RICE UNIVERSITY

Experimental and Numerical Investigations of Novel Architectures Applied to Compressive Imaging Systems

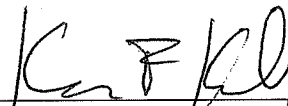by

Matthew Adam Turner

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

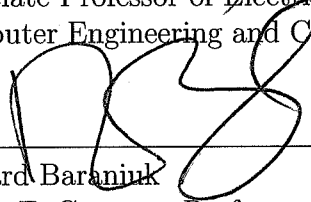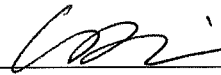## Master of Science

APPROVED, THESIS COMMITTEE:

Kevin F. Kelly, Chair
Associate Professor of Electrical and
Computer Engineering and Computer Science

Richard Baraniuk
Victor E. Cameron Professor of Electrical and
Computer Engineering and Computer Science

Wotao Yin
Assistant Professor of Computational and
Applied Mathematics

Houston, Texas

April, 2012

ABSTRACT

Experimental and Numerical Investigations of Novel Architectures Applied to Compressive
Imaging Systems

by

Matthew Adam Turner

A recent breakthrough in information theory known as *compressive sensing* is one component of an ongoing revolution in data acquisition and processing that guides one to acquire less data yet still recover the same amount of information as traditional techniques, meaning less resources such as time, detector cost, or power are required. Starting from these basic principles, this thesis explores the application of these techniques to imaging. The first laboratory example we introduce is a simple infrared camera. Then we discuss the application of compressive sensing techniques to hyperspectral microscopy, specifically Raman microscopy, which should prove to be a powerful technique to bring the acquisition time for such microscopies down from hours to minutes. Next we explore a novel sensing architecture that uses partial circulant matrices as sensing matrices, which results in a simplified, more robust imaging system. The results of these imaging experiments lead to questions about the performance and fundamental nature of sparse signal recovery with partial circulant compressive sensing matrices. Thus, we present the results of a suite of numerical experiments that show some surprising and suggestive results that could stimulate further theoretical and applied research of partial circulant compressive sensing matrices. We conclude with a look ahead to *adaptive sensing* procedures that allow real-time, interactive optical signal processing to further reduce the resource demands of an imaging system.

First and foremost, I must thank my family still on this Earth, and those passed on to the next life. I am deeply endebted to their unending encouragement, dedication, and love. Indeed this achievement owes much to my parents who have always driven me to persevere.

Next, I must express my deepest appreciation to my research adviser Kevin F. Kelly who encouraged me to apply for the NSF IGERT fellowship, provided space for me to work and grow as a researcher, and supported my ventures out into side projects and allowing me to find my own research path. Thanks also to my Kelly Lab-mates, especially Lina Xu, Yun Li, and Ting Sun from the camera lab. Cheers to Chad Byers for drinking coca tea and many beers with me in Bogotá. Thanks to Chaitra Rai for being a great friend who always had an open ear. Thanks to Corey Slavonic for many good conversations on walks for coffee, at lunch, or any of the many times I locked myself out of the office or lab.

Thanks very much to Woato Yin whose class taught me how important high-dimensional geometry is to understanding compressive sensing and sparse recovery. Indeed, his class caused me to retreat at first because it was so difficult, but in that difficulty I regained a grip and was able to view anew the beauty of geometry, which in high school first attracted me to mathematics.

Thanks to Richard Baraniuk for being such a positive character in the Rice Electrical and Computer Engineering Department. Because of him we have an all-star lineup of speakers every semester. His group members have been immensely helpful to me in the process of my Master's work.

I must also acknowledge Roger Moye of the Rice Shared Computing Grid who patiently answered my questions on Unix programming, batch scripts, parallel computing, and for his timely responses when MATLAB or something else was not working on the cluster. If not for his help, Chapter 4 would not exist.

Finally, a very huge thank you to my wonderful friends who have also been bandmates, mentors, roommates, and psychiatrists throughout these past four years. Life would be hell without you.

Contents

Illustrations

Tables

Chapter 1

Motivation and Introduction

1.1 Overview

The physical sciences, like the all aspects of society, are strained under the effects of the "data deluge," which could be described as our species' overwhelming ability to acquire or create data coupled with our relatively weak skill in extracting useful information from that data. In a raster scanning, hyperspectral microscope, such as a darkfield, Raman, or Fourier-transform infrared microscope, large amounts of data are collected, and oftentimes acquisition is either time-intensive, or in the case of infrared microsocopy and spectroscopy, monetarily expensive because exotic, non-silicon based detectors are required. This thesis is a description, exploration, and exposition, of how a new theory in signal processing, known as *compressed sensing*, can be applied to microscopy and other imaging systems in order to minimize these and other costs. Compressed sensing (CS) is a mathematical jewel itself. A groundswell of mathematical and engineering work has risen exploring the implications and theoretical applications of this theory. Based on the applications it has already found, CS could prove to be one of the most useful developments in mathematics so far this century.

In this chapter, we introduce some notation we will need to describe compressive imaging systems, followed by the introduction of our compressive infrared imaging system where we will further illustrate the connection between the measurement formalism and the physical camera system. We will see that compressive imaging reduces to solving an underdetermined set of equations, unsolvable by elimination methods. In order to solve these equations, we

regularize the problem, in other words, we apply *a priori* information so that the deficit of information is sufficiently reduced, and we recover an image as we would have otherwise.

1.2 Introduction to the Mathematics of Compressive Sensing

1.2.1 Matrix representation of a system of equations

In order to describe our measurement systems mathematically, we need to be able to efficiently write large systems of equations. In the sequel, *signal* and *image* will be used interchangably. Although there are philosophical differences, mainly that a signal can be exactly known and recovered, but an image will always be an approximation to "true reality," reality itself being an point of philosophical debate. Generally, we use the word 'signal' when referring to an arbitrary $x \in \mathbb{R}^n$ and 'image' when referring specifically to a signal approximated by an imaging system.

We now demonstrate how to compactly write large systems of equations in matrix notation by means of a simple example. Consider the system of equations

$$3x_1 + 2x_2 = -5$$

$$3x_1 - 9x_2 = 3$$

$$\text{with } x_1, x_2 \in \mathbb{R}.$$

This is represented in matrix form by

$$y = \Phi x \tag{1.1}$$

where

$$y = \begin{bmatrix} -5 \\ 3 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \text{ and } \Phi = \begin{bmatrix} 3 & 2 \\ 3 & -9 \end{bmatrix}$$

This allows us to efficiently describe an arbitrary number of equations and arbitrary number of unknowns,

$$
y \in \mathbb{R}^m \left\{ \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \underbrace{\begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m1} & \varphi_{m2} & \cdots & \varphi_{mn} \end{bmatrix}}_{\Phi \in \mathbb{R}^{m \times n}} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right\} x \in \mathbb{R}^n
$$

(1.2)

is still just $y = \Phi x$. Using this matrix notation for a system of equations, we can describe the acquisition of data that results in an image. For pixel-array or raster scan imaging,

$$
\Phi = I = \delta_{ij}, \ \text{with} \ \delta_{ij} = \begin{cases} 1 \ \text{if} \ i = j \\ 0 \ \text{otherwise} \end{cases},
$$

or

$$
I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}
$$

By simply reshaping the resulting vector y to be a rectangular 2D array, we recover an image of the scene. Each row of Φ, which we denote $\varphi^{(i)}$, probes the ith discretized point of x. If we regard x as the scene to be imaged, note that x is not discrete until we impose some grid on it. The value $y_i = \langle \varphi^{(i)}, x \rangle$ measures the brightness corresponding to each pixel. We call $\varphi^{(i)}$ is the ith *measurement vector*. If we write a column of Φ as $\varphi_i^c \in \Phi = (\varphi_1^c, \ldots, \varphi_n^c)$, then

we get the representation of the signal x,

$$y = \varphi_1^c x_1 + \varphi_2^c x_2 + \ldots + \varphi_n^c x_n = \sum_{i=1}^{n} \varphi_i^c x_i. \tag{1.3}$$

If $\Phi = I$,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} x_2 + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} x_n. \tag{1.4}$$

Thus we see that the measurements y may be viewed as a weighted sum of the coumns of Φ where the weights are the discretized points of the scene, x.

In this work, compressive imaging is achieved through techniques deriving from *transform imaging*, where more than one pixel is probed at a time, or, in other words, each measurement vector $\varphi^{(i)}$ has many 1s in it. In the example to follow, $\sum_{j=1}^{n} \varphi_j^{(i)} = \frac{n}{2}$, $i = 1, \ldots, m$ and $\varphi^{(i)} \in \{0, 1\}^n$. For compressive imaging, however, we do not have $\Phi \in \mathbb{R}^{n \times n}$, but instead $\Phi \in \mathbb{R}^{m \times n}$ with $m < n$. In other words we have less equations than unknowns. To quantify the amount of undersampling, define the undersampling, or equivalently compression, ratio

$$\delta = \frac{m}{n}.$$

The magic of compressive sensing is that we are still able to recover at least a good approximation to x, if not x exactly, from the underdetermined set of equations arising from our measurements $y = \Phi x$. Before explaining how to recover an image from these underdetermined equations, here is an example of compressive imaging in action that also serves to further solidify the notation to be used throughout the rest of this thesis.

1.3 Infrared Imaging via Compressive Sampling

As an example to prepare for the sequel in which we explore more advanced imaging systems, we introduce the 'Rice single-pixel camera.' Specifically this camera is an infrared camera, one of the useful applications of compressive sensing, especially for remote sensing and data-fusion applications of compressive sensing where detector cost is not the only one to be mitigated. Infrared imaging has important applications in missile technologies as is well-known, and also for night vision and surveillance techniques. Predator drones, for example, routinely acquire their targets for assassination with highly advanced infrared cameras [1]. A more interesting application may be to combine multiple views for enhanced video sensing via multiple compressive streams [2], and perhaps even do some sort of compressive data fusion with radar imaging systems on drones to develop a full view of a scene, both indoors via "through-the-wall radar imaging" [3] and outdoors. A further bonus is that data collected via compressive imaging is naturally encrypted as well as copmressed with no on-board computing. This will become more clear once we better understand compressive imaging.

1.3.1 Single-Pixel Camera General Setup

Here we begin our introduction of the infrared camera system by introducing the optical element that displays the measurement vectors $\varphi^{(i)}$. This is the digital micromirror device (DMD) from Texas Instruments, Inc., shown in Figure 1.1 displaying one measurement vector with an illustration of a series of m measurement vectors, again corresponding to the m rows of the measurement vector Φ. The white pixels correspond to $\varphi_{ij} = 1$, or we say this is an 'on' pixel, and black corresponds to $\varphi_{ij} = 0$, or an 'off' pixel.

As in Figure 1.2, we focus our target scene, in this case the blue painted card as in Figure 1.3, onto the digital micromirror device (DMD). The DMD, has 1024×768 mirrors with a diagonal length of $13.6\,\mu\text{m}$, so the maximum resolution in pixels one can attain in
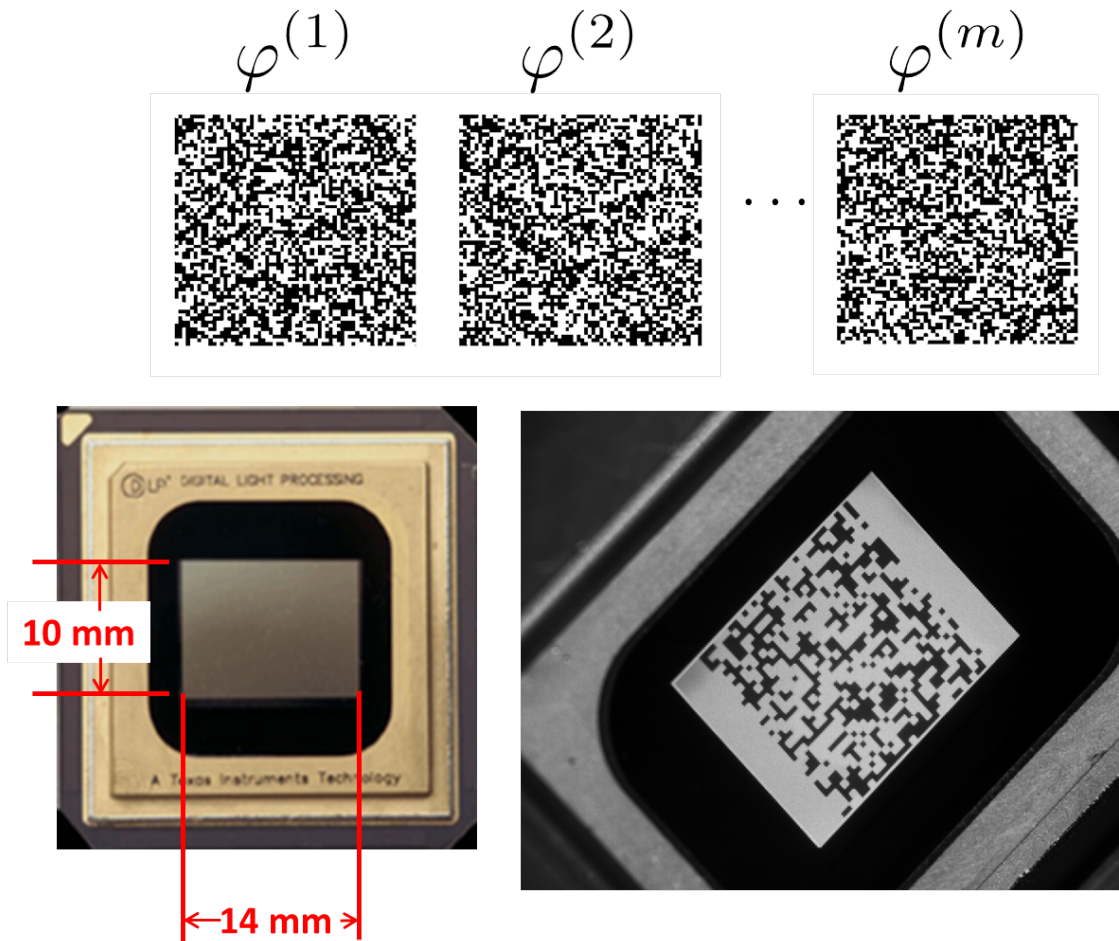
Figure 1.1 : Detail of the DMD along with a photograph of a DMD displaying a 32×32 permuted Walsh-Hadamard measurement vector, and a series of M 128×128 measurement vectors above it.

this setup is $n = 768 \times 1024$. The size of the mirrors is compared with an ant leg in Figure 1.2. To acquire images at other resolutions, we operate mirrors in blocks to constitute one pixel. The light from the scene is collected by a lens and focused onto the DMD. The DMD displays reshaped rows of the measurement matrix Φ, the photodetector measures converts light intensity to a voltage that is sent and stored on computer. The ith voltage measurement serves as the inner product $y_i = \langle \varphi^{(i)}, x \rangle$ to yield the set of measurements $y = \Phi x$. An 'on' pixel, or $\varphi_{ij} = 1$, directs light towards the detector, and an 'off' pixel directs light away from the detector. The mirrors are fixed to only flip $\pm 12°$ away from parallel with the face of the DMD. Thus, as in the transform imaging case described above, the DMD encodes the scene, and the photodetector measures the total intensity of light reflected towards the detector from the DMD for each measurement vector.

Our measurement matrix is a partial permuted Hadamard measurement matrix, which we write as

$$\Phi = R_\Omega \, P S_n \tag{1.5}$$

where

$$S_n = f(H_n) \tag{1.6}$$

where H_n is the Hadamard matrix of order n and f is a function on $A \in \mathbb{R}^{n \times n}$, a_{ij} being the elements of A such that

$$f(a_{ij}) = \begin{cases} 1 \text{ if } a_{ij} = 1 \\ \\ 0 \text{ otherwise} \end{cases} \tag{1.7}$$

This follows, but is not identical to, the construction of "S matrices" in Harwit and Sloane [4]. The purpose is to allow us to use Hadamard matrices with a single detector, meaning our measurement system can only implement measurement matrices Φ with elements $\varphi_{ij} \in \{0, 1\}$. P is an operator that permutes the columns of S_n and R_Ω selects a set of rows

Photodetector-
converts incident photons to voltage

Image

Digital Micromirror
Device (DMD)

Lens

Object
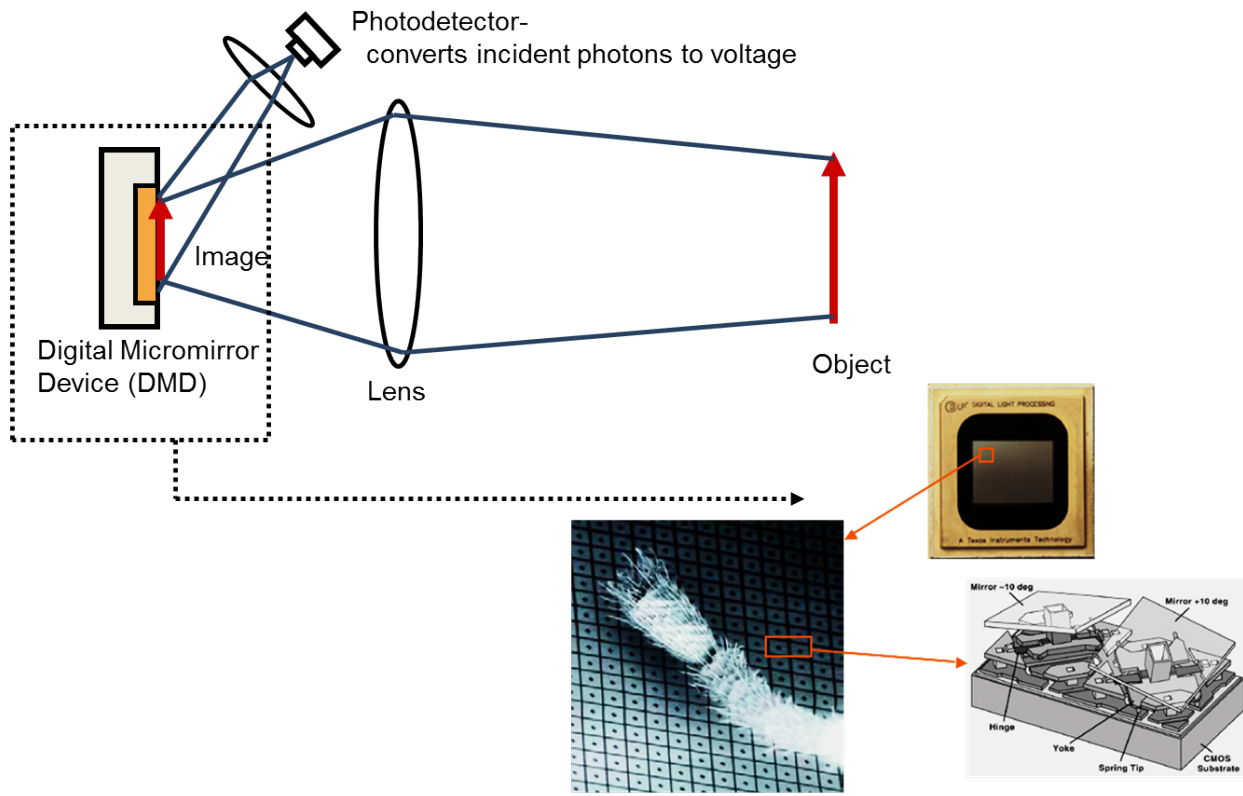
DMD resolution: 1024x768 mirrors = 786,432 mirrors. Dimensions: .56"x.42"

Figure 1.2 : Illustration of imaging system with DMD detail

indexed by the entries of the set Ω. We take $|\Omega| = m$ which determines the number of measurements we acquire, and traditionally we have set Ω to be m integers taken at random from the set of natural numbers less than or equal to n without replacement. Compressive sensing theory so far yields stronger guarantees for signal recovery when there is some element of randomness in the measurement vectors, thus the random permutations of the columns and then random selection of the rows [5, 6].

Once we have the measurement matrix Φ we sequentially display each row, or measurement vector, from Φ, collecting a measurement y_i for each measurement vector. Once we have displayed all m $\varphi^{(i)}$ to acquire the set of equations $y = \Phi x$, we employ the TVAL3 reconstruction algorithm to recover an image. We implement this scheme as shown in Figure 1.3 and as illustrated in Figure 1.2.

As in Figure 1.2, we focus our target scene, in this case the blue painted card as in Figure 1.3, onto the digital micromirror device (DMD). The blue square on the card is acrylic paint. Below the acrylic paint are the letters "IR" in charcoal. We illuminate the card with an array of 1450nm infrared light emitting diodes. Some of that light is able to penetrate the acrylic paint to either reflect off the card or be absorbed by the charcoal letters. The light reflecting from the scene is imaged onto the DMD, then collected by another lens and the total intensity of light coming from the DMD is measured for m different measurement vectors by a Hamamatsu photodetector (model no. G6122) sensitive to wavelengths from 1–2 μm and peak sensitivity at $\lambda = 1.95\,\mu$m.

The results of some compressive measurements for $\delta = .1$ and $\delta = .075$ are shown in Figure 1.4.
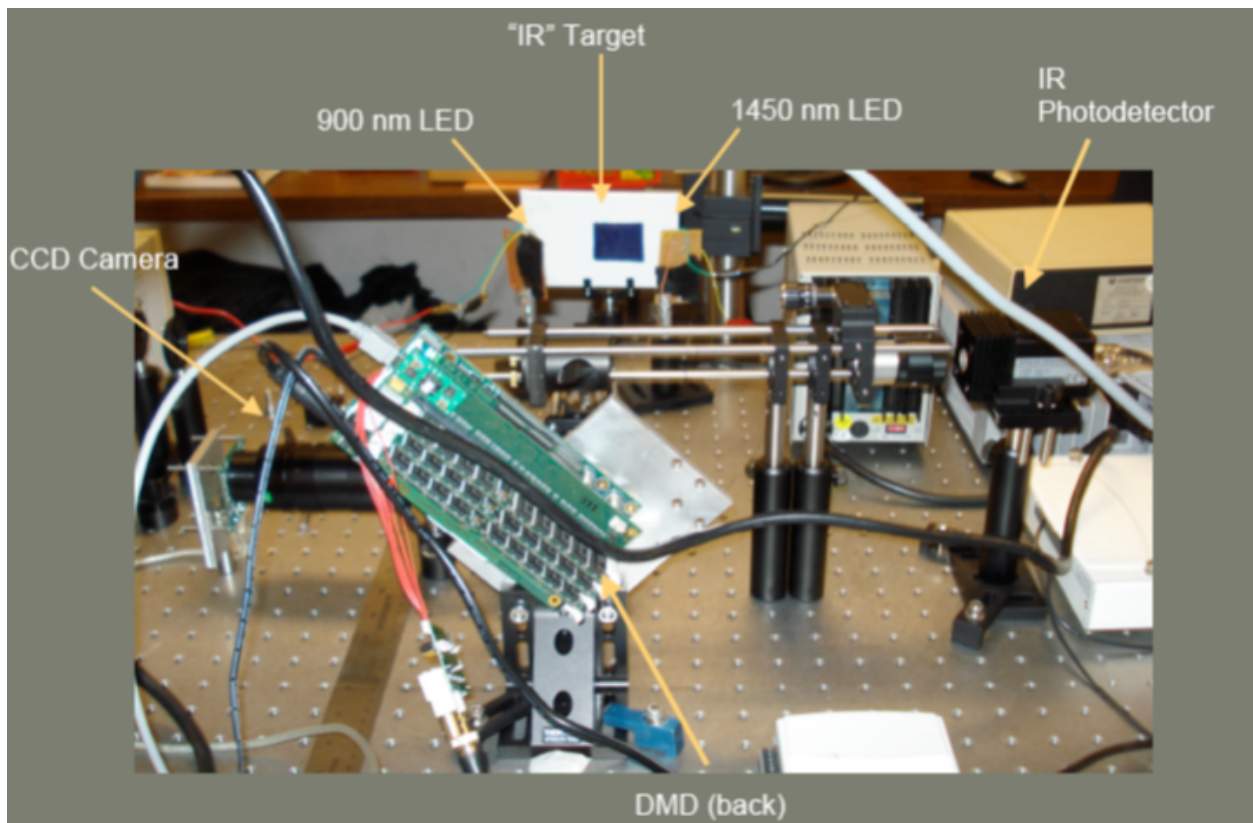
Figure 1.3 : Infrared camera setup on the optical table

$\delta \approx .1, m = 6550$ $\delta \approx .075, m = 4900$
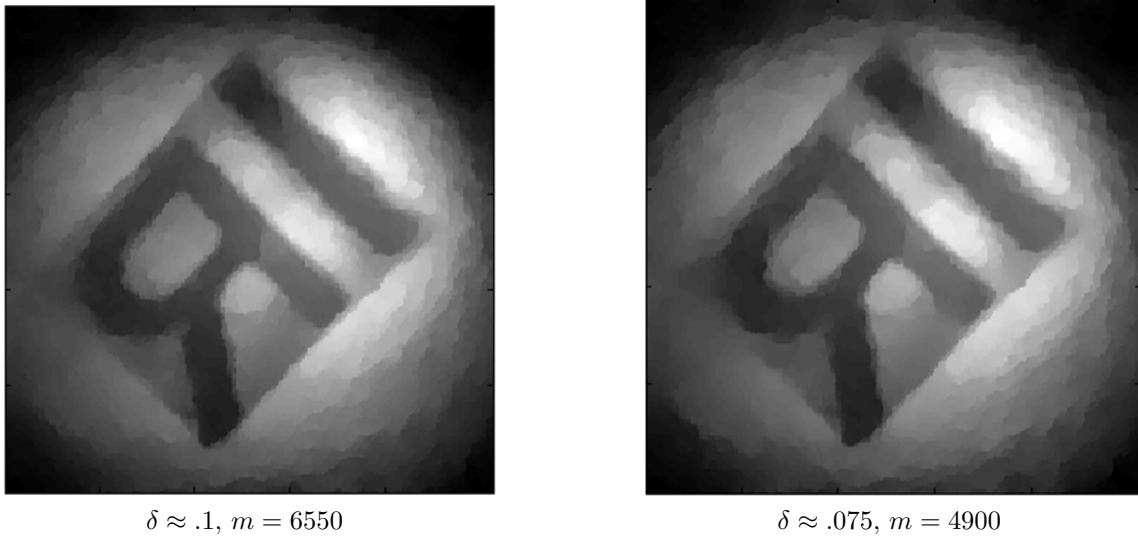
Figure 1.4 : $n = 256 \times 256$ images of charcoal-painted "IR" behind acrylic paint as imaged by our compressive imaging system with $\lambda = 1450\,\text{nm}$

1.3.2 Image Recovery with TVAL3

After we acquire the measurements, y, the challenge is to recover an image x^* that represents the scene x. To do so we *regularize* the problem, in this case by solving

$$(\text{TV}) \quad \min_x \sum_{i=1}^n \|D_i x\|_2 \text{ subject to } y = \Phi x \tag{1.8}$$

where D_i is a "local finite-difference operator" such that $D_i x \in \mathbb{R}^2$. Regularization is the mathematical process of applying *a priori* information, in this case knowledge that the solution x^* should have a small total variation (TV), $\sum_{i=1}^n \|D_i x\|_2$. This is one of many possible methods of regularization. We will explore two alternative methods in Chapter 4.

