardest times of graduate school, without your friendship and support I could not have made it through graduate school. Scott, thank you for being a person I could always talk to in lab and vent about my feelings. While there were times you may not have understood me, like when we would talk in Spanish, you always let me get my emotions out. Kevin, thank you for being not just a great friend, but also for being like family to me. I will always remember fondly all of the countless dinners, trips, and adventures we had together. To the rest of the Simpson lab, I thank you as well.

And of course, I would like to thank my parents who helped me through all of my time in higher education. Though, we had lived on opposite sides of the country from each other, I never felt far from home. Whether I had a new issue that I needed help with, or I was just feeling lonely, a simple phone call home would always make me feel safe and happy.

## TABLE OF CONTENTS

TABLE OF CONTENTS.....	5
LIST OF TABLES.....	9
LIST OF FIGURES .....	10
LIST OF ABBREVIATIONS.....	13
ABSTRACT.....	15
CHAPTER 1. INTRODUCTION .....	17
1.1 Pharmaceutical and therapeutic product development .....	17
1.2 Measurement of <i>in vitro</i> and <i>in vivo</i> properties .....	18
1.3 Advancements in optical instrumentation.....	19
1.4 Data analysis and security.....	21
1.5 Dissertation overview .....	22
1.6 References.....	24
CHAPTER 2. STOCHASTIC DIFFERENTIAL SCANNING CALORIMETRY .....	27
2.1 Introduction.....	27
2.2 Methods.....	29
2.2.1 SHG microscopy.....	29
2.2.2 DSC measurements.....	30
2.2.3 Data analysis .....	31
2.2.4 Sample preparation .....	31
2.3 Results and discussion .....	31
2.3.1 SDSC of urea crystals.....	31
2.3.2 IRF determination with SDSC.....	34
2.3.3 SDSC of trehalose dihydrate crystals .....	35
2.4 Conclusion .....	40
2.5 References.....	41
CHAPTER 3. FOURIER TRANSFORM FLUORESCENCE RECOVERY AFTER PHOTOBLEACHING WITH DIFFRACTIVE OPTICAL ELEMENTS .....	46
3.1 Introduction.....	46
3.2 Theory .....	49

3.2.1	Normal diffusion.....	49
3.2.2	FT diffusion .....	50
3.2.3	Signal power .....	52
3.2.4	Simultaneous conventional FRAP and FT-FRAP .....	53
3.3	Experimental methods .....	53
3.3.1	Sample preparation .....	53
3.3.2	Dot – array FRAP .....	54
3.3.3	Data analysis .....	55
3.4	Results and discussion .....	56
3.4.1	FT-FRAP of fluorescently labelled IgG in HA/collagen.....	56
3.4.2	Simultaneous conventional FRAP of IgG in HA .....	58
3.5	Conclusion .....	59
3.6	References .....	59
CHAPTER 4. POLARIZATION DEPENDENT DISCRIMINATION OF CRYSTALLINE AND AGGREGATE PROTEIN.....		62
4.1	Introduction.....	62
4.2	Theoretical framework.....	65
4.2.1	Polarization dependence of TPE-UVF .....	65
4.2.2	Circular/Linear dichroism.....	66
4.3	Methods.....	67
4.3.1	Sample preparation .....	67
4.3.2	Simulations .....	67
4.3.3	Data acquisition .....	68
4.3.4	Data analysis .....	69
4.4	Results.....	69
4.5	Conclusion .....	71
4.6	References .....	71
CHAPTER 5. CALIBRATION-FREE IMAGE ANALYSIS FOR QUANTIFICATION OF CRYSTALLINITY OF ROD-LIKE PARTICLES.....		73
5.1	Introduction.....	73
5.2	Theoretical Framework.....	76

5.2.1	High Aspect-Ratio Particle (HARP) Analysis.....	76
5.2.2	Auto-calibration.....	80
5.3	Experimental and computational methods.....	81
5.3.1	Image processing .....	81
5.3.2	Simulations .....	83
5.3.3	Experimental methods .....	83
5.4	Results.....	84
5.5	Experimental implementation.....	87
5.6	Future work.....	90
5.7	Conclusion .....	90
5.8	References.....	91
CHAPTER 6. ADVERSARIAL ILLUMINATION .....		94
6.1	Introduction.....	94
6.2	Methods.....	96
6.2.1	Acquisition and illumination .....	96
6.2.2	Image classification .....	97
6.3	Results.....	97
6.3.1	Adversarial illumination of 2D object .....	97
6.3.2	Adversarial illumination of 3D objects.....	99
6.4	Conclusion .....	99
6.5	References.....	100
CHAPTER 7. ADVERSARIAL SPECTROSCOPY .....		101
7.1	Introduction.....	101
7.2	Theory .....	104
7.3	Methods.....	105
7.3.1	Raman spectra.....	105
7.3.2	Linear discriminate analysis .....	106
7.3.3	Decoy generation .....	106
7.4	Results.....	107
7.5	Conclusion .....	109
7.6	References.....	109

VITA.....	111
PUBLICATIONS.....	112



## LIST OF TABLES

Table 4.1. Linear and circular responses for polarization studies of TPE-UVF for solutions of tryptophan. ....	70
--	----

## LIST OF FIGURES

Figure 1.1. Drug discovery pipeline. The pipeline is broken down into four major phases: drug discovery, pre-clinical tests, clinical tests, and marketing. Thousands of potential drug candidates are tested in early phases and are eventually eliminated to only a couple of final drugs. Pre-clinical testing focuses on tests such as: drug stability, phase behavior, bioavailability, and molecular mobility.....	17
Figure 1.2. Energy diagrams for commonly used nonlinear optical techniques: A) second harmonic generation (SHG), and B) two-photon excitation ultraviolet fluorescence (TPE-UVF) where $\omega$ is the excitation frequency, $2\omega$ is the doubled-frequency for SHG, and $\omega_{em}$ is the frequency emitted from TPE-UVF. ....	20
Figure 2.1. Schematic of the integrated SHG-DSC microscope. A pulsed laser beam (800 nm, 80 MHz) was scanned with a galvo-resonant scan pair, which was 4f-coupled to the back of an objective and focused onto a DSC sample pan. The SHG signal was collected in the epi-direction, isolated with a dichroic mirror and detected with a photomultiplier tube. ....	30
Figure 2.2. A) SDSC analysis of urea by combined SHG microscopy and DSC. The solid black line is the DSC trace of the melting of urea with endothermic direction being down (left-axis). The vertical colored lines show the temperatures in which the normalized SHG intensity of individual urea particles (acquired concurrently with DSC measurements) drops below a value of 0.1. DSC and SHG measurements both indicate a structural transformation in the same temperature range. The inset shows a zoom-in of the temperature range 135.5–136.5 °C where the color lines correspond to the normalized SHG intensity of individual urea particles. B) SHG microscope images of urea crystals during SDSC analysis. SHG microscopy reveals the stochastic nature of the melting transition and enables analysis on a per-particle basis. ....	33
Figure 2.3. Comparison of the experimental DSC trace and the DSC trace generated from melting events. The best-fit double-exponential IRF (yellow) is convolved with impulsive melting events (blue) of single-particles observed in SHG images to generate the purple trace, which is in good agreement with the experimental DSC data (red). The time constant and offset of double-exponential IRF were optimized to minimize difference between experimental and generated DSC traces. Best-fit values of $1.161 \pm 0.019$ s, $0.435 \pm 0.013$ s, and $3.088 \pm 0.013$ s were computed for the double-exponential time constants and the offset respectively.....	35
Figure 2.4. SDSC analysis of trehalose by SHG microscopy. The colored lines show the percentage of pixels in a trehalose particle above an SHG intensity threshold overlaid with the black line of the DSC trace with the endothermic direction being down. SDSC images were summed for the temperature range of 140 – 220 °C and the logarithm of the summed image was cropped to the measured region of each particle (each cropped image points to the representative SHG trace). 36	
Figure 4.1. Schematic depicting bleaching patterns with A) a comb pattern, which Fourier transforms to B) a one-dimensional set of equally spaced puncta, and C) a dot-array pattern, which Fourier transforms to D) a two-dimensional set of equally spaced puncta.....	51

Figure 4.2. Instrument schematic of a commercial FRAP system that was modified for FT-FRAP with DOEs. Modifications are indicated by red dashed boxes. A fiber-coupled laser replaced the original photobleaching laser for higher bleaching power, an automated shutter was added to control bleach times and a DOE was added after the first dichroic mirror to produce a dot-array photobleach pattern. HWP = half-wave plate, DCM = dichroic mirror, SPF = short-pass filter, TPEF = two-photon excited fluorescence, PMT = photomultiplier tube, DOE = diffractive optical element, CCD = charge-coupled device camera, LED = broadband green light emitting diode. 55

Figure 4.3. FT-FRAP with dot array bleach of IgG dissolved in PBS in HA matrix. A) First image after photobleaching with the dot-array pattern. B) 2D Fourier transform of (A). Blue and red squares indicate peaks that are equidistant from the center peak C) Fluorescence recovery of pooled peaks with best-fit curves for FT-FRAP. The best-fit value for the diffusion coefficient was  $26.5 \pm 0.07 \mu\text{m}^2/\text{sec}$ . D) Fluorescence recovery using conventional FRAP, and the best-fit value for the diffusion coefficient was  $23.21 \pm 0.09 \mu\text{m}^2/\text{sec}$ . The reported uncertainty is the standard deviation of the fit for both fits. 57

Figure 4.1. Polarization modulation pattern of light as a function of position on the  $\mu\text{RA}$  from Thorlabs. 64

Figure 4.2. Skeletal (bond-line) structure of the indole ring, and the electronically optimized structure. 67

Figure 4.3. Instrument schematic for the SONICC instrument. Images were collected in the epi direction a photomultiplier tube (PMT). The  $\mu\text{RA}$  and quarter-wave plate (QWP) were placed in the rear-conjugate plane of the beam path. 69

Figure 4.4. Polarization modulated TPE-UVF images with and without the  $\mu\text{RA}$  and QWP placed into the beam path. Images were collected for solutions of tryptophan as well as crystalline tryptophan and lyophilized lysozyme. 70

Figure 5.1. Graphical representation of a cuboidal particle's projection onto a 2D plane. Inset shows the particle's SHG projection. 77

Figure 5.2. Workflow for determination of aspect ratio from multiple FOVs. For a specific FOV, a conventional particle-counting algorithm finds all particles as well as their respective feret diameter and minimum feret diameter. The particle with the largest length is then used to calculate it's observed aspect ratio. This is then averaged over multiple FOVs. 80

Figure 5.3. A) Rod-like particle simulated SHG images at varying levels of crystallinity, B) experimentally acquired SHG images of ritonavir. Scale bar 300  $\mu\text{m}$ . 83

Figure 5.4. Recovered percent crystallinity for simulated data using the HARP algorithm (blue squares), and the spheroidal limit (red triangles). The ground truth percent crystallinity for each simulated image is shown as the solid black line. 86

Figure 5.5. Time evolution of percent crystallinity in 15% ritonavir HME, rotovap, and SDD ASDs stressed in situ temperature and relative humidity-controlled platform at 50OC/75% RH. HARP algorithm is represented by the blue squares, the spheroidal limit by red triangles, and integrated intensity by the purple circles. 88

Figure 6.1. A demonstration of misclassification using the GoogLeNet and image of a panda taken from ImageNet. By addition of an imperceptible attack pattern, intentional misclassification of the image to that of a gibbon is shown. .... 95

Figure 6.2. Diagram of the acquisition and illumination set-up. Images of a target object were collected using a Pi Camera attached to a Raspberry Pi. After calculation of the desired attack pattern, the projector also connected to the Raspberry Pi illuminated pattern over the same object. The same camera, or a second camera attached to another Raspberry Pi acquires a second image. .... 97

Figure 6.3. Adversarial illumination results for a 2D object (a strawberry) printed on a poster. Initial classification of the initial image collected using the Pi Camera (A) classified the poster as either a strawberry or a pan. The calculated illumination pattern is then shown (B), and the imaged acquired after illumination of this pattern on the poster (C). After acquisition of a second image, the new image was classified as a jigsaw puzzle. .... 98

Figure 7.1. Conceptual illustration of adversarial attack on spectral data. A) and B) represent the projection of spectra of two different classes onto LDA space. The thin black arrows represent major features of a given class that is used for class separation in LDA, and the thin red arrows represent the natural noise background for a given spectrum. The point in which spectrum is located is represented by the thick black arrow, and each spectrum can be seen to be located on different sides of the decision boundary allowing for classification. C) Demonstrates intentional perturbation of the spectral projects with patterned noise that moves the projection point past the decision boundary for intentional misclassification. .... 103

Figure 7.2. Workflow for GALDA. A) Raman spectral for forms I (red), and II (blue) of clopidogrel bisulfate and background (background). B) Projection of Raman spectra onto LDA space. C) Projection of decoy data generated from randomly inputs. D) Projection of decoy data in LDA space after addition of patterned noise. E) Separation of decoy data after training of LDA with decoy data as a separate class. Steps C – E are repeatable for until desired LDA discrimination of decoy data. .... 105

Figure 7.3. A) Projection of experimental Raman spectra for each class (form I – red, form II – blue, and background – black) in LDA-space. B) Mean spectrum for each class using 84 class measurements (spectra were displaced vertically for improved visualization). .... 107

Figure 7.4. Example of a spectroscopic adversarial attack, in which a Class 3 spectrum is optimally perturbed such that it is misclassified as Class 2. A) Optimal perturbation of the initial class 3 spectrum (black trace) to that of the mean target spectrum (light blue trace) of class 2 is added to the initial spectrum to get the corresponding perturbed spectrum (dark blue trace). B) Shows the applied perturbation which simply appears as noise..... 108

## LIST OF ABBREVIATIONS

API	active pharmaceutical ingredient
AR	aspect ratio
ASD	amorphous solid dispersion
BSA	bovine serum albumin
CNN	convolution neural network
DOE	diffractive optical element
DOF	depth-of-field
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
FIJI	FIJI is just ImageJ
FOV	field of view
FRAP	fluorescence recovery after photobleaching
FT	Fourier transform
GALDA	generative adversarial linear discriminant analysis
GAMESS	general atomic and molecular electronic structure system
GAN	generative adversarial network
GC	gas chromatography
HA	hyaluronic acid
HARP	high aspect-ratio particle
HME	hot melt extrusion
HT	high-throughput
IR	infrared
IRF	impulse response function
LC	liquid chromatography
LCP	lipidic cubic phase
LDA	linear discriminate analysis
LOD	limit of detection
mAb	monoclonal antibody
MHz	megahertz
ML	machine learning
MLE	maximum likelihood estimator

μRA	microretarder array
MS	mass spectroscopy
NLO	nonlinear optical
NMR	nuclear magnetic resonance
NN	neural network
PBS	phosphate-buffered saline
PCA	principal component analysis
PMT	photomultiplier tube
pXRD	powder X-ray diffraction
QWP	quarter-wave plate
rotovap	rotary evaporation
RH	relative humidity
RHF	restricted Hartree Fock
SDD	spray dried dispersion
SDSC	stochastic differential scanning calorimetry
SHG	second harmonic generation
SONICC	second order nonlinear imaging of chiral crystals
SNR	signal-to-noise ratio
ssNMR	solid-state nuclear magnetic resonance
TASC	thermal analysis by structural characterization
TDDFT	time-dependent density functional theory
TGA	thermal gravimetric analysis
TPE-UVF	two-photon excited ultraviolet fluorescence
TMR-IgG	tetramethyl rhodamine-labeled immunoglobulin G
XFEL	X-ray free-electron laser

## ABSTRACT

A thorough understanding of pharmaceutical and therapeutic products and materials is important for an improved quality of life. By probing the complex behaviors and properties of these systems, new insights can allow for a better understanding of current treatments, improved design and synthesis of new drug products, and the development of new treatments for various health conditions. Often, the impact of these new insights are limited by current technology and instrumentation and by the methods in which existing data is processed. Additionally, current standards for characterization of pharmaceuticals and therapeutics are time-consuming and can delay the timeline in which these products become available to the consumer. By addressing the limitations in current instrumentation and data science methods, faster and improved characterization is possible.

Development and improvement in optical instrumentation provides potential solutions to the current limitations of characterization methods by conventional instrumentation. Limitations in speed can be addressed through the use of nonlinear optical (NLO) methods, such as second harmonic generation (SHG) and two-photon excited ultraviolet fluorescence (TPE-UVF) microscopy, or by linear methods such as fluorescence recovery after photobleaching (FRAP). For these methods, a high signal-to-noise ratio (SNR) and a nondestructive nature decrease the overall sample size requirements and collection times of these methods. Furthermore, by combination of these optical techniques with other techniques, such as thermal analysis (e.g. differential scanning calorimetry (DSC)), polarization modulation, or patterned illumination, the collection of more complex and higher quality data is possible while retaining the improved speed of these methods. Thus, this modified instrumentation can allow for improved characterization of properties such as stability, structure, and mobility of pharmaceutical and therapeutic products.

With an increase in data quantity and complexity, improvements to existing methods of analysis, as well as development of new data science methods, is essential. Machine learning (ML) architectures and empirically validated models for the analysis of existing data can provide improved quantification. Using the aforementioned optical instrumentation, auto-calibration of data acquired by SHG microscopy is one such method in which quantification of sample crystallinity is enabled by these ML and empirical models. Additionally, ML approaches utilizing generative adversarial networks (GANs) are able to improve on identification of data tampering in

order to retain data security. By use of GANs to tamper with experimentally collected and/or simulated data used in existing spectral classifiers, knowledge of adversarial methods and weakness in spectral classification can be ascertained. Likewise, perturbations in physical illumination can be used to ascertain information on classification of real objects by use of GANs. Use of this knowledge can then be used to prevent further data tampering or by improving identification of data tampering.



## CHAPTER 1. INTRODUCTION

### 1.1 Pharmaceutical and therapeutic product development

The development of new pharmaceutical or therapeutic products is a time-consuming process in which a single drug product can take upwards of 10 years before reaching the consumer.<sup>1</sup> From start to finish of this arduous process (shown in Figure 1.1), a drug must go through design, characterization, and clinical testing before obtaining FDA approval. Once approved, demonstrating that the drug is safe for use, finally production and marketing of the drug can begin. Overall, the operation of all steps in this pipeline requires millions of researchers and health professionals, contributing to this multibillion-dollar industry.<sup>2</sup> For this reason, the ability to more efficiently design, characterize, and produce a drug product can significantly lower the costs and speed of these stages of development.

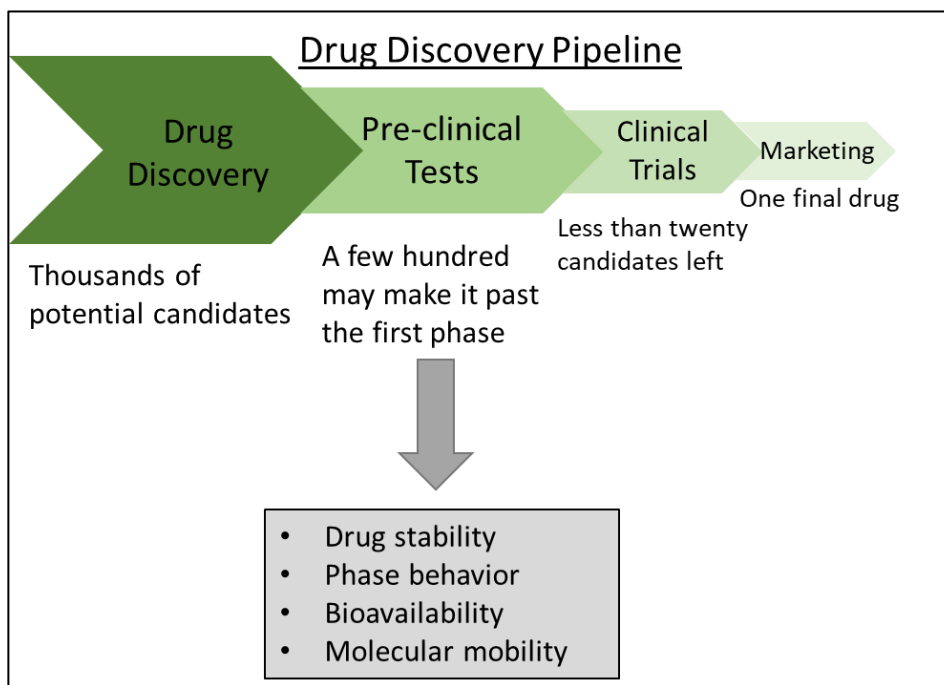


Figure 1.1. Drug discovery pipeline. The pipeline is broken down into four major phases: drug discovery, pre-clinical tests, clinical tests, and marketing. Thousands of potential drug candidates are tested in early phases and are eventually eliminated to only a couple of final drugs. Pre-clinical testing focuses on tests such as: drug stability, phase behavior, bioavailability, and molecular mobility.

Focusing on pre-clinical testing, research during this stage primarily looks at both *in vitro* and *in vivo* characterization of a potential pharmaceutical or therapeutic product. Furthermore, this stage of testing heavily focuses on the use of chemical instrumentation and other analytical techniques in order to measure these desired properties.<sup>3</sup> Product stability and phase behavior are commonly tested *in vitro* properties that use various analytical techniques to track changes in physical form before administration. During the storage of a drug product, if the active form of the drug is unstable or a higher energy form, the molecule can rearrange itself to the thermodynamically favored form. Likewise, variations in storage environment such as fluctuations in temperature, light, and humidity can cause a drug product to be altered into another unstable/metastable form or speed up conversion to the thermodynamically favored form.<sup>4</sup> After rearrangement to this new form, the original functionality of the drug product can be lost and lower the overall drug efficacy. Molecular rearrangement can also cause adverse effects by lowering the apparent solubility or diffusivity and as a result lower the overall bioavailability.<sup>5</sup> Commonly tested *in vivo* properties include stability as well as product mobility. Product mobility looks at the diffusivity and permeability of a drug product in a specific form which gauges the bioavailability. Depending on this measured bioavailability, the overall amount of drug loading and the method in which the drug is administered are altered accordingly.<sup>6-7</sup> Due to the effects *in vitro* and *in vivo* properties can have towards the efficacy and safety of a drug product, preclinical testing of these properties is essential before a drug can move towards clinical testing.

## **1.2 Measurement of *in vitro* and *in vivo* properties**

In order to measure the *in vitro* and *in vivo* properties of a pharmaceutical or therapeutic product, various analytical techniques and instrumentation have become common practice during pre-clinical testing due to the extensive research done to understand the basic principles and fundamental science behind them. As these techniques are well characterized, the data and results given by such methods and instrumentation are generally less contested in the scientific community. Likewise, common practices exist as well for how the data acquired by these techniques are analyzed. Deviations from these common practices though, whether by new and improved instrumentation or by use of new methods of data analysis, often times are reviewed more skeptically. Therefore, research improving the understanding of new methods and techniques can increase the confidence in their use.

For *in vitro* and *in vivo* characterization, chromatography,<sup>8</sup> spectroscopy,<sup>9</sup> and thermal analysis are all examples of commonly used benchtop techniques.<sup>10-11</sup> In the 1900s, extensive research was done on the development and characterization of many of these techniques. In 1900 for example, chromatography was first devised by Mikhail Tsvet,<sup>12</sup> and mass spectroscopy (MS) was devised both in 1918 and 1919.<sup>13</sup> After decades of research on such techniques, the data acquired from methods such as MS, liquid chromatography (LC), or gas chromatography (GC) are considered reliable for many chemical and biological applications. Specifically looking at *in vitro* stability and phase behavior, powder X-ray diffraction (pXRD), infrared (IR) spectroscopy, Raman spectroscopy, nuclear magnetic resonance (NMR), and differential scanning calorimetry (DSC) are commonly utilized techniques.<sup>11</sup> Characterization of *in vivo* properties such as mobility and permeability on the other hand are commonly analyzed with methods such as fluorescence recovery after photobleaching (FRAP) or various spectroscopic methods.<sup>14-15</sup>

While these methods have become common, various limitations for these methods do exist that limit their overall applicability and efficiency. Methods such as DSC, MS, or pXRD are often times highly destructive in nature, which prevents further testing done on the used sample and can also increase the amount of production time and cost to synthesize the needed amount of sample for adequate testing.<sup>11</sup> Thus, the amount of time a potential drug candidate stays in the pre-clinical testing stage, and the total cost of testing is increased. For applications detecting small amounts of crystallinity in a sample such that the active form of a drug in the non-crystalline form, conventionally methods such as pXRD, infrared (IR) or Raman spectroscopy often lack high enough sensitivity for trace detection.<sup>11</sup> Therefore, a sample may need to be tested for a longer time before a measurable amount of crystallinity is present. Additionally, many of these conventional methods can suffer from low signal-to-noise, small sample perturbations, and other adverse effects.<sup>11, 14-15</sup> Thus, improvements to these techniques as well as development of new techniques and instrumentation can significantly lower the amount of time a potential pharmaceutical or therapeutic product remains in the pre-clinical testing stage, lowering the operation costs as well.

### **1.3 Advancements in optical instrumentation**

While common benchtop technologies have various limitations that can affect efficiency and operation cost, advancements in optical instrumentation offer new and improved methods for

pharmaceutical and therapeutic characterization. Nonlinear optical (NLO) microscopy is one such field that offers improvements to characterization as well as introduces new information content not attainable by other methods. In general, NLO techniques provide nondestructive, chemical specific detection, enabling high signal-to-noise ratios (SNR), low limits of detection (LOD), and shorter acquisition times.<sup>16-17</sup> Thus, NLO methods are well suited for *in vitro* measurements to assess drug stability, structural analysis, and other molecular properties.

Second harmonic generation (SHG) is among the simplest NLO process, in which the frequency of light used to illuminate the sample is “doubled” after its interaction with the sample.<sup>17-18</sup> This process of frequency doubling is shown in Figure 1.2, where  $\omega$  is the excitation frequency and  $2\omega$  is the frequency-doubled light. Classically, SHG can be thought of as a “sloshing” of electrons within a molecule in response to a driving optical field. When this optical field is sufficiently high enough (e.g. focusing of an ultrafast laser pulse), the anharmonic potential energy surface can be accessed, which results in frequency-doubling of light. Due to this doubling process and the symmetry requirements that arise from use of this technique, SHG is highly selective for noncentrosymmetric materials, proving for a high SNR and low LOD, but is forbidden for centrosymmetric/higher symmetry materials.<sup>18</sup> Since many pharmaceutical and therapeutic materials are noncentrosymmetric themselves (e.g. protein crystals, small-molecule drugs), this limitation is generally not a hindrance.<sup>19</sup>

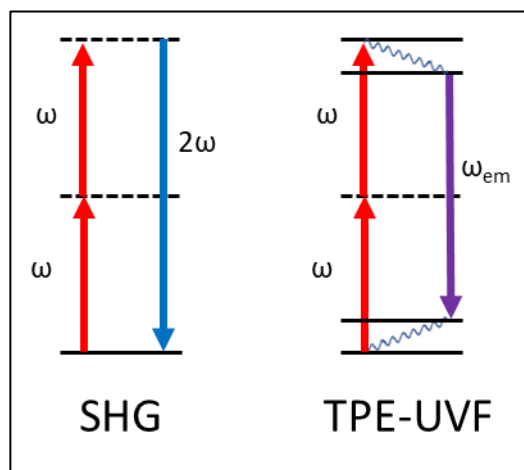


Figure 1.2. Energy diagrams for commonly used nonlinear optical techniques: A) second harmonic generation (SHG), and B) two-photon excitation ultraviolet fluorescence (TPE-UVF) where  $\omega$  is the excitation frequency,  $2\omega$  is the doubled-frequency for SHG, and  $\omega_{em}$  is the frequency emitted from TPE-UVF.

For samples that are centrosymmetric, two-photon excitation ultraviolet fluorescence (TPE-UVF) is an additional NLO technique for sample characterization (Figure 1.2). In general, a sample will produce a TPE-UVF signal if an aromatic residue (most commonly tryptophan) is present.<sup>18</sup> Thus, TPE-UVF is commonly used to investigate proteins, since most proteins contain at least one aromatic residue.<sup>18</sup> Unlike SHG, TPE-UVF lacks information on a samples structure, but can be modified to increase the overall information content attainable by this method. One such modification is polarization modulation (for TPE-UVF and SHG measurements) which allows for structural information to be measured for various of samples such as protein crystals. Therefore, TPE-UVF is a promising technique for *in vitro* characterization of biopharmaceuticals and biotherapeutics.

For *in vivo* characterization, FRAP is an optical technique that can measure molecular mobility and bioavailability, but has its own set of limitations that can be overcome with improvements to the conventional method. In general, conventional FRAP is assessed by photobleaching a fluorescent sample with an intense beam of light. Subsequent measurements of the fluorescent recovery of the bleached area and fitting of these data to the expected trend allows for recovery of molecular diffusion in a sample.<sup>14-15</sup> The value recovered for diffusion can then be used as a gauge for overall mobility of a pharmaceutical or therapeutic product and can be used to assess the overall bioavailability. Unfortunately, conventional FRAP is limited by its SNR and is highly sensitive to variations in the expected recovery behavior.<sup>14-15</sup> Thus, improvements can be made to this optical technique, such as Fourier analysis that will be discussed in later chapters and allow for improved characterization of *in vivo* properties.

## 1.4 Data analysis and security

With advancements in instrumentation and the development of new measurement techniques, improvements to data analysis methods are important as well. Many forms of data analysis rely on building various analytical or empirical models with a basis in fundamental physics as well as fundamental statistics.<sup>20</sup> While the models function well, extension of a model built for a specific system may have difficulty being used for another problem. Machine learning (ML) is one such developing field of data science that enables improvements to current analysis techniques and can easily be transitioned to other problems as well. In general, ML architectures work by building a statistical model that reflects a specific set of training data. Since the only

human intervention needed in these methods is the selection of data, ML techniques can utilize high dimensional data or subtle data correlations that may be missed by human identification.<sup>21</sup> Furthermore, with current advances in the field of ML, these models can accomplish tasks such as image classification better than humans and empirical/analytical models. Thus, with access to sufficient training, the trained model can be used for problems such as sample quantification important for the assessment of *in vitro* stability or improved curve fitting important for characterization of *in vivo* mobility.

Contrary to this, ML architectures can additionally be used to purposefully “trick” image and spectral classifiers. Specifically looking at the use of generative adversarial networks (GANs), improvements to these neural networks (NNs) have allowed for creation of generative data that can purposefully be classified a specific way by an image or spectral classifier. While work towards being able to trick such classifiers seems nefarious in nature, the resulting information from such attacks can be used to improve overall data security. One such method for improving current classification architectures is training future classifiers with this generative data. With sufficient amounts of this “decoy” data, the improved classifier can then recognize such attacks with minimal human supervision. Thus, the overall security and trust in given data is improved by the use of these ML models.

