ABSTRACT

ZHOU, YOU (JOE). Pixel Diffraction in Passive Millimeter Wave Compressive Imaging and Compressive Sensing Algorithm Development. (Under the direction of David Ricketts).

Compressive sensing has attracted increasing interest in the last two decades. In the field of compressive sensing, novel methods based on different theoretical models, e.g., convex optimization and Bayesian graphical methods, have been successfully developed, so that when accurate prior knowledge on the unknown input is available, novel compressive sensing algorithms can drastically reduce the number of measurements when estimating the input without compromising on the fidelity of estimates.

In this dissertation, we explore the applications of compressive sensing in two directions. In the algorithmic direction, a novel compressive sensing algorithm is extended for two data science tasks; in the hardware development direction, a unique diffraction source, namely pixel diffraction, in compressive imaging system is analyzed. While pixel diffraction limits depend on the specific spatial arrangements, our simple yet insightful method can be readily applied to any compressive imaging system that follows the same compressive sensing principle. Our result eradicates a potentially crucial source of uncertainties in millimeter wave compressive imaging systems.

Pixel Diffraction in Passive Millimeter Wave Compressive Imaging and
Compressive Sensing Algorithm Development

by
You (Joe) Zhou

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Electrical Engineering

Raleigh, North Carolina
2024

APPROVED BY:

| | |
|---|---|
| Chau-Wai Wong | Paul Franzon |
| David Aspnes<br>External Member | David Ricketts<br>Chair of Advisory Committee |

# BIOGRAPHY

You (Joe) Zhou received his B.Eng. in Mechatronics (Hons 1st Class) from the University of Sydney, Australia. From 2011 to 2014, he worked at Rio Tinto Alcan as an Electrical and Control Systems Engineer. From 2014 to 2015, he was an Automation Systems Commissioning Engineer at Origin Energy, Australia. From 2016 to 2017, he was a M.Sc. student in electrical and computer engineering at NC State University, and since 2017 has been working towards his Ph.D. degree in electrical and computer engineering at NC State University. His research interests lie in the interactions between hardware systems, first principles in physics and algorithms, Bayesian inference methods on signal processing problems, and machine learning in general.

# ACKNOWLEDGEMENTS

I am deeply indebted to my advisor Prof. David Ricketts for his extraordinary patience in guiding me to the end of my Ph.D. journey. Without his meticulous guidance, I would not be able to complete the program. I am always inspired by David's strong curiosity over a wide range of research fields in electrical engineering and his kindness to his students. He is my role model in life. I thank Prof. Dror Baron for introducing me to the intriguing topic of compressive sensing. I would like to thank Prof. Joel Trussell for advising me both as a senior faculty and as a true friend when I first arrived at NC State University. I thank Prof. Chau-Wai Wong, Prof. Rhett Davis, Prof. Paul Franzon, and Prof. David Aspnes for advising me on different research topics throughout the journey.

I would like to thank the Department of Electrical and Computer Engineering for being flexible and supportive during this difficult time. I have been very impressed with the friendly and supportive atmosphere in the department.

Finally, I would like to thank my friend Dr. Feifei Han for her continuous encouragement to me as my soulmate. Last but not least, it is my honor to thank my family, my mother, my father, and in particular, my grandfather Yunxin Dai for their selfless love and mental support.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1

# Introduction to Compressive Sensing

## 1.1 Motivation of Compressive Sensing

Linear inverse problems are ubiquitous in scientific research across disciplines such as wireless communications, medical and security imaging. Consider a linear inverse problem defined as the following:

$$y = Ax + z, \tag{1.1}$$

where $y \in \mathbb{R}^M$, $A \in \mathbb{R}^{M\times N}$, $x \in \mathbb{R}^N$, and $z \in \mathbb{R}^M$ are the noisy measurements, the known measurement matrix, the unknown signal to be resolved, and measurement noise respectively. The domain $\mathbb{R}$ can readily be extended to $\mathbb{C}$ as required. When prior knowledge, e.g., prior distributions or structural properties of $x$ is accessible, algorithms can be designed to incorporate the prior information in solving $x$ more accurately than the default pseudoinverse approach (2). Compressive sensing (CS) refers to the study of novel methods to systematically incorporate prior knowledge of $x$ into solving the inverse problem. CS is particularly useful for applications where measurements $y$ are expensive to obtain. CS enables robust reconstruction of $x$ with fewer measurements, i.e. $M < N$, hence the name compressive sensing. Sparsity is the most common prior in the CS literature, which refers to data structures such that the number of nonzero elements in $x \ll N$ in certain domains. For example, natural image data are often sparse in the wavelet or DCT domain (3; 4).

## 1.2 Compressive Sensing Algorithms

There are three types of algorithms for solving compressive sensing problems, namely the greedy method, convex optimization, and approximate message passing. In the classical greedy method, namely matching pursuit (5; 6; 7; 8), assuming $x$ is sparse, the indices and values of nonzero elements in $x$ are iteratively estimated by evaluating the correlation between $y$ and the nonzero elements in $x$. As nonzero elements are identified iteratively, the reconstruction is not optimal (9; 10). Based on the rapid development in convex optimization research during the last three decades, the problem of reconstructing the sparse $x$ subject to measurements (1.1) is formulated as a convex optimization problem of $\underset{\hat{x}}{\operatorname{argmin}} \|A\hat{x} - y\|_2^2 + \lambda \|\hat{x}\|_1$ (2; 10; 11; 12; 13; 14). The $l^1$ regularized formulation is valid due to its equivalence with the $l^0$ regularized optimization problem under very general conditions (15; 16), which promotes sparsity through $\lambda$. Unlike matching pursuit, convex optimization produces optimal estimates $\hat{x}$ as all entries in $\hat{x}$ are optimized simultaneously. Furthermore, priors other than sparsity can be readily incorporated by different regularization terms. In the example of CS image reconstruction, the total variation is the default regularization term that suppresses large intensity gaps across adjacent pixels in natural images (17; 18; 19). During the last 15 years, a Bayesian graphical method, namely approximate message passing (AMP), has become dominant in CS research. AMP is a low-complexity iterative algorithm intended for the high-dimensional linear inverse problem (1.1) (20; 21; 22; 23; 24). Over iterations, it improves the estimate of $\mathbb{E}[X|y, p]$, where $p$ is the prior distribution of the random variable $X$. For the standard sparse CS reconstruction problem, AMP-based algorithms are known to perform no worse than convex optimization methods, e.g., IHT, IST, and ADMM, at significantly faster speed (20; 21; 22).

## 1.3 Contributions and Organization of the Thesis

This thesis explores directions in extending the applications of compressive sensing both algorithmically and in practical hardware systems. Chapter 2 is dedicated to the algorithmic development of the novel AMP CS algorithm. It is first extended to incorporate additional side information optimally in a wireless communication channel estimation model. Then AMP is adapted to approximate an exemplary nonlinear function, a first application of AMP in the nonlinear function approximation domain.

Chapters 3, 4, and 5 provide a complete analysis of pixel diffraction in CS imaging applications, which is particularly prevalent in millimeter wave CS imaging systems. A dedicated analysis of such pixel diffraction is absent from previous millimeter wave imaging research, possibly because of the difficulties in isolating different distortion sources from the impact of pixel diffraction in a physical CS imaging system. Instead, our hypothetical slit experiment provides a simple yet insightful method for identifying the pixel diffraction limit, which is supported by full-scale electromagnetic (EM) simulations (Ansys HFSS) of the CS imaging experiment. The EM simulation approach, whose performance is cross-validated by rigorous slit diffraction theory, eliminates all undesired uncertainties in CS imaging experiments so that the impact of pixel diffraction on CS imaging performance may be evaluated independently.