To solve this we use TVAL3 of Li, Yin, and Zhang, which stands for *total variation minimization by augmented lagrangian and alternating direction algorithms* [7]. TVAL3 traces its roots back to a seminal paper by Courant in 1943 on the quadratic penalty method

[8]. Physicists will be familiar with the general principle of TVAL3, namely the augmented Lagrangian, which is virtually identical to the Lagrangian formalism of dynamics. The augmented Lagrangian method is used to solve the problem of Equation 1.8, (TV), as follows. First, rewrite (TV) in the equivalent form

$$\min_{x,w_i} \sum_{i=1}^{n} \|w_i\|_2 \text{ subject to } y = \Phi x \text{ and } D_i x = w_i. \tag{1.9}$$

The augmented Lagrangian for the rewritten problem is given by

$$\mathcal{L}(x, \lambda, \nu) = \sum_{i=1}^{n} \left(\|w_i\|_2 - \nu_i^T (D_i x - w_i) + \frac{\beta_i}{2} \|D_i x - w_i\|_2^2 \right) - \lambda^T (\Phi x - y) + \frac{\mu}{2} \|\Phi x - y\|_2^2. \tag{1.10}$$

TVAL3 solves this problem through a so-called "alternating direction algorithm" developed by Wang, Yang, Yin, and Zhang specifically designed to solve TV minimization problems in imaging [9]. For further discussion of TVAL3 including the augmented Lagrangian formalism, see [10] and [11]. Briefly, the algorithm finds a new approximation for x given Lagrange multipliers λ and ν, and then finds optimal Lagrange multipliers with the updated x, which then are used in another iteration to find a new approximation for x, and so on until $\nabla \mathcal{L}(x, \lambda, \nu) < \varepsilon_{tol}$, where ε_{tol} is a user-defined tolerance parameter, which says that $\mathcal{L}(x, \lambda, \nu)$ has reached its global minimum, guaranteed by a global convergence theorem [11].

We used the TVAL3 algorithm with the default options for the coefficients β_i and μ in Equation 1.10, and default stopping tolerance ε_{tol} to obtain the 256×256 images in Figure 1.4. We achieve very high compression ratios, with subsampling ratios $\delta = .1$ and $\delta = .075$, corresponding to $m = 6550$ and $m = 4900$ respectively.

1.4 Other Applications and Outline of the Thesis

By way of the above example, it should be clear that compressive sensing is a powerful method to reduce the resources required to acquire information. There are many applications for compressive sensing outside of imaging. Much of the groundwork for compressive sensing had been laid down before the breakthroughs by Candès, et al [5] and Donoho [6]. The paradigm shift came out of the recognition that successful recovery of a signal depends on a quantifiable dependence between how we acquire a signal and the structure of that signal. Compressive sensing grew out of efforts to find the simplest accurate representation of a signal.

Since CS guides one to acquire more information with less resources, it has found a wide range of applications in science and engineering. Because measurements are already in a compressed form no on-board compression is required. As such, there have been proposed CS systems for astronomy [12] and hyperspectral remote sensing [13, 14]. Other examples in physics include radio astronomy-based cosmology [15], and quantum state tomography [16], which allows one to more accurately determine the state of a collection of, for example, electron spins, which has applications in quantum computing.

Compressive sensing has also been applied to medicine and biology. CS holds much promise to improve magnetic resonance imaging because of the greatly reduced acquisition time it allows [17, 18]. Thus, when the patient is less able to control their movements, as with children, or if time is of the essence as it often is in medicine, CS techniques could be of immense benefit.

Shental, et al [19], identify carriers of "rare variants" of disease via CS techniques applied to group testing. Erlich, et al [20], applied a similar method to identify genetic disease in Ashkenazi Jews, and then extended this work to "Compressed Genotyping" to identify genetic variation of any sort in any number of individuals [21]. Machines known as DNA

microarrays determine the sequence of base pairs in DNA. Dai, et al., [22] and Sheikh, et al., [23] both explore the application of compressive sensing to DNA sequencing with DNA microarrays. One of the most interesting applications of compressive sensing straddles the disciplines of biology and physics. AlQuarishi and McAdams suggest a method for using compressive sensing to effectively learn a physical model for protein-DNA interactions, which could have important applications in drug delivery, disease treatment, and fundamental genomics [24].

In the sequel, we explore the application of compressive sensing to microscopic imaging in Chapter 2, introduce a novel compressive imaging system based on circulant matrices in Chapter 3, and then study numerically the efficacy of circulant matrices for general compressive sensing in Chapter 4. The final chapter will be dedicated to final thoughts and future directions with some preliminary data from *adaptive* sensing experiments, where the ith measurement vector depends on the result $y_{i-1} = \langle \varphi^{(i-1)}, x \rangle$. Although the work here is dedicated to imaging, many of the results might have parallel applications or provide insight into problems from the varied fields of application mentioned above, espeically Chapter 4.

Chapter 2

Compressive Microscopy

Now we turn our focus from novel imaging architectures and their characterization to concrete applications to real-world imaging systems. Raman imaging is a prototypical example of a measurement system that benefits substantially from reducing the time required to acquire a Raman microscopic image. In our lab, and as is common in other laboratories as well, high-resolution raster scan Raman images require a few to tens of hours of acquisition time. Each location must be probed individually, resulting in a measurement akin to that described in . Below we will discuss in more detail the nature of acquiring data for Raman microscopy.

2.1 Compressive Microscopy

In this section, we describe some various forms our microscope system could take, as well as the measurement formalism we'll need to describe them. To do that we first introduce the standard, non-compressive raster scanning system. Then we will introduce compressive microscope systems, including a simple method to choose the best parameters for TVAL3, first introduced in Section 1.3.2, that can easily be extended to other imaging systems.

2.1.1 Raster Scanning Microscope Systems

A typical, simplified raster scanning microscope setup is shown in Figure 2.1. A laser is sent through a beamsplitting mirror through the back of an objective lens which focuses the laser light ideally to a diffraction-limited point, reflected and/or scattered light is collected by the
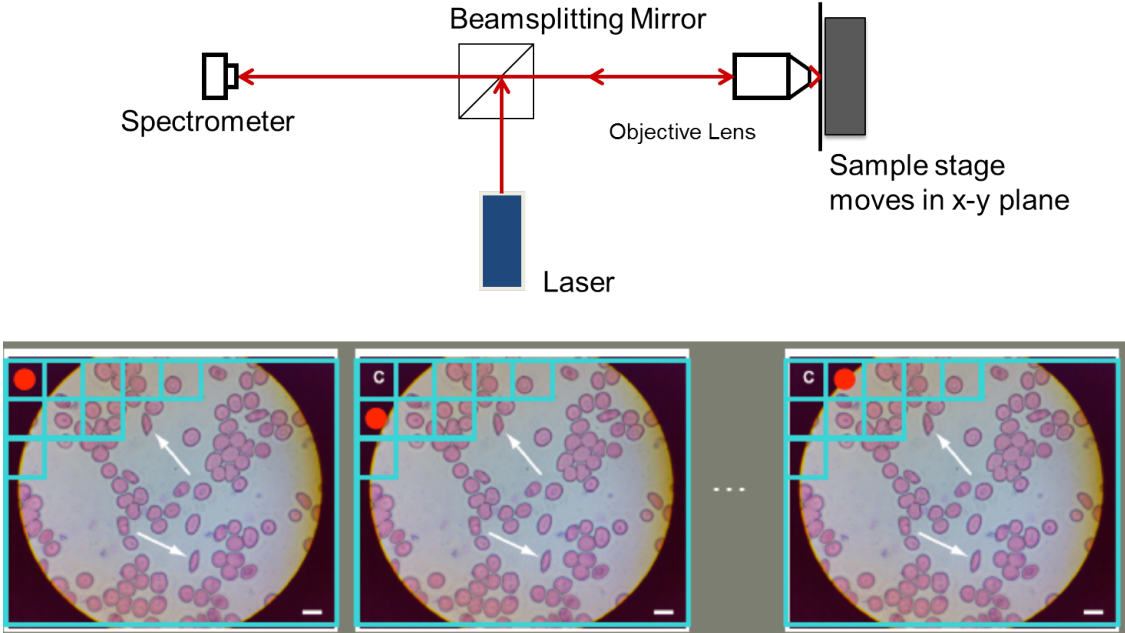
Figure 2.1 : Raster Scanning Hyperspectral Microscopy Setup

same objective lens, then directed back through the beamsplitter to either a photodetector in the case of standard imaging, or a spectrometer for hyperspectral (i.e. beyond only three, red, green, and blue, color channels) imaging. We use the same idealized measurement formalism as before. In the simple raster scan imaging system we have the measurements, y, given as

$$y = Ix, \tag{2.1}$$

where I is the $n \times n$ identity matrix. As in a pixel array, each discretized point in the scene is sampled by the measurement vector $\Phi = I$, the measurement y_i is proportional to the number of photons registered by the detector during the *acquisition time*, t_{acq}. In reality it is stored on computer as the voltage reading from a photodetector caused by the dislocation of electrons by incident photons. To physically sample each point in the image, the sample stage moves in the "x-y" directions (not to be confused with x and y vectors

from our measurement formalism) in discrete steps, the size of which determine the spatial resolution of the image. All that is needed to recover an image from the set of measurements $y = \Phi x$ when Φ is the identity is a proper accounting of what point was illuminated when.

If instead we are performing hyperspectral microscopy, instead of $y \in \mathbb{R}^n$, we have $y \in \mathbb{R}^{n \times n_{\text{spec}}}$, where n_{spec} is the resolution of the spectrometer. The spectrometer works like a prism to disperse the incoming light scattered or reflected by the sample, and then measure how much light at a set of discrete wavelengths is present at each point. However, in fact the light is dispersed by a *diffraction grating*, a reflective optical element cut with grooves. Each wavelength of light reflects at a different angle, causing the prism-like dispersion. A highly rectangular CCD pixel array is calibrated so that the light striking a portion of it is registered as, for example, $\lambda = 632.5$nm. The Ocean Optics USB4000 spectrometer we use is sensitive to light from 200-1100 nm with a Toshiba TCD1304AP Linear CCD array that has a resolution of 3648 pixels, or 3648 possible wavelength *bins*. Although the spectrometer is sensitive to light in this wide wavelength range, oftentimes a higher resolution in λ is exchanged for a smaller range of wavelength values to be probed. This technique is useful for Raman imaging, discussed below, but also for fluorescence microscopy, a popular technique used extensively in biological and medical research that allows for identification of various parts of a cell by functionalizing fluorophores, proteins that absorb then emit at specific wavelengths, to attach to specific parts of a cell. Before we discuss the Raman effect and its usefulness in more detail, we describe how we modify a white-light illumination microscope to employ compressive sensing, as well as some calibration data. We will introduce two options for a microscope setup for laser-illumination in Section **??**.
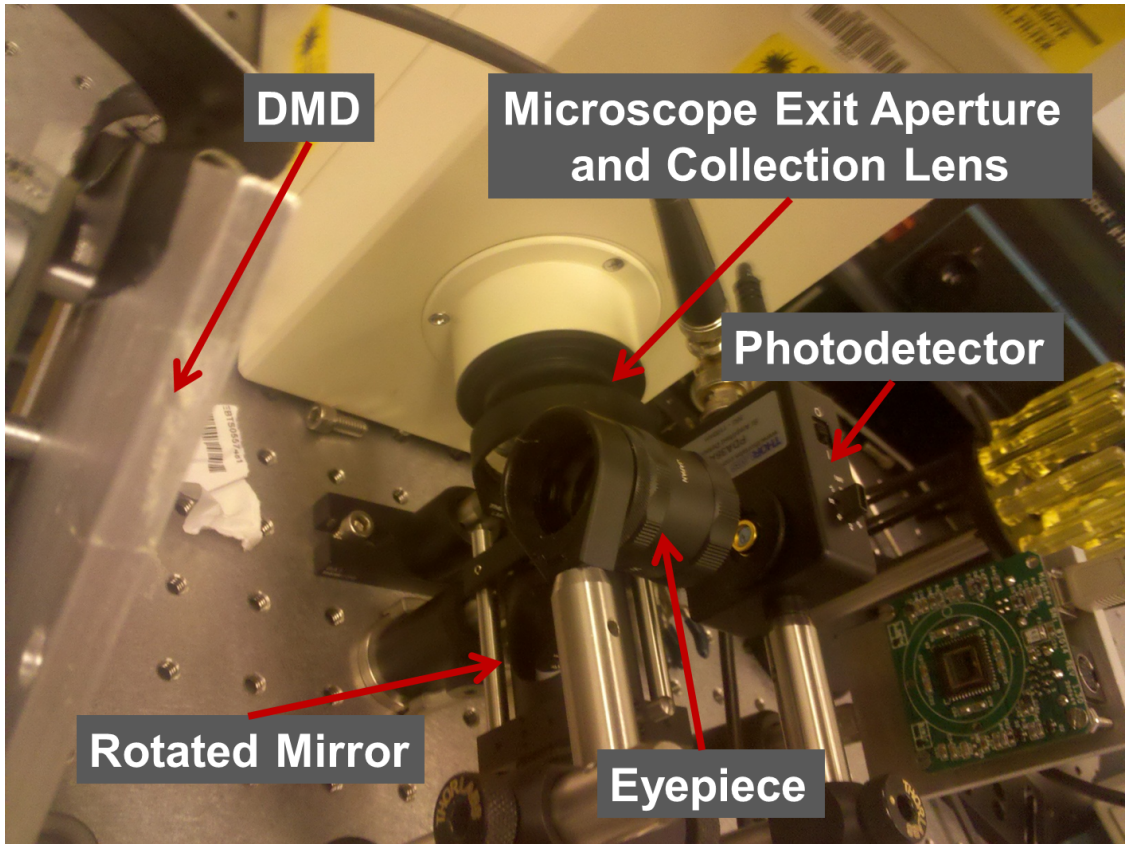
Figure 2.2 : A simple addition of our single-pixel camera to a standalone Zeiss microscope system.

2.1.2 Compressive Microscopy Setup

As introduced in the introduction, we will acquire measurements $y = \Phi x$ with Φ being a randomly permuted and subsampled Hadamard matrix as given in Equation 1.5. The experimental setup is essentially the same, except our image is collected using microscope optics. The simplest method for implementing compressive microscopy is to simply collect light from a prebuilt microscope, as shown in Figure 2.3. Here, an image collected by the internal optics of the microscope are projected out from the exit aperature, collected by a lens, then sent to the DMD via a rotated mirror. Then an eyepiece collects the light

corresponding to $y_i = \langle \phi^{(i)}, x \rangle$ and focuses it down the photodetector, which, as discussed previously, converts light intensity to an analog voltage signal, which is then converted to digital at the analog-to-digital converter (ADC), and saved on the computer.

Some calibration images of the standard Air Force test target AF-1951 taken with this setup are shown below. With $\delta = .95$ and again the default parameters for β and μ we acquired the image in Figure **??**. The smallest bars are 2.2μm wide, which sets the field of view to about 40μm^2. The image is fair quality, at best. In fact, especially in low-light situations, the default, or even the recommended range, of parameters is not optimal. To determine the best choice of β and μ, we solve and plot a series of solutions x^* to the optimization problem (TV) in Equation 1.8. Then, either using some mathematical heuristic or simply by visual inspection, one may choose the optimal parameters.

2.2 Raman Imaging

2.2.1 The Raman Effect

The Raman effect was first discovered by Indian physicist C.V. Raman in 1929. Briefly, the Raman effect is an optical, quantum mechanical effect that probes the vibrational properties of a material, be it in the gas, liquid, or solid phase. As is common knowledge, some of the light reflected from a material is the same wavelength as the incident light; the technical term is Rayleigh scattering. However, if the material is 'Raman-active,' there will also be other wavelengths of light 'reflected' as well. More accurately, this light is *scattered* instead of reflected, and we say it is Raman-scattered light. For readers familiar with fluorescence, this may sound like fluorescence, but it is not the same effect. The only true similarity is that both effects are mediated by the absorption of quanta of light, photons, by the negative charge carriers of a material, electrons.

One of the most basic principles of physics, the conservation of energy, is the route by
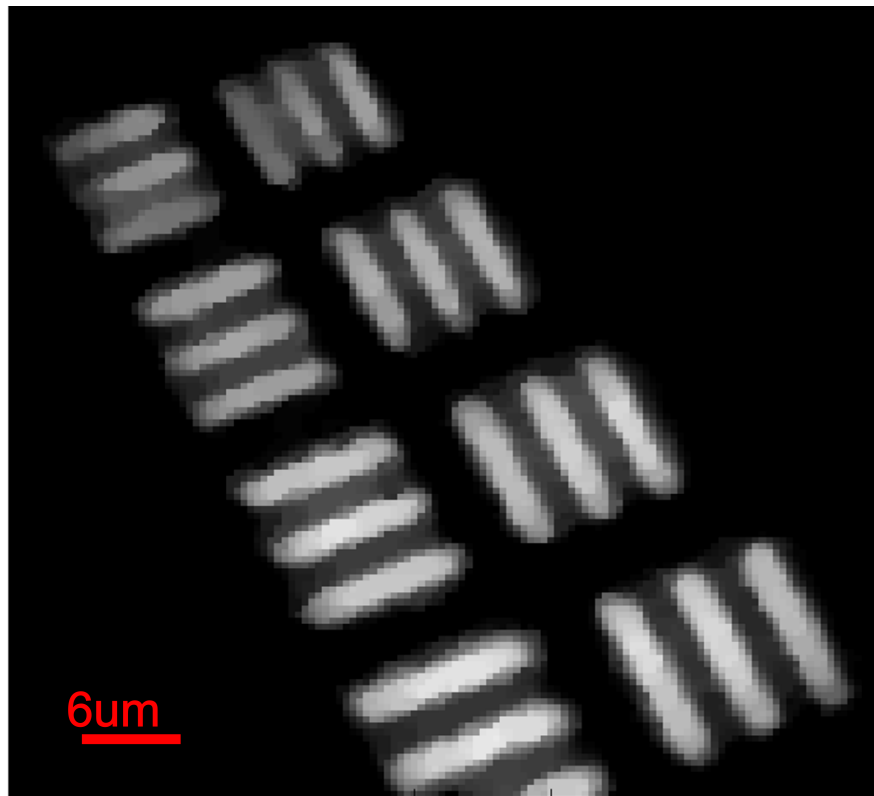
Figure 2.3 : Reconstruction of $n = 128 \times 128$ image of AF Test Target 1951 with $\delta = .95$ and default TVAL3 parameters β, μ.
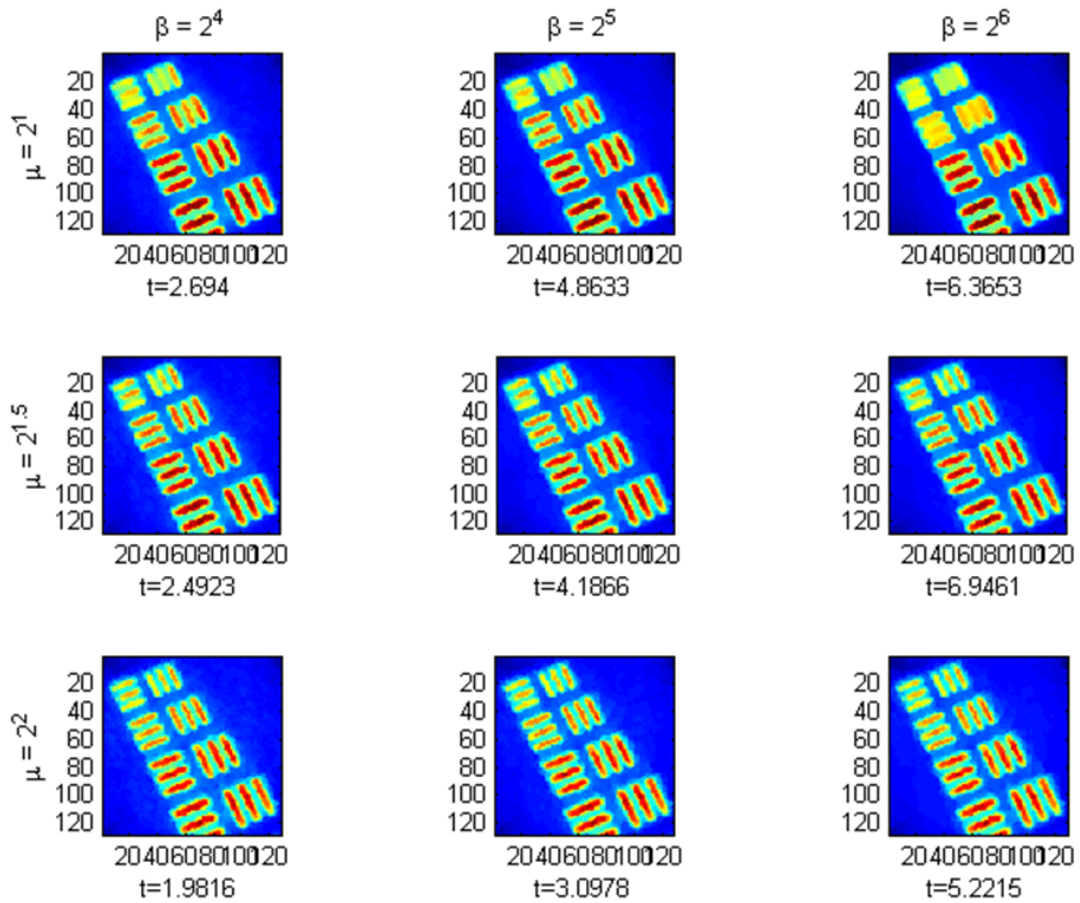
Figure 2.4 : Method for determining best parameters for TVAL3 reconstruction. We see that various choices for β, μ result not only in different quality reconstructions, but also different reconstruction times (in seconds) for each, as indicated below each reconstruction.

which Raman scattering occurs. Photons of varying wavelength correspond to photons of different energy. A beam of light consisting of 100 ultraviolet photons is more energetic than a beam of light with 1,000 infrared photons. However, in optics we would say that the infrared beam is more *intense*, since there are more photons. The energy E of a single photon is directly proportional to the frequency of the photon, ω,

$$E = \hbar\omega \tag{2.2}$$

In the Raman effect, we have that a beam of incident light of a certain energy, i.e. 'color', enters the material and light of a different energy enters. Therefore, energy was lost or gained. With the Raman effect it is more likely to observe a lower energy, or red-shifted, photon, so let's assume energy was lost. Where did this energy go?