## 1.5 Dissertation overview

In this dissertation, improvements to existing instrumentation and the development of new techniques for *in vitro* and *in vivo* characterization of pharmaceutical and therapeutic products will be addressed. Additionally, the challenges in data science for the analysis of the data from these new and improved techniques will be discussed, including the use of ML architectures for improved data security.

Chapter 2 details the integration of NLO microscopy with thermal analysis techniques to monitor phase transitions of pharmaceutical products on a per-particle basis. An optically transparent DSC instrument was integrated into an existing home-built second harmonic generation (SHG) microscope. This multimodal approach enabled simultaneous single-particle analysis for distinguishing the impact of crystal size and shape on phase transitions in a single experiment. Proof-of-concept experiments for this combined analysis were done on a simple organic molecule, urea, which contained only a single melting transition. The more complex phase

transitions of trehalose dihydrate, which can transition to a myriad of solid forms during dehydrations, were then investigated by this method. From this combined thermal/microscopy analysis, the presence of a “cold phase transformation” (i.e. phase transformation to the SHG-active crystalline form occurring immediately prior to melting) was observed.

Chapter 3 discusses modifications to a commercially available FRAP instrument that enables improvements to conventional methods of FRAP. Through integration of a diffractive optical element (DOE) into the existing beam path, patterned bleaching was enabled. Analysis of the data collected using the patterned bleaching in the spatial frequency domain allowed for improvements to SNR as well as improvements in confidence in the recovered diffusion coefficient. Additionally, patterned bleaching using a dot-matrix pattern enabled simultaneous conventional FRAP and Fourier transform FRAP (FT-FRAP). Likewise, investigation of anisotropic samples was possible using the dot-matrix bleach by measurement of the diffusion tensor and coordinate transformations.

Chapter 4 proposes a new polarization-based method for the discrimination of crystalline protein from aggregate protein. Through integration of a micro-retarder array ( $\mu$ RA) and a quarter-wave plate (QWP) into an existing TPE-UVF microscope, the polarization state of the incident light was modulated across the entire field-of-view (FOV). Fourier analysis of a given image after modulation of the polarization states was used to calculate the ratio between how a sample interacts with linearly polarization light and circularly polarized light. This analysis was demonstrated for a uniform solution of tryptophan and was compared with simulated results for the indole ring of tryptophan. The method is proposed for the use on samples of crystalline protein to demonstrate the discrimination capabilities.

Chapter 5 focuses on the development of an empirical model capable of quantifying the amount of crystallinity present in sample from an SHG image. To this end, a simulation package was developed in MATLAB capable of producing simulated SHG images for rod-like and spherical particles. The simulated data was then used to develop a new model, high aspect-ratio particle (HARP) analysis, that relies on calculation of particle widths and the average aspect ratio in a given FOV. This model was then applied to experimentally collected SHG images of ritonavir processed by hot melt extrusion (HME), rotary evaporation (rotovap), and spray dried dispersion (SDD) for quantification of sample crystallinity during varying storage times.

Chapter 6 details the work done on the intentional misclassification of objects by means of adversarial external illumination, as a demonstration of a concept extensible to chemical analysis

of high-dimensional data sets. For this work, a simple image classifier was trained that was capable of recognizing a small subset of real objects from an image. By use of a simple camera, images of a specific trained object were collected and classified. A pattern was then computationally generated for a specific image that, when added to the image data, causes purposeful misclassification while this perturbation remains unnoticeable to the human eye. Using a projector, the pattern was then externally illuminated over the sample, and imaged once more. Said image was then ran through the trained classifier to demonstrate misclassification by this external illumination. By identifying the patterns that induce misclassification through adversarial illumination, improved classifiers less prone to “overfitting” artefacts can be produced.

Chapter 7 describes follow-up work using generative adversarial linear analysis strategies to improve chemical analyses of high-dimensional spectral data, following a similar idea to the work presented in Chapter 6. Using an experimentally collected set of Raman data for clopidogrel bisulfate, classification of the two possible forms was done using linear discriminate analysis (LDA). The same classified data were then used to design a generative adversarial linear discriminate analysis (GALDA) approach in order to produce attacked spectra such that the resulting spectra appear as a specific form but are spectrally misclassified as another. Generation of this data was shown to greatly reduce overfitting artefacts associated with conventional LDA for dimension reduction of high-dimensional data.

## 1.6 References

1. Chow, S. C.; Pong, A., An overview of the regulatory approval process in drug development. *Drug Inf. J.* **1998**, *32*, 1175S-1185S.
2. Mullin, R., Drug development costs about \$1.7 billion. *Chem. Eng. News* **2003**, *81* (50), 8-8.
3. Clemens, G. R.; Wasinska-Kempka, G.; Hildebrand, H.; Sperzel, M.; Hartnagel, R. E.; LeBel, E.; Richter, A., Preclinical in vitro and in vivo testing comparing toxicities of camptothecin glycoconjugates with camptothecin, topotecan and irinotecan. *Clin. Cancer Res.* **1999**, *5*, 3863S-3863S.
4. Song, Z. T.; Sarkar, S.; Vogt, A. D.; Danzer, G. D.; Smith, C. J.; Gualtier, E. J.; Simpson, G. J., Kinetic Modeling of Accelerated Stability Testing Enabled by Second Harmonic Generation Microscopy. *Anal. Chem.* **2018**, *90* (7), 4406-4413.



5. von Moos, R.; Stolz, R.; Cerny, T.; Gillessen, S., Thalidomide: from tragedy to promise. *Swiss Med. Wkly.* **2003**, *133* (5-6), 77-87.
6. Jindal, A. B., The effect of particle shape on cellular interaction and drug delivery applications of micro-and nanoparticles. *Int. J. Pharm.* **2017**, *532* (1), 450-465.
7. Jahromi, L. P.; Ghazali, M.; Ashrafi, H.; Azadi, A., A comparison of models for the analysis of the kinetics of drug release from PLGA-based nanoparticles. *Heliyon* **2020**, *6* (2), 9.
8. Neupane, R.; Bergquist, J., Analytical techniques for the characterization of Antibody Drug Conjugates: Challenges and prospects. *Eur. J. Mass Spectrom.* **2017**, *23* (6), 417-426.
9. Behnke, G.; Quinones, R.; Iuliucci, R., Using inductively coupled plasma atomic spectroscopy and solid-state NMR spectroscopy to increase characterization speed of polymorphic drugs. *Abstr. Pap. Am. Chem. Soc.* **2018**, *255*, 1.
10. Khattab, F. I.; Hassan, N. Y. M.; Amer, M. M., Thermal-analysis of pharmaceutical compounds 3. Characterization of sulfonamides by thermal-analysis. *J. Therm. Anal. Calorim.* **1981**, *22* (1), 41-51.
11. Sherman, A. M.; Geiger, A. C.; Smith, C. J.; Taylor, L. S.; Hinds, J.; Stroud, P. A.; Simpson, G. J., Stochastic Differential Scanning Calorimetry by Nonlinear Optical Microscopy. *Anal. Chem.* **2020**, *92* (1), 1171-1178.
12. Sakodinskii, K. I., Discovery of chromatography by TSVET, M.S. *J. Anal. Chem.* **1993**, *48* (8), 897-904.
13. Sharma, K. S., Mass spectrometry-The early years. *Int. J. Mass Spectrom.* **2013**, *349*, 3-8.
14. Geiger, A. C.; Smith, C. J.; Simpson, G. J., Multi-photon excited Fourier-transform fluorescence recovery after photobleaching (FT-FRAP) with patterned illumination. In *Multiphoton Microscopy in the Biomedical Sciences Xx*, Periasamy, A.; So, P. T.; Koenig, K., Eds. Spie-Int Soc Optical Engineering: Bellingham, 2020; Vol. 11244.
15. Geiger, A. C.; Smith, C. J.; Takanti, N.; Harmon, D. M.; Carlsen, M. S.; Simpson, G. J., Anomalous Diffusion Characterization by Fourier Transform-FRAP with Patterned Illumination. *Biophys. J.* **2020**, *119* (4), 737-748.
16. Hanson, K. M.; Bardeen, C. J., Application of Nonlinear Optical Microscopy for Imaging Skin. *Photochem. Photobiol.* **2009**, *85* (1), 33-44.
17. Schmitt, P. D., Recent Advances in Nonlinear Optical Analyses of Pharmaceutical Materials in the Solid State. *Mol. Pharm.* **2017**, *14* (3), 555-565.

18. Simpson, G. J., *Nonlinear Optical Polarization Analysis in Chemistry and Biology*. Cambridge Univ Press: Cambridge, 2017; p 1-485.
19. Nunez, M. C.; Garcia-Rubino, M. E.; Conejo-Garcia, A.; Cruz-Lopez, O.; Kimatrai, M.; Gallo, M. A.; Espinosa, A.; Campos, J. M., Homochiral Drugs: A Demanding Tendency of the Pharmaceutical Industry. *Curr. Med. Chem.* **2009**, *16* (16), 2064-2074.
20. Hao, Z. X.; Qu, X. L.; Liu, Y. C., *Algorithm of Support Vector Machines Based on Statistics Learning Theory*. Sci Res Publ, Inc-Srp: Irvin, 2009; p 303-+.
21. Ranjan, R.; Sankaranarayanan, S.; Bansal, A.; Bodla, N.; Chen, J. C.; Patel, V. M.; Castillo, C. D.; Chellappa, R., Deep Learning for Understanding Faces Machines may be just as good, or better, than humans. *IEEE Signal Process. Mag.* **2018**, *35* (1), 66-83.

## CHAPTER 2. STOCHASTIC DIFFERENTIAL SCANNING CALORIMETRY

"Reprinted with permission from [Sherman, A. M.; Geiger, A. C.; Smith, C. J.; Taylor, L. S.; Hinds, J.; Stroud, P. A.; Simpson, G. J., Stochastic Differential Scanning Calorimetry by Nonlinear Optical Microscopy. *Anal. Chem.* **2020**, 92 (1), 1171-1178.]. Copyright [2020] American Chemical Society."

Integration of second harmonic generation (SHG) imaging with differential scanning calorimetry (DSC) connected single-particle analysis of stochastic phase transformations to the ensemble-averaged DSC transients. The SHG-activity of a crystal is highly sensitive to the specific molecular packing arrangement within a noncentrosymmetric lattice, providing access to information otherwise unavailable by conventional imaging approaches. Consequently, lattice transformations associated with dehydration/desolvation events were readily observed by SHG imaging and directly correlated to the phase transformations detected by the DSC measurements. Following studies of a model system (urea), stochastic differential scanning calorimetry (SDSC) was performed on trehalose dihydrate, which has a more complex phase behavior. From these measurements, SDSC revealed direct evidence of a "cold phase transformation" process not observable by the DSC measurements alone.

### 2.1 Introduction

In pharmaceutical drug development, drug substances and formulations with long-term physical and chemical stability ensure full optimization of a drug when administered to the patient. During the time consuming process of formulations development (estimated >3 months and \$1 million) it is thus critical to employ methods that inform on possible avenues of reduced bioavailability.<sup>1</sup> Identifying possible failure points early can reduce the chances of late-stage failure, which can be costly and time consuming to remediate. For active pharmaceutical ingredients administered as crystalline formulations, phase transitions to alternate crystal forms can have deleterious effects on aqueous dissolution rates, affecting both oral and parenteral bioavailability.<sup>2-4</sup> Administering these drugs as amorphous solid dispersions or as liquid formulations does not altogether avoid the problem, as spontaneous crystallization can occur

during storage or even in vivo under certain conditions.<sup>5-11</sup> Phase transitions can dictate both direct physical depletion of bioavailable drugs through transitions to non-soluble crystal forms, as well as susceptibility to chemical depletion through transitions to more reactive crystalline, liquid or glassy states. There is therefore a need in the pharmaceutical industry to thoroughly map the phase space of potential drug candidates, excipients, and mixtures.

Several thermal methods exist for characterizing phase transitions within pharmaceutical powders. Differential thermal analysis applies equal heat to reference and sample pans and measures the temperature differential between the two pans to detect thermal events associated with phase transitions. Thermal gravimetric analysis (TGA) continuously measures the mass of a sample while varying temperature. TGA is especially useful for dehydration and decomposition. Differential scanning calorimetry (DSC) enables the detection of phase transitions as a function of the energy differential between reference and sample pans heated to the same temperature. DSC is a widely used technique for phase transition characterization because of its sensitivity, ease-of use, and short measurement time.<sup>12</sup>

Despite these advantages, DSC probes the energy transfer into the full ensemble of the sample, averaging over all particles in the formulation. Phase transitions for ensembles of crystals are predicted to be stochastic, with nucleation being the rate-limiting step on a per-particle basis.<sup>13-</sup><sup>15</sup> If phase transformation is stochastic, individual transient events may be obscured or missed in ensemble-averaged analysis. Additionally, single-particle measurements can enable the accurate modelling of kinetics for phase transitions, allowing the distinction between different mechanisms of transition (i.e., concerted or continuous).<sup>16</sup>

Optical microscopy measurements of changes in gross morphology coupled with DSC can recover information on single-particle phase transformations. Reading and coworkers have developed an image analysis algorithm dubbed thermal analysis by structural characterization (TASC) to detect phase transitions and the spatial distribution of phases from images acquired using hot-stage microscopy.<sup>17-21</sup> TASC measures the subtle difference between an original image and subsequent images to quantify morphological changes within a region-of interest as temperature is varied. The algorithm also accounts for the possibility of translation of the region-of-interest within the field of view (FOV). TASC is a fast and inexpensive method for characterizing the spatial distribution of phase transformations. However, TASC methods rely exclusively on morphological changes in the optical response, which are only intrinsically related

to a subset of all possible phase transformations. Solid / solid phase transformations are largely inaccessible by TASC, and glass/solid phase transformations can be challenging to reliably detect without perturbing the sample. Desolvation (including dehydration) or polymorph transitions in the solid state are typically not accompanied by obvious changes in the gross morphology of a particle but can profoundly influence physico-chemical characteristics that are accompanied by thermal transients in DSC measurements. In addition, structural re-arrangements between polymorphs with similar bulk free energies are quite challenging to detect by either conventional DSC or bright-field microscopy. Coupling DSC with concurrent imaging capabilities that are more directly tied to crystal form would help address these ambiguities associated with conventional bright-field microscopy and DSC.

In this work, we demonstrated the integration of second harmonic generation (SHG) microscopy with DSC analysis to monitor phase transitions on a per-particle basis. This multimodal approach enabled simultaneous single-particle analysis for distinguishing the impact of crystal size and shape on phase transitions in a single experiment, complementing the ensemble-averaged information obtained from DSC. Furthermore, the strict symmetry requirements for SHG allow the mapping of complex desolvation and crystal form transitions with sensitivity in the ppm regime.<sup>22</sup> Previous work has utilized the exquisite sensitivity of SHG microscopy to internal structure of the lattice in non-centrosymmetric crystals as a standalone tool for polymorph discrimination, calibration free quantification of trace crystallinity, and protein crystal centering at X-ray beamlines.<sup>23-25</sup> Raman spectroscopy, terahertz spectroscopy, and SHG have all previously been used to characterize crystallinity, polymorphism, and phase transitions.<sup>26-33</sup> The merits of integrated SHG and DSC measurements are explored for a model system to then explore the complex phase transitions in trehalose dihydrate.

## **2.2 Methods**

### **2.2.1 SHG microscopy**

The experimental apparatus is depicted in Figure 2.1 and consists of a home-built SHG microscope integrated with an optical DSC stage (Linkam, DSC450). A tunable 80 MHz, Ti:sapphire, femtosecond laser (Spectra-Physics, Mai Tai) was used for the incident light source. The fundamental beam was raster-scanned across the sample using a resonant scanning mirror at

8.8 kHz (Electro-Optical Products Corporation) for the fast-scan axis and a galvanometer mirror (Cambridge-Tech) for the slow-scan axis. A 4x, 0.1 NA objective (Nikon) was used to focus the beam onto the sample, and the SHG signal was collected in the epi direction through the same objective used for delivery of the fundamental beam. The laser was tuned to 800 nm with a power of 80–120 mW at the sample. Two long-pass dichroic mirrors (Chroma, 650DCXR) and a band-pass filter (Chroma HQ400/20M-2P) were used to isolate the 400 nm SHG signal before it was detected by a photomultiplier tube (PMT) (Hamamatsu, H7422P-40 MOD). Responses of the PMT were digitized synchronously with the laser pulses by using a digital oscilloscope card (Alazar Tech, ATS9350) and mapped onto  $512 \times 512$  images via custom software written in-house (MATLAB).<sup>34</sup> The SHG videos were recorded at 17 frames per second for urea samples and at eight frames per second for trehalose samples.

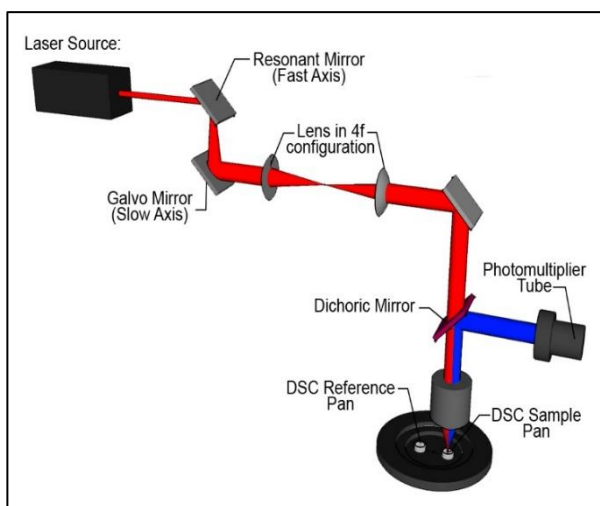


Figure 2.1. Schematic of the integrated SHG-DSC microscope. A pulsed laser beam (800 nm, 80 MHz) was scanned with a galvo-resonant scan pair, which was 4f-coupled to the back of an objective and focused onto a DSC sample pan. The SHG signal was collected in the epi-direction, isolated with a dichroic mirror and detected with a photomultiplier tube.

### 2.2.2 DSC measurements

Stochastic differential scanning calorimetry (SDSC) measurements were acquired by integration of the home-built SHG microscope with a Linkam Optical DSC450 stage. The DSC temperature ramp range was 125–145 °C for the urea samples and 80–250 °C for the trehalose samples with ramp rates of 10 °C/min and 20 °C/min, respectively. Before the data collection, an

isothermal hold time of two minutes for both urea and trehalose ensured a reliable starting temperature for all experiments to reduce the effects of initial temperature variation. Temperature and input power were measured by the DSC at a rate of five Hz. Standard aluminum sample pans were used in all DSC experiments, and were not sealed to allow for optical access of the sample. The DSC was calibrated using an indium standard in an open pan.

### **2.2.3 Data analysis**

Supersaturated solutions of urea (Sigma-Aldrich) were prepared in DI water solution and rapidly recrystallized in an ice bath. The water was decanted, and the crystals were left to dry overnight. Crystals were then passed through a 63  $\mu\text{m}$  mesh into the aluminum DSC sample pan. D-(+)-trehalose dihydrate (Sigma Life Sciences) was added to a DSC pan as received.

### **2.2.4 Sample preparation**

The experimental apparatus is depicted in Figure 2.1 and consists of a home-built SHG microscope integrated with an optical DSC stage (Linkam, DSC450). A tunable 80 MHz, Ti:sapphire, femtosecond laser (Spectra-Physics, Mai Tai) was used for the incident light source.

## **2.3 Results and discussion**

### **2.3.1 SDSC of urea crystals**

Initial proof of concept studies for SDSC were performed using urea, which undergoes a simple single-stage solid/liquid phase transformation. Figure 2.2 A shows the DSC trace for the melting transition of a urea sample (black line) overlaid with vertical lines corresponding to the normalized SHG areas of individual particles dropping below a threshold of 0.1.

The DSC trace indicates that the melting transition occurred at  $\sim 135^\circ\text{C}$ . This temperature is slightly higher than the literature value of  $133^\circ\text{C}$ . This difference is attributed to a non-negligible heat transfer time to the sample. The need for optical access required the use of an unencapsulated sample pan, resulting in slower heat transfer.<sup>35</sup> Despite some baseline drift in the per-particle SHG activity from particle motion, the single-particle SHG area shows much sharper transitions from the SHG-active crystalline form to the SHG-inactive molten liquid of urea indicated by the DSC trace, consistent with independent, stochastic melting events. Single-particle analysis yields a

mean phase transition duration of  $1.0 \pm 0.6$  s/particle while the DSC melting peak has a full-width-half-maximum of 3.3 s. This difference indicates that the rate of phase transformation was dictated largely by the rate for seed formation of the liquid-state within individual particles, followed by rapid phase transformation within a given particle. Figure 2.2 B shows representative frames from a video acquired with the SHG microscope during the melting of urea. From the SHG results, the nucleation rate at the phase transformation temperature can be estimated. In this work, a nucleation event is defined as the phase transformation of a single crystal (in this case, from solid to liquid). Nucleation rates were determined from the loss rate of SHG-active particles, measured as the intensity of each crystal decreased below a threshold (close to zero). Approximations for the per-particle mass using the density of urea and volume estimated from particle cross-sectional area yields a peak nucleation rate of  $1.6 \pm 0.2 \times 10^4$  nuclei per second per milligram (based on the observed rate of 6.13 nuclei/s in a FOV with an estimated 0.39  $\mu\text{g}$  of particulate mass; standard deviation was determined based on Poisson statistics for the number of crystals within the FOV).



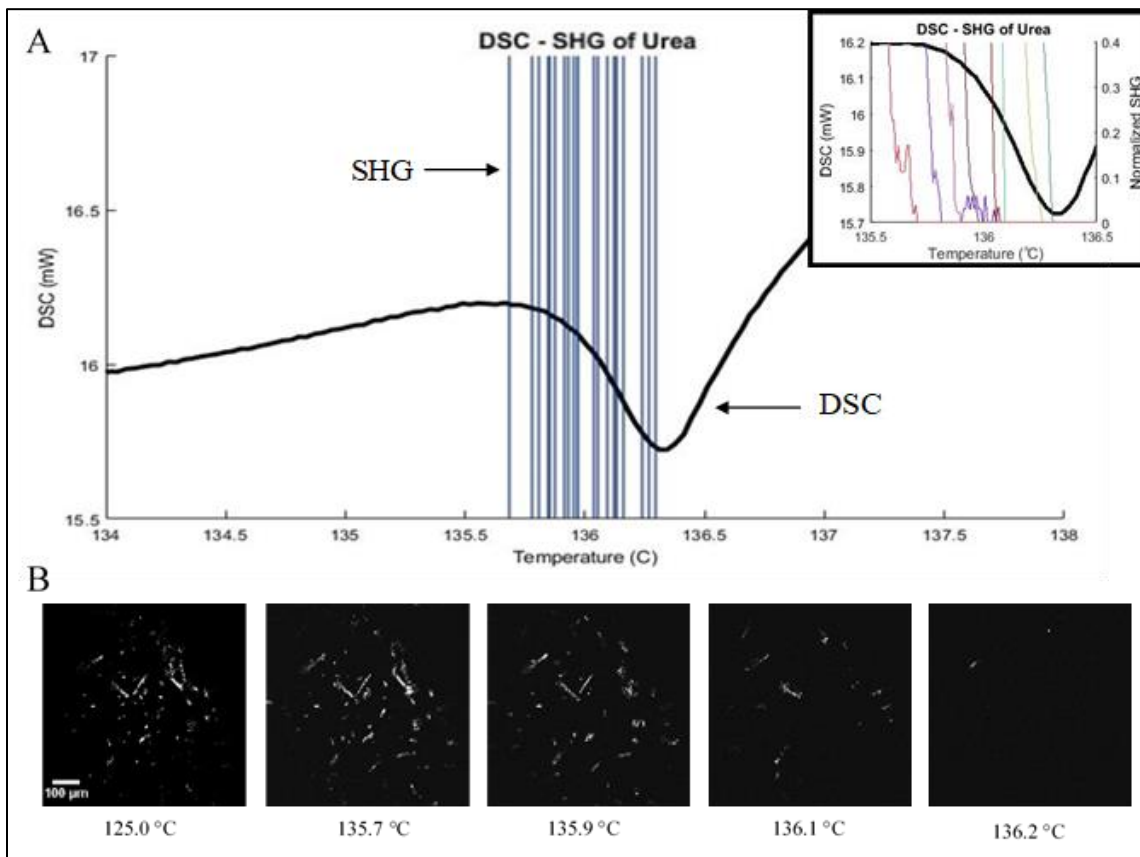


Figure 2.2. A) SDSC analysis of urea by combined SHG microscopy and DSC. The solid black line is the DSC trace of the melting of urea with endothermic direction being down (left-axis). The vertical colored lines show the temperatures in which the normalized SHG intensity of individual urea particles (acquired concurrently with DSC measurements) drops below a value of 0.1. DSC and SHG measurements both indicate a structural transformation in the same temperature range. The inset shows a zoom-in of the temperature range 135.5–136.5 °C where the color lines correspond to the normalized SHG intensity of individual urea particles. B) SHG microscope images of urea crystals during SDSC analysis. SHG microscopy reveals the stochastic nature of the melting transition and enables analysis on a per-particle basis.

Consistent with numerous previous DSC studies of powders, SDSC measurements were performed exclusively as the temperature was increased to interrogate phase transformations.<sup>36-37</sup> While in principle, additional information can be obtained from measurements of thermal events upon cooling, the phase transformations induced by heating were generally not easily reversible (e.g., dehydration). Even in the absence of composition changes, melting of many individual isolated particles results in coalescence to a homogeneous melt, while the reverse process will not

recover isolated crystalline particles upon cooling. Furthermore, SDSC measurements upon cooling are complicated in practice by wicking of the liquid, which moves much of the sample out of the FOV. Although not the focus of the present study, it is worth noting that SHG microscopy has a rich history of informing on crystallization kinetics in accelerated stability assessments for active pharmaceutical ingredients.<sup>38</sup>

### 2.3.2 IRF determination with SDSC

The simultaneous SHG measurements were also used to determine the impulse response function (IRF) of the DSC instrument. As described in a preceding paragraph, the thermal events induced in the sample produced an instrument response with a temporal delay associated with heat transfer times. The macroscopic DSC endotherm arose from the net collective contributions from many such stochastic events. As such, the DSC observables were given by the convolution of the impulsive phase transformations with the IRF of the DSC system. Assuming a double-exponential IRF consistent with heat flow through a thermal resistor, the measured set of impulsive phase transformations by SHG  $\mathbf{x}$  can be combined with the recorded DSC trace  $\mathbf{y}$  to recover the maximum likelihood estimate (MLE) for the impulse response function  $\hat{f}_{IRF} = f_{IRF}(\hat{\mathbf{a}})$  described by the set of parameters  $\mathbf{a}$  through  $\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \|\mathbf{y} - \mathbf{x} \otimes f_{IRF}(\mathbf{a})\|^2$ . The results of an MLE fit of the measured DSC transient using the single-crystal phase transformation data from SHG is shown in Figure 2.3, recovering a phase-lag of  $3.088 \pm 0.013$  s and rising and falling exponential time constants of  $1.161 \pm 0.019$  s and  $0.435 \pm 0.013$  s, respectively, for the double-exponential IRF. The phase-lag between the DSC event and the mean of the SHG events is attributed to delays from diffusive heat transfer from the individual particles to the pan and sample mount of the DSC instrument. The two time constants are tentatively attributed to the thermal diffusion times from the sample to the sample stage, and from the stage to the heat sink.<sup>39</sup> Even in a relatively simple melting transition, SHG provides the capability of mapping phase transformations on individual particles with temporal resolution much faster than achievable from conventional ensemble-averaged DSC measurements. The pairing of both methods provides information inaccessible by DSC alone; namely, that phase transformation proceeds through slow nucleation followed by rapid growth on a per particle basis with an intrinsic rate of  $1.0 \pm 0.6$  s/particle (variance is dominated by particle-to-particle stochasticity).

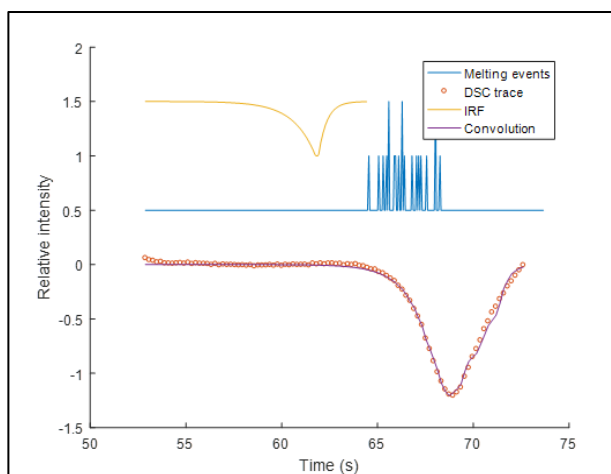


Figure 2.3. Comparison of the experimental DSC trace and the DSC trace generated from melting events. The best-fit double-exponential IRF (yellow) is convolved with impulsive melting events (blue) of single-particles observed in SHG images to generate the purple trace, which is in good agreement with the experimental DSC data (red). The time constant and offset of double-exponential IRF were optimized to minimize difference between experimental and generated DSC traces. Best-fit values of  $1.161 \pm 0.019$  s,  $0.435 \pm 0.013$  s, and  $3.088 \pm 0.013$  s were computed for the double-exponential time constants and the offset respectively.

### 2.3.3 SDSC of trehalose dihydrate crystals

The performance of SDSC using the integrated SHG-DSC microscope was further tested with trehalose dihydrate, a molecule with complex phase behavior, including the existence of multiple transient crystal forms accessible under varying conditions.<sup>40-43</sup> Since significant differences in crystal structure can exist between polymorphs, hydrates, and solvates, SHG has the potential to aid in distinguishing transitions due to its sensitivity to noncentrosymmetric crystal forms.<sup>23</sup> Figure 2.4 summarizes the SDSC analysis of trehalose dihydrate. Each colored line represents the percentage of pixels in an individual trehalose crystalline particle that exceeds a threshold. Four individual particles were monitored in this study. The black line is the DSC curve, in which dips correspond to endothermic events. For the temperature range of 140 - 220 °C, the collected SDSC images were summed, and the logarithm of this summed image was cropped to the measured region of each particle. Each of the cropped images points to their representative SHG trace.