In chapter 3, a new mechanism of performance degradation in CS imaging, namely pixel diffraction, is introduced. Explicitly, after incorporating diffraction into the standard CS signal model, the nonuniformity of diffraction effects due to pseudorandom coded aperture patterns becomes a limiting factor for image reconstruction quality. Through a hypothetical slit diffraction experiment, the pixel diffraction limit is determined for a particular CS imaging system arrangement. Lastly, the analysis further implies that the pixel diffraction effect may be removed by increasing the size of the receiver aperture in CS imaging.

In chapter 4, the Fraunhofer diffraction theory used in the slit diffraction experiment is replaced by rigorous diffraction theory. The rigorous diffraction model is critical because the Fraunhofer model is only an approximation of Kirchhoff or Rayleigh-Sommerfeld diffraction theory, whose assumptions are readily satisfied by conventional visible-spectrum applications but may not be met in the millimeter wave band (25). The analysis is a direct adaptation of (26). Last but not least, numerical results from the rigorous diffraction theory is used to validate the subsequent EM simulation study.

In chapter 5, the CS imaging results based on EM simulation validate the pixel diffraction limit proposed in chapters 3 and 4. Compared to physical experiments, electromagnetic (EM) simulations eliminate all unwanted imprecision and uncertainties, so that the reconstructed image quality degradation could be attributed solely to diffraction. Furthermore, we introduced the concept of equivalent SINR so that CS imaging performance under pixel diffraction may be evaluated without the highly computation-intensive full-scale EM simulation.

Chapter 6 concludes the dissertation and provides potential directions for future research.

# Chapter 2

# Algorithmic Development of AMP CS Algorithm

This chapter covers two projects in extending AMP CS algorithms for different signal models (27; 28). First, the standard AMP algorithms are summarized. The standard AMP algorithm (29) solves the linear inverse problem (1.1) by iteratively updating the estimates of the unknown input signal, with $x^t \in \mathbb{R}^N$ being the estimate at iteration $t$. The algorithm follows the three update equations each iteration, with $x^0 = \mathbf{0} \in \mathbb{R}^N$, a zero vector:

$$r^t = y - Ax^t + \frac{r^{t-1}}{\delta} \left\langle \eta'_{t-1}(x^{t-1} + A^T r^{t-1}) \right\rangle, \tag{2.1}$$

$$v^t = x^t + A^T r^t, \tag{2.2}$$

$$x^{t+1} = \eta_t(v^t). \tag{2.3}$$

Here $r^t \in \mathbb{R}^M$ is called residue, $\delta = \frac{M}{N}$ the measurement rate, and $\langle z \rangle = \frac{1}{N} \sum_{i=1}^N z_i$ refers to the ensemble average. In the Bayesian setting, in which prior knowledge of the unknown input is accessible as prior distributions, the update equation $\eta_t : \mathbb{R} \to \mathbb{R}$ is the element-wise denoiser function that reduces the error in estimates of $x$ at iteration $t$, and $\eta'_t(z) = \frac{\partial}{\partial z} \eta_t(z)$ is applied element-wise to each entry of vector $z$. For the common mean square error metric, the minimum mean square error (MMSE) denoiser is defined as

$$\eta_t(v^t) = \mathbb{E}[X|v^t], \tag{2.4}$$

where random variable $X$ follows the known prior distribution of input $x$, and pseudo-data $v^t$ is a realization from $V^t \sim X + \mathcal{N}\left(0, \sigma_t^2\right)$. The ”noise” variance $\sigma_t^2$ can be estimated using state evolution or empirically within the AMP algorithm following (20; 21; 22), and it decreases as $t$ increases until convergence. In standard AMP theory, the entries in the measurement matrix A are independent and identically distributed (i.i.d.) samples from $\mathcal{N}(0,1/M)$. When the measure matrix $A$ does not satisfy the condition, it is known that the AMP algorithm often becomes unstable. The divergence issue may be alleviated by damping (30) or other modifications to the AMP updates (31).