The answer is that the light stimulated vibrations in the material. To understand this better, let's consider the diatomic molecule, say O_2 for example, as a pair of weights on a spring. The nuclei are the weights, and the chemical bond (sharing of two electrons between the nuclei) is the spring. The spring has a stiffness, K, corresponding to the strength of the chemical bond. Assume that light, written as an electric field

$$\mathbf{E}(t) = E_0 \cos(\omega t + \delta_k), \tag{2.3}$$

is incident on the molecule. In general the polarization vector of a diatomic molecule may be written as

$$\mathbf{p} = \alpha\mathbf{E} \tag{2.4}$$

where α is the polarizability, or susceptibility to a change in polarization. In general it is an arbitrary tensor, however in this case we assume it acts in only one dimension and we take

it to only linear order,

$$\alpha(x) = \alpha(0) + \left(\frac{d\alpha}{dx}\right)\bigg|_{x=0} x \tag{2.5}$$

We assume that x is the solution of a simple harmonic oscillator, so that

$$x(t) = A\cos(\omega_1 t) \tag{2.6}$$

where

$$\omega_1 = \sqrt{\frac{K}{\mu}} \tag{2.7}$$

and μ is the reduced mass of the two nuclei. Combining Equations (2.3), (2.4), and (2.5), we get

$$\mathbf{p}(t) = \alpha_0 E_0 \cos(\omega t + \delta_k) + A\alpha'(0)E_0 \cos(\omega t + \delta_k)\cos(\omega_1 t) \tag{2.8}$$

Using a trigonometric identity we write

$$\mathbf{p}(t) = \alpha_0 E_0 \cos(\omega t + \delta_k) + \frac{1}{2}A\alpha'(0)E_0\{\cos\left((\omega - \omega_1)t + \delta_k\right) + \cos\left((\omega + \omega_1)t + \delta_k\right)\} \tag{2.9}$$

Thus the polarization changes as a function of time, meaning that we have the acceleration of charge with at two natural frequencies, ω, corresponding to the Rayleigh scattering (common reflection) and the Raman scattered light at frequency $\omega_R = \omega \pm \omega_1$ [25]. The minus sign says that light is scattered at a frequency less than the incident frequency, so with energy

$$E = \hbar(\omega - \omega_1).$$

Since in terms of color, this is a shift towards the 'red' end of the electromagnetic spectrum, we call this photon red-shifted. The plus sign corresponds to a blue-shift, or equivalently a

gain in energy for the incident photons,

$$E = \hbar(\omega + \omega_1).$$

The photons red-shifted via the Raman effect are called Stokes-shifted, and the blue-shifted photons are called anti-Stokes, after the famous English nineteenth century physicist Sir George Stokes. There is a roughly 10^{-6} chance that the incident photon will be Stokes-scatterd, and about a 10^{-8} chance that an incident photon will be anti-Stokes scattered. Thus, the fact that Raman originally observed this previously anomolous effect with modest optics and sunlight is quite amazing. Today, Raman spectroscopy and microscopy is performed with nearly single-frequency laser light, an array of precision optics, and cooled CCD spectrometers. Although it is such a weak effect, it is a very powerful method for probing the chemical structure of a sample, as we will see in the following section.

2.2.2 Determining Chemical Structure from the Raman Spectrum of a Material

The astute and informed reader will note that in order for the above derivation to be valid, the first derivative of the polarizability, $\alpha'(0) = \frac{d\alpha}{dx}|_{x=0} \neq 0$. This defines a so-called *selection rule* for Raman scattering—the first derivative of the polarizability with respect to space must be nonzero. In the case of vibration of the O_2 molecule, $\alpha'(0) = 0$ and so there actually is no Raman scattering from vibrations of O_2. However, one of the first important applications of Raman scattering was to show that there are other carbon dioxide, CO_2, had hitherto unknown vibrational *modes*, or in terms of Equation 2.9, unknown ω_1 [26, 27]. Our above derivation accounted only for vibrations in one dimension, but as also mentioned above, polarizability, α is a tensor in general, so there could be nine total nonzero partial first derivatives with respect to space, and rotational modes are also allowed, so that if an incident photon causes a molecule to rotate, that might also be reflected in the Raman

spectrum through a blue- or red-shifted peak in the Raman spectrum.

The example of CO_2 is also important for us because before the Raman spectrum was acquired for that molecule, its infrared (IR) spectrum, which also tells us about vibrational modes, but for a different set of selection rules, was known. The vibrational modes discovered by Raman spectroscopy differed from those discovered by infrared spectroscopy, but the vibrations were on the same order of energy. This illustrates how Raman spectroscopy allows one to use light in the *visible* part of the electromagnetic spectrum, with wavelength $\lambda \approx 400 - 700$nm, to probe energies in that would correspond to photons in the infrared, corresponding to $.8\,\mu$m $\leq \lambda \leq 100\,\mu$m and larger. To detect such photons requires more exotic light sources as well as more exotic detectors compared to visible light. Thus, Raman spectroscopy provides a simpler, complementary, method for investigating the vibrations of molecules.

Raman spectroscopy is not limited to probing vibrations of molecules. It is also possible to determine the vibrational modes of solids, known as *phonons*, as in quanta of sound, just like the photon is a quantum of light. We model a solid as a lattice of masses connected by springs instead of just two or three masses connected by springs in the case of molecules in either the gaseous or liquid state. When light is incident on the lattice, it either reflects as in Rayleigh scattering, or it creates a phonon, which again is a vibration that travels through the solid, and a lower-energy phton. It is sensible to call this a quanta of sound, because it is precisely vibrations of nuclei, transmitted through electron-electron interactions, that is responsible for the majority of thermal conductivity and transmission of sound in materials.

Among the many important applications of Raman scattering in solids are stress and strain analysis for silicon technologies and for investigating the properties of graphene and carbon nanotubes, including how many layers of graphene are present in a graphene sample and also how many layers comprise a nanotube or what diameters of nanotubes are present

in a sample. We choose these two examples because, as a proof of concept of the applicability of compressive sensing techniques to Raman microscopy, we show that compressive sensing is indeed effective for laser-illumination based microscopy, and that through simulations on actual Raman microscopy data, compressive sensing may be used to discriminate graphite from silicon, which suggests it might be effective for more complex samples.

2.2.3 Raster Scanning Raman Microscopy

Raman microscopy is a powerful experimental technique to determine the spatial distribution of substances in a sample. We will show a simple example of this with experimental data acquired on the commercial Renishaw Raman microscopy system, followed by two proposed architectures for laser-illuminated microscopy with the DMD. The example is graphite on silicon dioxide, which is a toy model for a more interesting system, graphene on silicon dioxide. Graphene is a single layer of carbon atoms in a 2D lattice, however the term graphene is also used to describe more than one layer stacked on top of one another. It seems that it is graphene until about ten layers, then graphene just becomes graphite. Graphite and graphene share similar features in their Raman spectrum since both are carbon allotropes, and in fact Raman spectroscopy is one method for determining whether a carbon sample is single-layer, double-layer, or more-layered graphene or just simply graphite. Graphene is a widely-studied material because of its novel conduction properties, strength, and overall novelty, and was the subject of a recent Nobel prize in Physics. Silicon also has two prominent peaks in its Raman spectrum, and both the Raman spectrum of silicon and of graphite are shown in Figure 2.5. We focus only on the so-called 'G' peak of graphite with wavenumber $k \approx 1590\,\mathrm{cm}^{-1}$, and the well-known silicon peak at $k = 520\,\mathrm{cm}^{-1}$.

We image the boxed region with graphite flake shown in Figure 2.6 by sensing how intense the Raman 'G' peak is. If the peak is not there, then we know that it is the silicon substrate
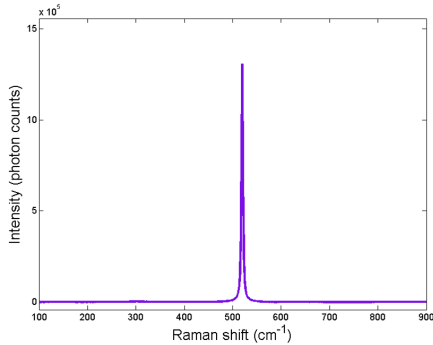
Silicon Raman spectrum, $k = 520\,\mathrm{cm}^{-1}$ peak
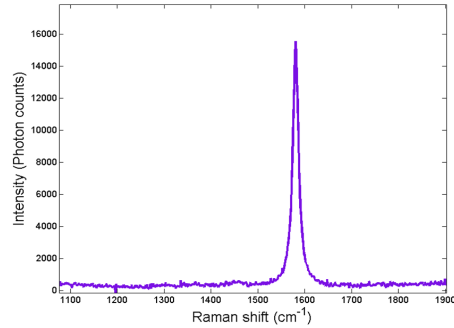
Graphite Raman spectrum, 'G' peak

Figure 2.5 : The Raman spectrum of silicon dioxide substrate and graphite. A typical measurement (compressive or not) in hyperspectral microscopy could be like either one of these, or have two peaks together, or possibly contain many any number of peaks corresponding to different materials present in the sample.

at that point, and if we observe the 'G' peak then we know it is graphite. By raster scanning over the sample and acquiring a spectrum as in Figure 2.5, as explained in Section 2.1.1, we can build a *Raman image* of the region of interest. The Raman image acquired by raster scanning is shown in Figure 2.7. The brighter the pixel, the larger the maximum value of the 'G' peak. To test the efficacy of compressive imaging for Raman imaging, we simulate compressive acquisition where $\Phi \neq I$ and $\Phi \in \mathbb{R}^{m \times n}$ where $m < n$. Instead of having Φ be randomly permuted Walsh-Hadamard vectors, Φ are partial circulant matrices as explained in Chapter 3. We hold off further discussion of the details of partial circulants for now. To reconstruct an image from measurements taken with partial circulant Φ, we use the Rec_PC algorithm of Yin, et al., [28], also to be futher discussed in Chapter 3. We only need to say here that we used default settings for user-defined parameters and that the reconstruction algorithm recovers the solution x^* according to the same problem (TV) in Equation 1.8. We look at the reconstruction of x^* the compressive measurements for a few values of δ in Figure 2.8. Data on convergence across a range of δ values is given in Figure 2.9. We see
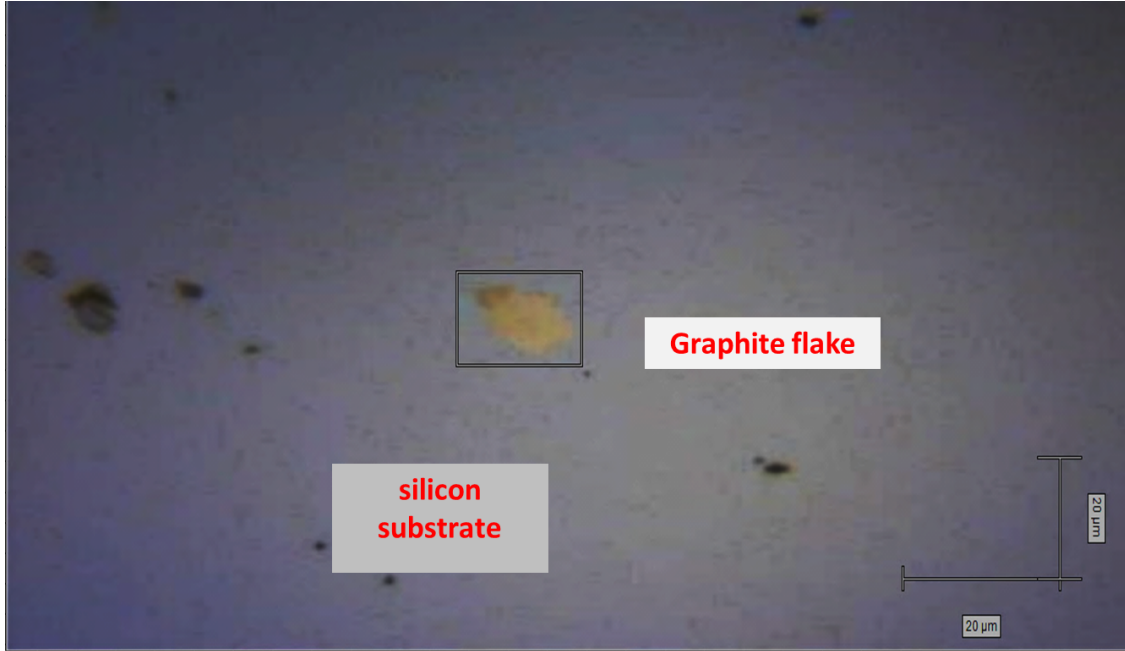
Figure 2.6 : Visible light view of the graphite flake on silicon substrate to be imaged via the Raman effect.

that computational resources increase non-linearly with δ, as expected.

2.2.4 Laser-illuminated Compressive Sensing Microscope System

To finish this chapter, we present preliminary data from our home built copmressive sensing microscope system, the optics on the table in Figure 2.10. With this setup we obtained the images shown in Figure 2.11.

With the proper equipment, it is clearly possible to acquire Raman microscopic images in a fraction of the time it would take with traditional methods. Because of some not yet understood spectral phenomena arising from the diffractive properties of the DMD, it may be better to construct the microscope as we have, shown in Figure 2.12, where instead of implementing the measurement vectors by patterning light coming from the scene, the DMD structures laser light sent to illuminate the scene, which is totally mathematically equivalent.
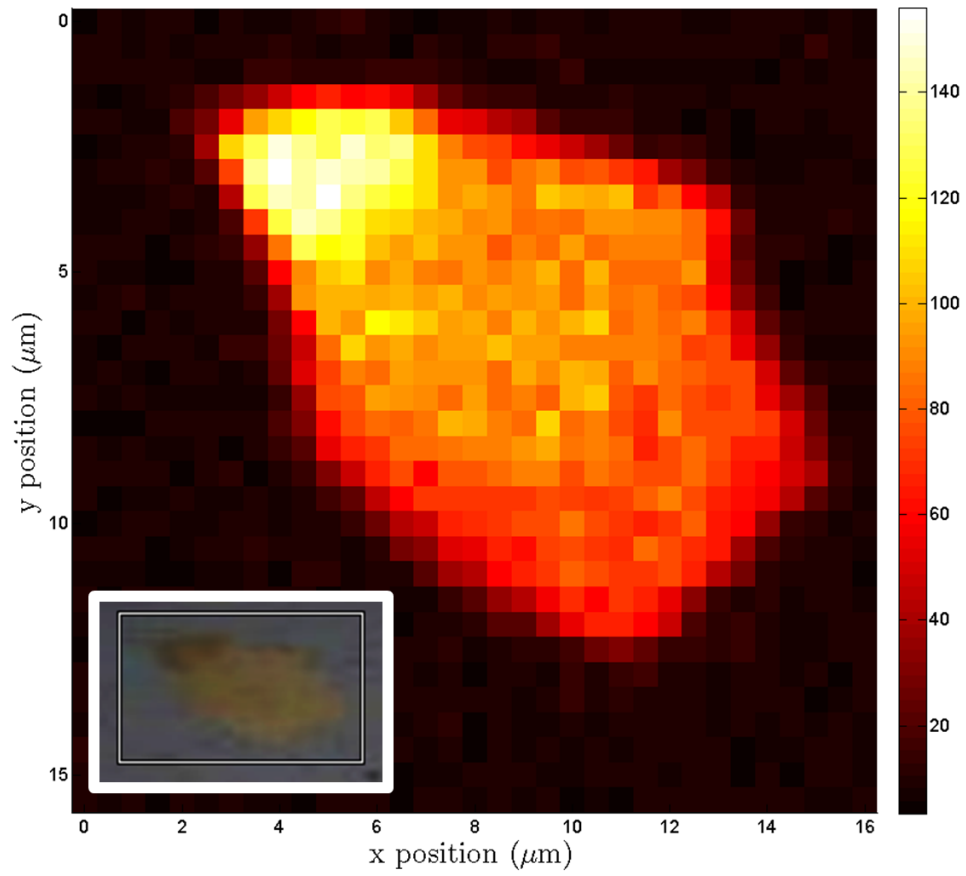
Figure 2.7 : Visible light view of the graphite flake on silicon substrate to be imaged via the Raman effect. The resolution is $n = 33 \times 33$
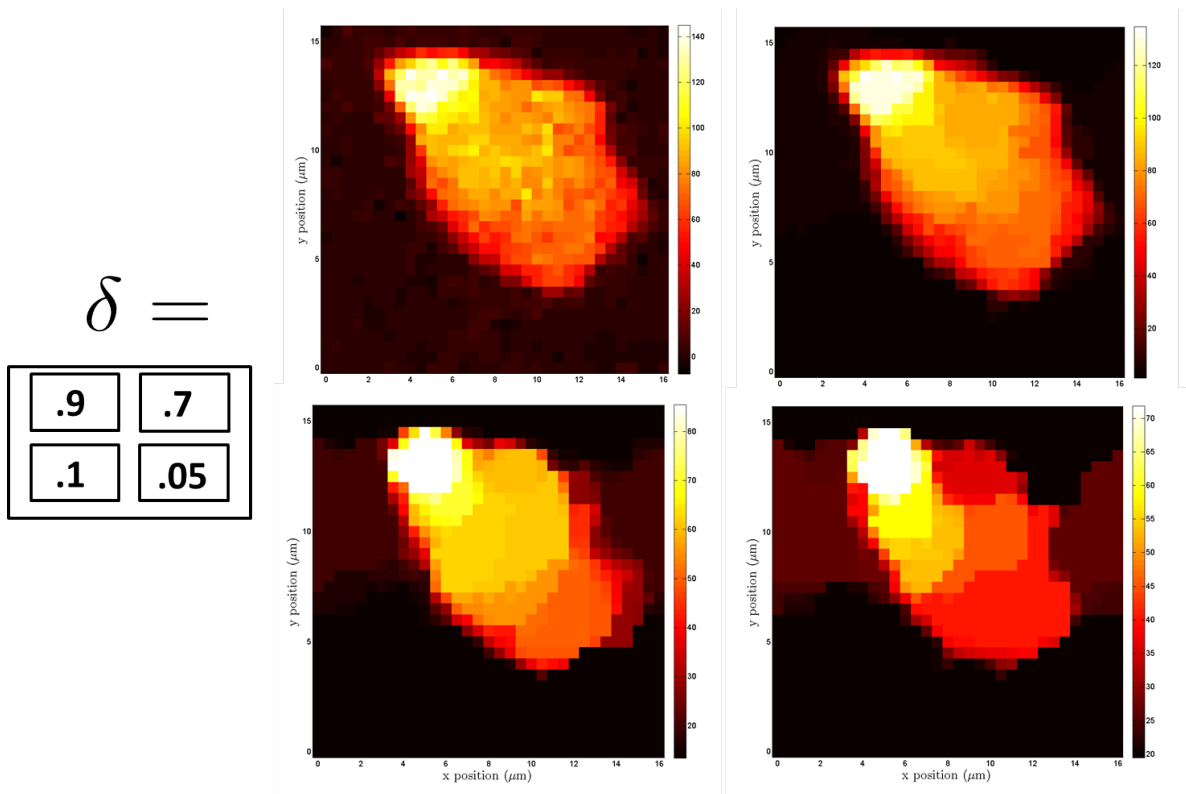
Figure 2.8 : Raman image recovered from simulated compressive measurements for some values of $\delta = m/n$.
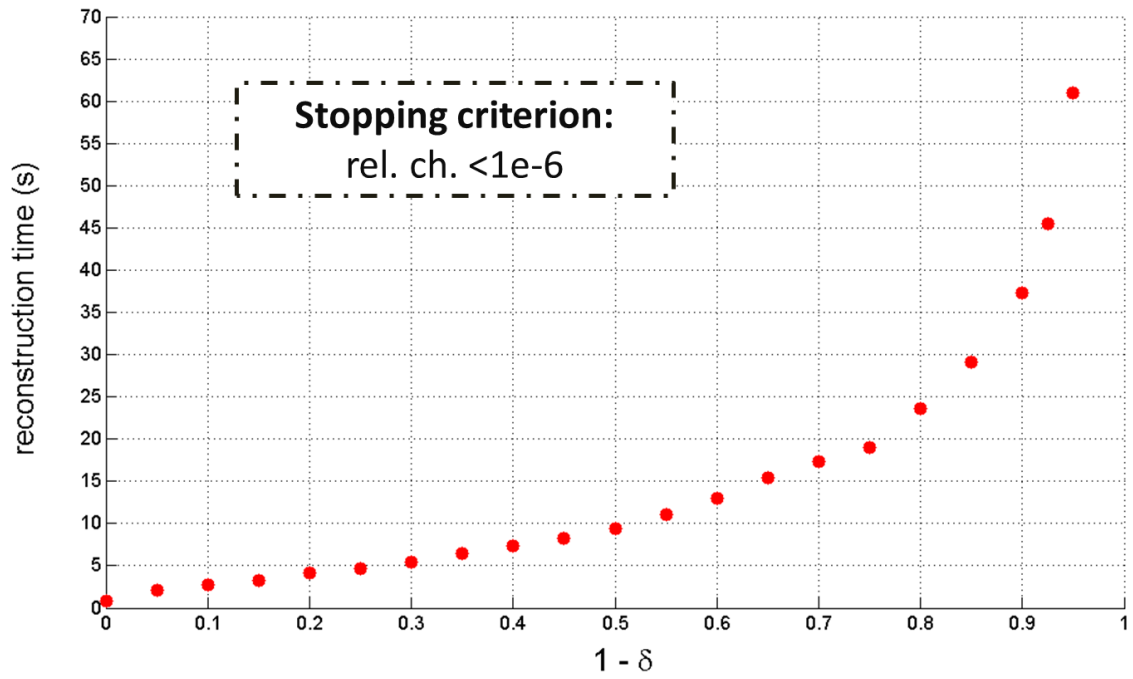
Figure 2.9 : Time for Rec_PC to recover solution x^* as a function of $1 - \delta$ where δ is the subsampling ratio. The dependence is non-linear.
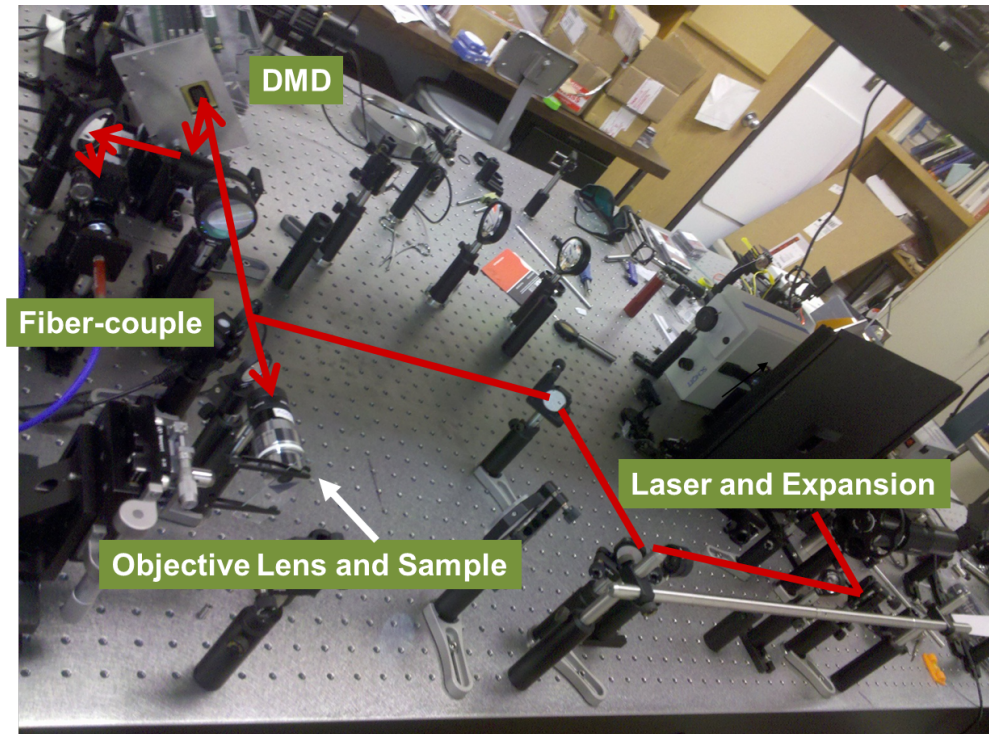
Figure 2.10 : Laser-illuminated compressive sensing experimental setup