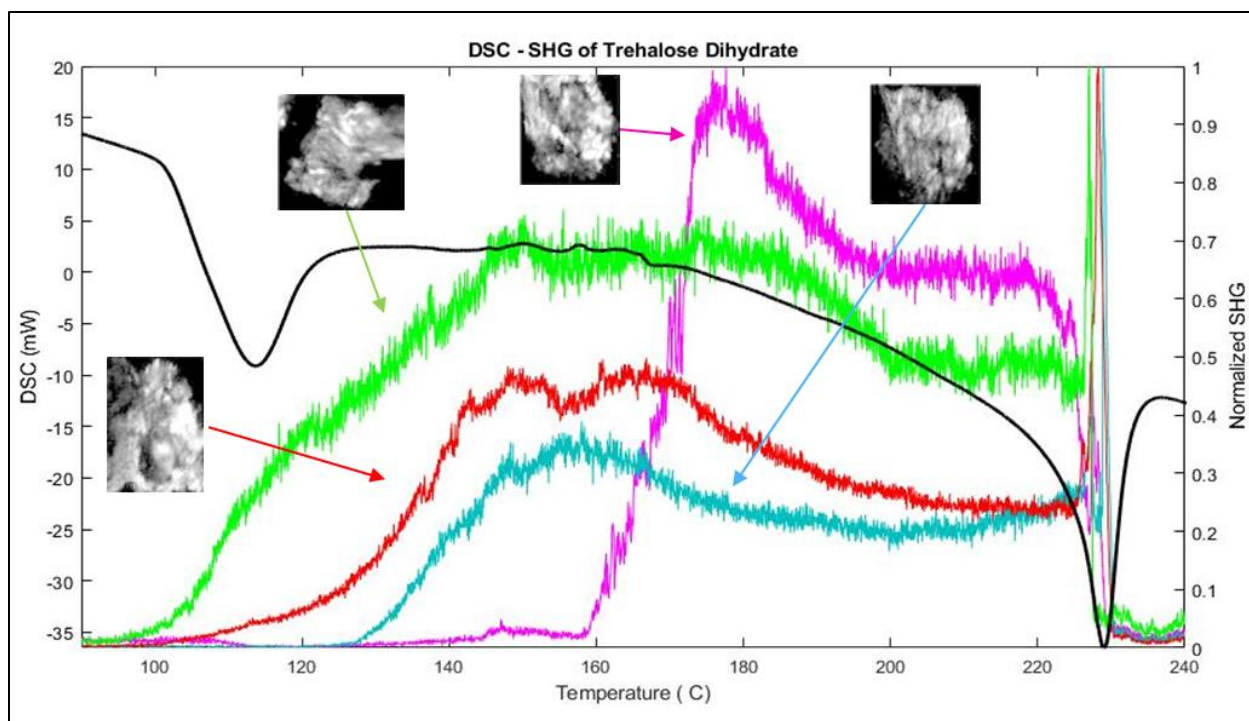


Figure 2.4. SDSC analysis of trehalose by SHG microscopy. The colored lines show the percentage of pixels in a trehalose particle above an SHG intensity threshold overlaid with the black line of the DSC trace with the endothermic direction being down. SDSC images were summed for the temperature range of 140 – 220 °C and the logarithm of the summed image was cropped to the measured region of each particle (each cropped image points to the representative SHG trace).

DSC and SHG data were interpreted using previously reported literature values to map the polymorphic transitions of the sample through the temperature ramp. At the initial 80 °C, the weak SHG signal arises from the trehalose dihydrate ( $T_h$ ) form that was initially added to the DSC pan before the temperature ramp.<sup>44</sup> The composition of this starting material was confirmed by powder X-ray diffraction (pXRD) to be primarily composed  $T_h$ , although the unstable anhydrous form ( $T_a$ ) may also be present at trace levels; previous reports indicate observation of the  $T_a$  form following room temperature storage at low relative humidity. The first signs of phase transformation activity occurred in the temperature range of 80–130 °C. The DSC curve exhibits a major endothermic peak in this range. This peak was previously reported to be the result of the dehydration of  $T_h$ .<sup>36</sup> Additionally, two distinct trends appear in the individual particle SHG data. Three of the trehalose particles increased in SHG activity while the other particle decreased over this same lower temperature range. The disparity in SHG activity between individual particles can be explained by the two major dehydration pathways available for  $T_h$  in this temperature range;  $T_h$  can either

partially dehydrate with rearrangement to the crystalline  $T_\gamma$  form or completely dehydrate to an amorphous form ( $T_{am}$ ).<sup>36, 44</sup> Similarly, the  $T_\alpha$  form can also transition to the  $T_{am}$  form upon heating.<sup>36-37, 44-45</sup> The  $T_\gamma$  polymorph is theorized to be a mixture of  $T_h$  encapsulated in a shell of the crystalline, anhydrous form ( $T_\beta$ ), which is SHG-active.<sup>36, 45</sup> Therefore, we conclude that the three particles that increased in SHG area in this temperature range likely transitioned to the SHG-active  $T_\gamma$  form, while the particle that reduced in SHG activity likely transitioned to the SHG-inactive  $T_{am}$  form at  $\sim 110$  °C from either the  $T_h$  or  $T_{am}$  form.

The SHG activity generally increased for individual particles over the subsequent temperature range of 130–160 °C. The DSC curve exhibited small but reproducible features in this temperature range, the magnitude of which varies between experiments, and has been reported to be due to the dehydration of  $T_h$ .<sup>36</sup> Measurements of the fraction of SHG-active pixels in individual particles show that the three particles that previously exhibited SHG signal continue to convert over this temperature range. This increase in highly SHG-active fraction can simply be attributed to crystal growth over this temperature range. The fourth particle that previously decreased in SHG area shows a sharp increase in SHG area starting at  $\sim 160$ °C. SHG-inactive  $T_{am}$  has been previously reported to transition to SHG-active  $T_\beta$  in the range of 150–200 °C.<sup>36, 44</sup> Likewise, the  $T_\alpha$  form has been reported to be able to transition directly to the  $T_\beta$  when held at a temperature of  $\sim 150$  °C.<sup>36-37, 44-45</sup> However, this  $T_\alpha$  transition has not been observed with the faster heating rates (20 °C/min.) used in the present study. Therefore, we conclude that this polymorphic transition from  $T_{am}$  to  $T_\beta$  is likely the cause of the sharp peak in SHG from the particle represented by the magenta trace.

Following the monotonic increases in transitioning to the  $T_\beta$  form from 130–160 °C, decreases in SHG were observed for all four particles in the range 160–200 °C. No peaks are observed in the DSC curve in this range, but the trehalose particles exhibit slowly varying and inhomogeneous loss in SHG activity. Several possible mechanisms for this loss in SHG activity were considered, two of which are described in detail. One possible explanation for loss in SHG activity with large crystal-to-crystal variability may be from interference effects as crystal sizes change during phase transformation.<sup>46</sup> Interference fringes from single crystals were reported previously in SHG microscopy measurements. However, interference effects were deemed unlikely for three reasons: i) interference effects would be anticipated to produce increases or decreases in SHG activity with equal probability, while only decays were observed experimentally,

ii) the backwards coherence length is short ( $\sim 100$  nm) relative to the sizes of the crystals, such that average activities would be expected, and iii) the interference is expected to vary with thickness, such that crystals with variance in dimensions would likely produce fringes over which the intensities are integrated. Alternatively, the decay in SHG signal can be explained by condensation of water vapor released by  $T_h$  onto the optical window of the DSC. Condensation was commonly seen in our experiments when working with hydrated samples because open crucible pans were used in the experiment to allow for imaging during DSC scans. Additionally, no purge gas was used to remove water vapor from the system, as addition of purge gas introduced measurement variance in the DSC analyses from variable heat transfer to the bath gas. Previously reported results by TGA indicate gradual water loss up to  $200^\circ\text{C}$ ,<sup>36</sup> consistent with the observed reduction in SHG signal that is proposed to arise from scattering losses from condensation. Furthermore, condensation is likely to impact some locations within the FOV to a greater extent than others, potentially providing an explanation for the large crystal-to-crystal variability in the loss. Thus, condensation is predicted to be the main cause of this loss in SHG signal.

Following the slow SHG signal loss from condensation, dramatic changes in the SHG-activity arose from  $200$ – $240^\circ\text{C}$ . In this range the DSC curve exhibited an endothermic peak at  $\sim 230^\circ\text{C}$ , and the SHG data displayed a sharp increase and subsequent decrease during the DSC peak. It is clear from the literature that the endothermic DSC peak arises from a melting transition.<sup>36, 45</sup> The final fall in SHG area is attributed to this melting event. However, the source of the sharp increase in SHG area prior to the melt has no precedent in reported DSC measurements. Two origins for the transient SHG area increase were considered.

First, significant mobility in the sample arose during the melting transition, which could alter the positions of the sample crystals relative to the focal plane. Crystals moving into the focal plane could result in an increase in SHG area, as SHG intensity scales with the squared power of incident light. Second, the increase could be explained by a change in the crystal form, degree of crystallinity, or crystal size of trehalose. To evaluate the first mechanism, a custom optic, designed in-house, was added to the beam path to extend the depth-of-field from  $\sim 20\ \mu\text{m}$  to  $\sim 100\ \mu\text{m}$  (detailed in a manuscript in preparation). The sharp peak in SHG prior to the melt was still observed while imaging with an extended depth-of-field, suggesting that crystal movement into the focal volume is not likely to be the major cause of the increase in SHG area.

Alternatively, the increase in SHG area could arise from a change in the crystal form, degree of crystallinity, or crystal size of trehalose. While it is possible that there exists a previously unknown polymorphic transition immediately prior to the melt of trehalose, this possibility was rejected on the principal of Occam's razor, in light of the extensive body of prior work done to characterize the polymorphism of trehalose. A more plausible origin for the spike in SHG area is from the final conversion of the remaining  $T_h$  core of  $T_\gamma$  polymorph to  $T_\beta$ , promoted by an increase in the energy in molecular diffusion immediately prior to the melting transition. Additionally, another closely related contribution to the increase in SHG area could be the rapid growth of  $T_\beta$  crystalline domains of trehalose from residual  $T_{am}$  due to increase mobility immediately prior to the melt. In summary, the particles observed in this study are theorized to undergo one of two general phase transition pathways:  $T_h \rightarrow T_{am} \rightarrow T_\beta \rightarrow \text{melt}$  or  $T_h \rightarrow T_\gamma \rightarrow T_\beta \rightarrow \text{melt}$ .

Notably, this “cold phase transformation” phenomenon, in which phase transformation to the SHG-active crystalline form occurs immediately prior to the transition to the liquid form, was clearly observable by SHG but undetectable in the DSC measurements. This latter proposed mechanism has precedent in observations of “cold crystallization” in DSC, in which transient crystallization arises upon transformation from a glassy state to a liquid.<sup>47</sup> In DSC, cold crystallization appears as an exothermic peak immediately preceding an endothermic peak. Cold crystallization has been observed in studies of  $T_h$  for the transition from a glassy material to the  $T_\beta$  form.<sup>45</sup> However, to our knowledge, this study represents the first observation of cold transformation between two different solid-state crystalline forms prior to the melt. Importantly, evidence supporting cold transformation was undetectable by DSC alone, presumably obfuscated by the comparatively longer response time and inherent ensemble averaging associated with the DSC measurements.

As discussed before, the reversibility of the SDSC was not investigated for trehalose dihydrate. Like the urea samples, wetting of the sample pan after heating of the sample presented challenges in maintaining sample position within the FOV following the melt. Likewise, reversibility after dehydration (< 220°C) would not be anticipated due to the large entropic barriers involved with rehydration of the sample under high humidity. Studies have shown that the dehydrated sample can rehydrate to the original material, but over timescales significantly longer than easily accessible by DSC.<sup>48-49</sup>

Analysis of the impulse response function from urea provides a route for connecting the cold phase transformation and melting processes to the DSC endotherm at  $\sim 230$  °C. The entire transformation to produce the SHG-active crystal form, followed by melting progressed all within  $\sim 2$  seconds, likely explaining the absence of prior work describing cold phase transformation. Neither of the resulting curves coincides directly with that of the DSC melting peak. This disparity likely arises from the statistics of small numbers, as only four particles in the FOV were available for analysis (as opposed to 58 crystals for urea).

Differences between single-particle and ensemble-averaged measurements are particularly noteworthy in the results shown in Figure 2.4. The integrated SHG intensity from one particular crystal indicated by the green trace nicely tracks the DSC trace and is likely to be representative of the major ensemble-averaged thermal event at  $\sim 110$  °C. However, it is clear that distinctly different phase transformation behaviors are observed within a subset of the population, exemplified by the red, cyan, and purple traces. This apparent incongruity is attributed to the stochastic nature of the sample, with multiple thermally accessible transformation pathways available.

Trehalose dihydrate has multiple competing pathways to various crystal forms, which are stochastically accessed over large temperature ranges and which are likely to produce significantly different heats of phase transformation. This rich landscape of crystal form space is largely obscured by ensemble-averaged measurements of DSC alone, but is clearly accessed by per-particle analysis using SHG microscopy.

## **2.4 Conclusion**

SDSC was shown to enable connection of internal structural rearrangements within crystalline materials measured by SHG on single particles to the heat flow recorded by DSC. SHG measurements were sensitive to variations in the molecular packing arrangements during the phase transitions of individual particles that are generally inaccessible in conventional microscopy. Coupling the structural data acquired from SHG imaging with DSC enabled single-particle measurements disentangled from the ensemble-averaged thermal transients recorded with DSC. A proof-of-concept SDSC experiment was performed on urea, which has a single solid/liquid phase transformation. From these measurements, the IRF of the DSC system was determined, showing that the melt proceeds through a slow nucleation step followed by rapid growth on a per particle



basis. Next, SDSC was used to characterize the phase behavior of a more complex system, trehalose dihydrate. The dehydration events of trehalose dihydrate were readily observed by SDSC and the higher temporal resolution of SHG measurements, compared to that of the temporal resolution of typical DSC measurements, allowed for detection of a rapid phase transformation not observable by DSC. Thus, SDSC is proposed as a novel technique for the pharmaceutical drug development pipeline for the characterization of single-particle phase transformations.

## 2.5 References

1. Strovel, J.; Sittampalam, S.; Coussens, N. P.; Hughes, M.; Inglese, J.; Kurtz, A.; Andalibi, A.; Patton, L.; Austin, C.; Baltezor, M., Early drug discovery and development guidelines: for academic researchers, collaborators, and start-up companies. **2016**.
2. Cheney, M. L.; Shan, N.; Healey, E. R.; Hanna, M.; Wojtas, L.; Zaworotko, M. J.; Sava, V.; Song, S.; Sanchez-Ramos, J. R., Effects of crystal form on solubility and pharmacokinetics: a crystal engineering case study of lamotrigine. *Crystal Growth & Design* **2009**, *10* (1), 394-405.
3. Maddileti, D.; Swapna, B.; Nangia, A., High solubility crystalline pharmaceutical Forms of blonanserin. *Crystal Growth & Design* **2014**, *14* (5), 2557-2570.
4. Yadav, A.; Shete, A.; Dabke, A.; Kulkarni, P.; Sakhare, S., Co-crystals: a novel approach to modify physicochemical properties of active pharmaceutical ingredients. *Indian journal of pharmaceutical sciences* **2009**, *71* (4), 359.
5. Newman, A.; Knipp, G.; Zografi, G., Assessing the performance of amorphous solid dispersions. *Journal of pharmaceutical sciences* **2012**, *101* (4), 1355-1377.
6. Baghel, S.; Cathcart, H.; O'Reilly, N. J., Polymeric amorphous solid dispersions: a review of amorphization, crystallization, stabilization, solid-state characterization, and aqueous solubilization of biopharmaceutical classification system class II drugs. *Journal of pharmaceutical sciences* **2016**, *105* (9), 2527-2544.
7. Jackson, M. J.; Kestur, U. S.; Hussain, M. A.; Taylor, L. S., Dissolution of danazol amorphous solid dispersions: supersaturation and phase behavior as a function of drug loading and polymer type. *Molecular pharmaceutics* **2015**, *13* (1), 223-231.
8. Konno, H.; Taylor, L. S., Ability of different polymers to inhibit the crystallization of amorphous felodipine in the presence of moisture. *Pharmaceutical research* **2008**, *25* (4), 969-978.
9. Shamblin, S. L.; Zografi, G., The effects of absorbed water on the properties of amorphous mixtures containing sucrose. *Pharmaceutical research* **1999**, *16* (7), 1119-1124.

10. Andronis, V.; Yoshioka, M.; Zografi, G., Effects of sorbed water on the crystallization of indomethacin from the amorphous state. *Journal of pharmaceutical sciences* **1997**, *86* (3), 346-351.
11. Rumondor, A. C.; Stanford, L. A.; Taylor, L. S., Effects of polymer type and storage relative humidity on the kinetics of felodipine crystallization from amorphous solid dispersions. *Pharmaceutical research* **2009**, *26* (12), 2599.
12. Clas, S.-D.; Dalton, C. R.; Hancock, B. C., Differential scanning calorimetry: applications in drug development. *Pharmaceutical science & technology today* **1999**, *2* (8), 311-320.
13. Oxtoby, D. W., Nucleation of first-order phase transitions. *Accounts of chemical research* **1998**, *31* (2), 91-97.
14. Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M., Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nature materials* **2011**, *10* (9), 693.
15. Sleutel, M.; Lutsko, J.; Van Driessche, A. E.; Durán-Olivencia, M. A.; Maes, D., Observing classical nucleation theory at work by monitoring phase transitions with molecular precision. *Nature communications* **2014**, *5*, 5598.
16. Liu, H.; Kwon, O.-H.; Tang, J.; Zewail, A. H., 4D imaging and diffraction dynamics of single-particle phase transition in heterogeneous ensembles. *Nano letters* **2014**, *14* (2), 946-954.
17. Reading, M., Thermal Analysis by Structural Characterization (TASC): Structural and Thermo-Rheological Information from Hot Stage Microscopy. *Microscopy Today* **2017**, *25* (5), 18-23.
18. Reading, M.; Morton, M.; Antonijevic, M.; Grandy, D.; Hourston, D.; Lacey, A., New methods of thermal analysis and chemical mapping on a micro and nano scale by combining microscopy with image analysis. *Microscopy: advances in scientific research and education. Formatex Research Center* **2014**, 1083-9.
19. Alhijjaj, M.; Belton, P.; Fabian, L.; Wellner, N.; Reading, M.; Qi, S., Novel Thermal Imaging Method for Rapid Screening of Drug–Polymer Miscibility for Solid Dispersion Based Formulation Development. *Molecular pharmaceuticals* **2018**, *15* (12), 5625-5636.
20. Alhijjaj, M.; Reading, M.; Belton, P.; Qi, S., Thermal analysis by structural characterization as a method for assessing heterogeneity in complex solid pharmaceutical dosage forms. *Anal. Chem.* **2015**, *87* (21), 10848-10855.
21. Alhijjaj, M.; Yassin, S.; Reading, M.; Zeitler, J. A.; Belton, P.; Qi, S., Characterization of heterogeneity and spatial distribution of phases in complex solid dispersions by thermal analysis by structural characterization and X-ray micro computed tomography. *Pharmaceutical research* **2017**, *34* (5), 971-989.

22. Kestur, U. S.; Wanapun, D.; Toth, S. J.; Wegiel, L. A.; Simpson, G. J.; Taylor, L. S., Nonlinear optical imaging for sensitive detection of crystals in bulk amorphous powders. *Journal of pharmaceutical sciences* **2012**, *101* (11), 4201-4213.
23. Chowdhury, A. U.; Dettmar, C. M.; Sullivan, S. Z.; Zhang, S.; Jacobs, K. T.; Kissick, D. J.; Maltais, T.; Hedderich, H. G.; Bishop, P. A.; Simpson, G. J., Kinetic trapping of metastable amino acid polymorphs. *Journal of the American Chemical Society* **2014**, *136* (6), 2404-2412.
24. Smith, C. J.; Dinh, J.; Schmitt, P. D.; Stroud, P. A.; Hinds, J.; Johnson, M.; Simpson, G., EXPRESS: Calibration-Free Second Harmonic Generation (SHG) Image Analysis for Quantification of Trace Crystallinity within Final Dosage Forms of Amorphous Solid Dispersions. *Applied spectroscopy* **2018**, 0003702818786506.
25. Scarborough, N. M.; Godaliyadda, G.; Ye, D. H.; Kissick, D. J.; Zhang, S.; Newman, J. A.; Sheedlo, M. J.; Chowdhury, A.; Fischetti, R. F.; Das, C., Synchrotron X-Ray Diffraction Dynamic Sampling for Protein Crystal Centering. *Electronic Imaging* **2017**, *2017* (17), 6-9.
26. Simon, F.; Clevers, S.; Dupray, V.; Coquerel, G., Relevance of the second harmonic generation to characterize crystalline samples. *Chemical Engineering & Technology* **2015**, *38* (6), 971-983.
27. Galland, A.; Dupray, V.; Berton, B.; Morin-Grognet, S.; Sanselme, M.; Atmani, H.; Coquerel, G., Spotting conglomerates by second harmonic generation. *Crystal Growth and Design* **2009**, *9* (6), 2713-2718.
28. Francis, A. T.; Nguyen, T. T.; Lamm, M. S.; Teller, R.; Forster, S. P.; Xu, W.; Rhodes, T.; Smith, R. L.; Kuiper, J.; Su, Y., In Situ Stimulated Raman Scattering (SRS) Microscopy Study of the Dissolution of Sustained-Release Implant Formulation. *Molecular pharmaceutics* **2018**, *15* (12), 5793-5801.
29. Heinz, A.; Strachan, C. J.; Gordon, K. C.; Rades, T., Analysis of solid-state transformations of pharmaceutical compounds using vibrational spectroscopy. *Journal of Pharmacy and Pharmacology* **2009**, *61* (8), 971-988.
30. Zeitler, J. A.; Kogermann, K.; Rantanen, J.; Rades, T.; Taday, P. F.; Pepper, M.; Aaltonen, J.; Strachan, C. J., Drug hydrate systems and dehydration processes studied by terahertz pulsed spectroscopy. *Int. J. Pharm.* **2007**, *334* (1-2), 78-84.
31. Zeitler, J. A.; Newnham, D. A.; Taday, P. F.; Threlfall, T. L.; Lancaster, R. W.; Berg, R. W.; Strachan, C. J.; Pepper, M.; Gordon, K. C.; Rades, T., Characterization of temperature-induced phase transitions in five polymorphic forms of sulfathiazole by terahertz pulsed spectroscopy and differential scanning calorimetry. *Journal of pharmaceutical sciences* **2006**, *95* (11), 2486-2498.

32. Strachan, C. J.; Rades, T.; Gordon, K. C.; Rantanen, J., Raman spectroscopy for quantitative analysis of pharmaceutical solids. *Journal of pharmacy and pharmacology* **2007**, *59* (2), 179-192.
33. Strachan, C. J.; Taday, P. F.; Newnham, D. A.; Gordon, K. C.; Zeitler, J. A.; Pepper, M.; Rades, T., Using terahertz pulsed spectroscopy to quantify pharmaceutical polymorphism and crystallinity. *Journal of Pharmaceutical Sciences* **2005**, *94* (4), 837-846.
34. Muir, R. D.; Sullivan, S. Z.; Oglesbee, R. A.; Simpson, G. J., Synchronous digitization for high dynamic range lock-in amplification in beam-scanning microscopy. *Review of Scientific Instruments* **2014**, *85* (3), 033703.
35. Gaisford, S.; Kett, V.; Haines, P., *Principles of thermal analysis and calorimetry*. Royal society of chemistry: 2016.
36. Raimi-Abraham, B. T.; Moffat, J. G.; Belton, P. S.; Barker, S. A.; Craig, D. Q. M., Generation and Characterization of Standardized Forms of Trehalose Dihydrate and Their Associated Solid-State Behavior. *Crystal Growth & Design* **2014**, *14* (10), 4955-4967.
37. Rani, M.; Govindarajan, R.; Surana, R.; Suryanarayanan, R., Structure in dehydrated trehalose dihydrate - Evaluation of the concept of partial crystallinity. *Pharmaceutical Research* **2006**, *23* (10), 2356-2367.
38. Song, Z. T.; Sarkar, S.; Vogt, A. D.; Danzer, G. D.; Smith, C. J.; Gaultier, E. J.; Simpson, G. J., Kinetic Modeling of Accelerated Stability Testing Enabled by Second Harmonic Generation Microscopy. *Anal. Chem.* **2018**, *90* (7), 4406-4413.
39. Höhne, G. W. H.; Hemminger, W.; Flammersheim, H.-J., Theoretical fundamentals of differential scanning calorimeters. In *Differential Scanning Calorimetry*, Springer: 1996; pp 21-40.
40. Singh, S. K., Sucrose and Trehalose in Therapeutic Protein Formulations. In *Challenges in Protein Product Development*, Springer: 2018; pp 63-95.
41. Taylor, L. S.; Williams, A. C.; York, P., Particle size dependent molecular rearrangements during the dehydration of trehalose dihydrate-in situ FT-Raman spectroscopy. *Pharmaceutical research* **1998**, *15* (8), 1207-1214.
42. Taylor, L. S.; York, P., Effect of particle size and temperature on the dehydration kinetics of trehalose dihydrate. *Int. J. Pharm.* **1998**, *167* (1-2), 215-221.
43. Taylor, L. S.; York, P., Characterization of the phase transitions of trehalose dihydrate on heating and subsequent dehydration. *Journal of pharmaceutical sciences* **1998**, *87* (3), 347-355.

44. Sussich, F.; Urbani, R.; Princivale, F.; Cesaro, A., Polymorphic amorphous and crystalline forms of trehalose. *Journal of the American Chemical Society* **1998**, *120* (31), 7893-7899.
45. Sussich, F.; Princivale, F.; Cesàro, A., The interplay of the rate of water removal in the dehydration of  $\alpha$ ,  $\alpha$ -trehalose. *Carbohydr. Res.* **1999**, *322* (1-2), 113-119.
46. Dow, X. Y.; DeWalt, E. L.; Sullivan, S. Z.; Schmitt, P. D.; Ulcickas, J. R.; Simpson, G. J., Imaging the nonlinear susceptibility tensor of collagen by nonlinear optical stokes ellipsometry. *Biophysical journal* **2016**, *111* (7), 1361-1374.
47. Wunderlich, B., Theory of cold crystallization of high polymers. *The Journal of Chemical Physics* **1958**, *29* (6), 1395-1404.
48. Furuki, T.; Kishi, A.; Sakurai, M., De- and rehydration behavior of alpha,alpha-trehalose dihydrate under humidity-controlled atmospheres. *Carbohydr. Res.* **2005**, *340* (3), 429-438.
49. Jones, M. D.; Hooton, J. C.; Dawson, M. L.; Ferrie, A. R.; Price, R., Dehydration of trehalose dihydrate at low relative humidity and ambient temperature. *Int. J. Pharm.* **2006**, *313* (1-2), 87-98.

## **CHAPTER 3.     FOURIER TRANSFORM FLUORESCENCE RECOVERY AFTER PHOTBLEACHING WITH DIFFRACTIVE OPTICAL ELEMENTS**

Fourier transform-fluorescence recovery after photobleaching (FT-FRAP) with diffractive optical elements (DOEs) is demonstrated for improved evaluation of normal diffusion in biologically relevant media. Characterization of diffusion is routine for the assessment of mobility in cell biology and lends insight into *in vivo* bioavailability of various drug products and biotherapeutics. Conventional point-bleach fluorescence recovery after photobleaching (FRAP) can measure diffusion over a micrometer scale and is noninvasive with low sample volume requirements. While conventional FRAP is complicated by sample heterogeneity, variations in the bleaching beam profile, and limitations in signal-to-noise, FT-FRAP was developed with patterned illumination to overcome these complications. Periodic “comb-bleach” illumination of lines was shown to reduce bias from sample heterogeneity and simplify the calculation of the diffusion coefficient by removal of the beam profile dependence. In the architecture described in this chapter, FT-FRAP was enabled through custom modification of a commercial instrumentation for conventional HT-FRAP to integrate DOEs supporting 2D patterned illumination for FT-FRAP directly compatible with high-throughput automated analyses.

### **3.1 Introduction**

FRAP is a technique commonly used for the characterization of mobility in various disciplines such as early-stage rejection of monoclonal antibody (mAb) candidates likely to exhibit low bioavailability upon subcutaneous injection.<sup>1-2</sup> Molecular mobility in FRAP is assessed by photobleaching a fluorescent sample with a conventionally Gaussian beam profile using a high-powered beam of light. For mobile samples, the molecules that have been photobleached will diffuse out of the bleaching region and be replaced by other non-photobleached fluorescent molecules. This replacement then produces a recovery in the fluorescence intensity in the bleached region that can be measured over time. With knowledge of the initial bleach shape, the fluorescent recovery can then be fit to a mathematical model allowing for determination of fluorescently labeled analyte within the medium.<sup>3</sup>

FRAP was first performed by Peters et al. in 1974 to measure the diffusion of protein in the membrane of red blood cell ghosts.<sup>4</sup> In studies more closely aligned with assessing subcutaneous mobility in an *in vitro* assay, FRAP has been used to study many other systems such as protein aggregates,<sup>5</sup> diffusivity in microporous tissue engineering scaffolds,<sup>6</sup> diffusion anisotropy in porcine ligaments,<sup>7</sup> and membrane protein mobility in lipidic mesophases.<sup>8</sup> Furthermore, FRAP has been used in the study of pharmaceutical materials to study molecular transport in hydrogels,<sup>9-12</sup> and extracellular matrices to improve drug delivery methods.<sup>13-14</sup>

Despite its broad applicability, FRAP has several limitations. One such limitation is that FRAP requires the molecule of interest to fluoresce, which is most often achieved by labeling. Labile positions for labeling are often at charge-carrying function groups, such that fluorescent labeling can change both long-range electrostatic interactions and short-range local interactions.<sup>15</sup> Consequently, mobility of the protein target may not be representative of the mobility of the unlabeled protein of interest. Additionally, complications can arise from FRAP measurements themselves such that FRAP models often require precise knowledge of the photobleaching profile for accurate assessment of diffusion. However, high laser power used for improved signal-to-noise enhancement can cause bleaching saturation affecting the overall photobleaching profile and can skew calculated values of the diffusion coefficient by over an order of magnitude.<sup>16</sup> Local heating effects can also arise from use of high laser power, which can also further skews diffusion measurements.<sup>17</sup> Moreover, heterogeneity of the fluorescent sample can likewise bias results from FRAP measurements. Since conventional FRAP measurements are most routinely performed by photobleaching a small Gaussian point for high signal-to-noise and fast recovery times, this region of the sample may not be representative of the overall diffusion characteristics of a heterogeneous or structured sample. Lastly, conventional, point-bleaching FRAP is commonly insensitive to anomalous diffusion (diffusion that deviates from normal Brownian diffusion) since the collected recovery curves can exhibit relatively subtle differences between samples that diffuse normally and anomalously.<sup>18</sup>

To overcome some of these limitations present in conventional FRAP, FT-FRAP has been developed by our group, in which a fluorescent sample is bleached by patterned illumination. By use of a periodic comb of lines photobleach pattern, analysis of the bleaching profile in Fourier space has been shown to allow high signal-to-noise recovery of the diffusion coefficient for both normal and anomalous diffusion.<sup>19-20</sup> Since patterned illumination in real-space photobleaches

over a larger field of view (FOV), many of the previously stated limitations and biases found in conventional FRAP are removed for FT-FRAP (such as heterogeneity and local heating effects).<sup>19-</sup>

<sup>20</sup> Likewise, since illumination and photobleaching was distributed over a larger region of the FOV by periodic pattern, the overall signal-to-noise ratio (SNR) in spatial frequency domain is greater allowing for higher statistical confidence in the recovered diffusion coefficients. Moreover, interrogation of higher harmonics in spatial frequency domain (corresponding to diffusion at varying length scales) allows for increased statistical confidence in the recovered diffusion coefficient for both normal and anomalous diffusion.