## 2.1 AMP with Side Information

This section is abridged from (32). I am the second author and my contribution lies in: i) the literature review on existing AMP literature with similar features; ii) performance testing of AMP-SI on the birth-death-drift model against a novel AMP algorithm that considers time dynamics; iii) evaluating birth-death-drift signal model in the channel estimation settings with Toeplitz measurement matrices and demonstrating the efficacy of AMP-SI for channel estimation tasks. Only materials related to my work are presented in this section.

### 2.1.1 Side Information

In practical applications, such as wireless communications, video decompression, etc., side information (SI) for the unknown input $x$ is often available in addition to measurements $y$. The side information on $x$, denoted as $\widetilde{x}$, can often be used to improve the quality of estimates. In this project, by observing the insight of Bayesian AMP, we replace (2.3) by

$$x^{t+1} = \eta_t(v^t, \widetilde{x}). \tag{2.5}$$

, where

$$\eta_t(v^t) = \mathbb{E}[X|v^t, \widetilde{x}]. \tag{2.6}$$

Given the probability density function (pdf) $f(x, \widetilde{x})$, the denoiser with side information (2.6) can be derived and incorporated into AMP iterations.

### 2.1.2 Prior Work

While integrating SI (or prior information) into signal recovery algorithms is not new (33; 34; 35; 36; 37; 38), the approach proposed in this work is a unified framework within AMP that supports arbitrary dependencies between the $(X_n, \widetilde{X}_n)_{n=1}^N$ pairs. Prior work using SI has been either heuristic, limited to specific applications, or outside the AMP framework. For example, Wang and Liang (33) proposed the Generalized Elastic Net Prior approach, which integrates SI into AMP for a specific signal prior density, but the method lacks Bayes optimality properties and is difficult to apply to other signal models. Our algorithmic framework overcomes these limitations through a generalized, Bayes optimal framework.

Ziniel and Schniter (1) developed an AMP-based signal recovery algorithm, namely DCS-AMP, for a time-varying signal model based on Markov processes for the support and amplitude. The Markov processes and corresponding dependencies between variables are captured by factor graph models. While our BDD model (details in Sec. 2.1.3) is closely related to their time-varying signal model, our emphasis is to introduce the AMP-SI framework and demonstrate how SI can be incorporated in AMP without needing to carefully craft factor graphs for every new signal model.

Manoel et al. implemented an AMP-based algorithm called MINI-AMP in which the input signal is repeatedly reconstructed in a streaming fashion, and information from past reconstruction attempts is aggregated into a prior, thus improving ongoing reconstruction results (38). Interestingly, the signal recovery approach of MINI-AMP resembles that of AMP-SI, in particular when Bernoulli-Gaussian signal model is used. Finally, Manoel et al. proved that MINI-AMP is MMSE-optimal (38).

### 2.1.3 Birth-Death-Drift Model

To demonstrate that AMP-SI is potentially useful for practical tasks, we investigate the application of AMP-SI to a stochastic signal model, namely the Birth-Death-Drift Model (BDD) model, which closely resembles the channel estimation problem in wireless communications. When the author joined this project, the conceptualization and theoretical components for applying AMP-SI to the proposed BDD model had completed. Based on the draft AMP-SI code, the author updated and completed the AMP-SI simulation experiments for the BDD model and benchmarked it against two novel AMP algorithms for similar tasks, the results of which are presented in the next section.

#### Connections to Channel Estimation

**BDD Motivation:** Our channel estimation scenario is illustrated in Fig. 2.1. Typical wireless devices transmit a pilot sequence and data payload in *batches*. In batch $b$, the pilot sequence $p$ is transmitted into the channel, where it is convolved with the channel response $x^b$, yielding noisy linear measurement,

$$y^b = \text{conv}(p, x^b) + z.$$

This convolution, $\text{conv}(\cdot,\cdot)$, can be expressed as the product of a Toeplitz matrix with a vector,