AF Target Bars, $n = 128 \times 128$ AF Target Bars, $n = 256 \times 256$
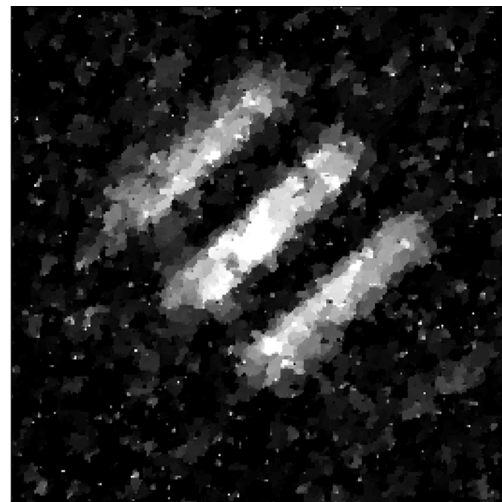
Figure 2.11 : Images of the smallest target on the AF Test Target 1951-A. The bars are $2.2 \, \mu$m wide. Images taken with 100x/.9NA Zeiss EC Epiplan/Neofluoar lens. Small field of view, high magnification.
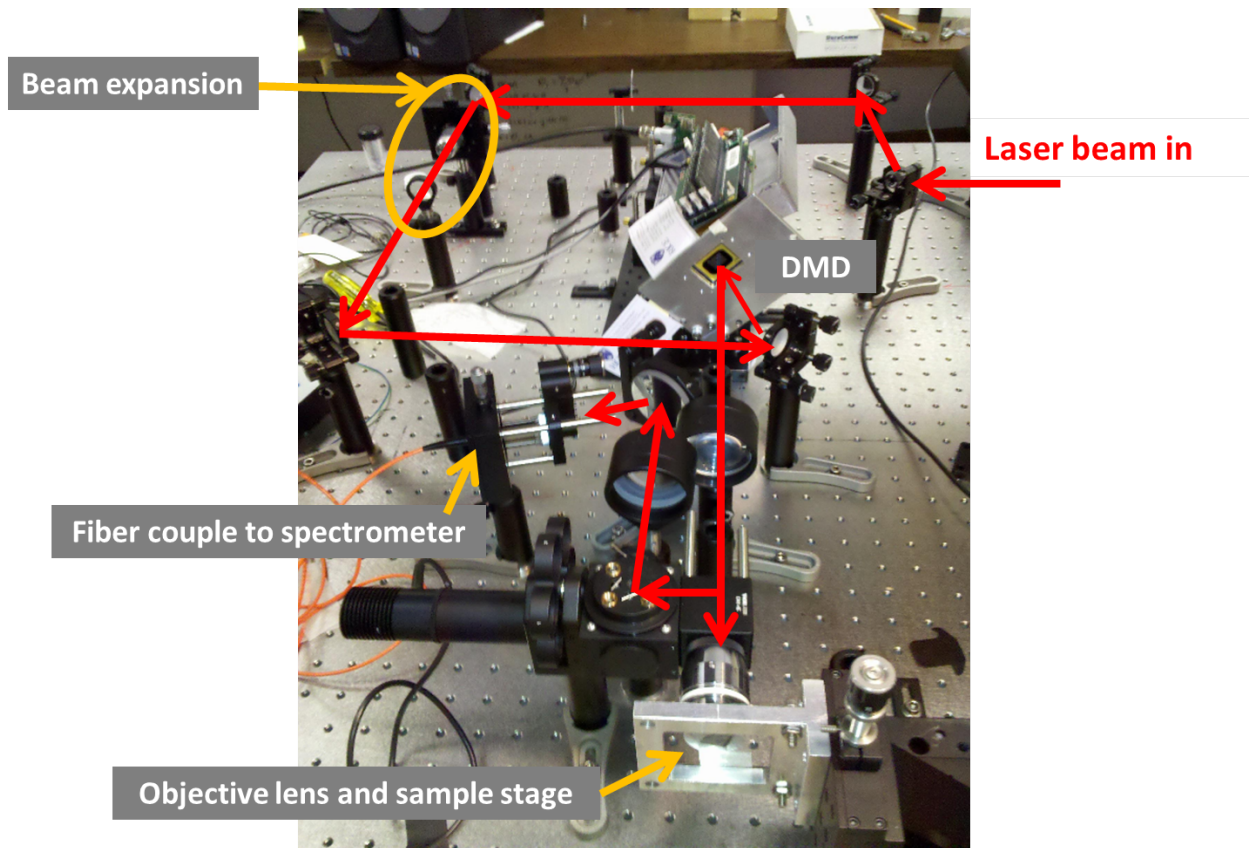
Figure 2.12 : An alternative laser-illuminated compressive sensing experimental setup

Studer, et al. [29], employ essentially the same setup for fluorescence microscopy for biology.

Chapter 3

Circulant Matrices for Compressive Imaging

In this chapter we explore the use of circulant matrices for imaging. By employing measurement matrices Φ whose m rows are taken from an $n \times n$ circulant or block-circulant matrix we will denote Φ°. We will define a circulant matrix mathematically, explain their utility, and show how circulant matrices can result in a more versatile, efficient imaging system. Much of this material will also serve as an introduction for Chapter 4. The work presented in Chapter 4 was motivated by the results of our imaging experiments presented in this chapter.

3.1 Theory of circulant matrices for imaging

3.1.1 Properties of circulant matrices

Circulant matrices "underpin elementary harmonic analysis" (Aldrovandi, 2001 [30]) because of their special relationship to the Fourier transform. This relationship enables us to more carefully design our measurement matrices, but maintain a fast matrix-vector multiply in the form of the Fourier transform. To see why, let us explicitly write a circulant matrix.

A circulant matrix is a matrix $C \in \mathbb{R}^{n \times n}$ with entries $t_i \in \mathbb{R}, \quad i = 0, 1, \cdots, n-1$ such

that

$$C = \begin{bmatrix} t_0 & t_{n-1} & t_{n-2} & \cdots & t_1 \\ t_1 & t_0 & t_{n-1} & \cdots & t_2 \\ t_2 & t_1 & t_0 & \cdots & t_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{n-1} & t_{n-2} & t_{n-3} & \cdots & t_0 \end{bmatrix}. \tag{3.1}$$

Such a matrix is also sometimes referred to as a *convolution matrix*. To see why, consider

$$c = (t_0, t_{n-1}, \ldots, t_1)$$

so that $c \in \mathbb{R}^n$. Define the (circular) convolution operator for vectors $a, b \in \mathbb{R}^n$ to be

$$(a * b)_k = \sum_{i=0}^{n-1} a_i b_{k-i}, \quad k = 0, 1, \ldots, n-1 \tag{3.2}$$

then we see that for $x \in \mathbb{R}^n$

$$Cx = c * x \tag{3.3}$$

This fact allows a fast matrix-vector multiply on a binary computer via the fast Fourier transform (FFT). Let $F \in \mathbb{R}^{n \times n}$ be the Fourier transform matrix

$$F_{jl} = \frac{1}{\sqrt{n}} e^{2\pi i (j-1)(l-1)/n} \quad (j, l = 1, 2, \ldots, n).$$

Now apply the identity $I = F^{-1}F$ to the right hand side of Equation 3.3 and use the convolution rule for the Fourier transform to obtain

$$Cx = c * x$$

$$= F^{-1}F(c * x)$$

$$= F^{-1}(Fc)(Fx)$$

$$= F^{-1}DFx$$

where $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \text{diag}(\lambda)$ is not only the Fourier transform of the vector c, but also the eigenvectors of the matrix C, which reveals one more remarkable property of circulant matrices, namely they are diagonalized by the Fourier transform, or, equivalently, the eigenvectors of a circulant matrix are the columns of the Fourier matrix, F. For more details see [30–32]. For an interesting application to machine multiplication for two numbers with arbitrary digits and precision, see Knuth, 1981 [33].

So, the utility of circulant matrices for compressive sensing should now be clear. All we need to do is define a *seed vector* c, compute its Fourier transform which gives us λ, and perform two FFTs modulated by the entries of λ. Thus, the computation of the matrix-vector multiply Cx will not take $O(n^2)$ operations, but instead $O(n \log(n))$ operations. Even just for $n = 128 \times 128 = 16384$, the resolution of the images we present below, we see a substantial decrease in the number of operations since $n^2 \approx 2.7 \times 10^8$ and $n \log(n) \approx 2.3 \times 10^5$, three orders of magnitude difference. This savings is essential for an efficient recovery algorithm.

3.2 Imaging with Partial Circulant Measurement Matrices

In this section we follow the structure of Chapter 1 and introduce the imaging system before diving too deeply into the mathematics. The advantage we gain from circulants for imaging is that we may pattern four copies of the first row, or *seed vector*, $\varphi^{(1)}$, of the measurement

Figure 3.1 : $\varphi^{(1)} \in \mathbb{R}^{1024}$ reshaped to 2D. White squares represent $\varphi_k^{(1)} = 1$, black squares represent $\varphi_k^{(1)} = 0, \quad k = 1, \ldots, n.$

matrix Φ onto an optical plate. An $n = 32 \times 32$ example of $\varphi^{(1)}$ is shown in Figure 3.1. By shifting the plate, we generate rows of a *block circulant* matrix, to be explained in more detail below.

3.2.1 Imaging Setup

The imaging setup is identical to those introduced in the previous chapters, except now instead of the DMD directing light towards or away from a photodetector via reflection, we have a optical plate patterned with a mask to either allow the light to pass through or block light from a pixel. This corresponds to $\varphi_{ij} \in \{1, 0\}$, again with a 1 being represented by white in Figures 3.1 and 3.3, and furthermore, as before, $\varphi_{ij} = 1$ is an element that allows light to pass to the detector, and $\varphi_{ij} = 0$ is an element that blocks light.

On a single optical plate we pattern four copies of $\varphi^{(1)}$ to make a $2N \times 2N$ pixel grid. By overlaying an $N \times N$ selection mask, shown in dashed green in Figure 3.2, we can generate all n measurement vectors by moving the selection mask. Each selected $N \times N$ square corresponds to a row of Φ°, as illustrated in Figure 3.2. In practice, we would not actually move the selection mask since this would also entail moving the photodetector. Instead we

move the optical plate itself, as illustrated in Figure 3.3.

$$\varphi^{(1)} \qquad\qquad \varphi^{(14)} \qquad\qquad \varphi^{(32)}$$
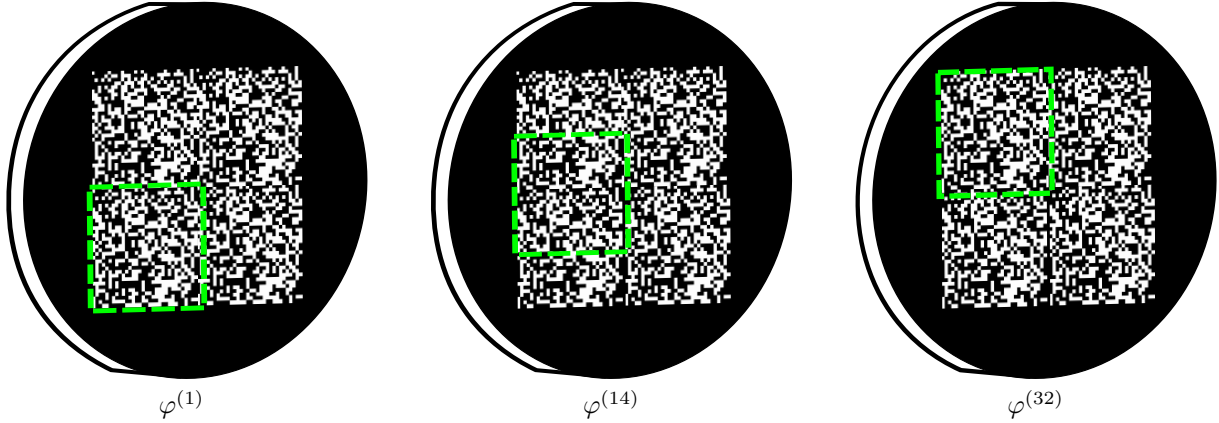
Figure 3.2 : Four copies of the seed vector $\varphi^{(1)}$ patterned onto an optical plate. By shifting a selection mask (represented by the red box) to select one measurement vector at a time, we generate, or 'select,' measurement basis vectors from Φ°, reshaped to 32×32.

We define

$$\Omega \subseteq \{0, 1, \ldots, N-1\} \times \{0, 1, \ldots, N-1\} = \Omega^{\circ} \qquad (3.4)$$

to represent the number of column shifts and row shifts of the optical plate used to acquire the measurements y. If we acquire m measurements $y_i = \langle \varphi^{(i)}, x \rangle$, then $|\Omega| = m$. There are four different methods of creating Ω that we explore in this work,

1. **Sequential:** Starting with the selection mask in the lower left corner, shift the selection mask one row at a time, $N-1$ times. Then shift by one column and again perform $N-1$ row shifts. Repeat until m unique regions are selected by the selection mask. This is illustrated in Figure 3.2.

2. **Box:** Do an equal number ($\approx \lceil \sqrt{m} \, \rceil$) of row and column shifts.

3. **Random:** Select m $\omega_i \in \Omega^{\circ}$ at random.
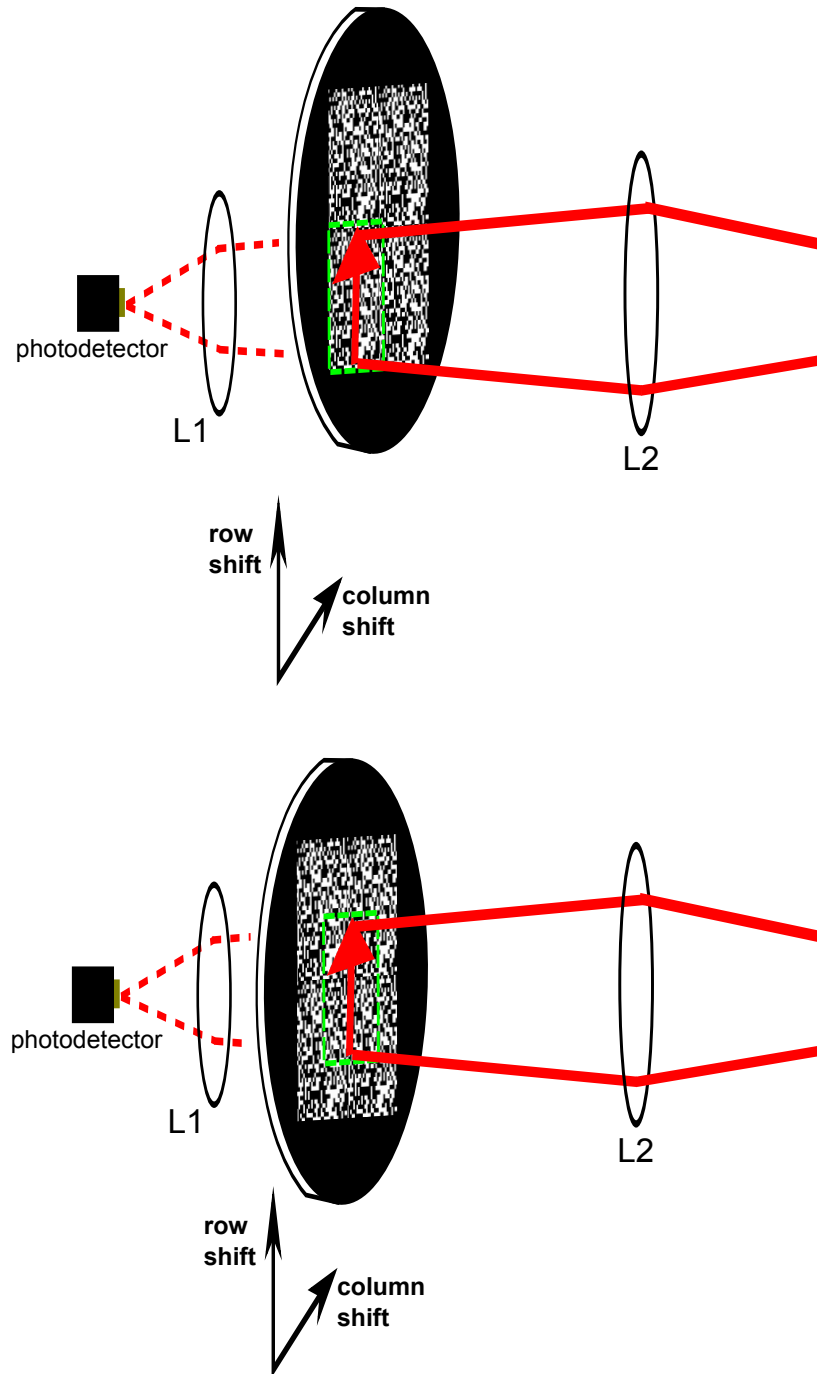
Figure 3.3 : By shifting one row or column of the mask at a time, we can generate all $n = N \times N$ rows of a block circulant matrix Φ°. The optical system is identical to the DMD-based setup, where a lens L2 focuses an image of the scene, represented by the arrow, onto the mask. The light that allows to pass, corresponding to an 'on' pixel, or $\varphi_{ij} = 1$, is collected by the lens L1 and directed towards the photodetector for measurement.

4. **Random Walk:** Restrict the plate to shift one row or column at a time, but take the step at random. The result is a random walk. At this time we do not optimize for self-crossings and require that m unique points are generated isntead of m steps taken.

3.2.2 Formally Describing How to Build Φ from Φ°

Recall the restriction operator R_Ω from Section 1.3 that selects the rows indexed by the set Ω, for example if $\Omega = \{1, 2\}$ then

$$R_\Omega A = R_\Omega \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{14} \\ a_{21} & a_{22} & \cdots & a_{24} \\ \vdots & \vdots & \ddots & \vdots \\ a_{41} & a_{42} & \cdots & a_{44} \end{bmatrix} = R_\Omega \begin{bmatrix} a^{(1)} \\ a^{(2)} \\ a^{(3)} \\ a^{(4)} \end{bmatrix} = \begin{bmatrix} a^{(1)} \\ a^{(2)} \end{bmatrix}. \tag{3.5}$$

Thus, we can compactly write building our measurement matrix $\Phi \in \mathbb{R}^{m \times n}$ from a circulant matrix $\Phi^\circ \in \mathbb{R}^{n \times n}$ as

$$\Phi = R_\Omega \Phi^\circ. \tag{3.6}$$

In our application where Ω corresponds to coordinates of row and column shifts of the optical plate, we cannot directly apply R_Ω as above. Here instead the coordinates in Ω define which row of the *block circulant* matrix Φ° will be taken. To see how this works, let us write the matrix

$$M = \begin{bmatrix} a & b & c & a & b & c \\ e & f & g & e & f & g \\ h & i & j & h & i & j \\ a & b & c & a & b & c \\ e & f & g & e & f & g \\ h & i & j & h & i & j \end{bmatrix}, \tag{3.7}$$

a 3×3 analogue of the pattern on the optical plate. In this toy example, then,

$$\varphi^{(1)} = \begin{bmatrix} a & b & c \\ e & f & g \\ h & i & j \end{bmatrix},$$

which we have indicated by coloring it red in Equation 3.8. Of course $\omega_1 = (0,0)$, no shifts. If we implement the sequential method beginning with a column shift, the selection mask would next select the red elements,

$$M = \begin{bmatrix} a & b & c & a & b & c \\ e & f & g & e & f & g \\ h & i & j & h & i & j \\ a & b & c & a & b & c \\ e & f & g & e & f & g \\ h & i & j & h & i & j \end{bmatrix}, \tag{3.8}$$

meaning

$$\varphi^{(2)} = \begin{bmatrix} b & c & a \\ f & g & e \\ i & j & h \end{bmatrix}$$

and $\omega_2 = (0,1)$. Continuing on we have

$$\varphi^{(3)} = \begin{bmatrix} c & a & b \\ g & e & f \\ j & h & i \end{bmatrix}.$$

If we want to reshape these as rows of Φ to implement $y = \Phi x$ we would have

$$\Phi = \begin{bmatrix} a & b & c & e & f & g & h & i & j \\ b & c & a & f & g & e & i & j & h \\ c & a & b & g & e & f & j & h & i \end{bmatrix}.$$

Define

$$\alpha = \begin{bmatrix} a & b & c \\ b & c & a \\ c & a & b \end{bmatrix}, \beta = \begin{bmatrix} e & f & g \\ f & g & e \\ g & e & f \end{bmatrix}, \text{ and } \gamma = \begin{bmatrix} h & i & j \\ i & j & h \\ j & h & i \end{bmatrix}$$

Then

$$\Phi = \begin{bmatrix} \alpha & \beta & \gamma \end{bmatrix}$$

Continuing like this we can write the matrix of all reshaped measurement vectors generated by such shifts as

$$\Phi^\circ = \begin{bmatrix} \alpha & \beta & \gamma \\ \gamma & \alpha & \beta \\ \gamma & \beta & \alpha \end{bmatrix}. \tag{3.9}$$

Just as C was called a convolution matrix in 1D, this matrix Φ° is a 2D convolution matrix. Φ° is not circulant as with 1D, but *block circulant*. We still have

$$\Phi^\circ x = F^{-1}DFx$$

as discussed above, however now F and F^{-1} are the 2D Fourier and inverse Fourier transform [34].
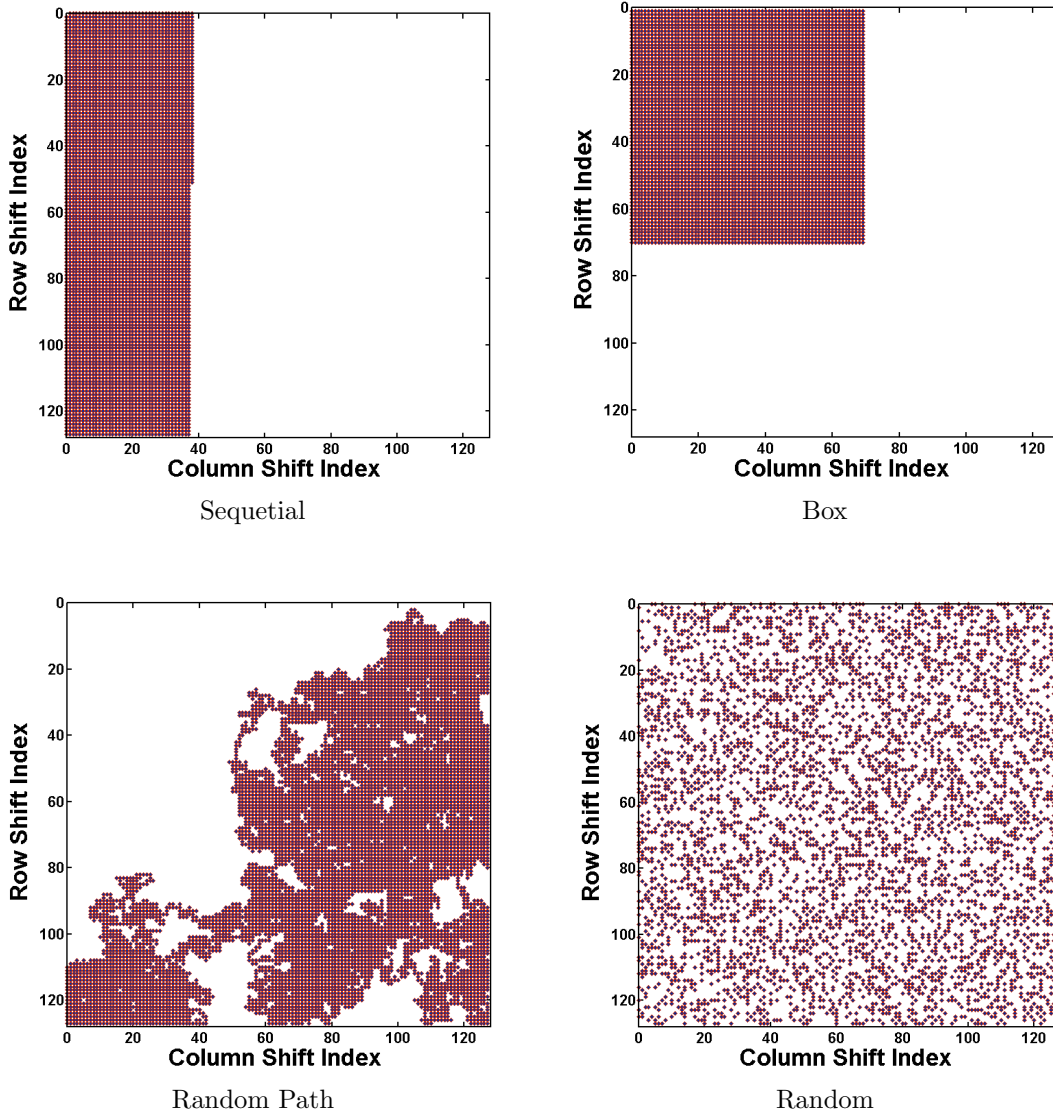
Figure 3.4 : Filled points in these plots indicate the location of the selection mask for individual measurements in terms of row and column shifts. Thus, there are more row shifts than column shifts for the sequential method and an equal number of row and column shifts for the box method. The random path shows some structure since the mask is only allowed to step one row or column shift to generate the next measurement basis vector in the sequence, and random is just that.