While FT-FRAP allows for the removal of many of the limitations found in conventional FRAP measurements, the initial system used previously to demonstrate FT-FRAP measurements is limited to periodic line bleaching and is not equipped for high-throughput measurements. Furthermore, since the system is restricted to periodic line bleaching (and the spatial frequency of this pattern is a comb of dots in one dimension), diffusion is only probed in one dimension. Thus, the analysis of anisotropic samples described by a diffusion tensor (i.e. samples in which diffusion in certain directions is limited) will only report on the component of diffusion along the comb axis with this initial architecture.

In this work, the design and implementation of FT-FRAP with DOEs is presented allowing for variations in bleach pattern allowing for investigation of anisotropic samples and high-throughput analysis. Briefly, DOEs were incorporated into a commercial FRAP instrument that modifies the conventional point-bleaching profile into patterned illumination. The DOEs are small plastic optics in which the surface was modified by lithographic methods such that the microstructure relief profile diffracts a beam of light into a predefined pattern in the spatial Fourier domain. To allow for measurement of diffusion by FT-FRAP in multiple directions, a dot-array bleach pattern was chosen for photobleaching. Performing the diffusion analysis in the spatial frequency domain following this patterned illumination allowed for the stated benefits of FT-FRAP (i.e. increased signal, reduced bias from heterogeneity, simplified the mathematical model by removing dependence on the photobleach profile, enabled single-shot characterization of anomalous diffusion), but also allows for analysis of anisotropic diffusion and high-throughput analysis.



## 3.2 Theory

### 3.2.1 Normal diffusion

As described in-detail previously,<sup>19-20</sup> Fick's Law of Diffusion (shown in Equation 4.1) in which  $C(\mathbf{p}, t)$  is the concentration of the analyte of interest as a function of position  $\mathbf{p}$  and time  $t$ , and  $\mathbf{D}$  is the three-dimensional diffusion tensor, is arguably simpler to solve after spatial Fourier transformation. This is shown in Equation 4.2, in which  $\tilde{C}$  is the spatial Fourier transform of  $C$  and  $\bar{\mathbf{v}}$  is the 3D spatial wavevector.

$$\frac{\partial}{\partial t} C(\mathbf{p}, t) = \nabla \mathbf{D} \nabla C(\mathbf{p}, t) \quad 4.1$$

$$\frac{\partial}{\partial t} \tilde{C}(\bar{\mathbf{v}}, t) = -4\pi^2 \bar{\mathbf{v}}^T \mathbf{D} \bar{\mathbf{v}} \tilde{C}(\bar{\mathbf{v}}, t) \quad 4.2$$

In the case of isotropic diffusion, the one-dimensional solution to Equation 4.2 is given by Equation 4.3.

$$\tilde{C}(\bar{\mathbf{v}}_x, t) = \tilde{C}(\bar{\mathbf{v}}_x, 0) e^{-(2\pi\bar{\mathbf{v}}_x)^2 D_{xx} t} \quad 4.3$$

For a given distance from the origin in the spatial frequency domain  $\bar{\mathbf{v}}$ , the time-dependent behavior is a simple exponential decay with a decay constant dependent on the diffusion coefficient and  $\bar{\mathbf{v}}$ . Note that for a single distance from the origin  $\bar{\mathbf{v}}$ , the decay constant has no dependence on the initial distribution of the photobleach  $\tilde{C}(\bar{\mathbf{v}}, 0)$ . Since multiplication by a Gaussian in the spatial Fourier domain corresponds to convolution with a Gaussian in real space, Equation 4.3 can further be simplified to Equation 4.4.

$$C(x, t) = C(x, 0) \otimes \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{x^2}{2\sigma_t^2}} \quad 4.4$$

The standard deviation of the spatial Gaussian distribution  $\sigma_t$  (equal to  $\sqrt{2Dt}$ ) increases with the square root of time, corresponding to convolution with a Gaussian broadening over time as diffusion proceeds.

Following the case of one-dimensional, isotropic diffusion, FRAP analysis can be extended to two-dimensional diffusion. First assuming isotropic diffusion (i.e. diffusion in the  $x$  and  $y$ -direction are dictated by the same diffusion coefficient), Equation 4.4 can be extended to Equation 4.5 such that the spatial Gaussian distribution  $\sigma_t$  is still equal to  $\sqrt{2Dt}$ .

$$C(x, y, t) = C(x, y, 0) \otimes \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(x^2+y^2)}{2\sigma_t^2}} \quad 4.5$$

For the case in which diffusion is not isotropic, the diffusion coefficient dictating diffusion along the x and y-direction is no longer a single scalar value and must be defined in terms of the diffusion tensor. In the limit of a thin sample, diffusion in the z-direction can be neglected and the diffusion tensor has only four terms ( $D = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$ ). Since diffusion from x to y, and y to x should be equivalent, the diffusion tensor has only three unique terms:  $D_{xx}$ ,  $D_{xy}$ , and  $D_{yy}$ .<sup>21</sup> Extending Equation 4.5 into these unique terms, FRAP analysis is extended to Equation 4.6.

$$C(x, y, t) = C(x, y, 0) \otimes \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_x\sigma_y}} e^{-\frac{2xy}{2\sigma_x\sigma_y}} \cdot \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{y^2}{2\sigma_y^2}} \quad 4.6$$

For this case, Where  $\sigma_x^2$  is equal to  $2D_{xx}t$ ,  $\sigma_y^2$  is equal to  $2D_{yy}t$ , and the cross-term  $\sigma_x\sigma_y$  is equal to  $2D_{xy}t$ .

### 3.2.2 FT diffusion

The simplified mathematical form to the solution for normal diffusion, given in Equation 4.3 enables the simplified analysis of FRAP experiments in the spatial frequency domain. To take advantage of this simplified form, bleach patterns were chosen to concentrate FRAP signal in a collection of discrete spatial frequencies, as the decay rate of each of these individual spatial frequencies is independent of point-spread function of the non-diffracted photobleach beam.<sup>19-20</sup> In previous work done by Geiger et al. on FT-FRAP,<sup>19-20</sup> use of a periodic comb bleach pattern was demonstrated. The comb pattern is represented by a series of evenly spaced photobleached stripes shown in Figure 4.1 A, and the spatial frequency of this pattern shown in Figure 4.1 B is a line of evenly spaced impulses. The dot-array pattern used in this work is represented by a set of evenly spaced photobleached spots in two dimensions shown in Figure 4.1 C. In the spatial frequency domain, the dot-array pattern transforms to another dot-array of evenly spaced impulses shown in Figure 4.1 D.

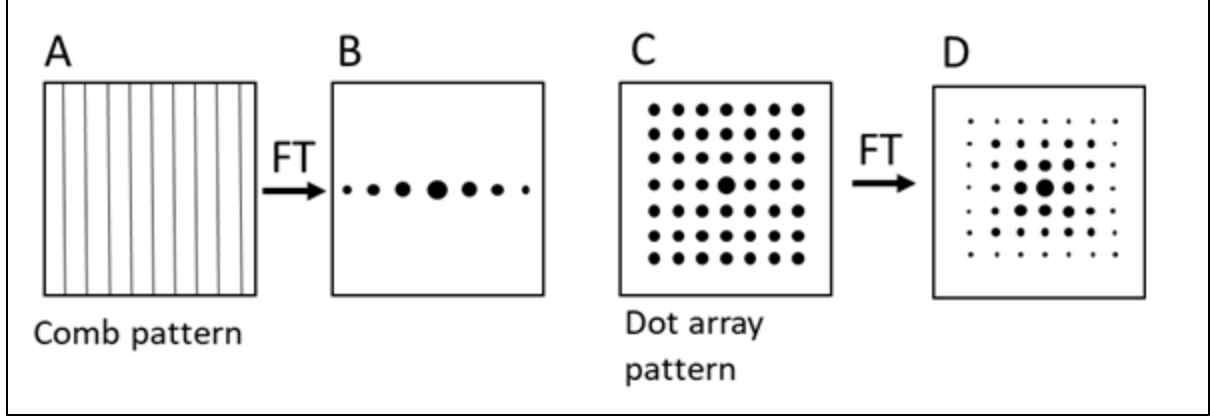


Figure 3.1. Schematic depicting bleaching patterns with A) a comb pattern, which Fourier transforms to B) a one-dimensional set of equally spaced puncta, and C) a dot-array pattern, which Fourier transforms to D) a two-dimensional set of equally spaced puncta.

For the comb bleach pattern, the time-dependent behavior of each point in the FT can be evaluated by an approach analogous to that of Equation 4.3 for the  $n$ th harmonic ( $n > 0$ ). For a given impulse in the spatial frequency domain, a single-exponential decay is expected irrespective of the initial bleaching pattern. The functional form of this signal decay is shown in equation 4.7 such that the initial photobleach corresponds to the series of puncta in the spatial Fourier domain positioned at  $(\bar{v}_x, \bar{v}_y) = (\pm n\bar{v}^0, 0)$ .

$$\widetilde{C}_n(\bar{v}_x, t) = \tilde{C}(\pm n\bar{v}_x^0, 0)e^{-4\pi^2(n\bar{v}_x^0)^2 D_{xx}t} \quad 4.7$$

Each of these puncta are initially scaled by the amplitude of the spatial Fourier transform of the beams point spread function, and additional amplitude at the origin puncta arises from the average fluorescence intensity of the sample. Fitting the single-exponential recovery for a given harmonic of the comb bleach then allows for recovery of the diffusion coefficient, while fitting to multiple harmonics allows for greater confidence in the recovered diffusion coefficient.

For the dot-array pattern, Equation 4.7 is further expanded to include diffraction peaks in two-dimensional Fourier space. These peaks are directly analogous to reflections off lattice planes in 2D periodic materials as measured by electron and X-ray diffraction, with the indices  $(h, k)$  of a given “reflection” in the FT image corresponding to the 1D slices through the 2D lattice. For a square lattice such that  $\bar{v}_x^0 = \bar{v}_y^0 = \bar{v}^0$ , each reflection is defined such that  $n^2 = h^2 + k^2$  (where  $h$  defines harmonics along one dimension, and  $k$  in the orthogonal dimension), Equation 4.7 becomes

$$\widetilde{C}_{h,k}(\bar{v}, t) = \tilde{C}(h\bar{v}_x^0, k\bar{v}_y^0, 0)e^{-4\pi^2(h^2 D_{xx} + 2hk D_{xy} + k^2 D_{yy})\bar{v}^0{}^2 t} \quad 4.8$$

where  $\bar{v}_0$  is the reciprocal of the spatial lattice constant for the periodic illumination. Thus, for each harmonic in the spatial frequency domain, the diffusion coefficient can be recovered by fitting to the single-exponential decay. For the case of anisotropic diffusion, or samples that were not expected to have normal diffusion, orthogonal harmonics [e.g.  $(h,k) = (0,1)$  and  $(h,k) = (1,0)$ ] can be fit to separate diffusion coefficients to recover the full diffusion tensor. For samples with the principal axes of diffusion not aligned with the angle of the dot-array pattern, diagonalization of the experimentally acquired diffusion tensor can allow for calculation of the diffusion tensor in the principal frame.

For samples that are expected to contain only normal, isotropic diffusion, peak intensities from analogous reflections were pooled together to calculate the diffusion coefficient with a higher statistical confidence. For this case of isotropic diffusion, or for a diffusion tensor defined relative to the principle moments, the off-diagonal elements are zero valued (e.g.  $D_{xy} = D_{yx} = 0$ ). Additionally, diagonal elements of the diffusion tensor are equivalent for isotropic diffusion such that Equation 4.8 can be simplified to Equation 4.9, and only a single diffusion coefficient is calculated.

$$\widetilde{C}_{h,k}(\bar{\mathbf{v}}, t) = \tilde{C}(h\bar{v}^0, k\bar{v}^0, 0)e^{-4\pi^2(h^2+k^2)\bar{v}^{0^2}D_{xx}t} \quad 4.9$$

### 3.2.3 Signal power

Bleaching with patterned illumination yield high SNR decay curves for diffusion characterization, which arises primarily by distribution of the photobleaching power across the entire FOV.<sup>19-20</sup> The signal power in conventional FRAP can be quantified by integrating over a typical Gaussian photobleach spot, represented by Equation 4.10, in which  $A_p$  is photobleach depth.

$$P_{Gaussian} \cong (C_0 A_p)^2 \int_{-\infty}^{\infty} e^{-\frac{x'^2}{\sigma^2}} dx' \int_{-\infty}^{\infty} e^{-\frac{y'^2}{\sigma^2}} dy' \cong \pi \sigma^2 (C_0 A_p)^2 \quad 4.10$$

In comparison, the signal power of a dot-array photobleach pattern of  $N \times N$  dots can be described by Equation 4.11 in which  $A_{n,m,p}$  is the photobleach depth of each dot of the dot-array pattern.

$$P_{dot-matrix} \cong \sum_{n=1}^N \sum_{m=1}^N \pi \sigma^2 (C_0 A_{n,m,p})^2 = N^2 P_{Gaussian} \quad 4.11$$

Assuming that the photobleach depth is equivalent for each spot of the dot-array pattern, the signal power advantage of this patterned illumination is approximately  $N^2$ -fold in this case (e.g. for the  $13 \times 13$  dot-array patterned used in this study, illumination power is increased 169-fold).

Conventional FRAP photobleaches only a small percentage of the total fluorophores in the FOV, which puts a ceiling on the total accessible signal. Patterned illumination enables access to greater FRAP signal by expanding the photobleaching region. Since white noise is uniformly distributed in both real-space and Fourier space, localization in the FT domain results in noise powers comparable to those produced by point-bleach measurements in real-space. From these combined considerations, patterned illumination provides substantial opportunities for signal to noise enhancement.

### **3.2.4 Simultaneous conventional FRAP and FT-FRAP**

Since a dot-array illumination pattern was chosen for bleaching, simultaneous conventional FRAP and FT-FRAP is possible. By isolation of the zero-order (non-diffracted) bleach spot of the dot-array pattern, a Gaussian fit using Equation 4.4 can be done on this spot for the collected image stack to calculate the diffusion coefficient. Simultaneously, transformation into the spatial frequency domain can be used to recover an independently calculated diffusion coefficient using FT-analysis. Conventional FRAP could likewise be performed on all 13 x 13 bleach spots of the dot-array bleach pattern. For a homogenous sample that exhibits only normal diffusion, each of the recovered diffusion coefficients would ideally yield statistically indistinguishable values. For heterogenous samples, or samples containing anisotropic or anomalous diffusion, recovery of the diffusion coefficient for each spot could lend insights into variations in local environment or local heating effects.

## **3.3 Experimental methods**

### **3.3.1 Sample preparation**

Fluorescently labeled goat anti-mouse immunoglobulin G (IgG) secondary antibody rhodamine (Invitrogen) and albumin from bovine serum (BSA) tetramethylrhodamine conjugate (Invitrogen) were dissolved in phosphate-buffered saline (PBS) at varying concentrations. Matrix solutions of hyaluronic acid (HA) (company) and collagen (company) were prepared in PBS and salt solutions of NaCl were prepared in PBS as well. The pH of these solutions was adjusted using NaOH and HCl to a range of pH 6.5 to 8. A 96 well microplate (Sigma Aldrich) was prepared using a Dragonfly® Crystal instrument (SPT Labtech) to create a solution plate with varying

amounts of salt, pH, and buffer. A second microplate was prepared by pipetting either of the buffered protein solutions into a single column of the 96 well plate, and pipetting the desired matrix solutions into their own columns as well. Both microplates were placed onto slots of a Mosquito® LV instrument (SPT Labtech) which was used to prepare a final 96 well lipidic cubic phase (LCP) screening plate (Hampton Research). Each well of the LCP plate corresponds to a fixed amount of protein solution dropped inside of varying amounts of buffer, matrix, and salt solutions. A top seal was placed onto the LCP plate to seal the plate. Finally, the plate was allowed to homogenize inside of a desiccator at 75% relative humidity (RH) for a day before imaging.

### **3.3.2 Dot – array FRAP**

A commercial FRAP microscope (Formulatrix, Bedford, MA) was modified in-house to support periodically patterned illumination during photobleaching for FT-FRAP experiments, as shown in Figure 4.2. A detailed description of the commercial microscope has been published previously.<sup>22</sup> A diffractive optical element (DOE) (DE-R 244) purchased from HOLOEYE Photonics AG (Berlin, Germany) was added to the excitation light path of the instrument immediately before the second dichroic mirror. The DOE diffracted the photobleach laser into a matrix of photobleach spots during the photobleaching phase of the FRAP experiment (i.e. a dot-array pattern). Distributing the photobleach power across many points in the sample enabled illumination with a higher power light source for photobleaching without saturation of the sample. Consequently, the ~10 mW, 532 nm photobleaching laser in the FRAP microscope was replaced with a 5 W, 532 nm, continuous-wave laser (SpectraPhysics Millennia SpectraPhysics, Santa Clara, CA), which was coupled into the microscope with a Thorlabs custom fiber optic patch cable at the location of the original photobleaching laser enabling ~100 mW of power at the sample plane. Since replacement of the original laser removed automated control of bleaching, an automated shutter was added to the instrument. The signal output from the original laser connector was used to open and close the automated shutter to allow for automated control of bleach times once more. The photobleach time was approximately five seconds and produced a dot-array pattern that could be analyzed in the spatial frequency domain for diffusion characterization. Post-bleach images were then collected for 60 – 120 seconds after bleaching at a rate of one frame per second.

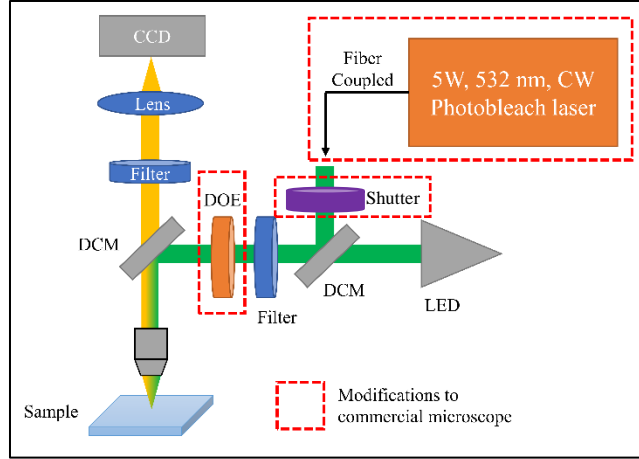


Figure 3.2. Instrument schematic of a commercial FRAP system that was modified for FT-FRAP with DOEs. Modifications are indicated by red dashed boxes. A fiber-coupled laser replaced the original photobleaching laser for higher bleaching power, an automated shutter was added to control bleach times and a DOE was added after the first dichroic mirror to produce a dot-array photobleach pattern. HWP = half-wave plate, DCM = dichroic mirror, SPF = short-pass filter, TPEF = two-photon excited fluorescence, PMT = photomultiplier tube, DOE = diffractive optical element, CCD = charge-coupled device camera, LED = broadband green light emitting diode.

### 3.3.3 Data analysis

Images from FT-FRAP experiments were analyzed using a custom image processing software written in-home in MATLAB. Diffusion properties were recovered from an image stack using a nonlinear least-squares fit to Equation 4.3 for normal diffusion. For the FT-FRAP with DOE analysis, image stacks were cropped to a square around the dot-array pattern centered on the center bleach spot. The image stack was then converted to the spatial frequency domain using the two-dimensional fast Fourier transform function in MATLAB. A nonlinear least-squares fits to Equation 4.10 was done for each harmonic. For the dot-array bleaching, simultaneous analysis of the from the zero-order beam using conventional FRAP and the higher-order diffractions using FT-FRAP was performed. Uncertainties in the fit for conventional FRAP and FT-FRAP were obtained from the covariance matrix,  $\mathbf{X}$  which was calculated from the Jacobian,  $\mathbf{J}$  and the variance in the residuals,  $v_R$ , as shown in Equation 4.12.

$$\mathbf{X} = v_R(\mathbf{J}^T \times \mathbf{J})^{-1} \quad 4.12$$

### 3.4 Results and discussion

#### 3.4.1 FT-FRAP of fluorescently labelled IgG in HA/collagen

Patterned photobleaching using the dot-array pattern was performed on IgG in matrices of HA or collagen on a 96 well LCP screening plate. Protein concentration was constant throughout the plate, while the concentrations of HA and collagen were constant for each column of the 96 well plate, but varied between columns of the plate. Figure 4.3 A shows the first post-bleach image of IgG in HA following exposure with the dot-array pattern. The dot-array pattern was rotated slightly to avoid overlap of the horizontal and vertical features of the image in the spatial frequency domain associated with the rectangular nature of digital images. The Fourier transform of the image after normalization and cropping to the bleach pattern in Figure 4.3 A is shown in Figure 4.3 B such that another dot-array pattern arises as expected in the spatial frequency domain. For each peak in Figure 4.3 B, Equation 4.10 was used to fit the peak intensity throughout the entire image stack to recover the diffusion coefficient using the appropriate  $h$  and  $k$  values.



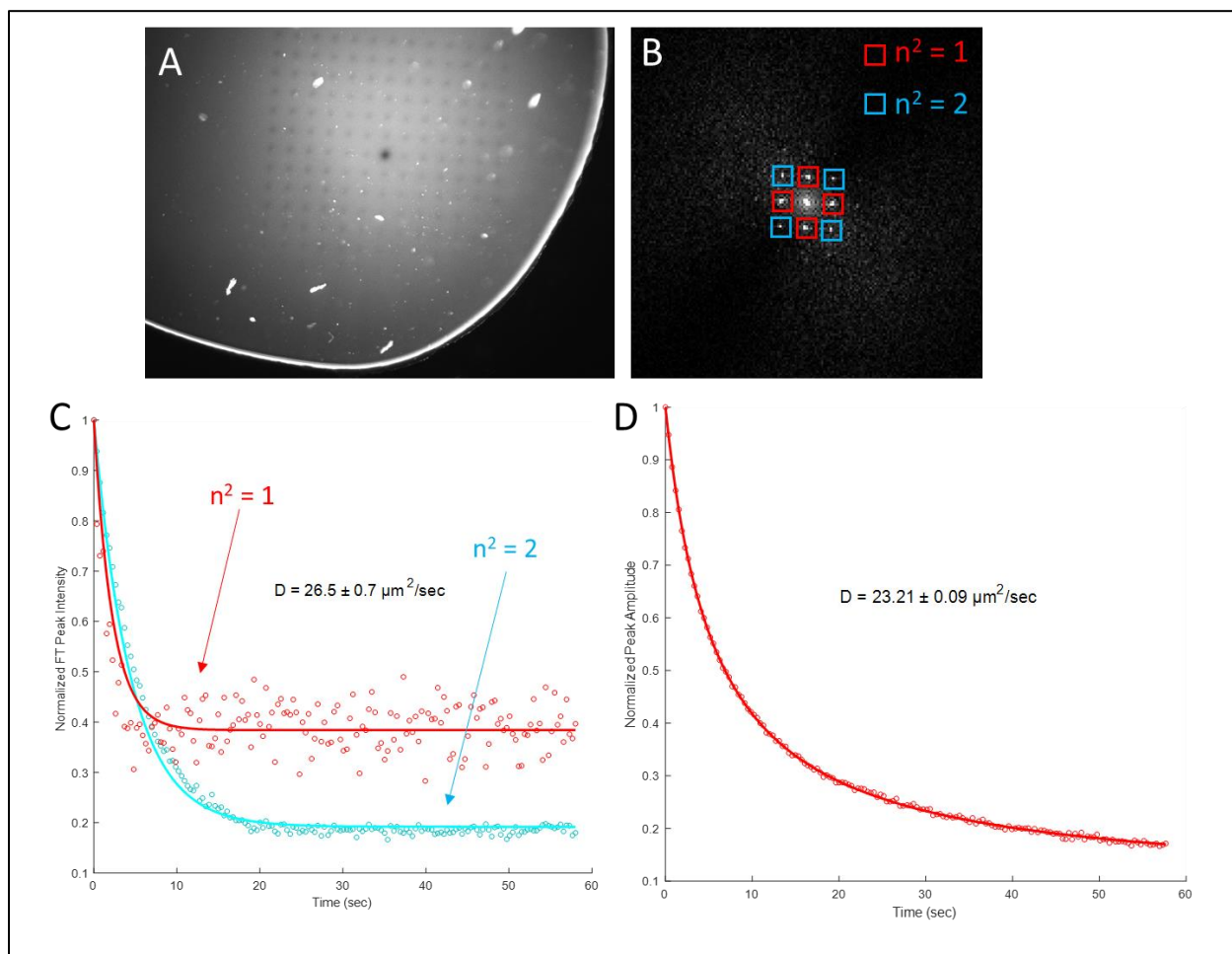


Figure 3.3. FT-FRAP with dot array bleach of IgG dissolved in PBS in HA matrix. A) First image after photobleaching with the dot-array pattern. B) 2D Fourier transform of (A). Blue and red squares indicate peaks that are equidistant from the center peak C) Fluorescence recovery of pooled peaks with best-fit curves for FT-FRAP. The best-fit value for the diffusion coefficient was  $26.5 \pm 0.07 \mu\text{m}^2/\text{sec}$ . D) Fluorescence recovery using conventional FRAP, and the best-fit value for the diffusion coefficient was  $23.21 \pm 0.09 \mu\text{m}^2/\text{sec}$ . The reported uncertainty is the standard deviation of the fit for both fits.

As in 2D diffraction patterns seen for X-ray imaging, the 2D peaks in spatial frequency domain corresponds to a lattice plane in reciprocal space defined by  $h$  and  $k$ . The zero-order non-diffracted peak represents the average fluorescence intensity of the image and contains no information on sample diffusion. Theory predicts that the decay in peak intensity for each of these peaks is dependent on both the diffusion coefficient and the Euclidean distance of this peak from

the zero-order non-diffracted peak hence the presence of the  $h^2 + k^2$  term in equation 4.10. Due to this prediction, peaks with identical values of  $n^2$  (equal to  $h^2 + k^2$ ) are expected to measure diffusion along the same interval of distance. Thus, analogous peaks are color coded in Figure 4.3 B (e.g. the peaks highlighted in red each represent the lowest frequency spacing of the dot-array bleach pattern in orthogonal directions and measure diffusion along the greatest distance interval).

For homogenous samples that are expected to contain only normal diffusion, peak intensities from analogous harmonics were pooled together to calculate diffusion from Equation 4.10. Time-dependent curves for harmonics corresponding to  $n^2 = 1$  or 2 are plotted in Figure 4.3 C as well as their corresponding fit. For IgG in HA (or collagen), the decay curves and fits were in good agreement with simple exponential trends as predicted by theory for normal diffusion. The diffusion coefficient calculated for this specific concentration of HA was  $26.5 \pm 0.7 \mu\text{m}^2/\text{second}$ , where the uncertainty is the standard deviation of the combined fit.

### **3.4.2 Simultaneous conventional FRAP of IgG in HA**

For the same sample of IgG in HA, conventional FRAP was simultaneously performed on the center bleach spot. Fitting this center bleach to a time-dependent Gaussian using Equation 4.4, the decay curve and corresponding fit are shown in Figure 4.3 D. The diffusion coefficient calculated for IgG in this specific concentration of HA was  $23.21 \pm 0.09 \mu\text{m}^2/\text{sec}$ . Thus, diffusion coefficients calculated by FT-FRAP and conventional FRAP were comparable to each other and this sample was determined to be both homogenous and exhibit normal diffusion. For other wells of the 96 well LCP plate, conventional FRAP and FT-FRAP yields statistically different diffusion coefficients. Possible origins of this disparity are variations in autofocus of the FRAP instrument, as well as potentially in image contrast or protein concentration due to sample prep error. Work is being done on determination of the source of this error and the addition of outlier rejection in the data analysis. Lastly, the intensity and bleach depth of the spots other than the zero-order non-diffracted peak were low. Thus, these spots were generally short-lived, and recovery of the diffusion coefficient yielded a low SNR.

### 3.5 Conclusion

Patterned bleaching was successfully demonstrated using DOEs in a commercial FRAP instrument allowing for high-throughput FT-FRAP of biologically relevant media. FT-FRAP greatly simplifies the mathematical analysis of diffusion by analysis in the spatial frequency domain. Furthermore, patterned illumination was shown to greatly increase the accessible bleach power and allow for measurement of diffusion over a greater portion of the FOV than with conventional FRAP. Thus, FT-FRAP with DOEs is less error prone to noise contributions due to sample heterogeneity, anisotropic/anomalous diffusion, and local heating and was demonstrated for IgG in HA and collagen matrices. Simultaneous FT-FRAP and conventional FRAP was likewise demonstrated for dot-array bleaching, but currently FT-FRAP produced a lower statistical certainty in the recovered diffusion coefficient than conventional FRAP. For samples exhibiting anisotropic diffusion, the dot-array pattern is predicted to allow determination of the diffusion tensor for a given sample. For samples not aligned with the angle of the dot-array pattern, coordinate transform should allow for calculation of the diffusion tensor in the desired frame. Thus, HT-FT-FRAP using DOEs is well suited for the assessment of mobility in cell biology and to lend insights into *in vivo* bioavailability of biotherapeutics.

### 3.6 References

1. Loren, N.; Hagman, J.; Jonasson, J. K.; Deschout, H.; Bernin, D.; Cella-Zanacchi, F.; Diaspro, A.; McNally, J. G.; Ameloot, M.; Smisdom, N.; Nyden, M.; Hermansson, A. M.; Rudemo, M.; Braeckmans, K., Fluorescence recovery after photobleaching in material and life sciences: putting theory into practice. *Q. Rev. Biophys.* **2015**, 48 (3), 323-387.
2. Reits, E. A. J.; Neefjes, J. J., From fixed to FRAP: measuring protein mobility and activity in living cells. *Nat. Cell Biol.* **2001**, 3 (6), E145-E147.
3. Beaudouin, J.; Mommer, M.; Bock, H.; Eils, R., Experiment Setups and Parameter Estimation in Fluorescence Recovery After Photobleaching Experiments: A Review of Current Practice. **2013**.
4. Peters, R.; Peters, J.; Tews, K. H.; Bähr, W., A microfluorimetric study of translational diffusion in erythrocyte membranes. *Biochimica et biophysica acta* **1974**, 367 (3), 282-94.
5. Babinchak, W. M.; Surewicz, W. K., Studying Protein Aggregation in the Context of Liquid-liquid Phase Separation Using Fluorescence and Atomic Force Microscopy, Fluorescence and Turbidity Assays, and FRAP. *Bio-protocol* **2020**, 10 (2).