$$y^b = \text{Toeplitz}(p)x^b + z, \tag{2.7}$$

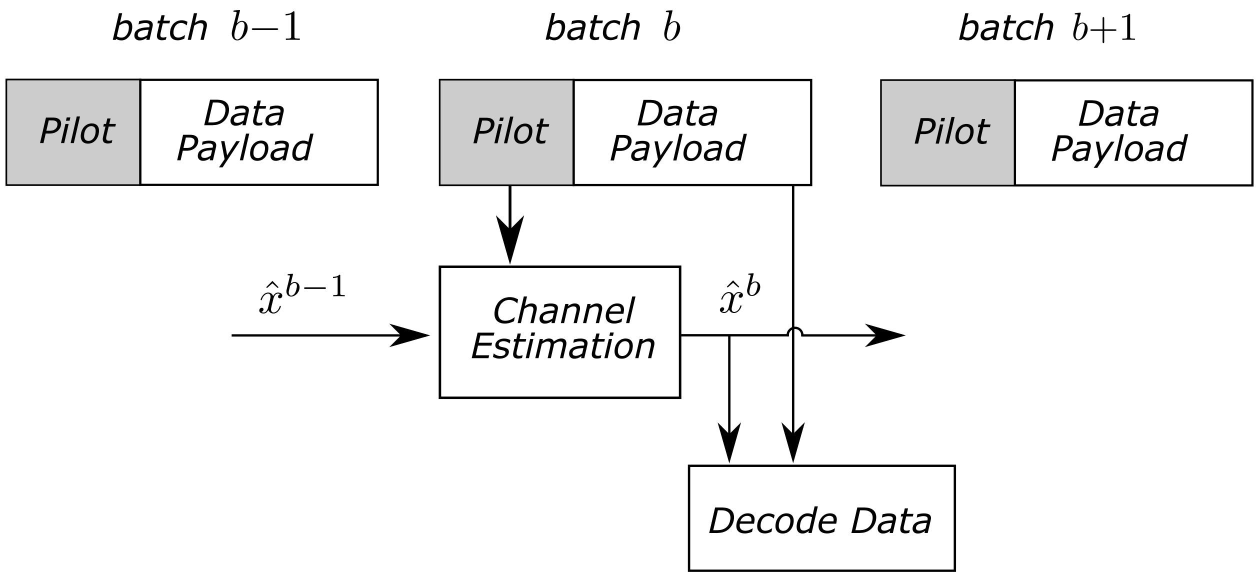



Figure 2.1: In batch $b$, the wireless device transmits a pilot and data payload. The channel filter $x^b$ is estimated using the channel response to the pilot together with SI $\widetilde{x} = \widehat{x}^{b-1}$, the channel filter estimated in the previous batch. The estimated $\widehat{x}^b$ is used to decode the data and as SI in the next batch to estimate $x^{b+1}$.

where $\text{Toeplitz}(p)$ is the Toeplitz matrix that corresponds to the pilot sequence $p$. To perform channel estimation using AMP-SI, we will consider (2.7) as a linear inverse problem (1.1), where $\text{Toeplitz}(p)$ is the measurement matrix. Our goal will be to estimate the channel response $x^b$ in batch $b$ using the noisy measurements $y^b$, matrix $\text{Toeplitz}(p)$, and $\widetilde{x} = \widehat{x}^{b-1}$, our estimate of the channel response in the previous batch, $b-1$ (Fig. 2.1). Our resulting estimate for the channel response, $\widehat{x}^b$, will then help us estimate the channel response in the next batch, $x^{b+1}$. To develop a conditional denoiser, we need a channel model that describes the channel response $x^b$, and especially its dependence on $x^{b-1}$, the channel response in the previous batch. We model the channel as an (unknown) finite impulse response (FIR) filter, whose taps correspond to the amplitude of the channel response at different delays. Many filter taps are close to zero, and this sparsity makes the channel estimation problem a sparse signal recovery task. Due to the slowly varying time dynamics of the channel, $x^b$ is not only sparse, but has strong dependencies with the channel response in adjacent batches. A possible model for changes from $x^b$ to $x^{b+1}$ involves ($i$) *birth* of new nonzeros in $x^{b+1}$ (corresponding to new wireless paths); ($ii$) *death* of nonzeros in $x^b$ that become zero in $x^{b+1}$ (existing paths are obscured as the user moves); and ($iii$) slow *drift* of existing nonzeros. We call these time-varying channel dynamics a *birth-death-drift* (BDD) model.