3.3 Imaging Results

In this section we show results from imaging with each of those four methods. Our imaging system is the same as in Section 1.3, except the illumination source is different. We use a broadband lamp here. The Hamamatsu detector is the same. We use the digital micromirror device (DMD) to simulate the mask motions as a proof-of-concept for this imaging scheme.

In order to recover an image from the measurements $y = \Phi x$, we use the Rec_PC (PC stands for 'partial circulant') algorithm of Yin, et al., [28]. Essentially it is the same as TVAL3, but with some modification, most notably to accommodate the circulant measurement matrices. One other difference is it offers an explicit handling of both total variation minimization and ℓ_1 minimization, along with the usual *fidelity constraint*. Rec_PC finds x^* such that

$$(\text{Rec\_PC}) \quad x^* = \min_x \alpha \sum_{i=1}^{n} \|D_i x\| + \frac{\mu}{2} \|\Phi x - y\|^2. \tag{3.10}$$

For the images below, we set $\alpha = 10^{-2}$ and $\mu = 1$. Again, the operator $D_i \in \mathbb{R}^{2 \times n}$ is a discrete gradient operator.

As shown in Figure 3.5, random Ω outperforms the sequential one. Furthermore, on comparison of all four methods introduced here, the relative error, defined as

$$\text{Relative Error} = \frac{\|x^*_{\delta=1} - x^*_{\delta}\|_2^2}{\|x^*_{\delta=1}\|_2^2}, \tag{3.11}$$

where $x^*_{\delta=1}$ is the solution when $m = n$, or $\delta = 1$, declines equally quickly for both random and random walk, converging nonlinearly, while sequential and box converge approximately linearly with increasing δ. Furthermore, it appears that the computational problem (Rec_PC) in Equation 3.10 is more difficult for box or sequential patterns than the randomized patterns since the algorithm takes longer to converge in these cases, as shown in Figure 3.7.
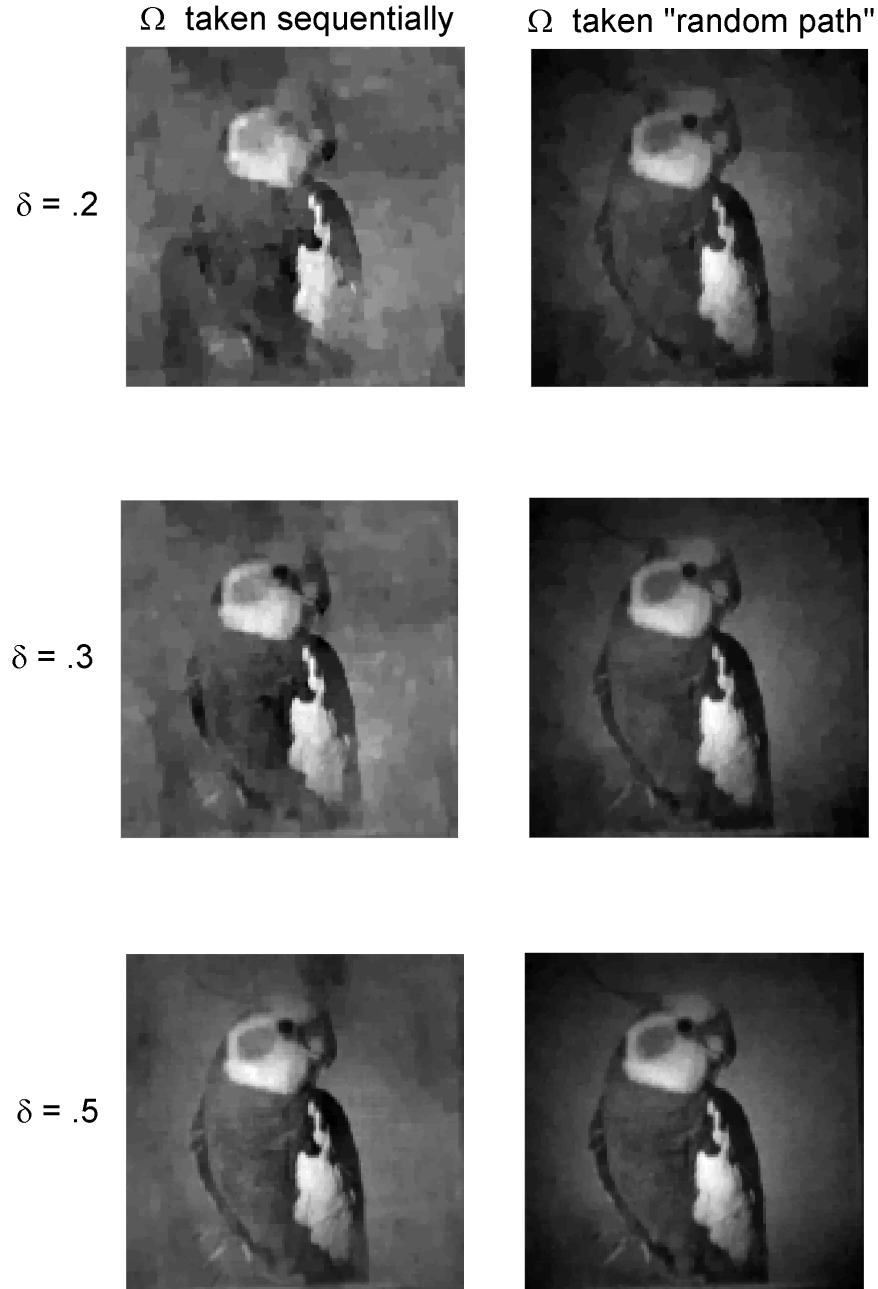
Figure 3.5 : Difference between taking measurement vectors from Φ° sequentially (left column) and according to a random path (right column) for a few subsampling ratios. Note the reconstruction with random path measurement vectors is relatively high quality even at a very low subsampling ratio δ. Data acquired by Lina Xu.
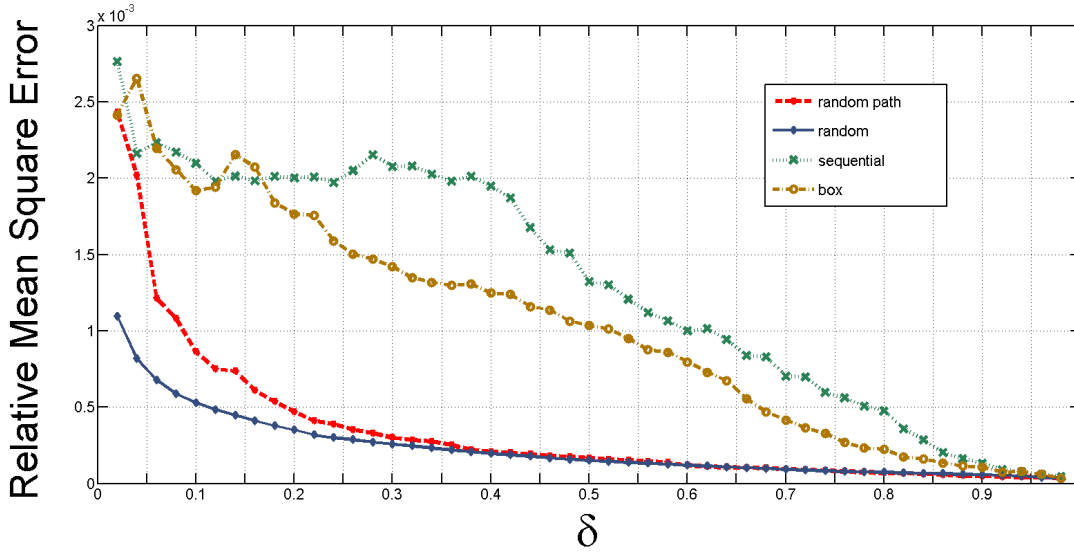
Figure 3.6 : Relative mean square error (normalized squared difference between reconstructed image for a given δ and the one reconstructed with $\delta = 1$) for the four methods of generating the measurement basis Φ
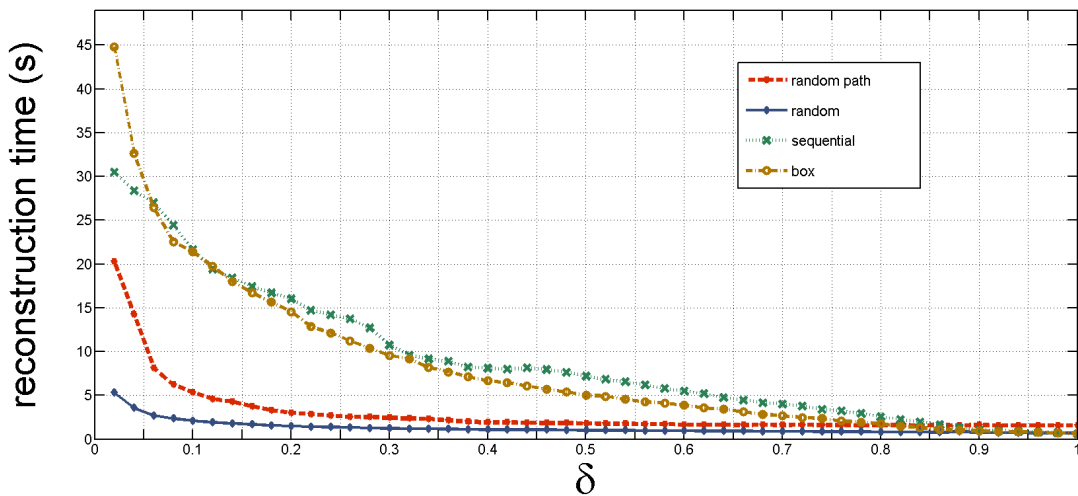
Figure 3.7 : Time to solve the underlying optimization problem and recover an image for various undersampling ratios, δ, for the four methods of generating the measurement basis Φ

3.4 Conclusion and Discussion

In this chapter we have demonstrated the feasibility and utility of circulant matrices for imaging. Apparently, when the measurement vectors are chosen in a sequential fashion, the recovery problem is more difficult than if the measurement vectors are chosen with an element of randomness. We may be tempted to generalize this phenomenon and say that in all cases, a sequentially-built Φ results in worse recovery than randomly-generated Φ. However results in Chapter 4 suggest this is not the case.

The dependence on shift type we see here may be due to the fact that our figure of merit is the magnitude of the derivative, the total variation. When we minimize a cost function that includes total variation, we are asking the recovery algorithm to maximize piecewise constancy. Thus, we acquire more information with a new measurement only if the measurement vector is not probing the same piecewise constant areas. By shifting sequentially, we probe the same piecewise constant area with the same set of measurement vector pixels $\varphi_{jk}^{(i)} = 1$, and thus acquiring less new information per measurement for most values of δ. The results shown in the next chapter support this claim.

Chapter 4

Experimental Investigation of Subsampled Circulant Matrices for Compressive Sensing

As we saw in the previous chapter, there is a marked difference in signal recovery when the measurement matrix Φ is subsampled via a restriction operator R_Ω where the index set Ω is chosen either randomly or sequentially. However in that chapter, we utilized a reconstruction algorithm based on total-variation (TV) minimization coupled with the usual least-squares condition. That sequentially-built Φ gather less information per measurement for the first measurements, and thus require more measurements for commensurate performance with randomly-built Φ is one possible and quite plausible explanation for this behavior. Plausible because TV minimization finds images with the largest possible regions of small derivative, so it favors piecewise constancy. An intuitive argument is that if a set of measurement vectors sample identical regions of near-constant light intensity from a scene, then it makes sense that the measurements acquired would be somewhat redundant.

If this were the end of the story, then for ℓ_1 minimization, which we reiterate serves as a proxy for ℓ_0 minimization, with strict equality constraints, we might expect that reconstruction would not be sensitive to how we build Φ. In this chapter we show results that defy that expectation. Furthermore these results suggest that the powerful *precise undersampling theorems* proposed, developed, and derived for gaussian random measurement matrices by Donoho and Tanner in a series of papers, [37–43] apply to partial circulant measurement matrices as well.

Our results indicate that partial circulant or partial block circulant matrices might well

belong to what Donoho and Tanner call the *universality class* of gaussian matrices for phase transitions with respect to signal recovery, or, as we will see shortly, phase transitions in the number of faces of certain *polytopes* after projection by a measurement matrix φ. These results show, perhaps surprisingly, that indeed, as Yin, et al., posited and demonstrated for a few test cases [28], partial circulant matrices, with entries taken from $\{0, 1\}$ no less, are as effective as Gaussian random measurement matrices. In the following we will be considering if, for a set of measurements $y = \Phi x$, the underlying signal x can be exactly recovered via alternate regularization techniques. We will define undersampling phase space as

$$(\delta, \rho) \in [0, 1]^2,$$

where

$$\delta = \frac{m}{n}, \text{ and } \rho = \frac{k}{m}.$$

Here k is the number of nonzeros in the vector x, or $\|x\|_0 = k$.

To close the introduction to this Chapter, we forward the following two 'research challenges' laid out by Donoho and Tanner [37],

"Characterize universality classes of Gaussian phase transitions. We have shown that many 'random' matrix ensembles yield phase transitions matching those of Gaussian matrices. Characterize the precise universality class of such matrices."

Not only do our experiments have a clear practical value, but the experiments we report here suggest that *most* of the partial circulant measurement matrices with seed vectors $\varphi^{(1)} \in \{0, 1\}$ in equal number follow the same phase transition in probability of success as the Gaussian measurement matrices as reported in the series of work by Donoho and Tanner. However we have found evidence suggesting that not all constructions yield Gaussian-like phase transitions.

When we built partial *block* circulant matrices as with a selection mask moving over a patterned optical plate in a sequential fashion, as explained in Section 3.2.2, and n is a perfect square such that $n = 2^w$, where $w \in \mathbb{Z}^+$, we see degradations in performance for three specific values of δ that arise for two different w we tested, $w = 10$, and $w = 8$. Thus, we may have found a candidate counterexample to gaussian measurement matrices. Investigation of this counterexample could yield insight into the connection between linear programming and polytope geometry. Perhaps we are seeing evidence that would address the second research challenge from [37],

> "Discover new transitions for (LP) and (BP)*. Many but not all matrix ensembles yield phase transitions matching those of Gaussian matrices. Discover more examples which do not, and which are also interesting matrix ensembles, either because the phase transition is better or because the matrix is explicit and deterministic"

Before we show our results, we introduce the convex programming problems we solve, the linear program (LP) and basis pursuit (BP). These problems are just two alternative methods of regularizing the problem of solving x in $y = \Phi x$. We also present some of the basic facts and concepts from polytope geometry. With that background in order, we present the main theoretical result from about seven years worth of work by Donoho and Tanner, and discuss how that result, which has no explicit connection to sparse recovery, can be used to predict the success rate for sparse recovery based on the properties of the choice of measurement matrix.

---

*Donoho and Tanner use the problem code '(P1)' instead of (BP).

4.1 Background

4.1.1 Linear Programming

LP stands for the linear program, which is to find an x^* such that

$$x^* = \min \mathbf{1}^T x \text{ subject to } \begin{cases} y = \Phi x \\ \\ x \geq 0 \end{cases} \tag{4.1}$$

The basis pursuit problem (BP) is, find x^* such that

$$x^* = \min \|x\|_1 \text{ subject to } y = \Phi x. \tag{4.2}$$

Clearly LP is just BP with the additional constraint that all entries of x be nonnegative.

What began as a curiosity into whether circulant matrices indeed yield Gaussian phase transitions for these problems may indeed meet both criteria of Donoho and Tanner's Challenge, namely, discover a phase transition that does not match the Gaussian ensemble and is an interesting ensemble because the matrix is explicit and deterministic.

4.1.2 Polytope Geometry

In order to understand the work of Donoho and Tanner, it is necessary to know only a few key definitions from convex polytope geometry. Although non-convex polytopes exist, we have no occasion to consider them here, so all future references to polytopes are understood to be references to convex polytopes.

A set $\mathcal{C} \in \mathbb{R}^n$ is convex if for any points $z_1, z_2 \in \mathcal{C}$,

$$t z_1 + (1 - t) z_2 \in \mathcal{C}, \quad t \in [0, 1]$$

as well, or in other words, a convex set contains the line segment between any two points in the set. If we have a set of points S, the *convex hull* of S is given by

$$\mathbf{conv}S = \{t_1 z_1 + \cdots + t_k z_k : z_i \in S, t_i \geq 0, i = 1, \ldots, k, t_1 + \ldots + t_k = 1\}$$

c.f. [10]. In words, the convex hull is the minimal surface that encloses all points in the set S.

A *polytope* is the convex hull of a finite set of points in $\mathbb{R}^n$. A polytope may also be defined as a polyhedron that is bounded, though we mention this only for completeness. For a more detailed and technical description of polytopes, see Ziegler's *Lectures on Polytopes* [44] and Grünbaum's more advanced *Convex Polytopes* [45].

The most basic polytope is the *standard n-simplex,*

$$T^{n-1} = \{x \in \mathbb{R}^n : \mathbf{1}^T x \leq 1, x \geq 0\}. \tag{4.3}$$

In two dimensions, the simplex is an isosceles triange, and in three dimensions, the simplex is a tetrahedron.

We also define the *crosspolytope* as

$$C^n = \{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i| \leq 1\}. \tag{4.4}$$

For $n = 3$, the crosspolytope is the octahedron. Already there are hints of a connection to (LP) and (BP) [46].

The most essential feature of polytopes we will be concerned with is the number of faces of dimension k. For a polytope P, we write the number of k-dimensional faces as $f_k(Q)$. $f_0(Q)$ is the number of vertices of Q, $f_1(Q)$ is the number of edges, $f_2(Q)$ the number of

two-dimensional faces, and so on. If $Q = C^3$, the cross-polytope (Equation 4.4), the convex

body with vertices at the unit vectors $\pm\mathbf{e}_1, \pm\mathbf{e}_2$, and $\pm\mathbf{e}_3$, we have

$$C^3 = \left\{ v \in \mathbb{R}^3 : \|v\|_1 = \sum_{i=1}^{3} |v_i| \le 1 \right\},$$

As can be counted in Figure 4.1 the face counts of C^3 are

$$\mathbf{C^3}: \quad f_0(P) = 6 \tag{4.5}$$

$$f_1(P) = 12 \tag{4.6}$$

$$f_2(P) = 8 \tag{4.7}$$

For short, we refer to a k-dimensional face as a *k-face*. It is no coincidence that we refer to

the k-sparsity of a signal and a k-face. Indeed, as we will see, the connection is exactly what

enables precise undersampling theorems.

4.1.3 The Connection between Polytope Geometry and Convex Optimization

Donoho and Tanner, 2010 [37], present the following theorem that is purely the result of

combinatorial geometry with no explicit connection to linear programming. We state it here

in full with some paraphrasing and modification of notation and note that in that paper

they cite their own work for the proof [42, 43, 47, 48], which is well beyond the scope of this

paper:

Theorem 4.1

Phase Transition for Face Counts of Gaussian Randomly Projected Polytopes

Let the $m \times n$ random matrix A have i.i.d. $N(0,1)$ Gaussian elements. Consider sequences

of triples (n, m, k) where $m = \delta n$, $k = \rho m$, and $n \to \infty$. There are functions $\rho_B(\delta; Q)$ for
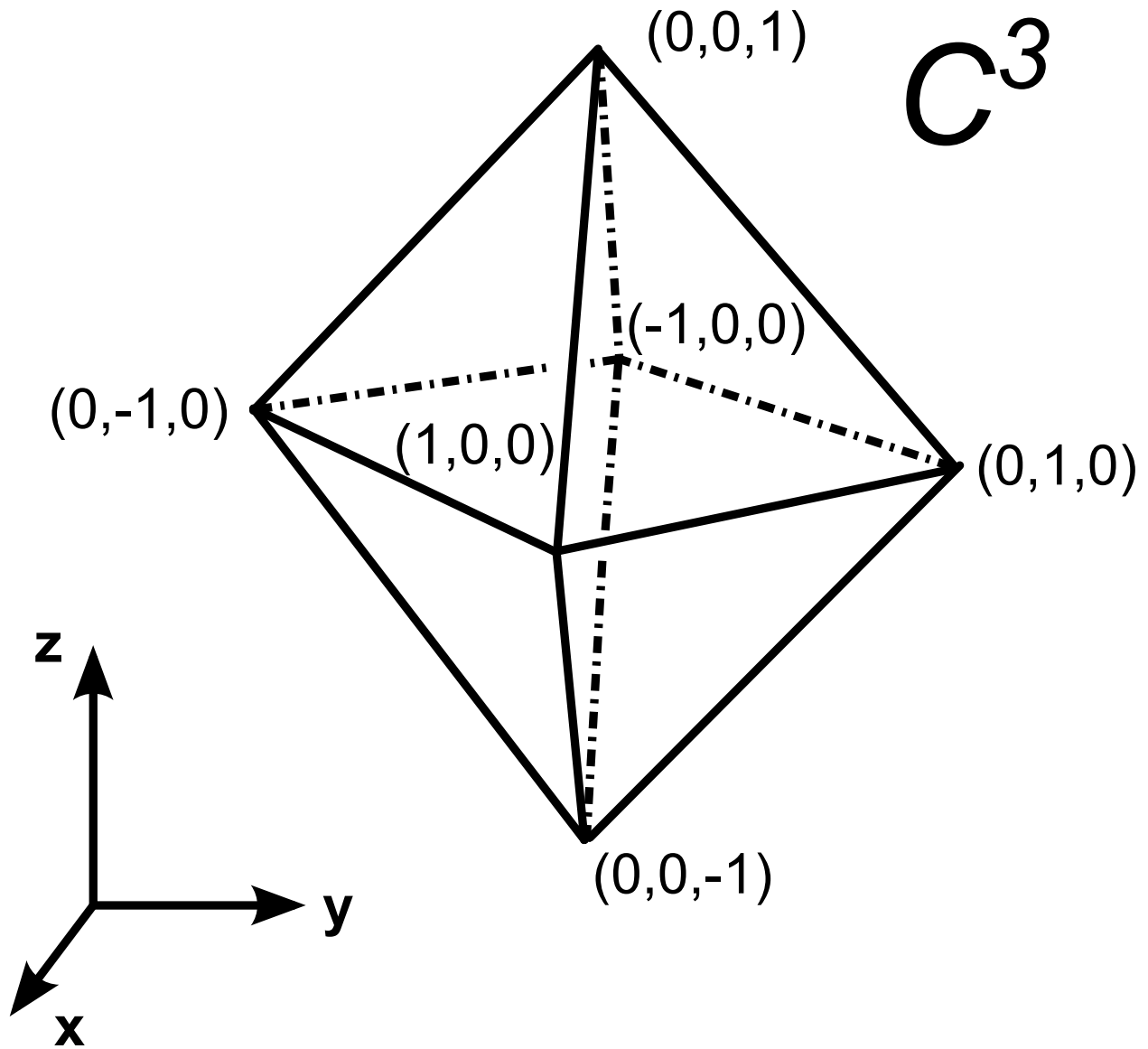
Figure 4.1 : The crosspolytope in three dimensions. There are six vertices, or 0-faces, twelve line segments, or 1-faces, and eight 2-faces, or what we commonly call a face.

$Q \in \{T, C\}$ *demarcating the phase transitions in face counts*

$$\lim_{N \to \infty} \frac{f_k(AQ)}{f_k(Q)} = \begin{cases} 1, & \rho < \rho_B(\delta; Q) \\ 0, & \rho > \rho_B(\delta; Q) \end{cases}$$

Amazingly, these boundaries defined by $\rho_B(\delta; Q)$ also predict the value of ρ where, for a given δ, gaussian and other types of random matrices transition from successfully recovering the underlying signal to failure as the number of nonzero elements in the test signal vector, $\|x\|_0 = k$ goes from 0 to m. Thus, if k is either referring to sparsity or k-face, whether or not $\rho = k/m$ is less than or greater than $\rho_B(\delta; Q)$ will predict whether or not (LP) or (BP) will recover the underlying signal x.