6. Offeddu, G. S.; Mohee, L.; Cameron, R. E., Scale and structure dependent solute diffusivity within microporous tissue engineering scaffolds. *J. Mater. Sci.-Mater. Med.* **2020**, *31* (5), 11.
7. Hashlamoun, K.; Abusara, Z.; Ramírez-Torres, A.; Grillo, A.; Herzog, W.; Federico, S., Fluorescence recovery after photobleaching: direct measurement of diffusion anisotropy. *Biomechanics and modeling in mechanobiology* **2020**, *19* (6), 2397-2412.
8. Cherezov, V.; Liu, J.; Griffith, M.; Hanson, M. A.; Stevens, R. C., LCP-FRAP Assay for Pre-Screening Membrane Proteins for In Meso Crystallization. *Cryst. Growth Des.* **2008**, *8* (12), 4307-4315.
9. Verheyen, E.; van der Wal, S.; Deschout, H.; Braeckmans, K.; de Smedt, S.; Barendregt, A.; Hennink, W. E.; van Nostrum, C. F., Protein macromonomers containing reduction-sensitive linkers for covalent immobilization and glutathione triggered release from dextran hydrogels. *J. Control. Release* **2011**, *156* (3), 329-336.
10. Branco, M. C.; Pochan, D. J.; Wagner, N. J.; Schneider, J. P., Macromolecular diffusion and release from self-assembled beta-hairpin peptide hydrogels. *Biomaterials* **2009**, *30* (7), 1339-1347.
11. Kosto, K. B.; Deen, W. M., Diffusivities of macromolecules in composite hydrogels. *Aiche J.* **2004**, *50* (11), 2648-2658.
12. Alvarez-Mancenido, F.; Braeckmans, K.; De Smedt, S. C.; Demeester, J.; Landin, M.; Martinez-Pacheco, R., Characterization of diffusion of macromolecules in konjac glucomannan solutions and gels by fluorescence recovery after photobleaching technique. *Int. J. Pharm.* **2006**, *316* (1-2), 37-46.
13. Remaut, K.; Sanders, N. N.; De Geest, B. G.; Braeckmans, K.; Demeester, J.; De Smedt, S. C., Nucleic acid delivery: Where material sciences and bio-sciences meet. *Mater. Sci. Eng. R-Rep.* **2007**, *58* (3-5), 117-161.
14. Peeters, L.; Sanders, N. N.; Braeckmans, K.; Boussery, K.; Van de Voorde, J.; De Smedt, S. C.; Demeester, J., Vitreous: A barrier to nonviral ocular gene therapy. *Invest. Ophthalmol. Vis. Sci.* **2005**, *46* (10), 3553-3561.
15. Toseland, C. P., Fluorescent labeling and modification of proteins. *J Chem Biol* **2013**, *6* (3), 85-95.
16. Braeckmans, K.; Stubbe, B. G.; Remaut, K.; Demeester, J.; De Smedt, S. C., Anomalous photobleaching in fluorescence recovery after photobleaching measurements due to excitation saturation- a case study for fluorescein. *J. Biomed. Opt.* **2006**, *11* (4), 13.

17. Heinze, K. G.; Costantino, S.; De Koninck, P.; Wiseman, P. W., Beyond Photobleaching, Laser Illumination Unbinds Fluorescent Proteins. *J. Phys. Chem. B* **2009**, *113* (15), 5225-5233.
18. Daddysman, M. K.; Fecko, C. J., Revisiting Point FRAP to Quantitatively Characterize Anomalous Diffusion in Live Cells. *J. Phys. Chem. B* **2013**, *117* (5), 1241-1251.
19. Geiger, A. C.; Smith, C. J.; Simpson, G. J., Multi-photon excited Fourier-transform fluorescence recovery after photobleaching (FT-FRAP) with patterned illumination. In *Multiphoton Microscopy in the Biomedical Sciences Xx*, Periasamy, A.; So, P. T.; Koenig, K., Eds. Spie-Int Soc Optical Engineering: Bellingham, 2020; Vol. 11244.
20. Geiger, A. C.; Smith, C. J.; Takanti, N.; Harmon, D. M.; Carlsen, M. S.; Simpson, G. J., Anomalous Diffusion Characterization by Fourier Transform-FRAP with Patterned Illumination. *Biophys. J.* **2020**, *119* (4), 737-748.
21. Shi, C. C.; Cisewski, S. E.; Bell, P. D.; Yao, H., Measurement of Three-Dimensional Anisotropic Diffusion by Multiphoton Fluorescence Recovery after Photobleaching. *Ann. Biomed. Eng.* **2014**, *42* (3), 555-565.
22. Fenalti, G.; Abola, E. E.; Wang, C.; Wu, B. L.; Cherezov, V., Fluorescence Recovery After Photobleaching in Lipidic Cubic Phase (LCP-FRAP): A Precrystallization Assay for Membrane Proteins. In *Membrane Proteins - Engineering, Purification and Crystallization*, Shukla, A. K., Ed. Elsevier Academic Press Inc: San Diego, 2015; Vol. 557, pp 417-437.

## **CHAPTER 4. POLARIZATION DEPENDENT DISCRIMINATION OF CRYSTALLINE AND AGGREGATE PROTEIN**

Polarization studies using two-photon excited ultraviolet fluorescence (TPE-UVF) enable a novel approach for discrimination between protein crystals and aggregate protein. For a given polarization state of light, the TPE-UVF response of aggregate protein and that of crystalline proteins is theorized to differ. This difference in response allows for discrimination of these two solid forms. Previous studies focused on the use of an electro-optic modulator (EOM) to produce varying polarization states of light, but the operation of EOMs often requires an advanced user. Likewise, EOM are often large and present challenges for retrofitting into existing instrumentation. By use of a commercially available microretarder array ( $\mu$ RA), this small optic modulated the polarization of light across the entire field of view (FOV) of a commercially available TPE-UVF instrument. Proof-of-concept polarization images were collected for solutions of tryptophan, and the circular to linear dichroism ratio was calculated for these images using Fourier analysis to differentiate between solid forms. Currently, work is being done to extend this work to crystalline samples of tryptophan and lysozyme before moving on to more interesting protein crystals.

### **4.1 Introduction**

In recent years, the use of proteins for various applications such as protein therapeutics,<sup>1-2</sup> biotechnology,<sup>3</sup> and bioinformatics<sup>4</sup> have attracted significant attention. By knowledge of a protein's structure and how said structure affects the array of protein functions (e.g. catalysis, DNA replication, transportation, etc.), modifications can be made to a protein (or synthesis of an entirely new protein) in order to perform new functions and tasks<sup>5</sup>. Thus, determination of the structure of a given protein using macromolecular crystallography allows for improvements to various protein applications.

In general, determination of a protein's composition is trivial,<sup>6</sup> but acquisition of a high-resolution structure of a protein is not.<sup>7</sup> In order to generate a high-resolution structure of a macromolecule such as a protein, the protein is commonly crystallized and probed by a sufficiently power X-ray source such as a synchrotron or X-ray free-electron laser (XFEL). High signal-to-noise diffractions patterns of the protein crystal are then collected and utilized in determination of

a protein's three-dimensional structure. Unfortunately, collection of sufficiently high signal-to-noise diffraction patterns is often highly destructive to the sample and multiple protein crystals are need for structural determination.<sup>8</sup> Thus, determination of a protein's structure is often dependent on the ability to reliably and efficiently crystallize a protein. For this reason, the ability to successfully detect and identify the presence of crystalline protein early in the crystallization process is crucial.

Currently, only a few methods exist that are able to detect and identify the presence of crystalline protein during crystallization trials. One such method for protein crystal identification is the use of second harmonic generation (SHG) microscopy. Due to the symmetry requirements of SHG, SHG can be used to accurately discriminate between crystalline or aggregate protein, as disordered aggregate protein will not produce an SHG signal while crystalline protein will.<sup>9</sup> SHG is however limited such that for centrosymmetric crystals are also generally undetectable due to the symmetry limitations, limiting the overall applicability of this method, but most proteins are homochiral and therefore cannot form a centrosymmetric crystal. TPE-UVF imaging on the other hand can be used as an alternative method to SHG imaging. In general, any molecule containing the aromatic residue tryptophan is TPE-UVF active such that most proteins can be detected by this method.<sup>10</sup> Unlike SHG imaging though, discrimination between crystalline and aggregate forms is generally not possible since both forms are TPE-UVF active. To overcome this limitation in TPE-UVF imaging, polarization methods are proposed to allow for the discrimination between solid forms. In brief, given a specific polarization state of light used for imaging, the TPE-UVF response to this polarization state should be different for crystalline forms and aggregate forms.<sup>11</sup> Thus, measuring this difference in responses for multiple polarizations should allow for discrimination of solid forms.

Commonly, modulation of the polarization state of light has been done using waveplates and polarizers, or an EOM allowing for modulation of phase, frequency amplitude, and/or the polarization state of light.<sup>12</sup> For the use of waveplates and polarizers, these optics are generally small and widely available, and placement of these optics into a commercial instrument or existing beam path is straight forward. Unfortunately though, these optics are commonly rotated manually inside of a mount to modulate the polarization of light desired for solid form discrimination.<sup>13</sup> Thus, collection of the TPE-UVF images can be time consuming and introduces  $1/f$  noise reducing the overall signal-to-noise.<sup>14</sup> Polarization modulation consisting of the placement of an EOMs in

the beam path has also been effective at controlling light polarization, but generally the large nature of these devices is difficult to fit into existing beam paths and commercially available instrumentation

Detailed in this work is the use of a  $\mu$ RA to modulate the TPE-UVF polarization across the FOV to enable single-shot discrimination between crystalline and aggregate protein. Using a commercially available, one-inch diameter liquid crystal polymer depolarizer (i.e. a  $\mu$ RA), this optic consists of an array of microretarding liquid crystals (or liquid crystal polymers) with the fast axis patterned to a specific angle. Thus incoming light is then affected by the orientation of this fast axis, and the polarization state is altered to another state. By placement of this optic in the rear-conjugate plane of a 4F coupled microscope, the polarization states of the imaging beam are subsequently patterned across the FOV.<sup>11</sup> Thus, multiple polarization states of light are able to be encoded over a single FOV. Figure 4.1 shows the changes in the orientation of modulation as a function of position on the  $\mu$ RA (figure taken from ThorLab's site). After collection of a single polarization-modulated image, Fourier analysis then allows for calculation of the circular and linear TPE-UVF responses over the entire FOV. Calculation of the ratio between these two responses can then be compared to experimentally determined responses for crystalline protein and aggregate protein for discrimination these forms. Proof-of-concept work was done for this proposed discrimination method for solutions of tryptophan before confirmation with actual protein crystals.

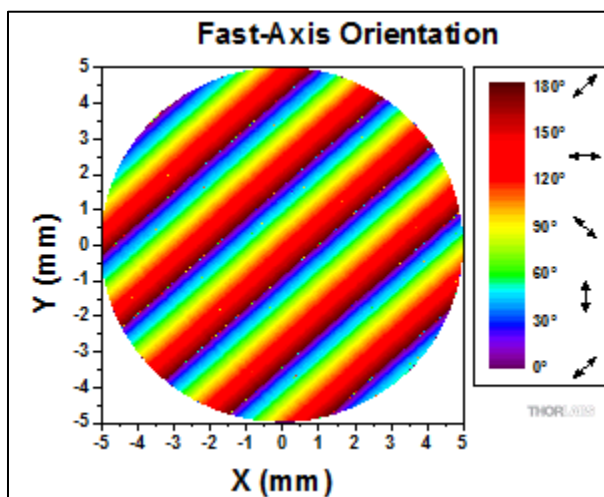


Figure 4.1. Polarization modulation pattern of light as a function of position on the  $\mu$ RA from Thorlabs.



## 4.2 Theoretical framework

### 4.2.1 Polarization dependence of TPE-UVF

For the theoretical framework for TPE-UVF polarization dependence, phase modulation of light can be related to a set of coefficients.<sup>15</sup> Assuming the molecular rotation is slow relative to the fluorescence lifetime, the polarization rotation angle ( $\gamma$ ) can be related using these coefficients to the TPE-UVF intensity ( $I_n$ ) where  $n$  is defined as either horizontal (H) polarization or vertical (V) polarization. Overall, ten unique coefficients can be calculated [five coefficients for horizontal (H) polarization, and five for vertical (V) polarization] that define this relationship in Equation 4.1 where A – E are the polynomial coefficients for the given polarization state (H or V).

$$I_n(\gamma) = A_n \cos^4(\gamma) + B_n \cos^3(\gamma) \sin(\gamma) + C_n \cos^2(\gamma) \sin^2(\gamma) + D_n \cos(\gamma) \sin^3(\gamma) + E_n \sin^4(\gamma) \quad 4.1$$

Equation 4.2 shows the calculation of these coefficients,  $\vec{A}_n$  is the vector of polynomial coefficients for polarization state  $n$ ,  $\mathbf{P}$  is the sparse matrix, and  $\chi_{nnn}$  is the Jones tensor for a given particle.

$$\vec{A}_n = \mathbf{P} \cdot (\vec{\chi}_J^* \otimes \vec{\chi}_J) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 4 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \chi_{nHH}^* \chi_{nHH} \\ \chi_{nHH}^* \chi_{nHV} \\ \chi_{nHH}^* \chi_{nVV} \\ \chi_{nHV}^* \chi_{nHH} \\ \chi_{nHV}^* \chi_{nHV} \\ \chi_{nHV}^* \chi_{nVV} \\ \chi_{nVV}^* \chi_{nHH} \\ \chi_{nVV}^* \chi_{nHV} \\ \chi_{nVV}^* \chi_{nVV} \end{bmatrix} \quad 4.2$$

For simplicity in the math, a particle is assumed to be uniaxial, and “rod-like” and thus have a single dominate element in the two-photon absorption coefficient ( $\beta_{zzz}$ ). Assuming the  $x$ -direction corresponds to H polarization and the  $y$ -direction to that of V polarization, the polynomial coefficients can then be calculated for a given angle of  $\varphi$  by Equations 4.3 and 4.4 (where  $\theta$ ,  $\varphi$ , and  $\psi$  are the three Euler angles corresponding to the polar, azimuthal, and twist angles of a given particle).

$$\vec{A}_H = \cos^2\varphi \begin{bmatrix} \cos^4\varphi \\ 4\cos^3\varphi\sin\varphi \\ 6\cos^2\varphi\sin^2\varphi \\ 4\cos\varphi\sin^3\varphi \\ \sin^4\varphi \end{bmatrix} |\beta_{zzz}|^2 \quad 4.3$$

$$\vec{A}_V = \sin^2\varphi \begin{bmatrix} \cos^4\varphi \\ 4\cos^3\varphi\sin\varphi \\ 6\cos^2\varphi\sin^2\varphi \\ 4\cos\varphi\sin^3\varphi \\ \sin^4\varphi \end{bmatrix} |\beta_{zzz}|^2 \quad 4.4$$

Since  $\beta$  is related to the its transition moment ( $\mu$ ) and its two-photon absorption cross section ( $\alpha$ ) (Equation 4.5), both of which are attainable from either computational calculations or reported values in literature, polynomial coefficients are calculatable for a given particle.

$$\beta = \mu \otimes \alpha \quad 4.4$$

Thus, TPE-UVF intensity can be related to the polarization modulation of light using Equation 4.1.

#### 4.2.2 Circular/Linear dichroism

The polynomial coefficients can be related to the response a system has to linearly polarized light versus that of circularly polarized light. To do so, the polynomial coefficients are first transformed to the related Fourier coefficients (a - e) by the relationship defined by Equation 4.4.<sup>15</sup>

$$\begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \frac{1}{8} \begin{bmatrix} 3 & 0 & 1 & 0 & 3 \\ 1 & 0 & -1 & 0 & 1 \\ 4 & 0 & 0 & 0 & -4 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} \quad 4.4$$

For TPE-UVF, the Fourier coefficients B and D can be reasonably assumed to be zero.<sup>15</sup> Furthermore, the A coefficient corresponds to parallel polarization with the incident fields, and E with cross-polarized light. Therefore, the linear response is equal to the sum of the A and E coefficients, and the circular contribution has contributions from all three non-zero coefficients scaling with a factor of 1/4, shown previously in Equation 4.1. From this, the circular and linear responses can be calculated from Equation 4.5.

$$\begin{bmatrix} \sigma_{LP} \\ \sigma_{CP} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1/4 & 1/4 & 1/4 \end{bmatrix} \begin{bmatrix} A \\ C \\ E \end{bmatrix} \quad 4.5$$

The ratio between  $\sigma_{LP}$  and  $\sigma_{CP}$  then has the potential to be used for the discrimination between crystalline and aggregate protein.

### 4.3 Methods

#### 4.3.1 Sample preparation

Solutions of tryptophan were prepared in ultrapure water at 15 mg/mL. The solutions were heated for 15 minutes at 80 °C until tryptophan was fully dissolved. The tryptophan solution was then allowed to cool to room temperature and placed on glass slides for imaging. Tryptophan crystals were crushed using a mortar and pestle to produce a distribution of crystals sizes, and were placed on glass slides for imaging. Lyophilized lysozyme (Sigma Aldrich) was placed on glass slides for imaging as well.

#### 4.3.2 Simulations

*Ab initio* calculations were performed on the indole ring of tryptophan, which was assumed to be the primary chromophore responsible for TPE-UVF. The indole molecule was optimized in Avogadro for an initial structural guess before optimization using the general atomic and molecular electronic structure system (GAMESS) software. A 6-311++G\* basis set was used with a B3LYP correlation function to optimize the structure of the indole ring, which was then used for further calculations.

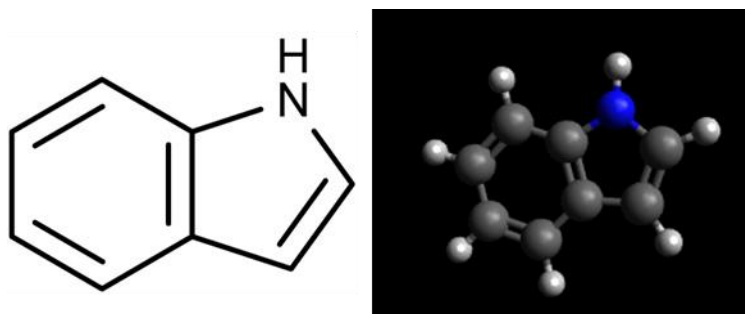


Figure 4.2. Skeletal (bond-line) structure of the indole ring, and the electronically optimized structure.

Once the structure was optimized, the isotropic two-photon absorption cross section ( $\alpha$ ) and transition moment ( $\mu$ ) were calculated using time-dependent density functional theory (TDDFT) with the same basis set for structure optimization and restricted Hartree Fock (RHF). The Kronecker product of the two-photon absorption cross section and transition moment was then used to determine the two-photon absorption coefficient ( $\beta$ ).<sup>15</sup>  $\beta$  was then used to calculate the Fourier coefficients for tryptophan by means of rotationally averaging and coordinate transformations. The linear to circular dichroism ratio from *ab initio* calculations was then calculated and compared to experimentally acquired values.

#### 4.3.3 Data acquisition

TPE-UVF images were acquired from the commercial SONICC system from Formulatrix in the epi direction. A  $\mu$ RA purchased from Thorlabs with workable wavelength range of 350 – 700 nm and a quarter-wave plate (QWP) were placed into the rear-conjugate plane of the existing beam path. The combined optics allowed for half-wave retardance of 532 nm light modulating polarization across the entire FOV. Calibration of the optics was performed by imaging the one-photon fluorescence of rhodamine 6G with a polarizer placed prior to the objective. Fluorescent signal was minimized by rotation of the polarizer and the acquired calibration image was used to calculate polarization states across the FOV. Images were then collected at max laser power (100 mW) for the tryptophan and lysozyme samples. The instrument schematic of this system is shown in Figure 4.3.

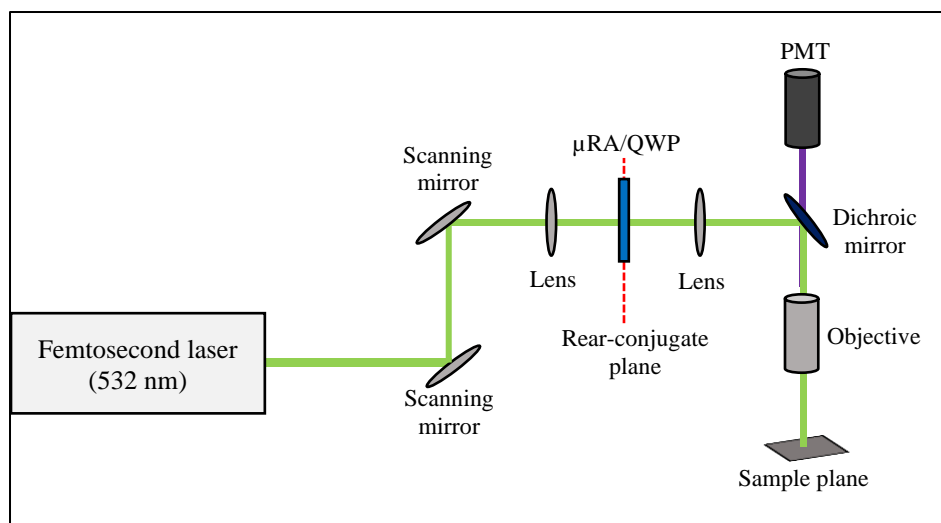


Figure 4.3. Instrument schematic for the SONICC instrument. Images were collected in the epi direction a photomultiplier tube (PMT). The  $\mu$ RA and quarter-wave plate (QWP) were placed in the rear-conjugate plane of the beam path.

#### 4.3.4 Data analysis

Analysis of TPE-UVF images was performed using a custom script (MATLAB). Images were first converted into Fourier space using the fast Fourier transform function, and the first and second harmonics produced by the line pattern of the  $\mu$ RA (see Figure 4.1) were isolated. These values were then used to calculate the Fourier coefficients of the sample and the circular and linear responses using Equation 4.4. The ratio of circular to linear dichroism was then calculated for a given sample. This ratio was then compared to simulated values for confirmation of the validity of this method.

#### 4.4 Results

TPE-UVF images were collected for samples containing solutions of tryptophan, crystalline tryptophan that had been milled to a uniform size, and lyophilized lysozyme. Images were acquired with the  $\mu$ RA and QWP installed into the SONICC system, and without for comparison (Figure 4.4). As seen for each of these samples, a clear modulation pattern could be seen across the entire FOV with this pattern being most clear for the uniform tryptophan solution. For the solid forms, the striping pattern is significantly harder to distinguish due to pre-existing variation in the TPE-UVF signal due to changes in sample structure.

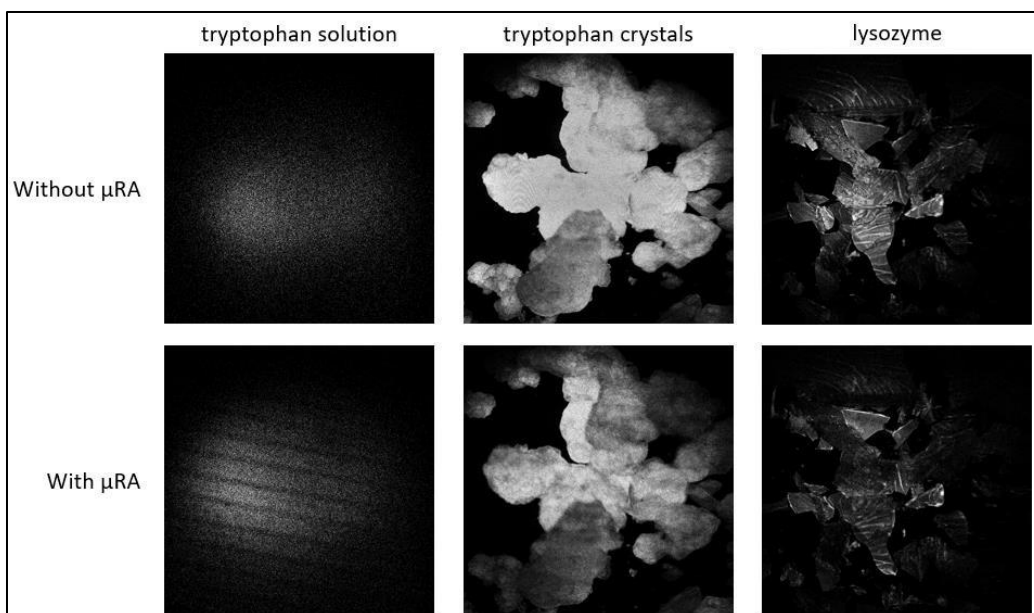


Figure 4.4. Polarization modulated TPE-UVF images with and without the  $\mu$ RA and QWP placed into the beam path. Images were collected for solutions of tryptophan as well as crystalline tryptophan and lyophilized lysozyme.

Due to these variations in TPE-UVF signal for crystalline tryptophan and solid lysozyme, the Fourier transform of these samples lacked sufficient signal-to-noise to isolate the first and second harmonics. Thus, currently only the uniform tryptophan solutions were able to be analyzed by this method. Table 4.1 shows the results of harmonic isolation and the calculation of the circular and linear responses for multiple FOVs of the same solution of tryptophan. In general, the variation in ratios between multiple different FOVs is reproducible for this specific sample.

Table 4.1. Linear and circular responses for polarization studies of TPE-UVF for solutions of tryptophan.

FOV	$\sigma_{LP}$	$\sigma_{CP}$	$\sigma_{LP} / \sigma_{CP}$
1	0.81	0.58	0.72
2	0.85	0.61	0.72
3	1.16	0.75	0.65
4	0.84	0.63	0.75
5	1.42	0.87	0.62
Average	0.73	0.69	0.69

## 4.5 Conclusion

Polarization modulation of TPE-UVF images was shown with the use of a commercially available  $\mu$ RA. Due to the small size and low cost of this optic, insertion of the  $\mu$ RA into an existing beam path was successfully demonstrated. Proof-of-concept studies were done on solutions of tryptophan which exhibits uniform TPE-UVF signal before insertion of the  $\mu$ RA. Fourier analysis of these samples was successfully demonstrated to extract the first and second harmonics allowing for calculation of the circular to linear dichroism ratio. Currently, work is being done on extending this method to non-uniform samples by means of improvement to the Fourier analysis. Once demonstrated for these samples, complex systems containing protein crystals and aggregate protein will be analyzed to demonstrate the method's ability to discriminate between solid forms.

## 4.6 References

1. He, M. Y., Advancing Protein Therapeutics: Current Challenges in Protein Therapeutics - CHI's Second Annual Conference. *IDrugs* **2006**, 9 (3), 177-178.
2. Neubert, H.; Xu, K. Y., New opportunities with quantification of protein therapeutics by LC-MS. *Bioanalysis* **2018**, 10 (13), 971-973.
3. Williams, K. L.; Gooley, A. A.; Packer, N. H., Analytical biotechnology and Proteome analysis. *Australas. Biotechnol.* **1996**, 6 (3), 162-&.
4. Akalin, P. K., Introduction to bioinformatics. *Mol. Nutr. Food Res.* **2006**, 50 (7), 610-619.
5. Wiedenheft, B.; Flenniken, M. L.; Allen, M. A.; Young, M.; Douglas, T., Bioprospecting in high temperature environments; application of thermostable protein cages. *Soft Matter* **2007**, 3 (9), 1091-1098.
6. Wang, H. C.; Xia, Y.; He, Y., The rapid identification of protein through its amino acid composition. *Acta Biochim. Biophys. Sin.* **1996**, 28 (3), 293-299.
7. Dhifli, W.; Diallo, A. B., ProtNN: fast and accurate protein 3D-structure classification in structural and topological space. *BioData Min.* **2016**, 9, 17.
8. Barends, T. R. M.; Foucar, L.; Botha, S.; Doak, R. B.; Shoeman, R. L.; Nass, K.; Koglin, J. E.; Williams, G. J.; Boutet, S.; Messerschmidt, M.; Schlichting, I., De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* **2014**, 505 (7482), 244-+.

9. Wampler, R. D.; Kissick, D. J.; Dehen, C. J.; Gualtieri, E. J.; Grey, J. L.; Wang, H. F.; Thompson, D. H.; Cheng, J. X.; Simpson, G. J., Selective Detection of Protein Crystals by Second Harmonic Microscopy. *J. Am. Chem. Soc.* **2008**, *130* (43), 14076-+.
10. Madden, J. T.; DeWalt, E. L.; Simpson, G. J., Two-photon excited UV fluorescence for protein crystal detection. *Acta Crystallogr. Sect. D-Struct. Biol.* **2011**, *67*, 839-846.
11. Ulcickas, J. R. W.; Ding, C. Q.; Deng, F. Y.; Simpson, G. J., Spatially encoded polarization-dependent nonlinear optics. *Opt. Lett.* **2018**, *43* (24), 5973-5976.
12. Robertson, N. A.; Hoggan, S.; Mangan, J. B.; Hough, J., Intensity stabilization of an argon-laser using an electrooptic modulator - performance and limitations. *Applied Physics B-Photophysics and Laser Chemistry* **1986**, *39* (3), 149-153.
13. Stoller, P.; Kim, B. M.; Rubenchik, A. M.; Reiser, K. M.; Da Silva, L. B., Polarization-dependent optical second-harmonic imaging of a rat-tail tendon. *J. Biomed. Opt.* **2002**, *7* (2), 205-214.
14. DeWalt, E. L.; Sullivan, S. Z.; Schmitt, P. D.; Muir, R. D.; Simpson, G. J., Polarization-Modulated Second Harmonic Generation Ellipsometric Microscopy at Video Rate. *Anal. Chem.* **2014**, *86* (16), 8448-8456.
15. Simpson, G. J., *Nonlinear Optical Polarization Analysis in Chemistry and Biology*. Cambridge Univ Press: Cambridge, 2017; p 1-485.



## **CHAPTER 5. CALIBRATION-FREE IMAGE ANALYSIS FOR QUANTIFICATION OF CRYSTALLINITY OF ROD-LIKE PARTICLES**

Statistical modeling and image analysis enabled calibration-free quantification of crystallinity within amorphous solid dispersion (ASD) samples imaged using second harmonic generation (SHG). Due to the impact of trace crystallinity on ASDs, quantification of the crystalline content during process manufacturing and storage allows for insight into drug stability, but detection of crystallinity below ~1% V/V presents challenges for conventional benchtop methods. The selectivity of SHG microscopy for homochiral active pharmaceutical ingredient (API) crystals allows for detection down to the ppm regime for quantifying crystallinity content in pharmaceutical formulations. In previous work, noise reduction in the calibration-free quantification of SHG images at trace crystallinity was demonstrated using particle counting methods, which relied on the explicit assumption of spheroidal crystalline particles. In this work, the use of a modified particle counting algorithm capable of correcting for the bias introduced by the particle-particle overlap of rod-like particles is demonstrated. Following validation of the reliability of this method using simulated SHG images, the algorithm was used on experimental data from the accelerated stability testing of rod-like ritonavir particles suspended in an ASD. Since the particle-counting algorithm can recover the SHG intensity and the percent crystallinity based on image analysis, it can be used as an internal calibrant for quantifying higher levels of crystallinity, enabling autocalibration based on the average SHG intensity per unit volume.

### **5.1 Introduction**

The number of potential drug candidates screened in small molecule drug discovery pipelines and the chemical complexity of those candidates are steadily increasing while the aqueous solubility and bioavailability are declining.<sup>1-4</sup> To overcome this issue of low drug solubility/low bioavailability, focus has been on presenting potential candidates in forms with greater apparent solubility without adversely affecting the drug potency.<sup>4-6</sup> One such method comes from the incorporation of the API in a polymeric carrier matrix to make an ASD. While the bioavailability and apparent solubility are significantly increased for an ASD, the ASD is typically metastable, in which the thermodynamically favored condition of API crystallization is under

kinetic control.<sup>4, 6-7</sup> Thus, crystals within an ASD have the potential to seed and grow.<sup>8-10</sup> With conversion of some API back to crystalline form, the benefits to the apparent solubility and bioavailability gained from suspension in an ASD can be diminished. Since the presence of trace crystallinity can greatly affect growth kinetics of the sample, the ability to monitor and quantify in this low crystallinity regime can allow for better insight into drug stability and kinetic modeling of crystallization of ASDs.<sup>11</sup>

Common benchtop techniques for monitoring the crystalline content present in drug formulations include powder X-ray diffraction (pXRD),<sup>12-13</sup> infrared (IR) and Raman spectroscopy,<sup>13-14</sup> differential scanning calorimetry (DSC),<sup>13, 15</sup> and solid-state nuclear magnetic resonance (ssNMR).<sup>13, 16</sup> Unfortunately, these techniques often lack sensitivity below 1% total crystallinity for the detection of trace crystallinity using benchtop instrumentation. SHG is a nonlinear optical technique capable of probing this regime of low crystallinity. Due to the symmetry requirements of SHG, SHG microscopy can be used to selectively image non-centrosymmetric crystals at low limits of detection (down to the ppt regime in specific materials and ppm in powders).<sup>17-19</sup>

To quantify crystallinity present in a sample imaged by SHG microscopy, present work has led to two complementary image analysis methods capable of informing the overall percent crystallinity (e.g. integrated intensity and particle counting with spheroidal particles).<sup>20-21</sup> At regimes of low crystallinity with negligible particle overlap, particle counting algorithms can be used with spheroidal particles morphology to determine the percent crystallinity of the sample. From the particle cross-sectional area, volumes can be deduced directly from the images, negating need for independent calibration. At higher percent crystallinity, with a large amount of particle overlap in the images, integration of the SHG intensity can be used to determine the percent crystallinity instead but commonly makes use of a calibration standard.<sup>17, 20</sup>

In general, particle counting methods have a signal-to-noise advantage over integration of the SHG intensity.<sup>20</sup> For these two methods, both signal averaging and particle counting share a common source of variance due to the Poisson distribution of the number of particles in a given field of view (FOV). However, despite these measurement uncertainties, particle counting provides a signal to noise advantage in volume determination by suppression of three additional sources of measurement variance: i) dependence of the SHG activity on crystal orientation, ii) differences in SHG intensity with crystal thickness, and iii) dark counts from non-crystalline

regions within the FOV. The SHG-activity of a given particle is highly dependent on crystal orientation, approximated by an exponential probability density distribution for a uniform distribution in particle orientation.<sup>17</sup> Second, the coherent nature of SHG produces an SHG activity scaling nonlinearly with crystal thickness, ranging from a 6<sup>th</sup> power dependence on diameter in the small crystal limit to independent of thickness in the large crystal limit. Finally, thermal excitation of photoelectrons from the photocathode and ambient low levels of background light both contribute to dark counts on the detector, which are independent of location in the FOV. Integration over the measured intensity within the FOV will generally include variance from the Poisson number of dark counts. Thus, particle-counting approaches offer a signal-to-noise ratio (SNR) advantage by suppression of these three additional noise sources.

Although particle counting provides a signal to noise benefit, the dynamic ranges of these methods are complicated by bias introduced from particle overlap.<sup>20, 22-23</sup>. At higher levels of crystallinity, the probability of particle-particle overlap within the depth-of-field (DOF) increases, introducing bias in the recovered volume if not accounted. Since the depth of field is larger than the lateral resolution under the conditions in which SHG imaging is typically performed, the resulting images correspond to projection of the samples volume onto a single 2D image,<sup>17</sup> increasing the chances for overlap in the integrated image. This superimposition can frustrate particle-counting approaches, in which overlapping particles in an image are treated as a single particle; the measured cross-sectional areas of the overlapping particles will be smaller than if the particles were well separated.<sup>20</sup> Thus, depending on the degree of overlap of the two particles, the volume recovered through these conventional methods will be greater or smaller than the ground truth result.<sup>24-27</sup>

In previous studies, work was done on bridging the gap between the two methods of integrated SHG intensity and particle counting for overlapping spheroidal particles.<sup>20</sup> Based on the relative variance in the recovered volume, a statistical model was developed from analytical expressions for optimizing the signal-to-noise of particle counting methods to correct for the bias that originates from particle overlap. Likewise, with extraction of the SHG-activity per unit volume by the developed particle counting methods, this lower region of crystallinity was used as an internal standard for extension to signal integration. These methods were then supported by Monte Carlo simulations and used to interpret nucleation and growth rates of evacetrapib (a cholesteryl ester transfer protein inhibitor).<sup>28</sup> Despite these successes, previous work was limited exclusively

to low aspect ratio, spheroidal particles,<sup>29-30</sup> complicating application to the more general case of arbitrary aspect ratio, including rod-like and needle-like particles routinely encountered in pharmaceutical products.

In this work, a generalized particle counting algorithm was developed to bridge the gap between the two methods of quantification (particle counting and signal integration) for moderately overlapping particles of high aspect ratio. Through determination of individual particle widths and the average aspect ratio, an empirical model was developed in order to correct for bias introduced from particle-particle overlap. A computation package for simulating SHG images for rod-like particles was developed to evaluate the model. The developed model was then used on experimentally acquired data of an ASD of ritonavir to assess the particle counting model's applicability to sensitively recover crystallinity over a large dynamic range in accelerated stability testing of ASDs.

## **5.2 Theoretical Framework**

### **5.2.1 High Aspect-Ratio Particle (HARP) Analysis**

Before considering the bias introduced by particle-particle overlap, it is useful to consider the origins of the SNR improvement from particle counting methods. Since SHG images are the projection of the total sampled volume onto a single 2D plane<sup>17</sup>, individual particles in the DOF of the microscope appear as single 2D particles. This projection is illustrated in Figure 5.1, where the projection of a single rod-like particle onto a 2D-plane is shown, with the inset image showing the corresponding theoretical SHG image of this particle. In the absence of particle overlap and assuming that all particles fall within the total depth-of-field imaged (30.5  $\mu\text{m}$ , determined from SHG axial z-scan measurements of individual 500 nm particles), conventional particle-counting algorithms can approximate the total particle volume by use of the cross-sectional area of each of the particles.

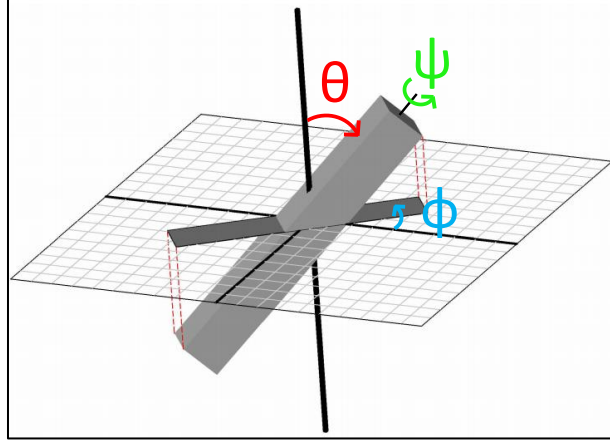


Figure 5.1. Graphical representation of a cuboidal particle's projection onto a 2D plane. Inset shows the particle's SHG projection.

In the limit of spherical particles, recovery of the volume is relatively simple. By use of Equation 5.1, the individual particle volumes ( $V_i$ ) can be determined from the cross-sectional area of each particle ( $A_i$ ) assuming each particle is perfectly spherical and well-separated.

$$V_i = \frac{4}{3} \pi \times \left(\frac{A_i}{\pi}\right)^{3/2} \quad 5.1$$

The total particle volume can then be determined by the summation of the contributions from each of these individual particles. For higher aspect ratio particles of rod-like and needle-like shape, the volume can easily be recovered by Equation 5.1 from the cross-sectional area (assuming this area corresponds to its bisection in a large depth-of-field). For rod-like particles, quantification is complicated due to variability of the particle's aspect ratio and the introduction of Euler angles ( $\theta$ ,  $\phi$ ,  $\psi$ ) needed to fully define a rod-like particle in real space. In this work, plate-like particles were not taken into consideration.

For rod-like particles (as opposed to plate-like or spheroidal), upon inspection of the three Euler angles (Figure 5.1), the polar tilt angle ( $\theta$ ) has the most significant effect on the projection of a rod-like particle onto the 2D object plane, while the azimuthal angle ( $\phi$ ) and twist angle ( $\psi$ ) do not significantly affect volumes based on area analysis. Since the azimuthal angle is defined relative to the image plane (Figure 5.1), this angle only affects the projection through an in-plane rotational transformation while leaving the cross-sectional area unaffected. Rod-like objects of  $C_\infty$  symmetry such as needle-like crystals will likewise be unaffected by the twist angle. Twist of high aspect ratio rectangular cuboids (except at polar angles of  $0^\circ$ ) will affect the projection of the minor axis of the particle (width) to its cross-sectional area. This affect will be greatest at a twist angle

of  $45^\circ$  in which the cross-sectional area will have a maximum width of  $\sqrt{2} * w$  if  $w$  is the width of the rod-like particle. Since the distribution of twist angles is reasonably assumed to be uniformly distributed, only a minor fraction of particles will be close to this maximum perturbation angle while the majority will only slightly be affected by the twist. Unlike the azimuthal and twist angles, the polar angle, defined perpendicular to the image plane (Figure 5.1), affects the projection of the major axis of the particle (length), and can significantly affect the observed cross-sectional area. The magnitude of these effects on the observed cross-sectional area can be shown by considering the aspect ratio (AR) of a particle, defined as the ratio of the particles length ( $l$ ) and width ( $w$ ) as given in Equation 5.2.

$$AR = \frac{l}{w} \quad 5.2$$

The ratio between the observed cross-sectional areas of particles parallel and perpendicular to the optical axis, is shown in Equation 5.3. Likewise, the observed aspect ratio is approximated in Equation 5.4 according to its polar angle,  $\theta$ .

$$\frac{A_{\parallel}}{A_{\perp}} = \frac{w * w}{l * w} = \frac{w}{l} = \frac{1}{AR} \quad 5.3$$

$$AR_{obs} \approx AR_{true} \times \cos \left( \theta - \frac{\pi}{2} \right) \quad 5.4$$

Unlike the twist angle, which has a maximum perturbation to the cross-section area of  $\sqrt{2}$ , the observed cross-sectional area can be as small as  $\frac{1}{AR}$  of the unperturbed cross-sectional area of the particle. Since the observed cross-sectional areas of high aspect ratio particles can significantly be affected by the polar angle, perturbation by the polar angle serves as a significant challenge for quantification of rod-like particles by particle counting techniques.

Assuming the long axis a particle is perpendicular to the optical axis, the volume calculated from Equation 5.1 generally overestimates the particle volume for rod-like particles (Equation 5.5).

$$\frac{V_{sphLimit}}{V_{true}} = \frac{\frac{4}{3}\pi \left( \frac{A_{obs}}{\pi} \right)^{3/2}}{w^2 l} = \frac{\frac{4}{3}\pi \left( \frac{w * l}{\pi} \right)^{3/2}}{w^2 l} = \frac{4}{3\sqrt{\pi}} \times \sqrt{\frac{l}{w}} = \frac{4}{3\sqrt{\pi}} \times \sqrt{AR} \quad 5.5$$

Factoring in the polar angle into Equation 5.5, the deviations from the true volume become more complicated (Equation 5.6) and the validity of the spheroidal limit in Equation 5.1 decreases.

$$\frac{V_{sphLimit}}{V_{true}} = \frac{4}{3\sqrt{\pi}} \sqrt{AR_{obs}} = \frac{4}{3\sqrt{\pi}} \sqrt{AR_{true} * \cos \left( \theta - \frac{\pi}{2} \right)} \quad 5.6$$

For the proposed HARP algorithm for quantification of the particle volume in this study, a modified particle-counting algorithm was developed to minimize bias from the distribution of polar angles and the corresponding bias from overlapping particles. To this end, the success of the HARP algorithm lies in determination of the average aspect ratio and particle width to calculate the volume.

Fundamentally, the HARP algorithm relies on image analysis to estimate the total volume in a FOV from the average aspect ratio and individual particle widths. The estimated volume from HARP ( $V_{harp}$ ) is calculated in Equation 5.7.

$$V_{harp} = w_{obs}^3 \times AR \quad 5.7$$

From Equation 5.7, this estimated volume approximates the true volume in a FOV when perturbations to the individual particle widths are small, and the true aspect ratios are narrowly distributed around the calculated average aspect ratio. As described previously, particle widths are affected by the twist angle, with a maximum perturbation of only  $\sqrt{2}$  for a small number of particles. Likewise, particle overlap is assumed to not have a significant bias on the recovered particle width. Thus, particle widths were recovered for all particles in the FOV using the Local Thickness plugin written for FIJI.<sup>31</sup> Since the main focus of this work is on the growth kinetics of crystals in an ASD, every particle is assumed to grow at the same rate, and have approximately the same average aspect ratio.<sup>32</sup> Thus,  $V_{harp}$  is assumed to be only slightly biased to larger values of particle volume due to effects on the particle widths.

The HARP algorithm determines the average aspect ratio by finding the particle with the largest aspect ratio in each FOV to provide an estimate for the true aspect ratio of each particle (Figure 5.2). To determine the average aspect ratio to use in calculation of the particle volume, multiple FOVs were used to find the particle for each FOV with the largest aspect ratio (presumed to be closest to lying flat in the image plane). Under the assumption that the effects from the twist angle were minimal, this selection is expected to provide a reasonable estimate for the ground truth. Since every particle is reasonably assumed to grow with the approximately same aspect ratio, this approximate aspect ratio was used in Equation 5.7 for the entire system. Additionally, this analysis was performed exclusively in the low crystallinity regime to negate the effects particle-particle overlap, since two or more overlapping particles in the FOV could have the largest aspect ratio (e.g. two rod-like particles with the same Euler angles, overlapped at the tails would have double

the observed aspect ratio). In practice, an average of particles from multiple FOV's was taken to improve the signal to noise ratio and minimize potential bias.

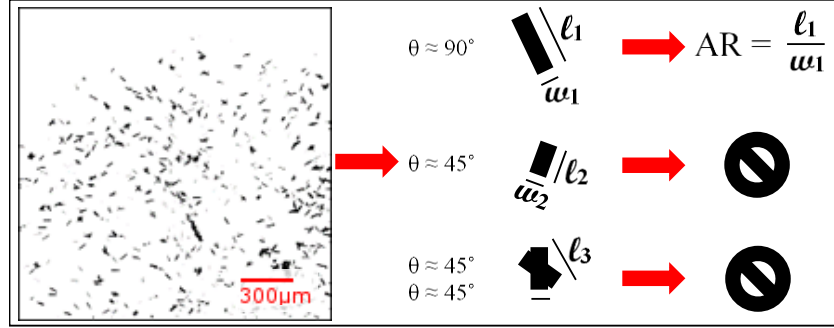


Figure 5.2. Workflow for determination of aspect ratio from multiple FOVs. For a specific FOV, a conventional particle-counting algorithm finds all particles as well as their respective feret diameter and minimum feret diameter. The particle with the largest length is then used to calculate it's observed aspect ratio. This is then averaged over multiple FOVs.

Lastly, an empirical correction factor ( $\alpha$ ) was introduced into the HARP algorithm to account for bias introduced in the image analysis. Specifically, due to limitations in the number of pixels measured in an SHG image affecting recovered particle values, and the underlying spread in aspect ratios for each particle, the values recovered by HARP were biased to higher volumes in the lower crystallinity range. By use of additional sets of simulated data with varying aspect ratios, a correction factor was determined by residual optimization between the known particle volume and the recovered volume. In this optimization, values used in the optimization were limited to the low percent crystallinity regime since past this low percent crystallinity regime volume recovery is optimally performed by integrating the SHG intensity.<sup>20</sup> In this regime, the residual between recovered volume and known volume given by the simulation package for every simulated image was minimized by determination of  $\alpha$ . Once determined, this value was multiplied by the initially recovered particle volume for each particle (see Equation 5.8).

$$V_{harp} = \alpha \times w_{obs}^3 \times AR \quad 5.8$$

### 5.2.2 Auto-calibration

Auto-calibration provides a means for connecting the volumes recovered from particle-counting to the integrated intensity. For kinetic studies, in which crystallinity within a system



increased over time, initial time points with low values of percent crystallinity were quantified using the HARP algorithm to recover their respective percent crystallinities and image intensities. Using these calculated values, the average integrated SHG intensity per unit volume for this specific system was calculated, which is equivalent to the slope of a conventional calibration curve. Using this value for the sensitivity, the particle volume from the integrated SHG intensity within the entire FoV was determined at long-time points, in which substantial particle overlap prevented particle-counting approaches. Since this calibration was made from images of the system taken at early time points, the specific particle shapes and crystalline forms of the system were captured in this calibration and should not adversely affect the analysis.

Relative to the integrated SHG intensity, particle counting removes measurement variance from orientation-dependent and size-dependent SHG-activity, but at the expense of providing a limited dynamic range. Consequently, there is a tipping point, below which particle counting is advantageous and above which the use of integrated SHG intensity reduces bias and measurement uncertainty. Extension of the previous analysis from [20] enabled identification of the point in which particle-counting methods were switched to integrated-intensity for SNR optimization. In brief, the anticipated SNR was evaluated for both particle counting with bias corrections to account for overlap and for signal integration over the entire FOV. For relatively small spheroidal particles, this transition arose when crystals occupied approximately  $\sim 2.5\%$  of the FOV<sup>20</sup>. Since the bias-corrected particle-counting method determined both particle volume and image intensity, samples with percent crystallinities smaller than the 2.5% allowed for calculation of the average integrated SHG intensity per volume. Past this point, percent crystallinity was quantified by integrating the overall SHG intensity, with the SHG per unit volume determined by autocalibration from the low crystalline regime. For rod-like particles using HARP analysis, the switchover point was identified to be 4% crystallinity, below which the recovered particle volumes by HARP remained linear and the analysis variance remained small.

## **5.3 Experimental and computational methods**

### **5.3.1 Image processing**

For both the simulated SHG images and the experimentally acquired images, a modified particle-counting algorithm was written as a macro for FIJI (Fiji is just ImageJ). This algorithm

consisted of four analyses combined to calculate the particle volume of an SHG image: i) conventional particle analysis, ii) calculation of the average particle aspect ratio, iii) determination of particle widths, and iv) rescaling by a scalar correction factor.

For conventional particle analysis, SHG images were set to binary to be able to run the particle analysis. Particle analysis was then set to measure the area of each particle, the feret diameter, minimum feret diameter, and the particle centroid. With these measurements set, the conventional volume was calculated using the recovered area of each particle. The feret diameter and minimum feret diameter were then used to calculate of the aspect ratio, and the centroid of each particle was used for identification of the particle width below.

For each particle identified from the conventional particle analysis, the width of each of these particles was determined using the centroid, minimum feret diameter, and the “Local Thickness” plugin.<sup>31</sup> To find the width of each particle, the output of the “Local Thickness” function was overlaid with the original image to create a false color image is output such that the intensities inside of a particle primarily corresponds to the width of the particle. The “Local Thickness” function works by finding the sphere with the largest diameter that can be placed inside of each particle. For regions of particle-particle overlap, the value recovered from this plugin can be larger than the individual widths, and the corners of rod-like particles will recover smaller values since only smaller spheres can fit in these regions.<sup>31</sup> For each particle, the recovered width can be determined by combination of the output from the “Local Thickness” plugin with the centroid measurement that gives x and y coordinates corresponding to the average position of the pixels in each particle, and the minimum feret diameter, which roughly corresponds to the smallest distance that can be traveled from the centroid while remaining on the particle. Using the centroid values and min feret diameters, the macro was written to scan the region around the centroid by a square with a width corresponding to twice the minimum feret diameter. The max value in this region is then assigned the width of the particle and used in Equation 5.7 to determine the particle volume after determination of the aspect ratio.

The final step for the recovery of the particle volume (before addition of the correction factor) is calculation of the average AR. Using the approach described in detail above, the longest particle in each FOV was found by finding the particle with the maximum feret diameter. The aspect ratio of this particle was then calculated by dividing the feret diameter by the minimum feret diameter. Since the AR can be affected by particle-particle overlap and out-of-plane particle

rotation, the calculated AR was averaged for multiple FOVs. With the average AR calculated, the volume of each particle (before correction by  $\alpha$ ) was then calculated using Equation 5.7.

### 5.3.2 Simulations

To evaluate the HARP algorithm, a simulation package was developed to generate images of rod-like particles similar to those routinely encountered experimentally. With controllable inputs for the particle aspect ratio, average particle lengths, the distribution of particle lengths, Poisson noise, and average particle intensity, Figure 5.3 A shows examples of the simulated data compared to experimentally acquired data of ritonavir in Figure 5.3 B. Particle lengths were defined as a truncated log-normal distribution with lognormal parameters  $\mu_L = 1.2$  and  $\sigma_L = 1$  (corresponding to the mean and standard deviation of the log of the distribution, respectively). Poisson noise was integrated into the simulated images on a per-pixel basis within each simulated particle. The log-normal distribution in particle sizes was capped to prevent placement of large particles not representative of the experimental observations and truncated in the lower size limit to prevent placement of particles of only a few pixels in size that may not be readily detectable by SHG imaging.

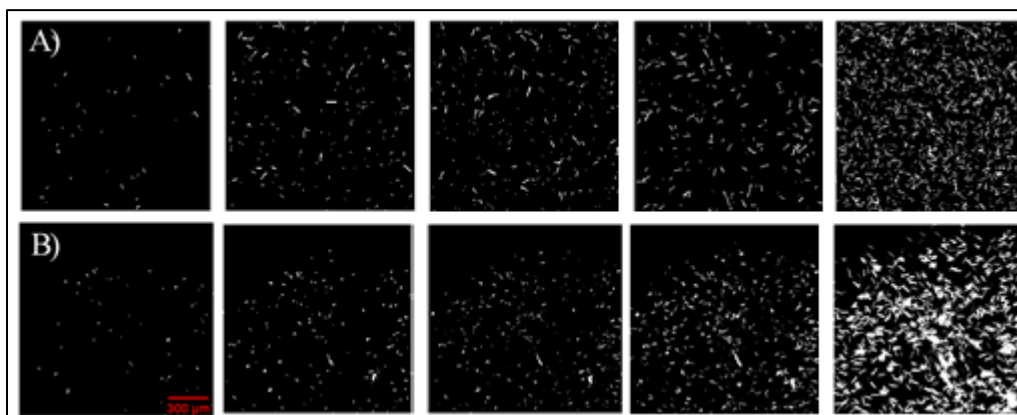


Figure 5.3. A) Rod-like particle simulated SHG images at varying levels of crystallinity, B) experimentally acquired SHG images of ritonavir. Scale bar 300  $\mu\text{m}$ .

### 5.3.3 Experimental methods

For experimental SHG images, ritonavir ASDs placed under accelerated stability testing conditions produced rod-like crystals appearing first with zero overlap and then with a varying

degree of overlap between the particles. An amorphous ritonavir extrudate was prepared with 15% (w/w) ritonavir (AbbVie), copovidone (BASF), plasticizer/surfactant and silicon dioxide by hot melt extrusion (HME) and then milled to a powder. The dry glass transition temperature of the milled extrudate was 56 °C, as measured by DSC. For the studies described herein, samples were prepared by placing powdered extrudates on the cover slip inside a custom temperature and relative humidity-controlled chamber with scheduled automated imaging capabilities as a thin layer (~300  $\mu\text{m}$ ) over a sample area of ~1  $\text{cm}^2$ . Samples were maintained at 50°C/75% relative humidity (RH). An SHG microscope system (Formulatrix SONICC) was used to quantify crystallization within the ASDs at every hour for 24 hours. SHG images were obtained using 350 mW excitation laser power and 894 ms exposure time. Images were collected in the transmission direction due to the high transparency of the sample through deliquescence of the powdered ASD at elevated temperature and humidity.

## 5.4 Results

As seen in Figure 5.3, the simulated data closely resembled the experimental ritonavir data. Pixelation effects in the simulations (simulated images were made to be  $512 \times 512$  to match the experimental image size) resulted in a significant variance in the simulated aspect ratios. For example, for a desired aspect ratio of ten: a horizontally oriented simulated particle ten pixels long would have a width of one pixel, and a particle nine pixels long would need a width of 0.9 pixels but was placed in the image with a width of one pixel. Thus, the true aspect ratio of simulated particles varied from five to ten with an average aspect ratio of approximately seven.

Additional sets of simulated data were created to explore the impact of varying aspect ratios. These data sets had average aspect ratios from approximately 4 to 11, each with a similar log-normal sized distribution with the rest of the simulation parameters kept the same. These extra data sets and the data presented in Figure 5.3 A were used in the validation of the HARP algorithm's recovery of the average aspect ratio for a given system and for optimization of the correction factor  $\alpha$ . For all sets of data, the recovered average aspect ratios were close to that of the true average aspect ratio of the simulations with a slight increase in the recovered value when images of high crystallinity (values past the switchover point) were used in the analysis. Using these multiple data sets, the empirical correction factor was optimized to a value of 0.61. With the coefficient

optimized and validation of the aspect ratio recovery, the HARP algorithm was run on the simulated and experimental data.

By creating a set of images with increasing number of particles, the HARP algorithm can be evaluated for the regimes of low and high crystallinity and the intermediate regime between the two. Figure 5.4 shows the results of both the spheroidal limit and the HARP algorithm in comparison to its ground truth for the data visualized in Figure 5.3 A. The red triangles above the line are the percent crystallinity values recovered for each simulated image, and the blue squares corresponds to the percent crystallinities recovered from the modified algorithm. For the low percent crystallinity regime (0% - 5%) both the spheroidal limit and the modified approaches recover values close to the ground truth values (black line), but vary from this point.

Past 5% crystallinity, the recovered particle volume of the spheroidal limit significantly increased. The source of this rapid increase in recovered values was the increasing probability of particle-particle overlap in these simulated images. Since overlapping particles resulted in larger recovered areas by particle-counting, the volume calculated by Equation 5.1 was greater than the true volume of the particles. Furthermore, as the probability of three or more particles overlapping increased, this bias to larger recovered particle volumes likewise increased. Thus, past the regime shown in Figure 5.4, a rapidly diverging trend was seen between the ground truth crystallinity and that recovered by application of the spheroidal limit analysis for images containing rod-like crystallites.

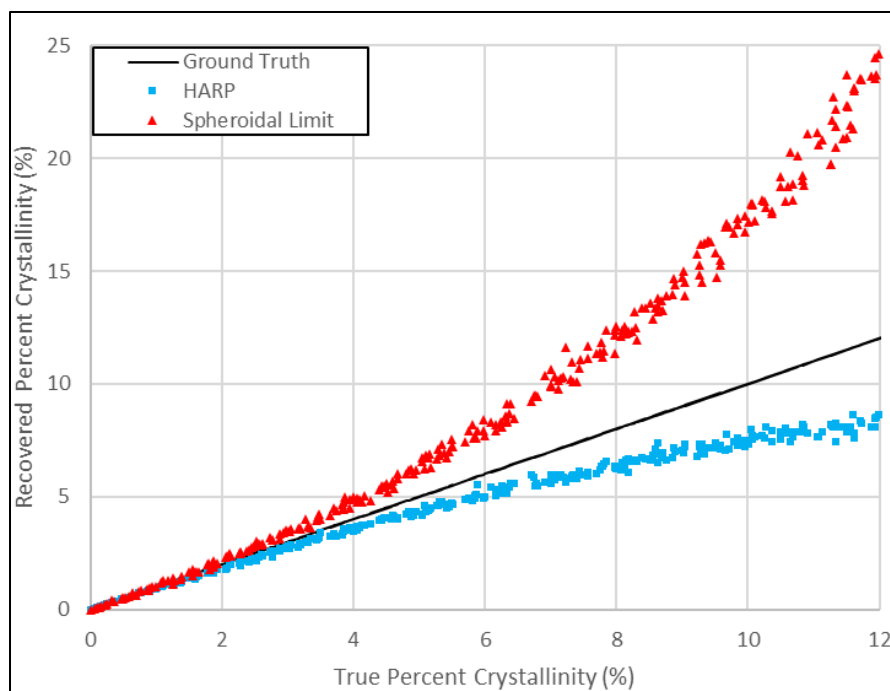


Figure 5.4. Recovered percent crystallinity for simulated data using the HARP algorithm (blue squares), and the spheroidal limit (red triangles). The ground truth percent crystallinity for each simulated image is shown as the solid black line.

The opposite trend was seen for the HARP algorithm, in which the recovered volumes were biased to smaller values than the ground truth by particle-particle overlap. Since this algorithm relies on volume determination by recovery of individual particle widths with a set aspect ratio determined in the low crystallinity regime, overlap can only affect the recovered volume by altering particle count or width. Though the recovered width will slightly increase due to particle-particle overlap, the significant reduction in recovered particle volume from a decreasing particle count is dominant. Eventually at much greater values of percent crystallinity past the dynamic range of the HARP algorithm, clusters of multiple overlapping particles were able to more significantly alter recovered particle widths and the recovered volume began to plateau. Overall past 5% crystallinity, the HARP algorithm had closer recovered values to the ground truth than that of the spheroidal limit method, but significant enough deviations that the calculation of the average integrated SHG intensity per volume for auto-calibration was limited to the linear regime of 0% to 4% crystallinity. Thus, past this linear regime, particle counting was switched to signal integration for quantification of crystallinity in the experimental data.

Not all features in the experimental images were recovered by the simulations as described. The most visually defining difference between the two types of data is the placement of particles in the FOV. For the simulated images, each rod-like particle was placed randomly in the FOV with a uniform distribution, while the particles observed experimentally were absent in the upper region of the FOV (presumably from vignetting). Incorporation of contributions from vignetting in the simulations can be performed by straightforward multiplication by a smoothly varying spatially fixed attenuating function. In addition, starburst-like patterns corresponding to radial growth habits tended to appear at high crystallinity depending on experimental conditions that were not included in the simulations. Modeling this grouping and starburst patterns was neglected, as the radial habits were only clearly obvious at higher percent crystallinities, corresponding to regions with optimal signal to noise from signal integration. In this regime, details related to growth habit do not substantively impact crystallinity assessments produced by integration over the entire FOV.