4.2 Experimental setup

For the images below, we solve 126×126 separate optimizations for

$$\delta = \rho = (.02\ .0275\ .035\ .0425\ldots\ .95\ .9575),$$

and generate a new problem for each (δ, ρ) pair for every one of N_{trials} trials. N_{trials} varies from diagram to diagram, from 10 to 30. Future work is to standardize N_{trials}. For each of these $126 \times 126 \times N_{trials}$ problems we generate a new test vector x and a new measurement matrix Φ according to the problem parameters (δ, ρ), and what type of problem is being solved, i.e. basis pursuit or linear program. We test two different methods We count the solution as success if $\|x^* - x\|_\infty < 10^{-6}$. The colors on the plots below indicate the percentage of trials which the solver, CVX [49], succeeded for a given (δ, ρ). We use CVX because all other solvers tested failed.

In the next section we discuss how, on a large scale, the phase transitions match those

for gaussian matrices as derived and supported experimentally in [39], but there are subtle differences and unexpected regions of failure, as well as isolated instances of failure even below the boundary that in [37] and the other related phase transition work, would be a region of 'perfect correspondence' or perfect recovery because the ℓ_1 solution is actually identical to the ℓ_0 solution.

4.3 Phase Diagrams for Select n with Explanations

To reiterate, the purpose of these numerical tests is to see if randomly generated Ω, which we denote Ω_r, outperform sequentially generated Ω, which we denote Ω_s. Thus, we build phase diagrams either with the measurement matrices for each trial generated by $\Phi^\circ = R_{\Omega_r}$ or $\Phi^\circ = R_{\Omega_s}$. Superimposed on each phase diagram is the curve $\rho_B(\delta; Q)$, with $Q = C$ for basis pursuit and $Q = T$ for the linear program.

4.3.1 Basis Pursuit Results

First we show results for solving basis pursuit with Ω_r and Ω_s. Because we have historically worked with square powers of 2 due to the use of Hadamard-based measurement vectors, we begin with $n = 32 \times 32$. The result for Ω_s is shown in Figure 4.2, and the result for Ω_r is shown in Figure 4.3. It appears that perhaps there is some evidence of diminished performance with Ω_s compared with Ω_r. There are three distinct collapses in the phase frontier, one at small δ, one for $\delta = .25$, and one for $\delta \approx .55$.

There are also some oscillations in the phase frontier whose origin is unknown. To ensure that our results were valid, we checked the result for Φ whose elements are each taken i.i.d. from a Gaussian distribution. The phase diagram is well-documented to not contain those oscillations. The phase diagram we obtained for Gaussian Φ still shows the oscillations, shown in Figure 4.4. So, although we don't know the origin of these, we know we are doing
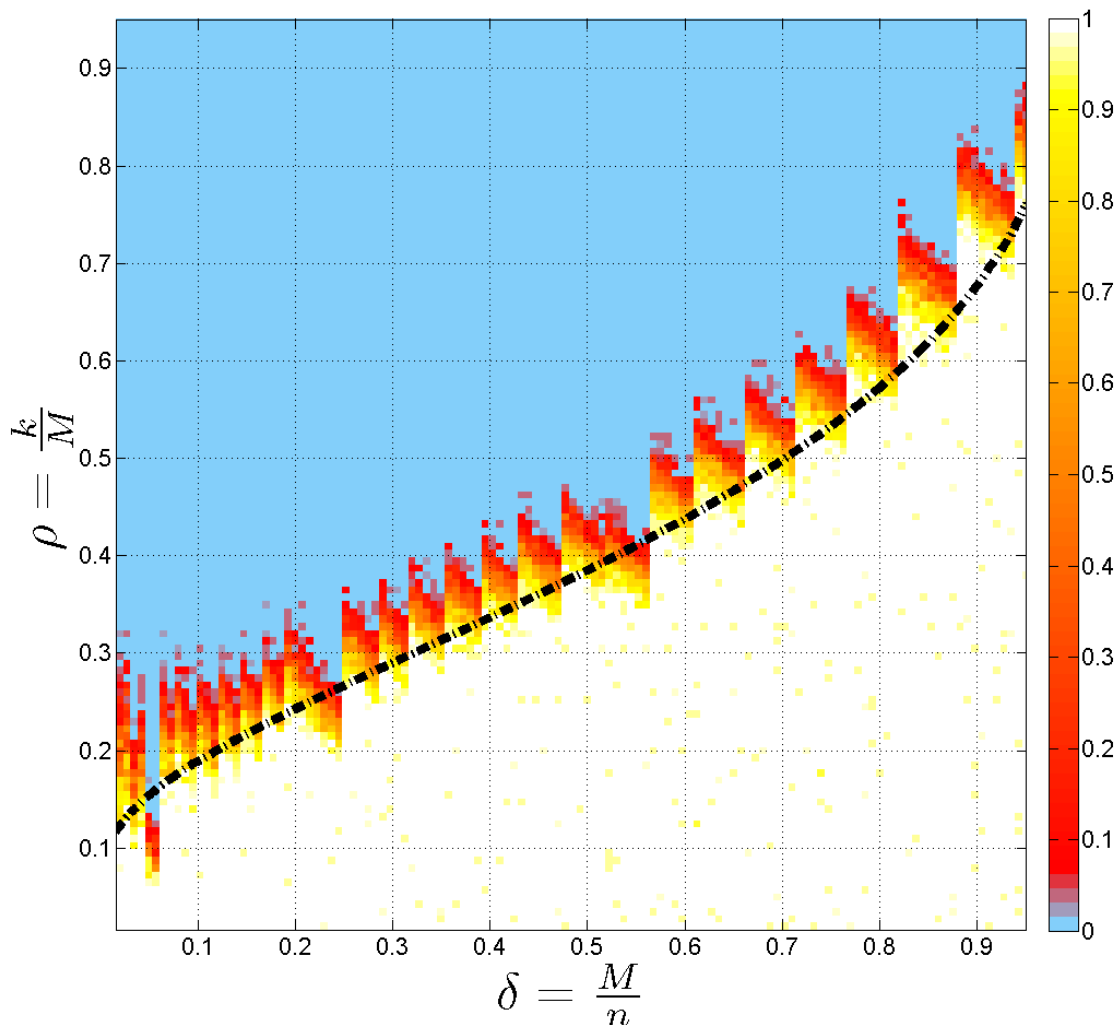
Figure 4.2 : (BP) $n = 32 \times 32, \Omega_s$

as well as we can since even the standard Gaussian Φ shows these oscillations.

4.3.2 Linear Program

The next natural question to ask is, "Is this collapsing phase frontier still present when solving LP?" The answer is yes, for $n = 32 \times 32$. In Figure 4.5, we see that again for $\Phi = R_{\Omega_s} \Phi^\circ$ there are collapses in the phase frontier at the same locations. However for Ω_r
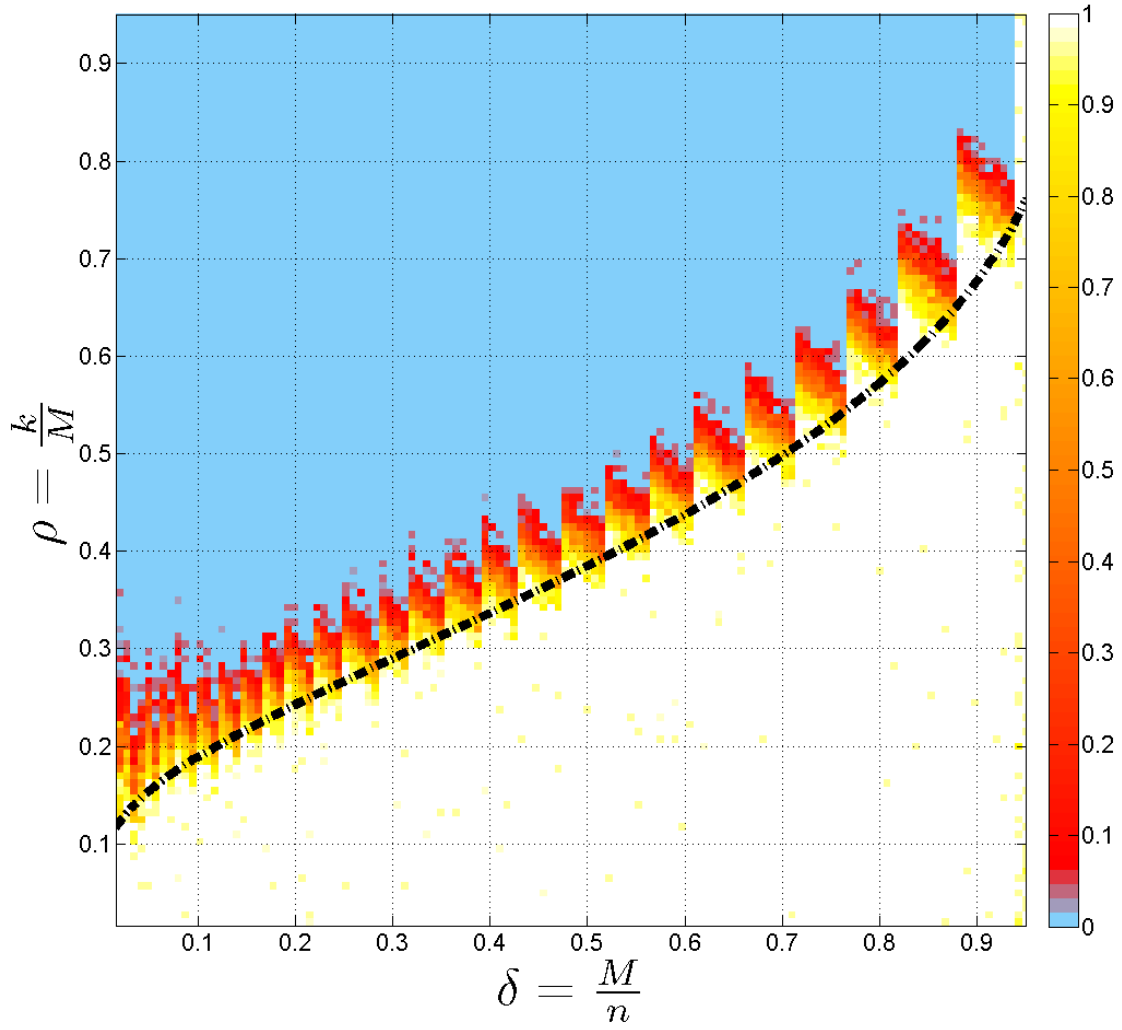
Figure 4.3 : (BP) $n = 32 \times 32, \Omega_r$
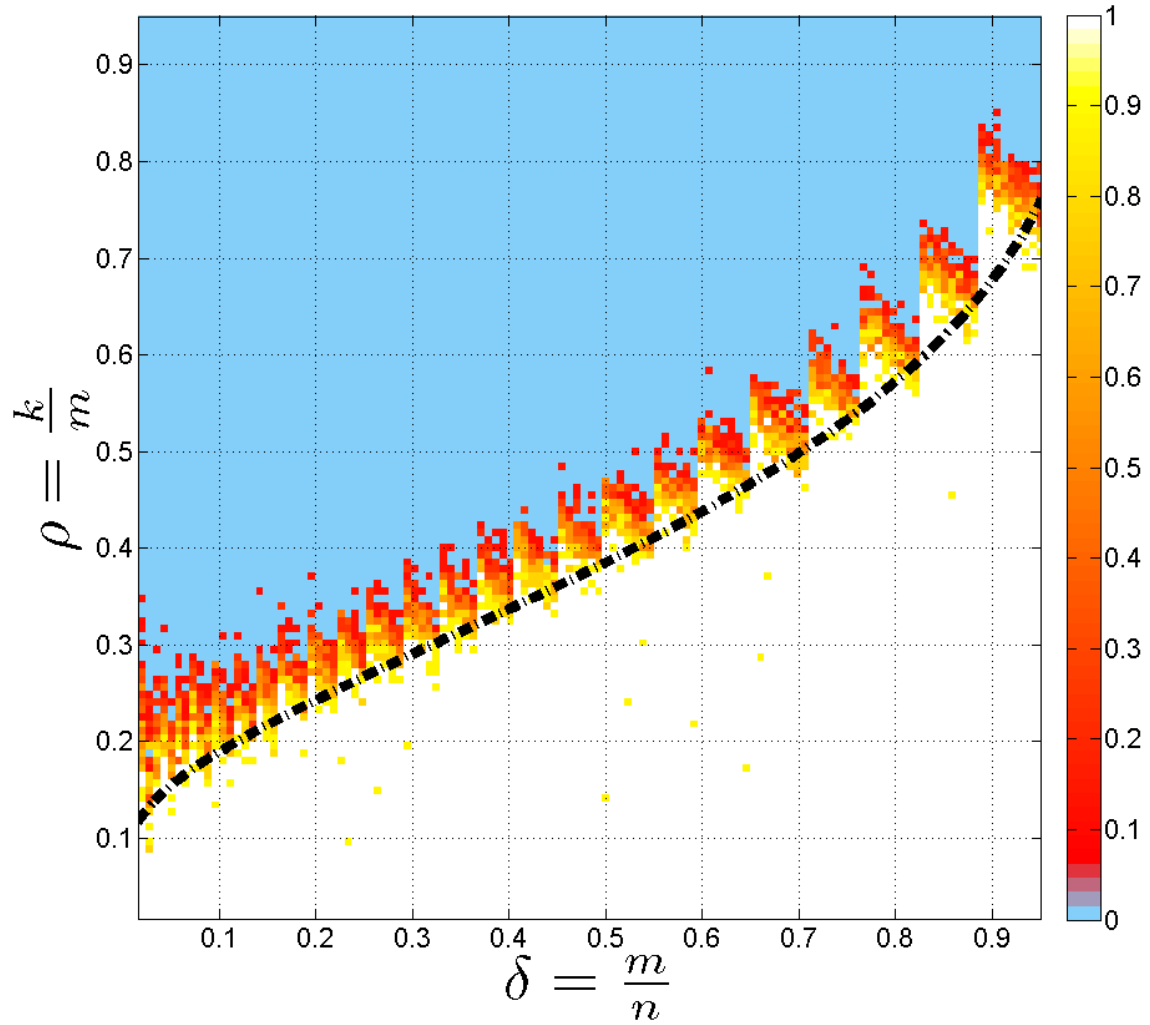
Figure 4.4 : (BP) $n = 32 \times 32, \Phi$ has Gaussian entries
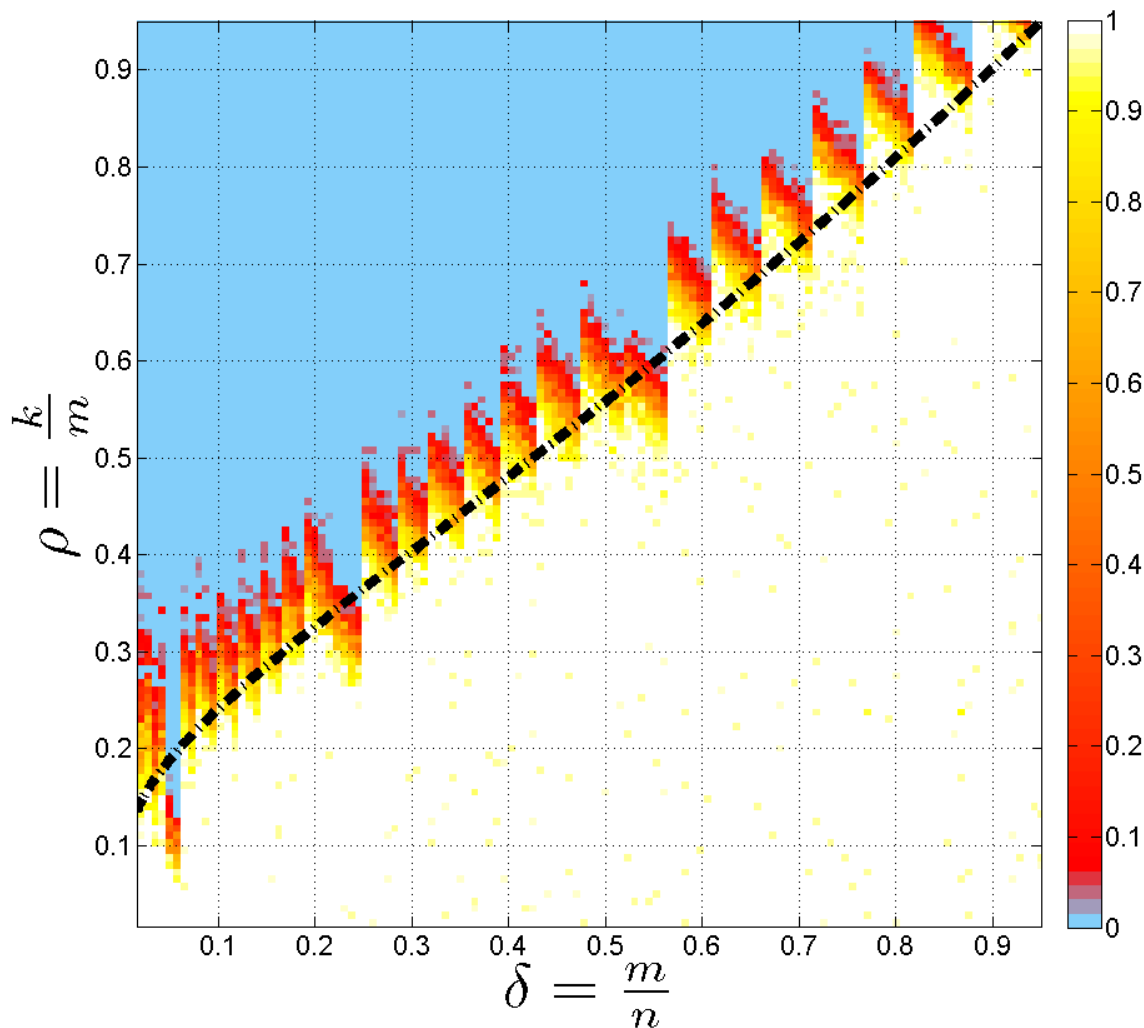
Figure 4.5 : (LP) $n = 32 \times 32, \Omega_s$

there are no such collapses, as shown in Figure 4.6. Then the next question we ask is, "Do we see this behavior at other resolutions?" The answers to that for $n = 33 \times 33$, $n = 31 \times 31$, and $n = 30 \times 30$ is no, as shown in Figure 4.7, Figure 4.8, and Figure 4.9 respectively.