## **5.5 Experimental implementation**

Needle-shaped ritonavir crystals were formed from 15% DL(w/w) ritonavir ASDs under accelerated stability testing conditions. ASD samples of ritonavir (15%), sorbitan monolaurate (10%), copovidone (74%), and colloidal silicon dioxide (1%) were prepared by HME, rotary evaporation (rotovap), and spray dried dispersion (SDD). About 5mg of each of powdered ASD samples were placed onto glass cover slips inside a custom-built *in situ* platform for accelerated stability testing as thin layers (~300  $\mu\text{m}$ ). SHG images were collected at four different fields of views. The conditions in the platform were maintained at an elevated temperature of 50°C and 75% relative humidity for accelerated stability testing conditions.

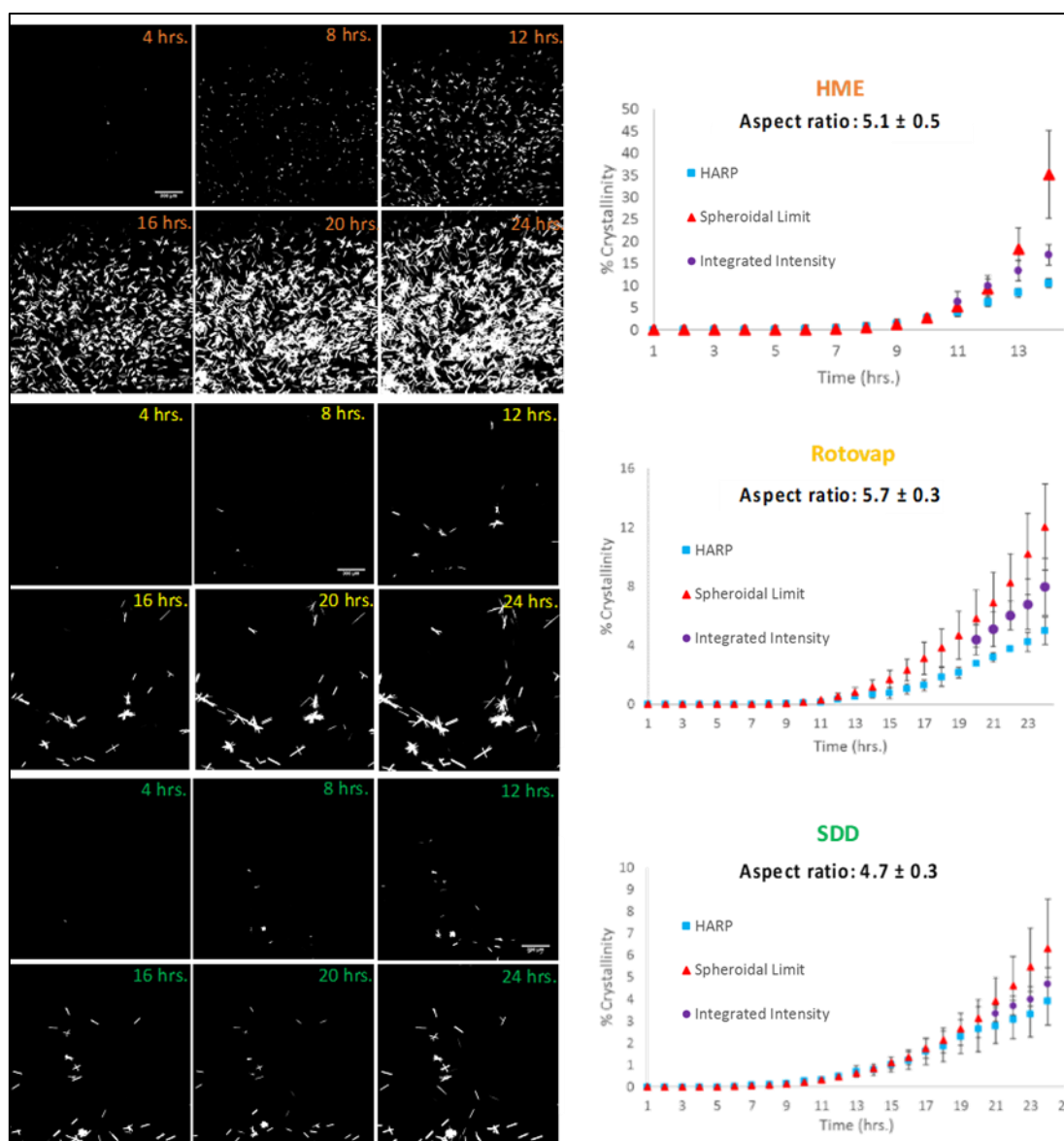


Figure 5.5. Time evolution of percent crystallinity in 15% ritonavir HME, rotovap, and SDD ASDs stressed in situ temperature and relative humidity-controlled platform at 50OC/75% RH. HARP algorithm is represented by the blue squares, the spheroidal limit by red triangles, and integrated intensity by the purple circles.

The amorphous polymer as well as other excipients present in the ASDs did not produce significant coherent SHG signal. However, ritonavir crystals formed during accelerated conditions adopt a non-centrosymmetric lattice upon crystallization that is SHG active, enabling selective detection of the crystalline particles using the calibration-free particle counting algorithm for rod-like particles. SHG micrographs of six representative time-points are shown in Figure 5.5 to illustrate the evolution of crystallinity in the ASDs during the accelerated stability testing. The



percent crystallinity of the new crystalline phase consisting rod-like ritonavir crystals was both calculated using the HARP algorithm and using the conventional particle counting algorithm assuming spherical ritonavir particles using FIJI. Though the respective aspect ratios for ritonavir crystals arising from HME, rotovap, and SDD ASDs were very similar, as shown in Figure 5.5, the crystallization kinetics are substantially different in all three. Overall, the HARP algorithm was shown to remove bias associated with the previous spheroidal particle counting approach applied to rod-like particles and recover crystallinities better aligned with those from the integrated intensity. Consistent with expectations, the HARP model provided significant improvements in the SNR in the low ( $<4\%$ ) crystalline regime relative to signal integration, consistent with the error bars in Figure 5.5. As discussed above, quantification was switched from particle-counting to integrated intensity by use of autocalibration past 4% crystallinity. As expected, the values recovered after switching to integrated intensity were between the two particle-counting methods as one biased above ground truth and the other below. Thus, the recovered volumes from switching to integrated intensity were believed to more closely predict the level of crystallinity for all three methods of sample preparation.

Consistent with expectations from the simulations, the HARP approach provides a signal-to-noise advantage over comparatively simple SHG intensity averaging over the FOV. Generally, this advantage comes from the suppression of additional sources measurement noise that affects signal averaging. Since dark counts arise from signal due to ambient light and the thermal excitation of photoelectrons, the signal from these events are randomly scattered throughout the FOV. Thus, only the variance from dark counts superimposed on individual particles can affect the HARP algorithm, while both the particle intensity variance and the background variance arising from dark counts affects signal averaging methods. Additionally, variances in the SHG intensity arising from distributions in crystal orientation and thickness are suppressed in the HARP approach, but not in signal integration. As discussed in the details of the HARP algorithm, particle volume is calculated by determined of the aspect ratio and width of rod-like particles, neither of which are negatively affected by variances in crystal orientation and thickness. In the HARP approach, crystal orientation has the potential to complicate determination of the average aspect ratio, but averaging over many FOVs mitigated this effect. Likewise, particle thickness was directly determined in the HARP algorithm and thus did not contribute substantial variance in the analysis.

These effects do not significantly affect the HARP approach for volume determination, but impact the integrated intensity methods.

In addition to the SNR advantages, HARP also supported auto-calibration to enable quantification of crystallinity by integrated intensity without the need for external standards. Below 4% crystallinity, the SHG intensity used by the HARP algorithm only serves to identify the borders of the crystal. Volume is determined from area and shape analysis. As such, HARP enables independent determination of both the integrated SHG intensity and the fractional crystallinity, the ratio of which yields the slope one would obtain from a calibration curve. Thus, this value used as the auto-calibration standard was determined with methods that suppressed the sources of measurement variance previously discussed and the SNR advantage is assumed to translate to particle volumes past this switchover point. As seen visually in Figure 5.5, the error bars of the integrated intensity method with use of auto-calibration were significantly smaller than that of the conventional spheroidal limit. Thus, the recovered volumes calculated using HARP and auto-calibration were shown to more closely predict the level of crystallinity for experimental data.

## **5.6 Future work**

Due to the limited use of the HARP analysis for just high-aspect ratio particles, machine learning (ML) architectures will be investigated for quantification of crystallinity in samples imaged by SHG regardless of particle shape. For this work, convolutional neural networks (CNNs) will be investigated due to their common use in image analysis. By training of these CNNs with simulated SHG images of crystalline particles of varying shapes and sizes from the developed simulation package, it is hypothesized that quantification of experimentally acquired SHG images will be possible with the trained CNN. Collection of ground truth data for a selection of experimentally acquired SHG images will then enable verification of the accuracy of the trained CNNs.

## **5.7 Conclusion**

A geometric model for the quantification of crystallinity in SHG images for particles of high aspect ratio was devised and evaluated in simulations, then applied in accelerated stability testing of ritonavir ASDs. The HARP algorithm allowed for calculation of the average integrated SHG

intensity per volume from the particle-counting regime, allowing for quantification by signal integration without the need for an external standard. The volume fraction of crystallinity critical for kinetic modeling in stability testing was recovered by HARP with significant improvements in SNR relative to signal integration in the low (<4%) crystallinity regime. Furthermore, HARP independently recovers both the integrated intensity and the crystalline volume fraction from image analysis, enabling auto-calibration for analysis in the high signal regime (>4%) without the need for external standards. These results support the broader application of SHG imaging for quantitative analysis of amorphous solid dispersions in pharmaceutical development pipelines.

## 5.8 References

1. Babu, N. J.; Nangia, A., Solubility Advantage of Amorphous Drugs and Pharmaceutical Cocrystals. *Cryst. Growth Des.* **2011**, *11* (7), 2662-2679.
2. Thayer, A. M., Finding Solutions. *Chem. Eng. News* **2010**, *88* (22), 13-18.
3. Ritchie, T. J.; Macdonald, S. J. F.; Young, R. J.; Pickett, S. D., The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discov. Today* **2011**, *16* (3-4), 164-171.
4. Leuner, C.; Dressman, J., Improving drug solubility for oral delivery using solid dispersions. *Eur. J. Pharm. Biopharm.* **2000**, *50* (1), 47-60.
5. Serajuddin, A. T. M., Salt formation to improve drug solubility. *Adv. Drug Deliv. Rev.* **2007**, *59* (7), 603-616.
6. Laitinen, R.; Lobmann, K.; Strachan, C. J.; Grohgan, H.; Rades, T., Emerging trends in the stabilization of amorphous drugs. *Int. J. Pharm.* **2013**, *453* (1), 65-79.
7. He, Y.; Ho, C., Amorphous Solid Dispersions: Utilization and Challenges in Drug Discovery and Development. *J. Pharm. Sci.* **2015**, *104* (10), 3237-3258.
8. Song, Z. T.; Sarkar, S.; Vogt, A. D.; Danzer, G. D.; Smith, C. J.; Gualtier, E. J.; Simpson, G. J., Kinetic Modeling of Accelerated Stability Testing Enabled by Second Harmonic Generation Microscopy. *Anal. Chem.* **2018**, *90* (7), 4406-4413.
9. Li, N.; Taylor, L. S., Microstructure Formation for Improved Dissolution Performance of Lopinavir Amorphous Solid Dispersions. *Mol. Pharm.* **2019**, *16* (4), 1751-1765.
10. Elkhaz, A.; Sarkar, S.; Dinh, J. K.; Simpson, G. J.; Taylor, L. S., Variation in Supersaturation and Phase Behavior of Ezetimibe Amorphous Solid Dispersions upon Dissolution in Different Biorelevant Media. *Mol. Pharm.* **2018**, *15* (1), 193-206.

11. Smith, C. J.; Griffin, S. R.; Eakins, G. S.; Deng, F. Y.; White, J. K.; Thirunahari, S.; Ramakrishnan, S.; Sangupta, A.; Zhang, S. W.; Novak, J.; Liu, Z.; Rhodes, T.; Simpson, G. J., Triboluminescence from Pharmaceutical Formulations. *Anal. Chem.* **2018**, *90* (11), 6893-6898.
12. Newman, A. W.; Byrn, S. R., Solid-state analysis of the active pharmaceutical ingredient in drug products. *Drug Discov. Today* **2003**, *8* (19), 898-905.
13. Shah, B.; Kakumanu, V. K.; Bansal, A. K., Analytical techniques for quantification of amorphous/crystalline phases in pharmaceutical solids. *J. Pharm. Sci.* **2006**, *95* (8), 1641-1665.
14. Yang, Z. Y.; Nollenberger, K.; Albers, J.; Moffat, J.; Craig, D.; Qi, S., The effect of processing on the surface physical stability of amorphous solid dispersions. *Eur. J. Pharm. Biopharm.* **2014**, *88* (3), 897-908.
15. Johari, G. P.; Ram, S.; Astl, G.; Mayer, E., CHARACTERIZING AMORPHOUS AND MICROCRYSTALLINE SOLIDS BY CALORIMETRY. *J. Non-Cryst. Solids* **1990**, *116* (2-3), 282-285.
16. Berendt, R. T.; Munson, E. J., Effect of enantiomeric ratio and preparation method on proline crystal form. *Crystengcomm* **2012**, *14* (7), 2479-2488.
17. Chowdhury, A. U.; Zhang, S. J.; Simpson, G. J., Powders Analysis by Second Harmonic Generation Microscopy. *Anal. Chem.* **2016**, *88* (7), 3853-3863.
18. Kestur, U. S.; Wanapun, D.; Toth, S. J.; Wegiel, L. A.; Simpson, G. J.; Taylor, L. S., Nonlinear optical imaging for sensitive detection of crystals in bulk amorphous powders. *J Pharm Sci* **2012**, *101* (11), 4201-13.
19. Newman, J. A.; Schmitt, P. D.; Toth, S. J.; Deng, F. Y.; Zhang, S. J.; Simpson, G. J., Parts per Million Powder X-ray Diffraction. *Anal. Chem.* **2015**, *87* (21), 10950-10955.
20. Smith, C. J.; Dinh, J.; Schmitt, P. D.; Stroud, P. A.; Hinds, J.; Johnson, M. J.; Simpson, G. J., Calibration-Free Second Harmonic Generation (SHG) Image Analysis for Quantification of Trace Crystallinity Within Final Dosage Forms of Amorphous Solid Dispersions. *Appl. Spectrosc.* **2018**, *72* (11), 1594-1605.
21. Kissick, D. J.; Gualtieri, E. J.; Simpson, G. J.; Cherezov, V., Nonlinear Optical Imaging of Integral Membrane Protein Crystals in Lipidic Mesophases. *Anal. Chem.* **2010**, *82* (2), 491-497.
22. Ilevbare, G. A.; Liu, H. Y.; Edgar, K. J.; Taylor, L. S., Understanding Polymer Properties Important for Crystal Growth Inhibition-Impact of Chemically Diverse Polymers on Solution Crystal Growth of Ritonavir. *Cryst. Growth Des.* **2012**, *12* (6), 3133-3143.

23. Mugheirbi, N. A.; Tajber, L., Crystal Habits of Itraconazole Microcrystals: Unusual Isomorphic Intergrowths Induced via Tuning Recrystallization Conditions. *Mol. Pharm.* **2015**, *12* (9), 3468-3478.
24. Shen, L. P.; Song, X. Q.; Iguchi, M.; Yamamoto, F., A method for recognizing particles in overlapped particle Images. *Pattern Recognit. Lett.* **2000**, *21* (1), 21-30.
25. de Langlard, M.; Al-Saddik, H.; Charton, S.; Debayle, J.; Lamadie, F., An efficiency improved recognition algorithm for highly overlapping ellipses: Application to dense bubbly flows. *Pattern Recognit. Lett.* **2018**, *101*, 88-95.
26. Kolehmainen, J.; Elfvengren, J.; Saarenrinne, P., Interference-based overlapping particle tracking velocimetry for fluidized beds. *Exp. Fluids* **2014**, *55* (10), 15.
27. Honkanen, M., Reconstruction of a three-dimensional bubble surface from high-speed orthogonal imaging of dilute bubbly flow. In *Computational Methods in Multiphase Flow V*, Mammoli, A. A.; Brebbia, C. A., Eds. Wit Press: Southampton, 2009; pp 469-480.
28. Pack, B. W.; Babayan, Y.; Schrad, M. A.; Stroud, P. A.; Sperry, D. C.; White, K. C.; Aburub, A., Development of an in vivo-relevant drug product performance method for an amorphous solid dispersion. *J. Pharm. Biomed. Anal.* **2017**, *142*, 307-314.
29. Zhang, Z. L., Particle overlapping error correction for coal size distribution estimation by image analysis. *Int. J. Miner. Process.* **2016**, *155*, 136-139.
30. Song, X. Q.; Yamamoto, F.; Iguchi, M.; Shen, L. P.; Ruan, X. D.; Ishii, K., A method for measuring particle size in overlapped particle images. *ISIJ Int.* **1998**, *38* (9), 971-976.
31. Hildebrand, T.; Ruegsegger, P., A new method for the model-independent assessment of thickness in three-dimensional images. *J. Microsc.-Oxf.* **1997**, *185*, 67-75.
32. Sarkar, S.; Song, Z.; Griffin, S. R.; Takanti, N.; Vogt, A. D.; Ruggles, A.; Danzer, G. D.; Simpson, G. J., In Situ Crystal Growth Rate Distributions of Active Pharmaceutical Ingredients. *Mol. Pharm.* **2019**.

## CHAPTER 6. ADVERSARIAL ILLUMINATION

A method for controlling the appearance of a 2D or 3D arbitrary object by external illumination is presented. The goal for such illumination is to alter the classification of the object by projection of a carefully determined compensation pattern. By use of a simple camera, an image of an object is collected and after accounting for variations in reflectivity, environmental lighting, and camera exposure, the image is appropriately classified. Post-classification, a compensation pattern is determined that when combined with the original image pushes classification into the decision boundary of another object classification. A simple projector connected to the camera then projects this compensation pattern over the object and a second image is acquired. Classification of this second image should then not recover the original object classification, and this has currently been demonstrated for a 2D image of a strawberry. Overall, this method has direct application to study of how such adversarial illumination attacks can affect systems such as self-driving vehicles and to formulate methods in which these attacks can be prevented. Likewise, this illumination approach has analog in chemical analysis, in which a calibration file could intentionally be manipulated to impact classification of a high dimensional dataset.

### 6.1 Introduction

The appearance of an object is highly dependent on factors such as lighting, color, and texture, and variations to this can result in object misclassification. A simple example of such a case would be identification of a chemical compound by use of Raman spectroscopy. For this type of chemical analysis, the spectral peaks of a compound are identified and matched with specific vibrational modes of a molecule. By subtle manipulation of the apparent noise in the spectrum, or alteration to its shape or size, identification of the compound can be obscured. Thus, the ability to control the appearance of an object has implications in many fields of science ranging from chemical analysis (discussed further in Chapter 7) to applications such as camouflage,<sup>1</sup> or the ability to negatively affect image classification and systems that rely on that classification. Self-driving vehicles are an example of such a system that relies on accurate object classification.<sup>2-3</sup> While driving, the cameras that monitor the road and the surrounding environment of the vehicle is on heavily relies on accurate classification to safely navigate a road. Misclassification of objects

such as street signs or other objects on the road is a significant issue in the safety of such systems.<sup>4</sup> Therefore, having an understanding on how variations in appearance can affect proper classification is important.

In general, actual alteration of an object's appearance in order to induce misclassification has first been proven for post-acquisition data space.<sup>5</sup> The work done by Warde-Farley and Goodfellow is an example of such a post-acquisition alteration.<sup>5</sup> Figure 6.1 shows the work done by these authors in which an image of a panda is initially classified with moderate confidence by GoogLeNet. By alteration of this image by the addition of an imperceptible noise mask, the resulting image appears identical to the original by the human eye, but is now misclassified with significant confidence to be a gibbon. While this adversarial attack has been proven to work, the possibility of such an attack to an already acquired image is low. To perform such a task, an image must be accessible to overlay the desired attack pattern which is easily preventable with use of image encryption or other security measures.<sup>6</sup>

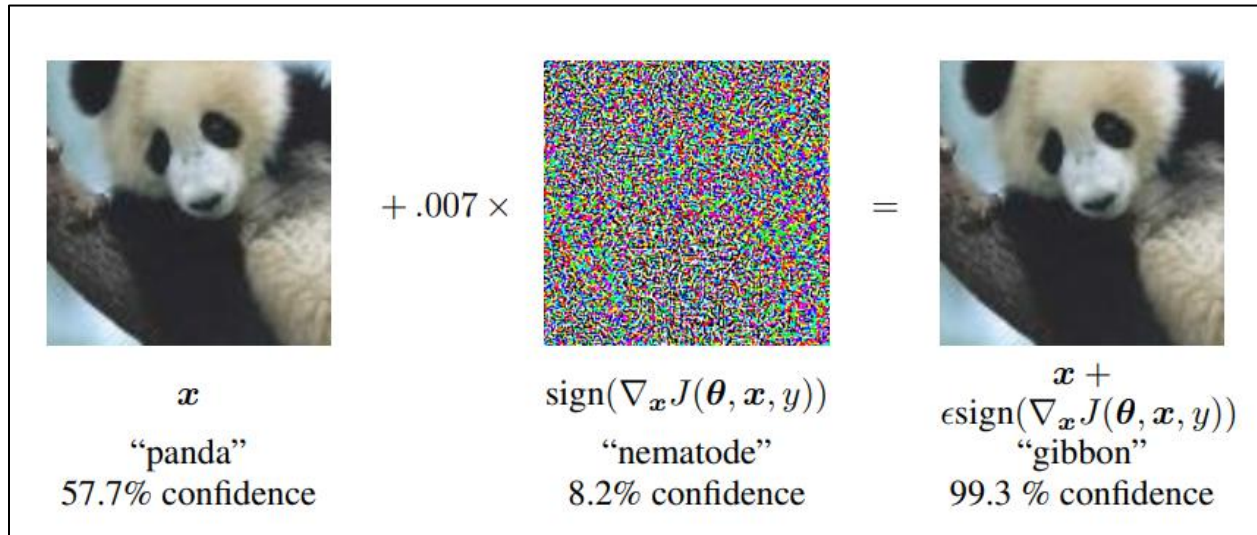


Figure 6.1. A demonstration of misclassification using the GoogLeNet and image of a panda taken from ImageNet. By addition of an imperceptible attack pattern, intentional misclassification of the image to that of a gibbon is shown.

While adversarial attacks in data space have been shown to be extremely potent, intentional misclassification by other methods or in other fields of science (e.g. chemical classification) has gained more traction in recent years. Due to the limited ability to access data post-acquisition, being able to physically and imperceptible alter the appearance of an object may be easier but

comes with challenges of its own. Common examples of this are the use of stickers/patches placed on physical objects like a road sign or a common household item that causes intentional misclassification,<sup>4, 7</sup> or glasses that allow a person to be recognized by facial recognition software as a famous individual.<sup>8</sup> While these physical attacks are generally easier to place on an object, variations in perception angle, lighting conditions can affect the efficacy of the attack, or an attacked object may be easier to detect.<sup>4, 7-8</sup>

In this work, a new physical adversarial attack technique is proposed in which a compensation pattern is externally illuminated using a simple projector. By using a camera attached to this projector, an image of an arbitrary object can be acquired. Using an image classifier pre-trained in detecting such an object, the decision boundary of the neural network (NN) for such an object can be determined and used to design a pattern that can move the decision point in this space onto another object class. This calculated pattern was then externally illuminated by the projector and overlaid with the physical object. After a second round of image acquisition and classification, verification of misclassification was done. As proof of concept, initial studies were done on 2D object printed on poster paper to demonstrate object misclassification, which was then extended to the intentional misclassification of 3D objects. With this foundation in place, extension to a broad set of classes of multi-dimensional measurement data could benefit from these adversarial approaches such as the work presented later in Chapter 7.

## **6.2 Methods**

### **6.2.1 Acquisition and illumination**

Image acquisition and pattern illumination were done using a simple camera/projector set-up shown for Figure 6.2. Images of an arbitrary object were imaged using a Pi Camera controlled using a Raspberry Pi. The camera and Raspberry Pi were held in place using a custom 3D printed mount attached to a breadboard. Images were transferred to another computer capable of running PyTorch to calculate the desired compensation pattern. This pattern was then transferred back to the Raspberry Pi which was also connected to a simple projector. The attack pattern was then illuminated over the initial object and fine adjustment of the image for registry with the imaged object was done using the key functionality offered by Pygame. A second image was then collected



using the same Pi Camera or a second Pi Camera attached to its own Raspberry Pi and sent to another computer for final image classification.

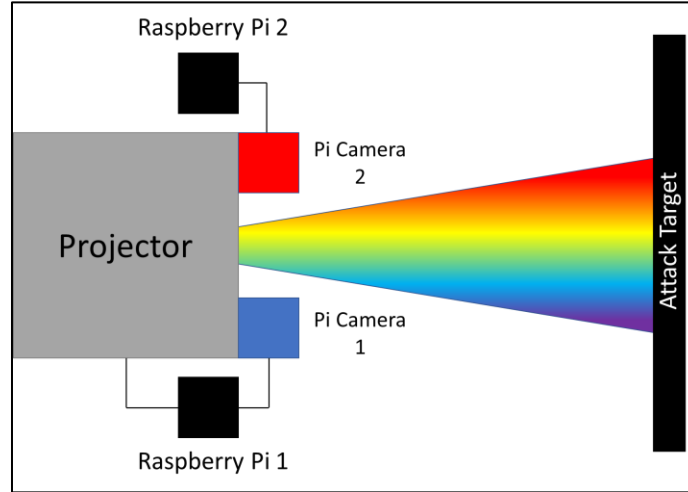


Figure 6.2. Diagram of the acquisition and illumination set-up. Images of a target object were collected using a Pi Camera attached to a Raspberry Pi. After calculation of the desired attack pattern, the projector also connected to the Raspberry Pi illuminated pattern over the same object. The same camera, or a second camera attached to another Raspberry Pi acquires a second image.

### 6.2.2 Image classification

A simple image classifier in PyTorch was trained on a small subset of images from ImageNet. For a given image, this pre-trained NN outputs a confidence percentage for each of the trained object classes. The five object classes with the highest confidence are then listed for the given image. Classification success is marked by whether the true object class is present in one of these five classes or not. If the true object class is not present in these five classes, the image is taken to be misclassified by the NN.

## 6.3 Results

### 6.3.1 Adversarial illumination of 2D object

Demonstration of this adversarial illumination using the simple camera/projector system was done for 2D objects printed on poster paper. Figure 6.3 shows the results of this for an image of a strawberry printed on a 32-inch x 32-inch poster. Figure 6.3 A shows the initially acquired image

of this poster which was classified with highest confidence as either a strawberry or a pan. Thus, initial classification of the poster was determined to be a success. Figure 6.3 B then shows the calculated attack pattern for this given image, which was then overlaid over the poster image. After registry of the attack pattern with the poster in x-y space, Figure 6.3 C shows the resulting image captured by the Pi Camera. Using the same NN trained for image classification, the object class with the highest degree of confidence was a jigsaw puzzle and strawberry was not present in any of the top five highest confident object classes. Thus, the object was successfully misclassified and intentional misclassification using external illumination was demonstrated.

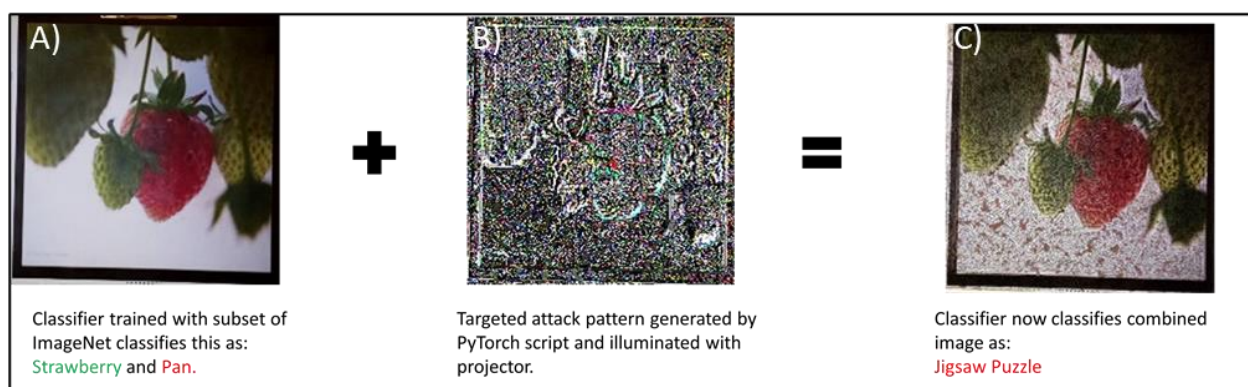


Figure 6.3. Adversarial illumination results for a 2D object (a strawberry) printed on a poster. Initial classification of the initial image collected using the Pi Camera (A) classified the poster as either a strawberry or a pan. The calculated illumination pattern is then shown (B), and the imaged acquired after illumination of this pattern on the poster (C). After acquisition of a second image, the new image was classified as a jigsaw puzzle.

While intentional misclassification was successful, challenges did arise for this given system. For 2D objects printed on poster paper, choice of paper and poster finish significantly affected system performance. Images printed on a glossy poster material produced images with glare that affected image classification. This was especially present when the attack pattern was illuminated with the projector. To circumvent this limitation, 2D objects were instead printed on matte poster material that had significantly less reflectivity. Likewise, the projector could instead illuminate the pattern at an angle to the poster since the highest percentage of reflection and glare occurs when the pattern is illuminated normal to the direction of the camera. This was not done experimentally though since the projector angle would have to be factored in the calculation of the desired attack pattern due to image morphing when projected at an angle. Additionally, variations lighting

conditions and camera exposure were seen to affect image classification as well as alter the observed color determination. Camera, projector, and lighting conditions were set according to try and minimize these effects.

### **6.3.2 Adversarial illumination of 3D objects**

Misclassification of 3D objects was done using a similar camera/projector system by the Chan group. Objects such as a basketball, coffee mug, and cardigan were imaged using a simple camera, and the desired attack pattern was calculated as done before. After illumination of this pattern using a projector, the resulting images were successfully misclassified as a buckler, a whiskey jug, and a poncho respectively. Thus, external illumination was successfully demonstrated for 2D and 3D objects.<sup>9</sup> Future work for this project looks to extend this adversarial illumination architecture to intentional misclassification of images collected in real-time.

## **6.4 Conclusion**

A method for controlling the appearance of a 2D and 3D objects by external illumination was successfully demonstrated. By use of a simple camera and projector system, a compensation attack pattern was calculated from an image acquired of a specific object. By illumination of this pattern over the imaged object, intentional misclassification by a NN pre-trained in classification of the initial object was demonstrated. This misclassification architecture was first demonstrated for a 2D poster image of a strawberry, but was later extended to a set of 3D objects. Overall, this method has direct application to study of how such external attacks can affect existing classification architectures and the corresponding systems (e.g. self-driving vehicles) that rely on proper classification. The results of this work will then be translated to help with improvement of existing image classifiers to prevent further attacks by these methods. Furthermore, extension of adversarial attacks to systems other than post-acquisition data space has been demonstrated and shows that adversarial attacks can be extended to a variety of other problems such as spectral classification.

## 6.5 References

1. Yang, J. J.; Zhang, X. F.; Zhang, X.; Wang, L.; Feng, W.; Li, Q., Beyond the Visible: Bioinspired Infrared Adaptive Materials. *Adv. Mater.*, 40.
2. Prabhakar, G.; Kailath, B.; Natarajan, S.; Kumar, R.; Ieee, Obstacle Detection and Classification using Deep Learning for Tracking in High-Speed Autonomous Driving. In *2017 Ieee Region 10 International Symposium on Technologies for Smart Cities*, Ieee: New York, 2017.
3. Stroescu, A.; Cherniakov, M.; Gashinova, M., Classification of High Resolution Automotive Radar Imagery for Autonomous Driving Based on Deep Neural Networks. In *2019 20th International Radar Symposium*, Knott, P., Ed. Ieee: New York, 2019.
4. Patel, N.; Krishnamurthy, P.; Garg, S.; Khorrami, F.; Ieee, Adaptive Adversarial Videos on Roadside Billboards: Dynamically Modifying Trajectories of Autonomous Vehicles. In *2019 Ieee/Rsj International Conference on Intelligent Robots and Systems*, Ieee: New York, 2019; pp 5916-5921.
5. Warde-Farley, D.; Goodfellow, I., *Adversarial Perturbations of Deep Neural Networks*. Mit Press: Cambridge, 2016; p 311-342.
6. Khan, M.; Shah, T., Literature Review on Image Encryption Techniques. *3D Res.* **2014**, 5 (4), 25.
7. Brown, T. B.; Mané, D.; Roy, A.; Abadi, M.; Gilmer, J. Adversarial Patch 2017, p. arXiv:1712.09665. <https://ui.adsabs.harvard.edu/abs/2017arXiv171209665B> (accessed December 01, 2017).
8. Sharif, M.; Bhagavatula, S.; Reiter, M. K.; Bauer, L.; Acm, Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition. *Assoc Computing Machinery*: New York, 2016; p 1528-1540.
9. Gnanasambandam, A.; Sherman, A.M., Optical Adversarial Attack. *International Conference on Computational Photography*. **Submitted**.

## CHAPTER 7. ADVERSARIAL SPECTROSCOPY

In recent years, more and more decision-making processes in science and everyday life have relied on increasingly larger amounts of data, and heavily depend on the validity of such data. For chemical decision-making processes, spectral classifiers are commonly used for sample identification but are growingly susceptible to intentional manipulation of the spectral data. The vulnerability of these classifiers to an adversarial attack is explored for a collected set of Raman data for clopidogrel bisulfate. To this end, an adversarial attack based on the addition of patterned noise in linear discriminant analysis (LDA) space was demonstrated. Following this, generation of “decoy” spectral data such that classification of the crystalline forms of clopidogrel bisulfate were intentionally misclassified by use of generative adversarial LDA (GALDA). Training of a new spectral classifier with this “decoy” data enabled a classification architecture immune to adversarial attacks based on this patterned noise as well as reduced overfitting artefacts associated with conventional LDA. Thus, the overall security and confidence in decisions made using dimension reduction and this improved classifier was increased.

### 7.1 Introduction

In everyday life, the average use and collection of data has significantly increased and will continue to increase more and more. Cell phone applications accumulate data continuously and process said data to improve the user experience,<sup>1</sup> smart home equipment real-time process household conditions,<sup>2</sup> and even modern scientific research has increased its dependence on data.<sup>3</sup> Improvements to existing instrumentation, and the development of novel techniques and instrumentation dramatically increase the amount of data collected by an average researcher. As the amount of data increases though, the number of researchers and scientific professionals to make informed decisions using this increased load of data has not adjusted accordingly.<sup>4</sup> Thus, developments in data science have allowed researchers to combat this increased data volume complexity.<sup>5</sup>

Focusing on the analysis spectral data collected from a Raman spectrometer, programs capable of spectral classification have enabled rapid analysis of this data.<sup>6</sup> In general, Raman spectroscopy measures the vibrational modes of molecule and produces a corresponding Raman

spectrum. Variations chemical structure, or local environment can alter the total amount of Raman scattering produced.<sup>6-8</sup> Thus, for a molecule such as clopidogrel bisulfate which has multiple crystalline forms, spectra collected from each form is unique to that form.<sup>6</sup> Since the bioavailability and efficacy of each form is varied and differentiation is important, spectral classification can enable detection and differentiation of the varying forms.<sup>6, 9</sup> Common methods for spectral classification or aid in improved classification include use of neural networks (NNs),<sup>10</sup> LDA,<sup>6, 11</sup> principal component analysis (PCA),<sup>12</sup> etc.

LDA and PCA are methods that by themselves are not spectral classifiers, but can enable easier spectral classification by dimension reduction. Both methods rely on determination of orthogonal basis sets that characterize the features of a spectrum.<sup>13-14</sup> Projection of a spectrum onto said basis sets reduces the overall dimension to a single point in this projection space (the dimension of this space is  $\mathbb{R}^n$ , where  $n$  is the number of basis sets projected onto). For PCA, calculation of the basis sets is unsupervised and maximizes the variance of the projected data.<sup>14</sup> LDA on the other hand is a supervised process that is better defined to enable post-dimension reduction spectral classification. By definition of spectral classes (e.g. spectra given by varying polymorphs of clopidogrel bisulfate), the calculated basis sets of LDA best separate these defined classes.<sup>6</sup> Once projected onto LDA space, spectral classification is possible by defining the boundaries regions between the separated classes. Classification of a given spectrum in which the class is unknown can then done by localizing where this spectrum lands in LDA space.

While LDA is well equipped to enable simple spectral classification, limitations for this method exist. Primarily, class separation is not guaranteed using LDA for classes with highly similar spectra (especially so for lower dimensional LDA space).<sup>15</sup> While class separation is more probable in high-dimensional LDA space, training of the LDA basis set is limited by overfitting to the training data used. Additionally, while LDA is a generally better dimension reduction method for classification than PCA, LDA needs a sufficient amount of data of each class for training for recovery of the LDA basis sets.<sup>6, 13</sup> Finally, LDA is susceptible to variations in spectral noise which can spread out the overall class region in LDA, and can result in overlapping classes.<sup>16</sup>

Focusing on this susceptibility to spectral noise, this work demonstrates an adversarial attack approach on spectral space for the intentional misclassification of spectrum. Taking the work done by Warde-Farley and Goodfellow in which images were intentionally misclassified by the intentional addition of patterned noise and extending this to spectral data instead,<sup>17</sup> this work

demonstrates the first attempt at adversarial attacks on chemical spectra. In brief, these adversarial attacks use knowledge the LDA basis sets to generate decoy spectra that visually appear as a specific spectral class but have patterned noise that causes misclassification. Visualization of this patterned noise can be seen in Figure 7.1. Figure 7.1 A and Figure 7.1 B show the projection of two varying spectral classes on the same LDA space in which the small black arrows denote major class features and the red arrows the natural noise background present in a given class. Since the point each class lands on is separated by the calculated decision boundary, each class is properly classified. Figure 7.1 C shows the idea behind the adversarial spectroscopy in which the spectral noise of a given class is instead patterned to move the projection location past the decision boundary and cause misclassification.

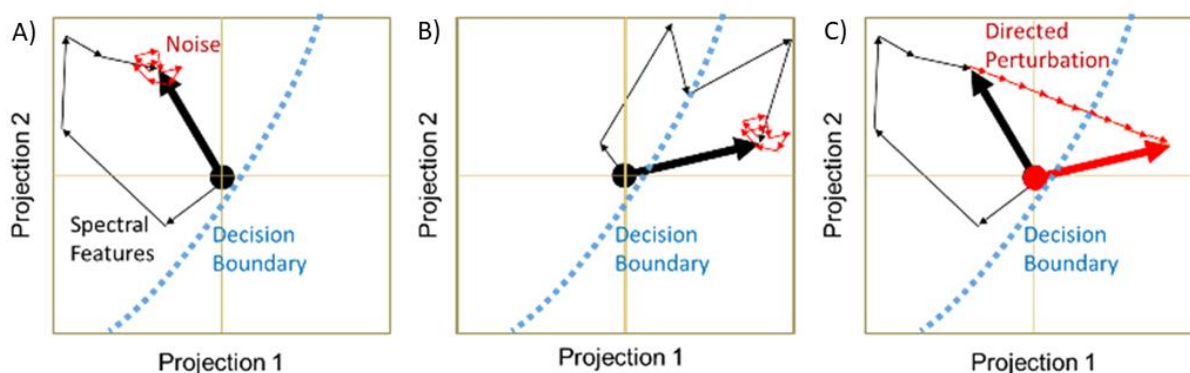


Figure 7.1. Conceptual illustration of adversarial attack on spectral data. A) and B) represent the projection of spectra of two different classes onto LDA space. The thin black arrows represent major features of a given class that is used for class separation in LDA, and the thin red arrows represent the natural noise background for a given spectrum. The point in which spectrum is located is represented by the thick black arrow, and each spectrum can be seen to be located on different sides of the decision boundary allowing for classification. C) Demonstrates intentional perturbation of the spectral projects with patterned noise that moves the projection point past the decision boundary for intentional misclassification.

Generated “decoy” data from adversarial spectroscopy can then be used in the training of a new classification architecture. By training of LDA with the decoy data as a separate class, spectra that have intentionally been misclassified are identifiable and able to be removed from future analysis. Therefore, it is hoped that this work will provide a test-bed for further development of these adversarial attacks in spectral space and that training of new spectral classifiers less prone to

malfeasance will increase the overall confidence in data validity and decisions made from analysis of these data.

## **7.2 Theory**

Figure 7.2 provides an overview of the workflow for GALDA. Raman spectra of form I (red), and form II (blue) clopidogrel bisulfate and the Raman background (black) were pre-classified and shown in Figure 7.2 A. The classified data was then reduced in dimension by LDA analysis and Figure 7.2 B shows the projection of multiple spectra onto the space. Generation of decoy data from randomly generated inputs and the resulting projection onto LDA space is shown in Figure 7.2 C. Patterned noise is then added to each decoy spectrum to induce spectral classification showing in Figure 7.2 D. Performing LDA a second time with the decoy data designated as a separate class allows for separation of this data from the real spectral data shown in Figure 7.2 E, and the resulting LDA is theorized to be less prone to adversarial attacks. Steps C - E are then repeatable until desired improvement to LDA.



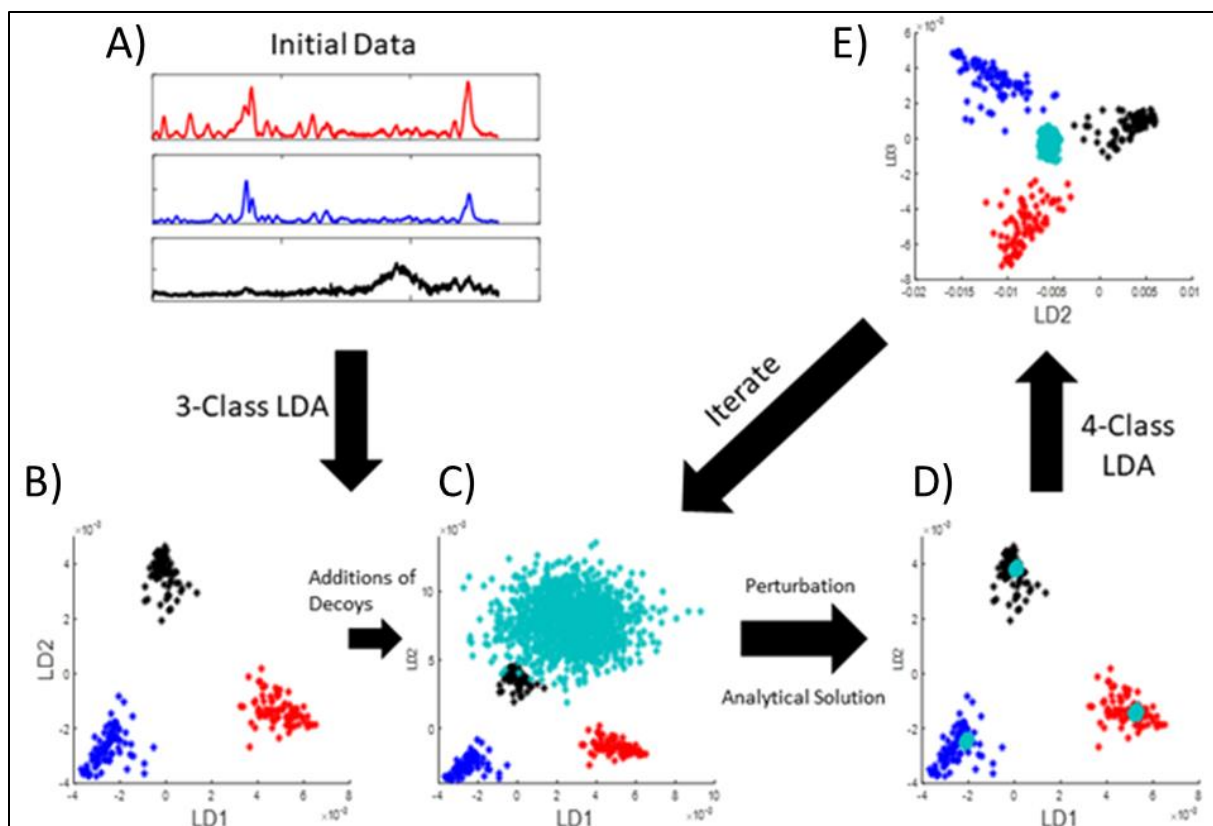


Figure 7.2. Workflow for GALDA. A) Raman spectral for forms I (red), and II (blue) of clopidogrel bisulfate and background (background). B) Projection of Raman spectra onto LDA space. C) Projection of decoy data generated from randomly inputs. D) Projection of decoy data in LDA space after addition of patterned noise. E) Separation of decoy data after training of LDA with decoy data as a separate class. Steps C – E are repeatable for until desired LDA discrimination of decoy data.

## 7.3 Methods

### 7.3.1 Raman spectra

Pure form I and form II of clopidogrel bisulfate were produced at Dr. Reddy's Laboratories and were used as received. Particles were spherical in nature with similar distributions in particle size. 252 Raman spectra were collected using a custom Raman microscope system from clopidogrel samples on glass slides containing both forms, and were separated into the three respective classes (form I, form II, and background).<sup>6</sup> Details of the Raman microscope are further discussed in [6]. After collection of Raman spectra, a Savitzky-Golay filter was applied to smooth the spectra, and a rolling ball filter was used to remove the fluorescent background. Spectra were then normalized using the respective integrated intensities.

### 7.3.2 Linear discriminate analysis

LDA was performed on the experimentally acquired Raman data using an in-house MATLAB code. For this Raman data, the Fisher's linear discriminate ( $J$ ) was calculated by Equation 7.1 where  $W$  is the matrix of optimal projection vectors to maximize class separation (i.e. the eigenvectors  $S_W^{-1}S_B$ ),  $S_W$  is the within-class variance, and  $S_B$  is the between-class variance.

$$J = \frac{W^T S_B W}{W^T S_W W} \quad 7.1$$

To calculate these variances,  $X_i$  is defined as the set of  $n$  spectra for a given class  $i$  for  $c$  number of classes. Given the centroid of  $X_i$  ( $\mu_i$ ), the within-class and between-class variance can be calculated from Equation 7.2 and Equation 7.3 respectively.

$$S_W = \sum_{i=1}^c \sum_{x \in X_i} (x - \mu_i)(x - \mu_i)^T \quad 7.2$$

$$S_B = n_{tot}^{-1} \sum_{i=1}^c n_i (\mu_i - \mu_{tot})(\mu_i - \mu_{tot})^T \quad 7.3$$

$\mu_{tot}$  is defined as the weighted mean spectrum of the entire data set containing  $n_{tot}$  number of spectra. Due to the lack in required number of spectral data for proper inversion of  $S_W^{-1}$ , inversion was performed using Gaussian elimination of the product  $S_W^{-1}S_B$ . This allowed for recovery of the inverse of  $S_W$  with far few spectra needed.

### 7.3.3 Decoy generation

Decoy spectra were generated from initial random inputs with a uniform distribution with the same dimension as the experimentally acquired Raman spectra. Bounds of  $6 \times 10^{-3}$  and 0 were placed on the random values for this decoy data to mimic the distribution of intensities for the experimentally data. Decoy data was then projected onto the reduced dimension LDA space using the LDA eigenvectors calculated for the experimental data. Projection of decoy data generally was closest to that of the background class due to the lack of spectral features.

For each spectrum, the perturbation  $\delta$  was calculated using Equation 7.4 which corresponds to the optimal noise that can reposition the given decoy data to a specific class boundary.

$$\delta = \underset{\delta}{\operatorname{argmin}} [\|D \cdot (x_s + \delta - x_t)\|_2^2 + \beta \|\delta\|_2^2] \quad 7.4$$

$D$  is comprised of the set of eigenvectors calculated from LDA,  $x_s$  is the initial decoy spectrum,  $x_t$  is the target spectrum (i.e. a randomly chosen spectrum of a given class), and  $\beta$  is a scalar

parameter that minimizes spectral perturbation. Calculation of  $\delta$  was analytically solved by optimization of partial derivatives to give Equation 7.5 where  $\hat{\delta}_i$  is the optimal perturbation.

$$\hat{\delta}_i = \frac{(\mathbf{D}^T \mathbf{D}(\mathbf{x}_s - \mathbf{x}_t))_i}{\text{Diag}(\mathbf{D}^T \mathbf{D})_i + \beta} \quad 7.5$$

## 7.4 Results

Figure 7.3 shows the initial LDA results for the experimentally acquired Raman data (Figure 7.3 A) and the corresponding mean spectrum of each class (Figure 7.3 B). In Figure 7.3 A, blue corresponds to form I of clopidogrel bisulfate, red to form II, and black to the glass slide background (blue and black color designations are swapped for Figure 7.3B). The mean spectrum for each class was calculated from 84 individual spectra for each class. After initial LDA, the three classes were well-separated in LDA space and simple linear decision boundaries between the classes could be determined. Visually, class 3 (background) can be summarized by a single large rolling peak at around channel number 950. Class 1 and class 2 are similar, since they are both from clopidogrel bisulfate, but can easily be distinguished by the relative peak intensities at channel numbers 363 and 385 corresponding to peaks found in class 1 but not class 2.

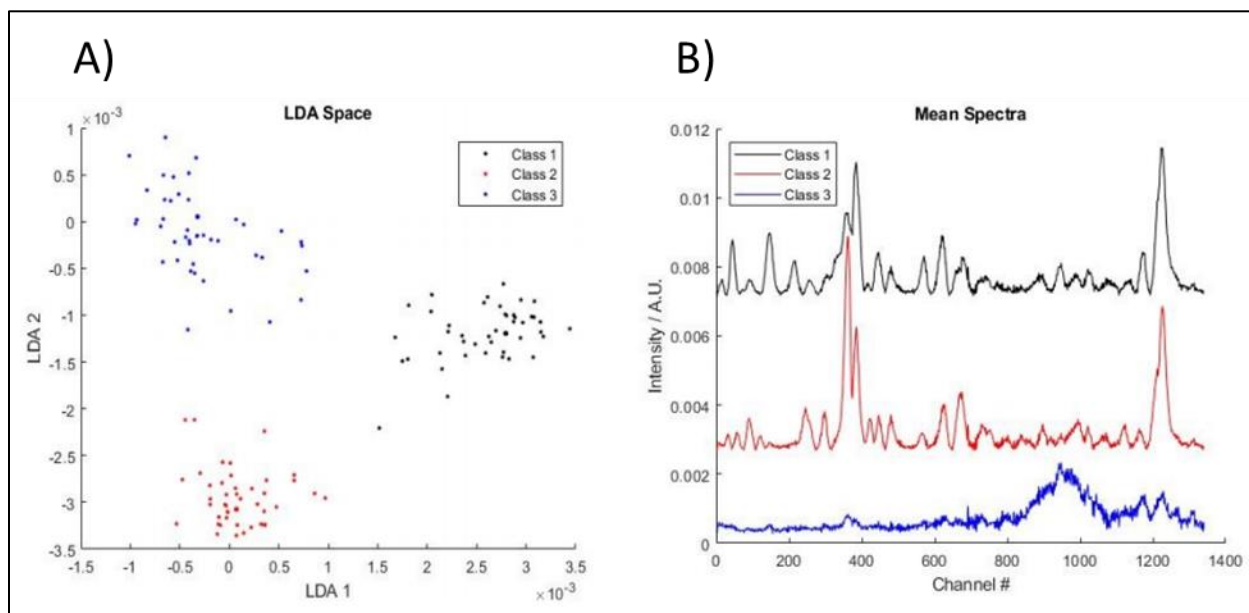


Figure 7.3. A) Projection of experimental Raman spectra for each class (form I – red, form II – blue, and background – black) in LDA-space. B) Mean spectrum for each class using 84 class measurements (spectra were displaced vertically for improved visualization).

Following the workflow shown in Figure 7.2, decoy data was generated as described above and projected onto LDA space. For each decoy spectrum, the optimal perturbation was calculated using Equation 7.5 and added to corresponding spectrum. Figure 7.4 A shows an example case for such a perturbation in which the initial spectrum (black trace) was a real Raman spectrum of class 3, and the perturbed spectrum (dark blue trace) was the spectrum corresponding to the addition of the optimal perturbation to the mean target spectrum (light blue trace) corresponding to class 2. For this case, the optimal perturbation is shown in Figure 7.4 B which as theorized simply appears as noise. Furthermore, for this case, visually the perturbed spectrum still appears to the human eye as background, but is now misclassified as class 2. Thus, this case shows the need for improved classification architectures that have been trained to detect such adversarial attacks.

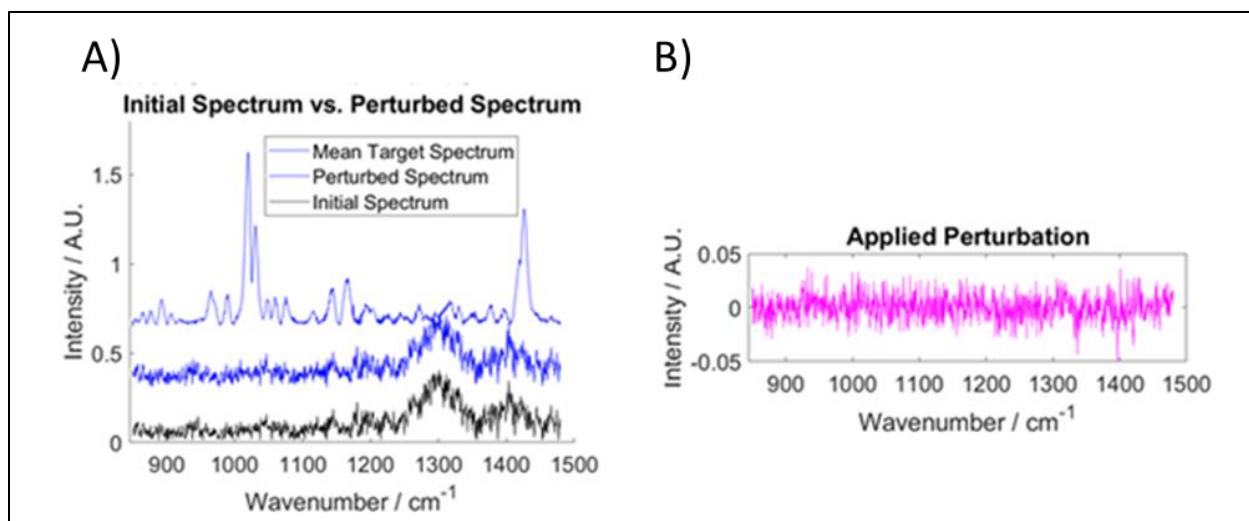


Figure 7.4. Example of a spectroscopic adversarial attack, in which a Class 3 spectrum is optimally perturbed such that it is misclassified as Class 2. A) Optimal perturbation of the initial class 3 spectrum (black trace) to that of the mean target spectrum (light blue trace) of class 2 is added to the initial spectrum to get the corresponding perturbed spectrum (dark blue trace). B) Shows the applied perturbation which simply appears as noise.

After perturbation of the decoy spectra, four class LDA was performed with this generated/perturbed data. The updated LDA basis set then allowed for separation of the decoy data from the experimental data as shown in the workflow diagram (Figure 7.2 E). Iterating this decoy generation, perturbation, and four-class LDA allowed for calculations of LDA eigenvectors less susceptible to adversarial attacks based on patterned noise. Current work on this project includes

the variation in the number of training data used to train the initial LDA step to reduce overfitting artefacts associated with conventional LDA for dimension reduction of high-dimensional data.

## 7.5 Conclusion

The first method for adversarial attacks of chemical spectra was demonstrated for Raman spectra of clopidogrel bisulfate. Using LDA as a dimension reduction method, experimental data was initially projected onto LDA space. Perturbation of randomly generated “decoy” spectra by GALDA and real spectra to a desired spectral class was then shown. Subsequent four-class LDA then allowed for discrimination of generated and real spectra, and the resulting LDA eigenvectors determined were shown to be less susceptible to further adversarial attacks by patterned noise. Thus, the overall confidence in spectral classification is significantly improved, and the validity of decisions made using this classification architecture is increased as well.

## 7.6 References

1. Gupta, A.; Kaushal, R.; Ieee, *Towards Detecting Fake User Accounts in Facebook*. Ieee: New York, 2017; p 137-142.
2. Dahmen, J.; Thomas, B. L.; Cook, D. J.; Wang, X. B., Activity Learning as a Foundation for Security Monitoring in Smart Homes. *Sensors* **2017**, *17* (4), 17.
3. Provost, F.; Fawcett, T., Data Science And Its Relationship To Big Data And Data-Driven Decision Making. *Big Data* **2013**, *1* (1), 51-59.
4. Anagnostopoulos, I.; Zeadally, S.; Exposito, E., Handling big data: research challenges and future directions. *J. Supercomput.* **2016**, *72* (4), 1494-1516.
5. Aparicio, S.; Aparicio, J. T.; Costa, C. J.; Ieee, Data Science and AI: trends analysis. In *2019 14th Iberian Conference on Information Systems and Technologies*, Ieee: New York, 2019.
6. Chowdhury, A. U.; Ye, D. H.; Song, Z. T.; Zhang, S. J.; Hedderich, H. G.; Mallick, B.; Thirunahari, S.; Ramakrishnan, S.; Sengupta, A.; Gualtieri, E. J.; Bouman, C. A.; Simpson, G. J., Second Harmonic Generation Guided Raman Spectroscopy for Sensitive Detection of Polymorph Transitions. *Anal. Chem.* **2017**, *89* (11), 5959-5966.
7. Yang, H. O.; Kim, J. H.; Kim, K. J., Study on the Crystallization Rates of beta- and epsilon-form HNIW in in-situ Raman Spectroscopy and FBRM. *Propellants Explos. Pyrotech.* **2020**, *45* (3), 422-430.

8. Dandeu, A.; Humbert, B.; Carteret, C.; Muhr, H.; Plasari, E.; Bossoutrot, J. M., Raman spectroscopy - A powerful tool for the quantitative determination of the composition of polymorph mixtures: Application to CaCO<sub>3</sub> polymorph mixtures. *Chem. Eng. Technol.* **2006**, 29 (2), 221-225.
9. Lira, L. M.; Lourenco, A. L.; Neves, G. F.; de Sousa, V. P.; Rodrigues, C. R.; Cabral, L. M., Assessment of analytical techniques for characterization of crystalline clopidogrel forms in patent applications. *Braz. J. Pharm. Sci.* **2014**, 50 (2), 229-242.
10. Yu, K.; Zhang, Y. S.; Jiang, D. S., The application of wavelet neural network in spectral data compression and classification. *J. Infrared Millim. Waves* **1998**, 17 (3), 215-220.
11. Yuan, H. L.; Tang, Y. Y.; Lu, Y.; Yang, L. N.; Luo, H. W., Spectral-Spatial Classification of Hyperspectral Image Based on Discriminant Analysis. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, 7 (6), 2035-2043.
12. Qin, D. M.; Hu, Z. Y.; Zhao, Y. H., Automated classification of stellar spectra based on PCA and wavelet transform. In *Object Detection, Classification, and Tracking Technologies*, Shen, J.; Pankanti, S.; Wang, R. S., Eds. Spie-Int Soc Optical Engineering: Bellingham, 2001; Vol. 4554, pp 268-273.
13. Sharma, A.; Paliwal, K., Linear discriminant analysis for the small sample size problem: an overview. *Int. J. Mach. Learn. Cybern.* **2015**, 6 (3), 443-454.
14. de Ketelaere, B.; Hubert, M.; Schmitt, E., Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent, High-Dimensional Data. *J. Qual. Technol.* **2015**, 47 (4), 318-335.
15. Wan, H.; Wang, H.; Guo, G. D.; Wei, X., Separability-Oriented Subclass Discriminant Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, 40 (2), 409-422.
16. Morais, C. L. M.; Lima, K. M. G.; Martin, F. L., Uncertainty estimation and misclassification probability for classification models based on discriminant analysis and support vector machines. *Anal. Chim. Acta* **2019**, 1063, 40-46.
17. Warde-Farley, D.; Goodfellow, I., *Adversarial Perturbations of Deep Neural Networks*. Mit Press: Cambridge, 2016; p 311-342.

## **VITA**

Alex M. Sherman was born on January 5, 1994 in Salt Lake City, Utah to Gary and Michelle Sherman. He graduated Great Falls High School in 2021. He then attended Montana State University (MSU), where he obtained a B.S in general chemistry and mathematics in 2016. After graduating from MSU, he began his graduate career in chemistry at Purdue University. Working under Dr. Garth Simpson, his thesis work focused on the development of optical methods with emphasis on nonlinear optical microscopy, and development of new data analysis methods focusing mainly on image analysis techniques.

## PUBLICATIONS

1. **Sherman, A.M.**; Geiger, A.C.; Smith, C.J.; Taylor, L.S.; Hinds, J.; Stroud, P.A.; Simpson, G.J., Stochastic Differential Scanning Calorimetry by Nonlinear Optical Microscopy. *Anal. Chem.* **2020**, 92 (1), 1171-1178.
2. **Sherman, A.M.**; Takanti, N.; Rong, J.; Simpson, G.J., Nonlinear optical characterization of pharmaceutical formulations. *Trends Analyt Chem.* **2021**, pre-proof.