However, what about another square power of 2? CVX is too slow to create phase diagrams for $n = 64 \times 64$, so we try $n = 16 \times 16$. It's exciting that for this resolution as well, there are collapses in the phase frontier at two of the same locations as $n = 32 \times 32$,
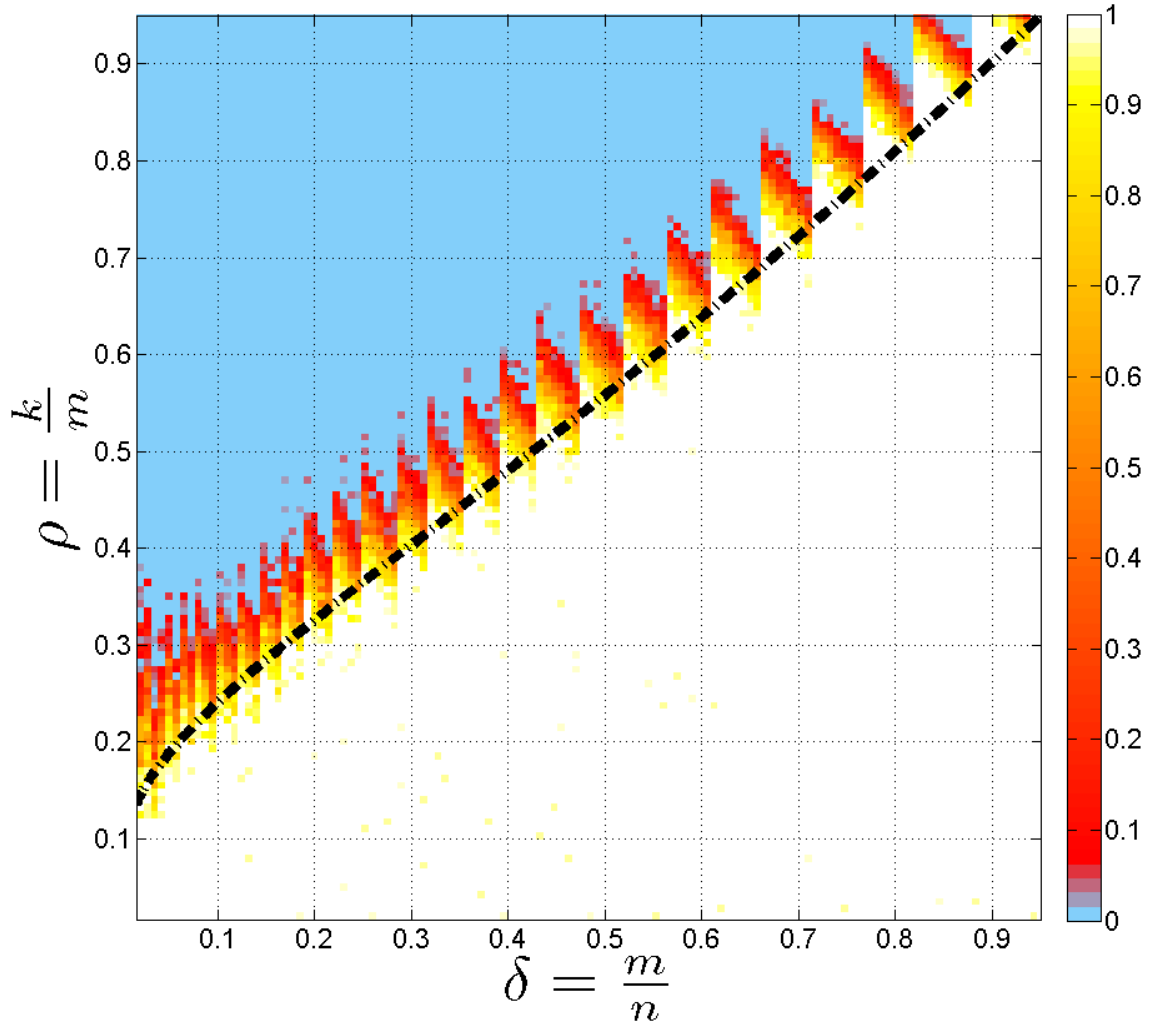
Figure 4.6 : (LP) $n = 32 \times 32, \Omega_r$
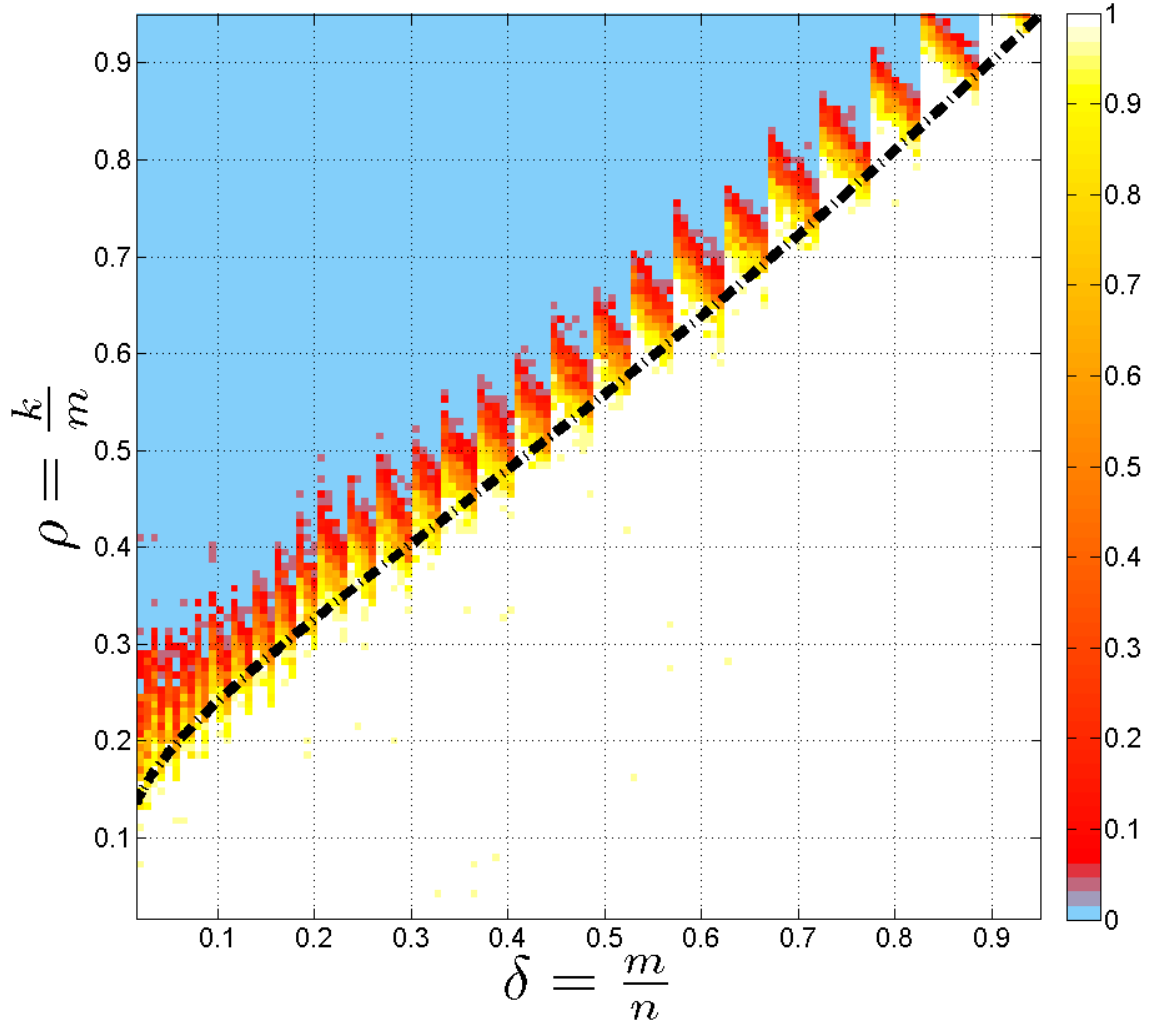
Figure 4.7 : (LP) $n = 33 \times 33, \Omega_s$
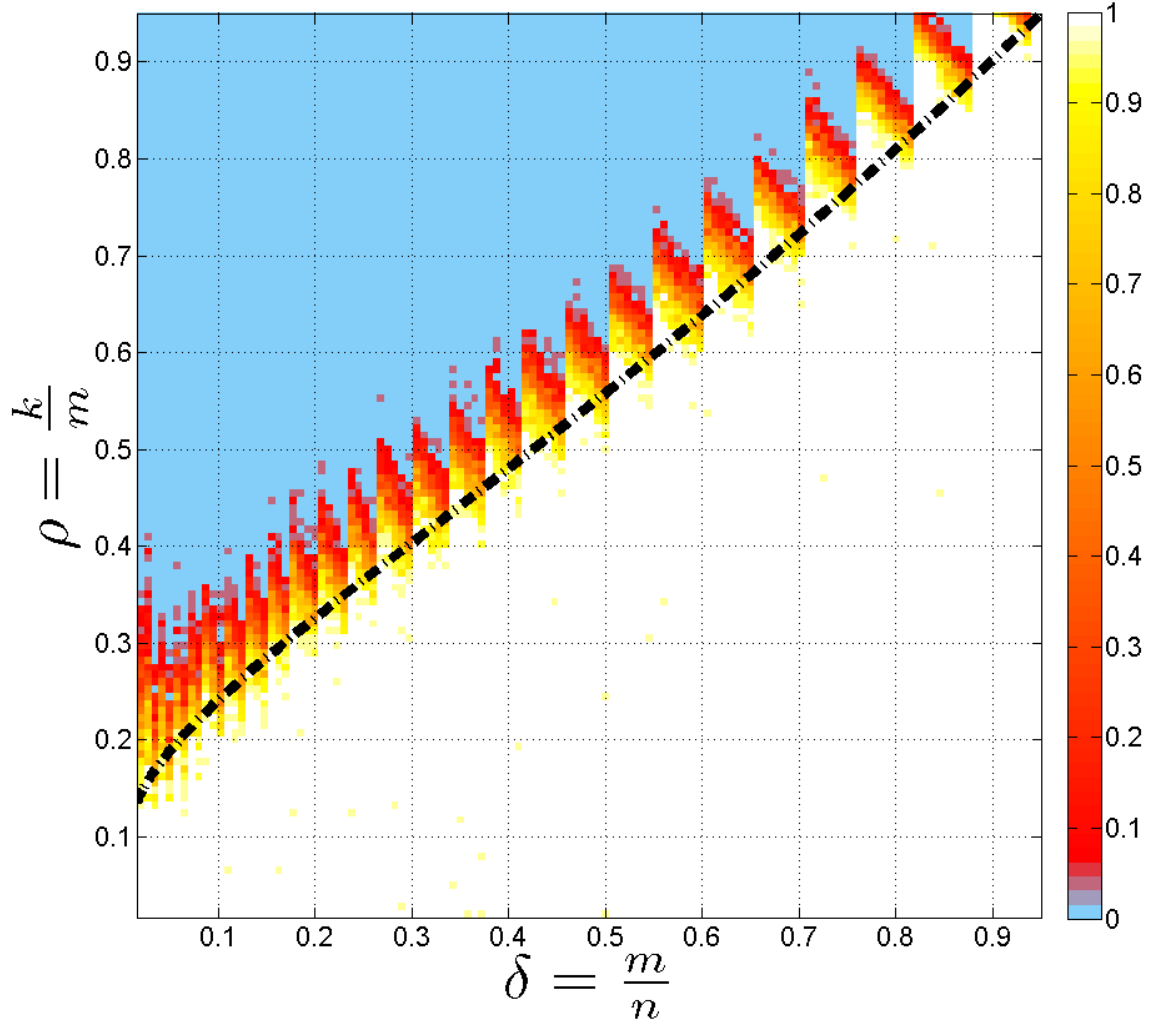
Figure 4.8 : (LP) $n = 31 \times 31, \Omega_s$
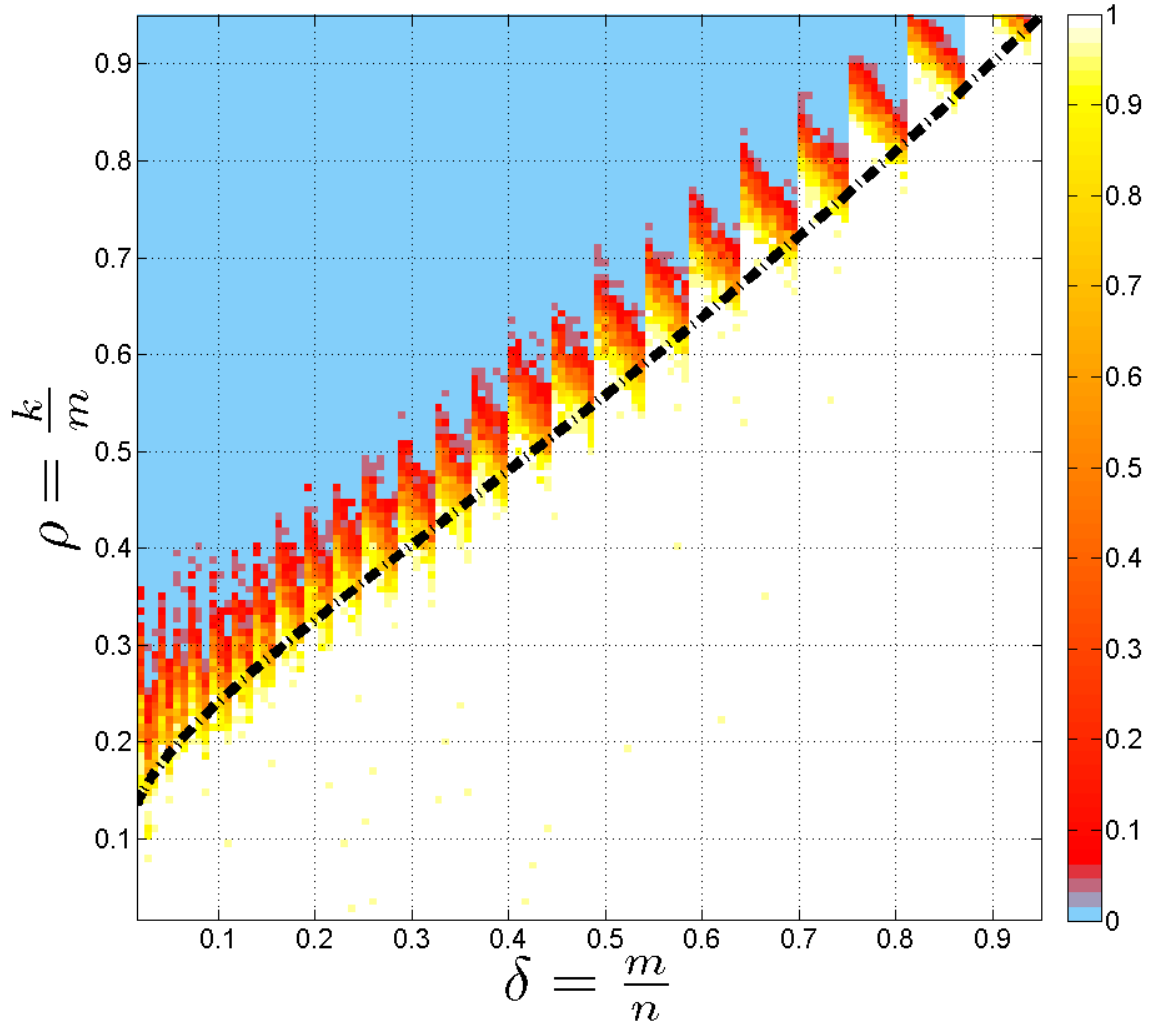
Figure 4.9 : (LP) $n = 30 \times 30, \Omega_s$
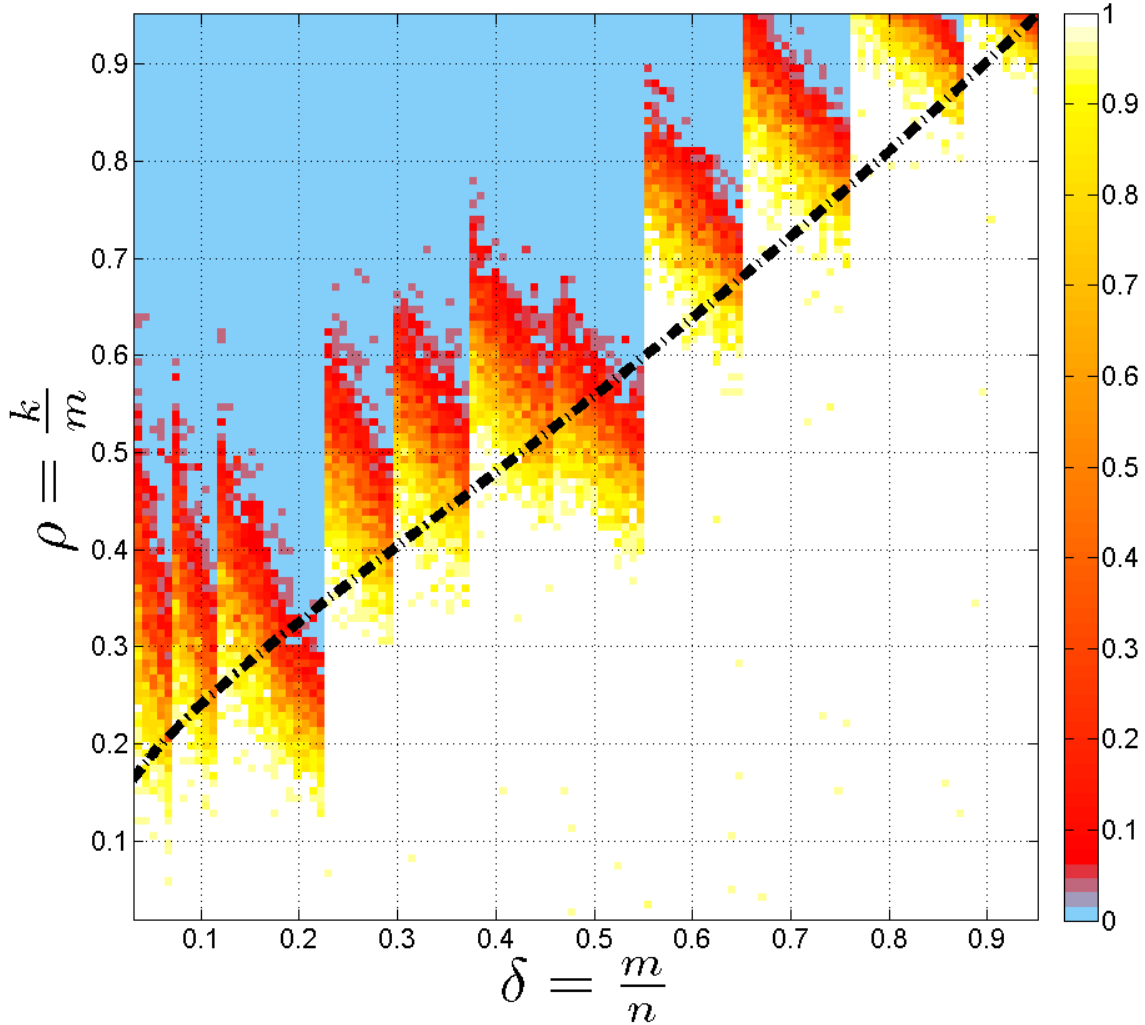
Figure 4.10 : (LP) $n = 16 \times 16, \Omega_s$

namely $\delta = .25$ and $\delta \approx .55$, as shown in Figure 4.10.

Operating on a hunch that this is related to resolutions that are not just square powers of two, but square powers of any prime number. The immediate test then is $n = 27 \times 27 = 3^6$, and the result is shown in Figure 4.11. Again we do in fact see a collapse in the phase frontier, but this time at different, and only two, values of δ.
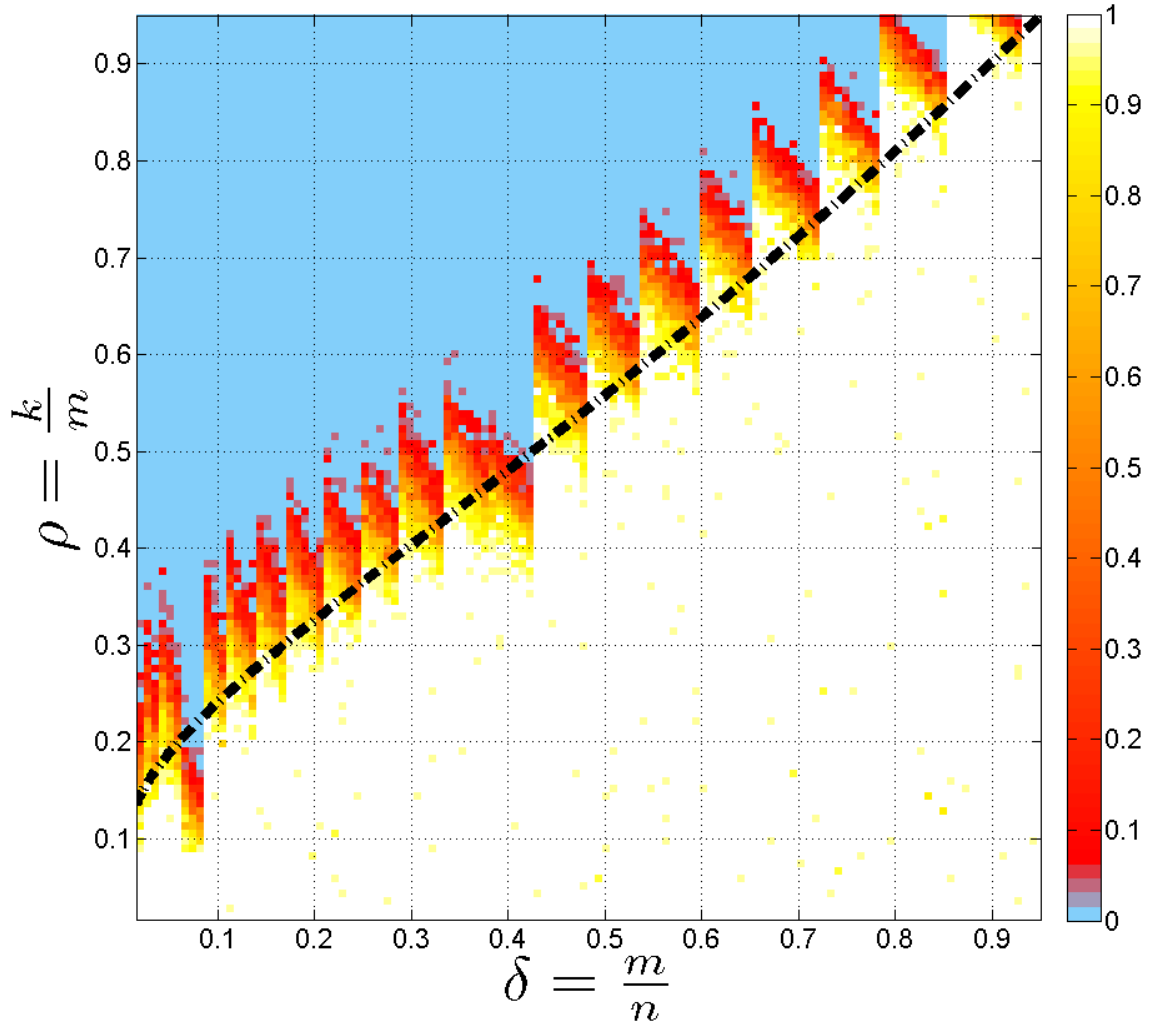
Figure 4.11 : (LP) $n = 27 \times 27, \Omega_s$

4.3.3 Some Coherence Statistics

As illustrated for pixel-array imaging in Chapter 1, Equation 1.4, the ith pixel is the coefficient on the ith column of Φ, i.e.

$$y = \varphi_1^c x_1 + \varphi_2^c x_2 + \ldots + \varphi_n^c x_n = \sum_{i=1}^{n} \varphi_i^c x_i. \tag{4.8}$$

Based on this fact, if columns φ_i^c of Φ are very similar, it will be difficult to determine the values of separate pixels x_i. A widely used statistic that has connections to the restricted isometry property (RIP), which says how close the operator Φ is to an isometry, is the coherence,

$$\mu(\Phi) = \max_{i \neq j} \left| \langle \varphi_i^c, \varphi_j^c \rangle \right|. \tag{4.9}$$

If this value is small, then Φ is more nearly an isometry. According to the theory surrounding the restricted isometry property, the smaller $\mu(\Phi)$, the larger $\|x\|_0$ may be to still recover x from the underdetermined set of equations $y = \Phi x$ [50, 51]. Given that coherence is an indicator of recovery performance, we must investigate whether or not the coherence $\mu(\Phi)$ is different for sequentially-built Φ and randomly-built Φ. Indeed we see a difference, but not as expected, and not a difference that supports the hypothesis that sequentially-built Φ would perform worse.

To test the coherence of Φ built from these two possibilities, we calculate $\mu(\Phi)$ explicitly for increasing values of δ for each case. We do this for three separate resolutions for which we obtained the phase diagrams above, namely $16 \times 16, 32 \times 32$, and 33×33. For each value of δ in each case, we generate $N_t rials = 100$ different partial circulant matrices. We calculate $\mu(\Phi)$ at twelve different values of δ, evenly spaced from $\delta = .05$ to $\delta = .9212$ for each trial. Then we find the average of these 100 trials, and plot that along with the standard deviation (error bars) as shown in Figure 4.12.
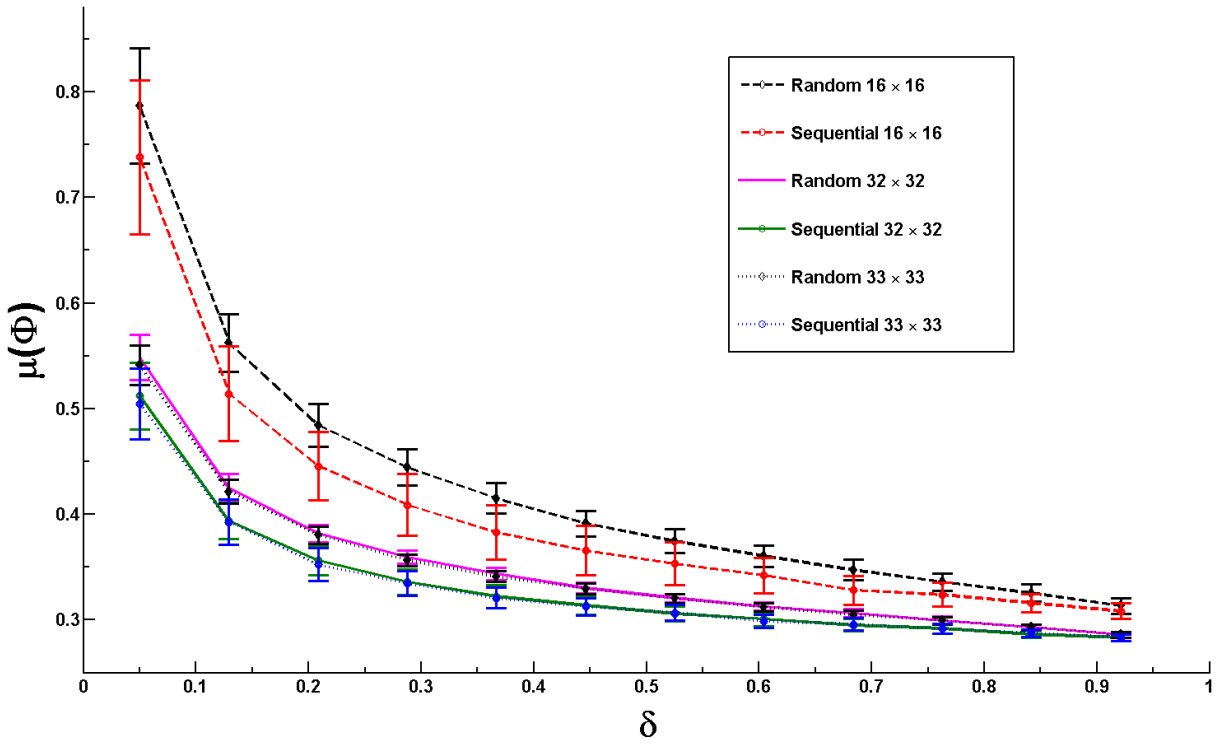
Figure 4.12 : $\mu(\Phi)$ for various resolutions and for random and sequential methods for a series of values δ. Perhaps unexpectedly, the coherence for sequential-type Φ is lower than for random-type, however the deviation from the mean is larger for sequential than for random.

Surprisingly we find that the coherence is larger on average for sequential-type partial circulant matrices than it is for random. The standard deviation, however, is larger with sequential than it is for random for most values of δ. As expected, the coherence decreases as n increases. Most importantly, there is no evidence that coherence properties of these matrices can explain the observed difference in performance as visualized by the collapses in the phase fronteirs above.

4.4 Discussion and Future Work

With the exception of collapses in the phase frontier as shown for $n = 2^8, 2^{10}$, and 3^6, the phase diagrams essentially match that of the theoretical curve $\rho_B(Q)$ as introduced in the theorem of Donoho and Tanner. This suggests that, indeed, partial circulant matrices are members of the same "universality class" as Gaussian matrices, meaning that there could exist a provably precise undersampling theorem for partial circulants as well. Then an immediate, though challenging, future step would be to prove that as $n \to \infty$, the fraction of successful recovery for a given problem instance is either 1 or 0 with the boundary in phase space between perfect recovery and failure given by $\rho_B(Q)$. Secondly, it would be interesting to know for what other resolutions n do we see collapses in the phase frontier as we did for those three, $n = 2^8, 2^{10}$, and 3^6. Whether or not there are more resolutions for which this occurs, what is the origin of the collapse?

Chapter 5

Future Directions and Conclusion

Here we present preliminary data on adaptive sensing, which represents one further step into incorporating signal processing into an imaging system. This should not be confused with adaptive optics, the technique in astronomy that uses a laser 'guide star' to monitor fluctuations in atmospheric conditions and deformable mirrors to compensate for that [52]. Maybe some combination of the two is possible, which could result in a very powerful imaging technique. With adaptive sensing, the adaptivity is not explicitly to account for environmental conditions, but instead adaptive in the sense that the ith measurement vector $\varphi^{(i)}$ depends on the outcome of the measurement before it. Of course such a scheme would not be applicable to measurement systems like CCD array imaging where the measurements are multiplexed in space and taken all at one time.

Finally, we conclude with some thoughts centered around the idea of building signal processing into imaging and measurement devices in general, how that can influence the science that is done with those sensors, and how the two thrusts, the drive to acquire data for scientific purposes and the drive to find new and powerful mathematics, sustain each other in this paradigm. Like neurons building up, branching out, finding each other, preparing to forge powerful new generalizations and analogies, I see the beginning of connections forming between science and powerful new sensing paradigms including adaptive and compressive sensing, and probably also using ideas for noise reduction like adaptive optics.

5.1 Adaptive Sensing

Although it was not emphasized in previous chapters, the measurements were *non-adaptive*, that is, the measurement vector $\varphi^{(i)}$ did not depend on $y_{i-1} = \left\langle \varphi^{(i-1)}, x \right\rangle$. In this section we report preliminary data on *optical adaptive sensing*, where the current measurement vector does in fact depend on the result of the measurement before it. Based on work by collaborators Indyk, et al., [53] entitled *On the Power of Adaptivity in Sparse Recovery*, we built an optical system to initially handle only the simplest case: the 1-sparse case. The experimental setup is simple: I simply focused a laser down to a single digital micromirror device (DMD) pixel and ran three different algorithms,

1. **Binary Search:** A portion of the DMD is either on or off for each iteration. The result of the current measurement informs the algorithm which portion of the DMD should be halved next. This continues until the portion of the DMD is only two pixels.

2. i **grayscale levels at the ith measurement:** Here we take advantage of the property that the DMD can half-tone (flip on and off quickly) to create grayscale levels inbetween 'off' and 'on'. With each new measurement vector, $\phi^{(i)}$, we add one more level of gray to the measurement system. Thus, if our detector is sensitive up to i levels, we can determine if the nonzero pixel is in one of i separate areas of the DMD.

3. 2^{2^i} **grayscale levels at the ith measurement:** Identical in concept to the i grayscale levels at the ith measurement, but with more levels. This is of course dependent on the DMD's ability to create that many grayscale levels, which as we show below is suspect, and the detector's ability to discriminate between that many grayscale levels.

First, let us show the binary search. As explained above, we search for the single non-zero pixel, as shown in Figure 5.1. With each new measurement vector, we see if the intensity is the same as for 'all on' (the top illustration in Figure 5.1), or if it is zero. If it is zero, as it is

for the first split-in-half of the DMD, then the portion of the DMD that was 'off' is split to half-on, half-off, and so on. Because it is easy and quick to visualize, and adaptive imaging may have applications for hyperspectral microscopy and astronomy, we use a spectrometer to acquire the result $y_i = \langle \phi^{(i)}, x \rangle$. We simply take the maximum reading of the spectrometer, or mathematically,

$$y_i = \|y_{ij}\|_\infty, \quad j = 1, 2, \ldots, n_{spec},$$

where n_{spec} is the spectrometer resolution. With the laser focused to the point above, we iterate through all the patterns in Figure 5.2 until we are down to only two pixels off or on. Since each measurement splits the DMD in half, it takes $m = \log_2 n$ measurements to find the 'on' pixel.

measurement index i	Expected Relative Intensity	Measured Relative Intensity
0	1.000	1.000
1	0.000	0.091
2	0.500	0.378
3	0.333	0.155
4	0.500	0.449

Table 5.1 : Table showing expected and actual measurement values for the ith measurement with the i gray levels scheme, taken as the maximum of the peak of a spectrometer reading.

The illustration of the i-grayscale levels for the ith measurement with the 2^{2^i} gray levels scheme is shown in Figure 5.3. In Table 5.1 we show the result of the ith measurement.

measurement index i	Expected Relative Intensity	Measured Relative Intensity
0	1.000	1.000
1	1.000	1.000
2	0.333	0.290
3	0.600	0.579

Table 5.2 : Table showing expected and actual measurement values for the ith measurement, taken as the maximum of the peak of a spectrometer reading.
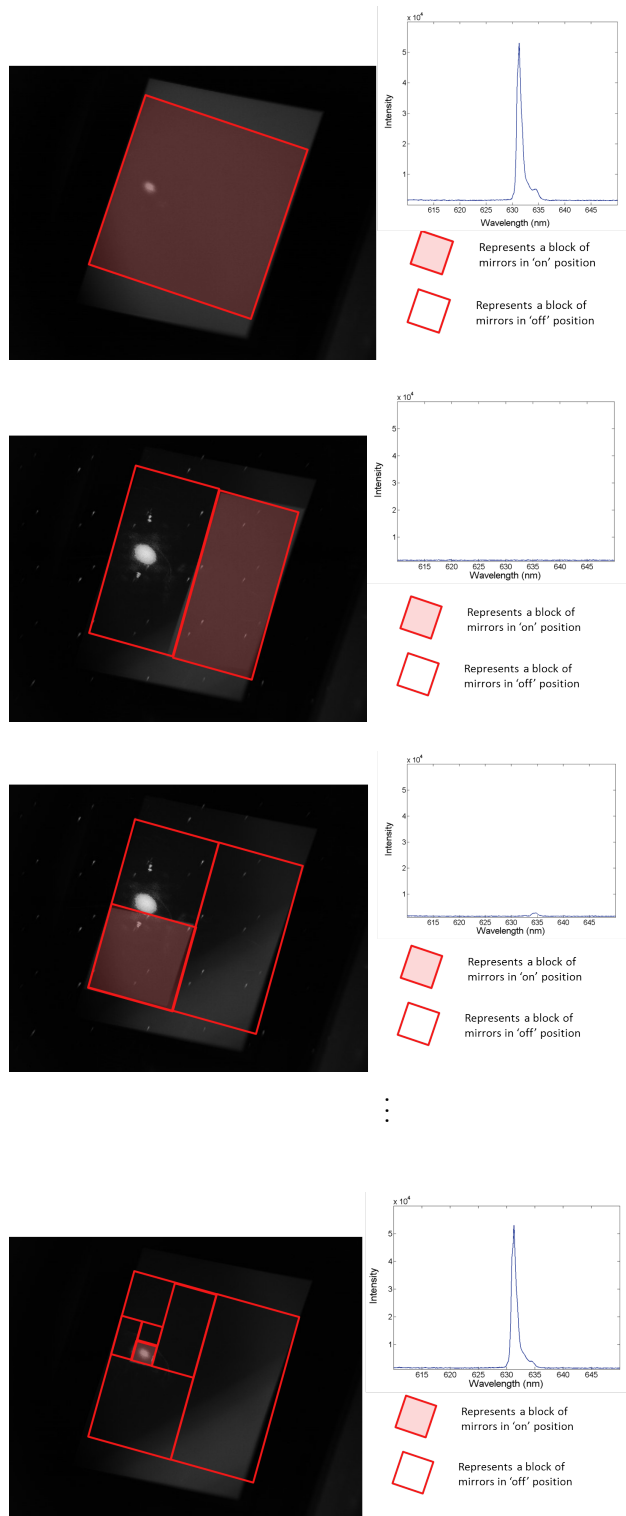
Figure 5.1 : A series of 1-bit adaptive measurements. An alternative description would be a binary search, where on each measurement we split the portion of the DMD where the 'on' pixel could be and ask 'Which half of this active space has the 'on' pixel?'.
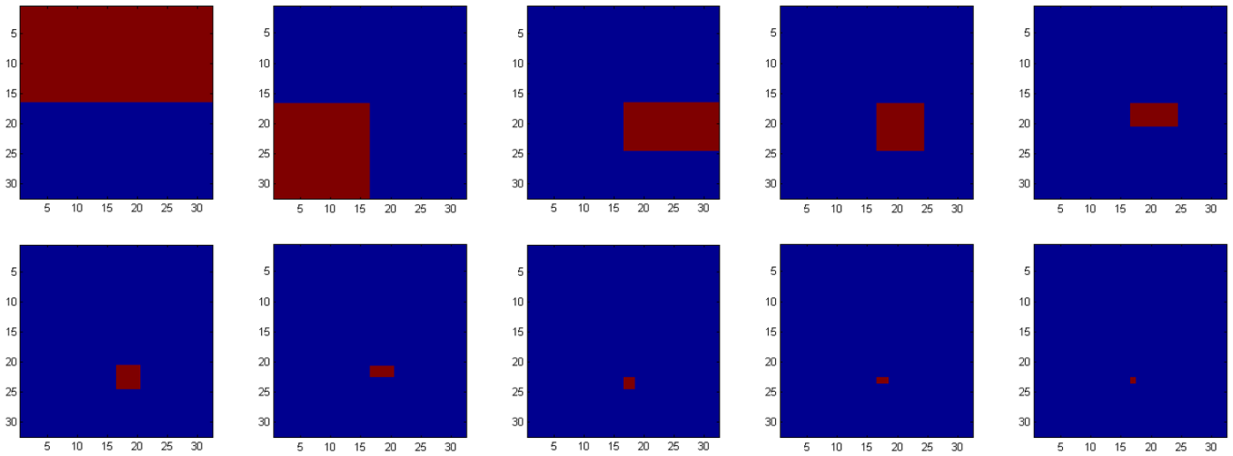
Figure 5.2 : Sequence of all $\log_2(1024) = 10$ 32×32 binary adaptive measurement vectors. The 0th measurement vector, where all the mirrors are 'on' is omitted.

And finally the 2^{2^i} grayscale levels for the ith measurement with actual measurement results shown in Table 5.2. The illustration of the corresponding DMD patterns we displayed is shown in Figure 5.4.

The measurements are not totally divergent from their expected values, however it does raise questions about whether the DMD/spectrometer combination is suitable for this application. Indeed, the grayscale functioning of the DMD is primarily for projection screens to project sequences of pictures that look like they are moving realistically to human beings. Of course, just the production of motion itself is an illusion, and human beings don't care that darker regions be exactly 30% darker than the brightest region—only the relative brightnesses of scenes matter. In that spirit, Figure 5.5 shows a series of 64 separate grayscale levels that we had the DMD implement along with the actual reading at the spectrometer while the DMD was displaying that grey level.

The graph shows that there may be only eight distinct gray levels the DMD can achieve. Of course before the gray-level adaptivity can be extensively developed, this phenomenon
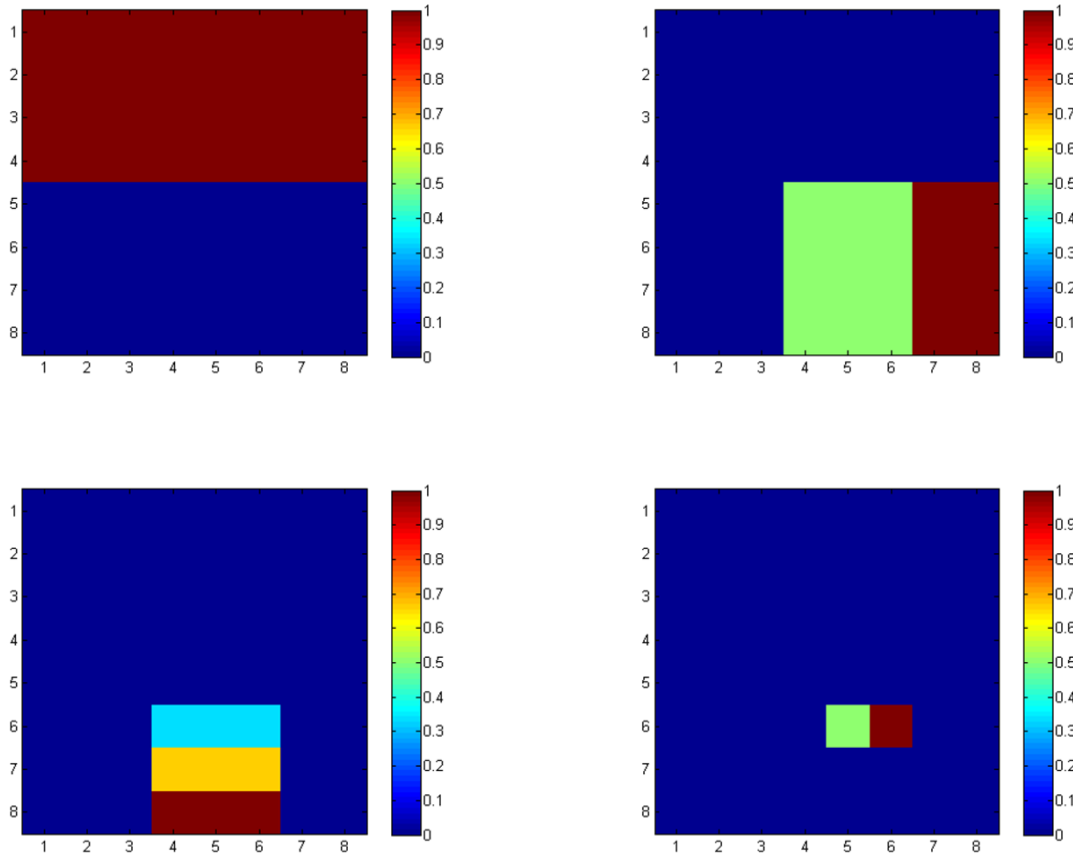
Figure 5.3 : Sequence of all 4 8×8 binary adaptive measurement vectors. The 0th measurement vector, where all the mirrors are 'on' is omitted. Theoretically $m \approx \sqrt{\log_2 n}$, however with only $n = 64$, we do not achieve this dramatic of an improvement.
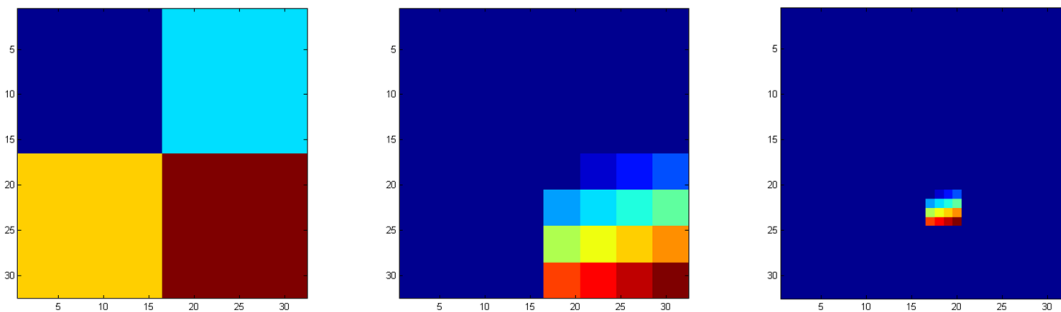
Figure 5.4 : Sequence of all $\log_2(\log_2 1024) = 3$ 8×8, 2^{2^i} grayscale levels for the ith measurement adaptive measurement vectors. The 0th measurement vector, where all the mirrors are 'on' is omitted.
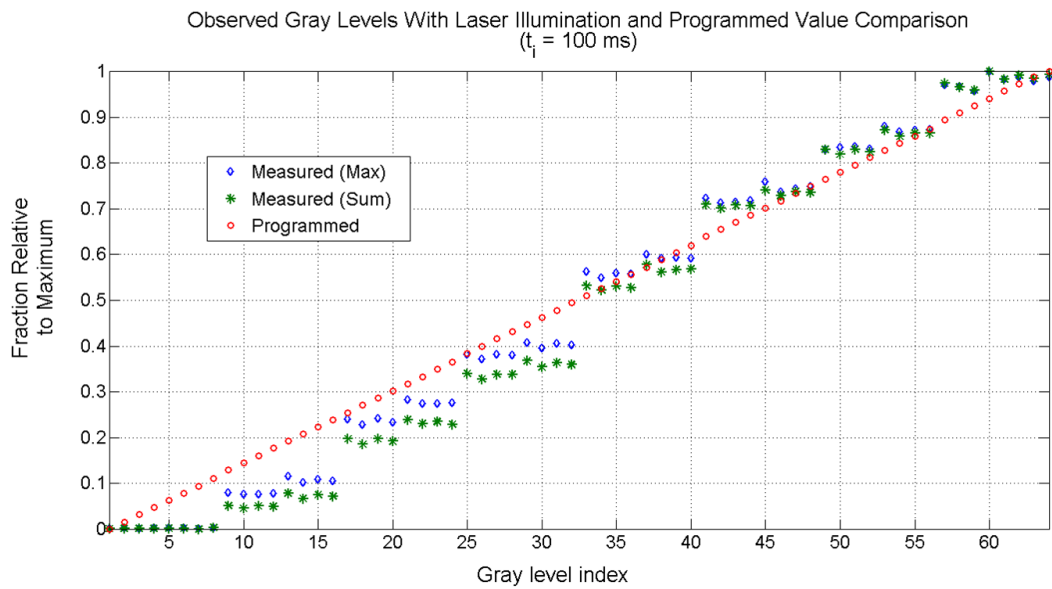
Figure 5.5 : Results from sending 64 different gray levels for the DMD to display along with the actual measurement recorded at the DMD. We quantified the spectrometer reading by 'sum' and 'max', 'sum' meaning we summed over all wavelength bins and divided by the number of pixels, and for 'max' we took the maximum value over all wavelength bins as we did for the measurements presented in the rest of the chapter.

needs to be understood and corrected for, if necessary.

5.2 Conclusion

This thesis itself weaved through signal processing, the cutting edge in imaging technology, microscopy and hyperspectral microscopy, and discrete geometry with connections to convex programming. Beginning with some basic notation, a simple example of copmressive imaging, and some intuition into the (not really) magic of compressive sensing, we transitioned into discussing more complex imaging systems, namely hyperspectral microscopy systems. In an effort to design and build a better imaging system, we encountered some interesting behavior, namely that the sequentially built partial circulant sensing matrices underperformed compared with randomly built partial circulant sensing matrices. By applying scientific methods to dig deeper into whether this performance gap was isolated to total variation (TV) minimization problem, we found suggestive evidence that said indeed it was isolated to TV minimization. However, that investigation itself led to new observations with attendant new questions, including, "Why are there collapses in the phase frontier for $n = 2^8, 2^{10}$, and 3^6?" Also, "Why is the coherence for randomly built measurement matrices, Φ, larger than that for sequentially built Φ?" And finally, the big question, "Does the same precise undersampling theorem that holds for random Gaussian sensing matrices also hold for partial circulant sensing matrices?"

It is my firm belief that this is a pattern that will repeat itself. The idea that how we acquire data can more efficiently lead to models of the data is a powerful one. When researchers outside of information theory, statistics, and mathematics begin to apply these theories to their own problems, they will see compressive sensing and sparse recovery through their own nuanced lens. The result will be a more parsimonious worldview where information and physical models are less and less distinguishable.

Bibliography

[1] D. Zucchino and D. S. Cloud, "U.s. deaths in drone strike due to miscommunication, report says," *Los Angeles Times*, October 14 2011.

[2] D. Reddy, A. Sankaranarayanan, V. Cevher, and R. Chellappa, "Compressed sensing for multi-view tracking and 3-d voxel reconstruction," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pp. 221 –224, oct. 2008.

[3] European Signal Processing Conference, *Multi-view through-the-wall radar imaging using compressive sensing*, August 2010.

[4] M. Harwit and N. J. Sloane, *Hadamard Transform Optics*. Academic Press, New York, 1979.

[5] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, 2006.

[6] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[7] C. Li, W. Yin, and Y. Zhang, "Tval3: Tv minimization by augmented lagrangian and alternating direction algorithms," November 2010.

[8] R. Courant, "Variational methods for the solution of problems with equilibrium and vibration," *Bulletin of American Mathematics*, vol. 49, pp. 1–23, 1943.

[9] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. University Press, Cambridge, 2009.

[11] C. Li, *An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing*. PhD thesis, Rice University, September 2009.

[12] J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, 2008.

[13] J. Ma, "Single-pixel remote sensing," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, pp. 199 –203, april 2009.

[14] C. Li, T. Sun, K. Kelly, and Y. Zhang, "A compressive sensing and unmixing scheme for hyperspectral data processing," *Image Processing, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2011.

[15] Y. Wiaux, G. Puy, and P. Vandergheynst, "Compressed sensing reconstruction of a string signal from interferometric observations of the cosmic microwave background," *Mon. Not. R. Astron. Soc.*, vol. 402, no. 2626, 2010.

[16] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," *Phys. Rev. Lett.*, no. 150401, 2010.

[17] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, December 2007.

[18] S. S. Vasanawala, M. T. Alley, B. A. Hargreaves, R. A. Barth, J. M. Pauly, and M. Lustig, "Improved pediatric mr imaging with compressed sensing1," *Radiology*, vol. 256, no. 2, pp. 607–616, 2010.

[19] N. Shental, A. Amir, and O. Zuk, "Identification of rare alleles and their carriers using compressed se(que)nsing," *Nucleic Acids Research*, vol. 38, no. 19, p. e179, 2010.

[20] S. N. A. A. Z. O. Erlich, Y., "Compressed sensing approach for high throughput carrier screen," pp. 539 –544, 30 2009-Oct. 2 2009.

[21] Y. Erlich, A. Gordon, M. Brand, G. Hannon, and P. Mitra, "Compressed genotyping," *Information Theory, IEEE Transactions on*, vol. 56, pp. 706 –723, feb. 2010.

[22] W. Dai, O. Milenkovic, M. Sheikh, and R. Baraniuk, "Probe design for compressive sensing dna microarrays," in *Bioinformatics and Biomedicine, 2008. BIBM '08. IEEE International Conference on*, pp. 163 –169, nov. 2008.

[23] M. Sheikh, O. Milenkovic, and R. Baraniuk, "Designing compressive sensing dna microarrays," in *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*, pp. 141 –144, dec. 2007.

[24] M. AlQuraishi and H. H. McAdams, "Direct inference of proteindna interactions using compressed sensing methods," *Proceedings of the National Academy of Sciences*, 2011.

[25] J. R. Ferraro and K. Nakamoto, *Introductory Raman spectroscopy.* Boston : Academic Press, c1994., 1994.

[26] C. R. Bailey, "The raman and infra-red spectra of carbon dioxide," *Nature*, vol. 123, p. 410, 1929.

[27] F. Rasetti, "Raman effect in gases," *Nature*, vol. 123, p. 205, 1929.

[28] W. Yin, S. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with toeplitz and circulant matrices," in *Proceedings of Visual Communications and Image Processing (VCIP)*, 2010.

[29] V. Studer, J. Bobin, M. Moussavi, E. Candes, and M. Dahan, "Compressive fluorescence microscopy for bioogical and hyperspectral imaging," *arXiv*, vol. arXiv:1201.0650v1, 2012.

[30] R. Aldrovandi, *Special Matrices of Mathematical Physics*. World Scientific, 2001.

[31] A. Papoulis, *Signal Analysis*. McGraw-Hill, New York, 1977.

[32] P. J. Davis, *Circulant Matrices*. Chelsea, New York, 1994.

[33] D. E. Knuth, *Seminumerical Algorithms*, vol. 2 of *The Art of Computer Programming*, section 4.3.3, pp. 290 – 295. Reading, Massachusetts: Addison-Wesley, second ed., 1981.

[34] Y. Xu, W. Yin, and S. Osher, "Learning circulant sensing kernels," vol. Rice University CAAM Report TR12-05, 2012.

[35] H. Rauhut, "Circulant and toeplitz matrices in compressed sensing," in *Proc. SPARS'09, Saint Malo*, 2009.

[36] H. Rauhut, J. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 242 – 254, 2012.

[37] D. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, pp. 913 –924, June 2010.

[38] D. Donoho and J. Tanner, "Exponential bounds implying construction of compressed sensing matrices, error-correcting codes, and neighborly polytopes by random sampling," *Information Theory, IEEE Transactions on*, vol. 56, pp. 2002 –2016, april 2010.

[39] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philosophical Transactions of the Royal Society A: Mathematical,Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4273–4293, 2009.

[40] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[41] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, pp. 617–652, 2006. 10.1007/s00454-005-1220-0.

[42] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. pp. 9446–9451, 2005.

[43] D. L. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 27, pp. 9452–9457, 2005.

[44] G. M. Ziegler, *Lectures on Polytopes*. Graduate Texts in Mathematics, Springer-Verlag, 1995.

[45] B. Grünbaum, *Convex Polytopes*. Graduate Texts in Mathematics, Springer, 2003.

[46] A. Schrijver, *Theory of Linear and Integer Programming.* John Wiley & Sons, 1986.

[47] D. L. Donoho and J. Tanner, "Counting faces of randomly projected polytopes when the projection radically lowers dimension," *Journal of the American Mathematical Society*, vol. 22, pp. 1–53, 2009.

[48] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete Comput. Geom.*, vol. 43, pp. 522–541, Apr. 2010.

[49] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," Feb. 2011.

[50] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9 - 10, pp. 589 – 592, 2008.

[51] J. A. Tropp, "Norms of random submatrices and sparse approximation," *Comptes Rendus Mathematique*, vol. 346, pp. 1271 – 1274, 2008.

[52] R. Davies and M. Kasper, "Adaptive optics for astronomy," *Annual Review of Astronomy and Astrophysics*, vol. 50, no. 1, p. null, 2012.

[53] P. Indyk, E. Price, and D. P. Woodruff, "On the power of adaptivity in sparse recovery," *CoRR*, vol. abs/1110.3850, 2011.