**Formal definition of the BDD model:** The BDD model may be considered as a discrete-time Markov chain, in which the signal evolves following one of the following four cases. For any entry $n \in \{1, 2, \ldots, N\}$,

**Case 1:** Zero entry remains zero, i.e., $[x_p]_n = 0$ and $[x_c]_n = 0$

**Case 2:** *Death* – nonzero entry becomes zero, i.e., $[x_p]_n \sim \mathcal{N}(0, \sigma_s^2)$ and $[x_c]_n = 0$

**Case 3:** *Drift* – nonzero entry remains nonzero, i.e., $[x_p]_n \sim \mathcal{N}(0, \sigma_s^2)$ and $[x_c]_n = \rho[x_p]_n + \mathcal{N}(0, \sigma^2)$

**Case 4:** *Birth* – zero entry becomes nonzero, i.e., $[x_p]_n = 0$ and $[x_c]_n \sim \mathcal{N}(0, \sigma_s^2)$.

The four cases fully determine the pdf $f(X_p, X_c)$, where $x_p$ denotes the previous signal, the estimate of which acts as the side information $\widetilde{x}$ for estimating the current signal denoted by $x_c$. By choosing $\rho > 0$ such that $\rho^2\sigma_s^2 + \sigma^2 = \sigma_s^2$ , we ensure $var(X_c) = \sigma_s^2$ so that the signal energy remains constant over batches. Within the channel vector $x_c$ to be estimated, the entries belong to Case k with probability $\varepsilon_k$, hence $\sum_{k=1}^4 \varepsilon_k = 1$. The mathematical derivation of the explicit expression of the denoiser is not included as it is not required for the subsequent sections and it is not done by the author.

### 2.1.4 Numerical Results for BDD

Fig. 2.2 presents the experimental result of recovering a signal $x_c$ over 15 time batches following the BDD model in Sec. 2.1.3 using AMP, AMP-SI, and DCS-AMP, another AMP-based algorithm for time-varying signals (1). In each time batch, the SI is the pseudo-data output of AMP-SI in the previous batch, except for Batch 1 where SI is unavailable and we resort to standard AMP. For DCS-AMP, we implement the algorithm in filtering mode to match our SI setting. The signal $x_c$ is of dimension $N = 10000$, the steady-state standard deviation is $\sigma_s = 1$, the correlation among nonzero entries is $\rho = 0.95$, and the measurement noise has standard deviation $\sigma_z = 0.077$, which corresponds to SNR $= 20$dB. The empirical MSE is averaged over 100 trials. For each batch, AMP-based approaches often converge in 10–20 iterations, and 30 iterations are used to be safe. We set $\varepsilon_1 = 0.80$, $\varepsilon_2 = \varepsilon_4 = 0.01$, and $\varepsilon_3 = 0.18$ so that there are approximately $K = N(\varepsilon_3 + \varepsilon_4) = 1900$ nonzero entries per signal. The measurement matrix has i.i.d. standard Gaussian entries in each batch, and the number of measurements is $M = 3000$. It can be seen that AMP-SI outperforms DCS-AMP in every batch (except Batch 1).

Inspired by (1), we investigated how transition probabilities between different BDD states affected the performance of AMP-SI. Within the problem setting of Fig. 2.2, we varied $\varepsilon_2$, $\varepsilon_4 \in [0.01, 0.11]$, and $\rho \in [0.05, 0.999]$. We found that AMP-SI out-performed AMP at Batch 15 for all configurations. The performance gap was largest for large $\rho$ and small $\varepsilon_2$ and $\varepsilon_4$; the gap narrowed

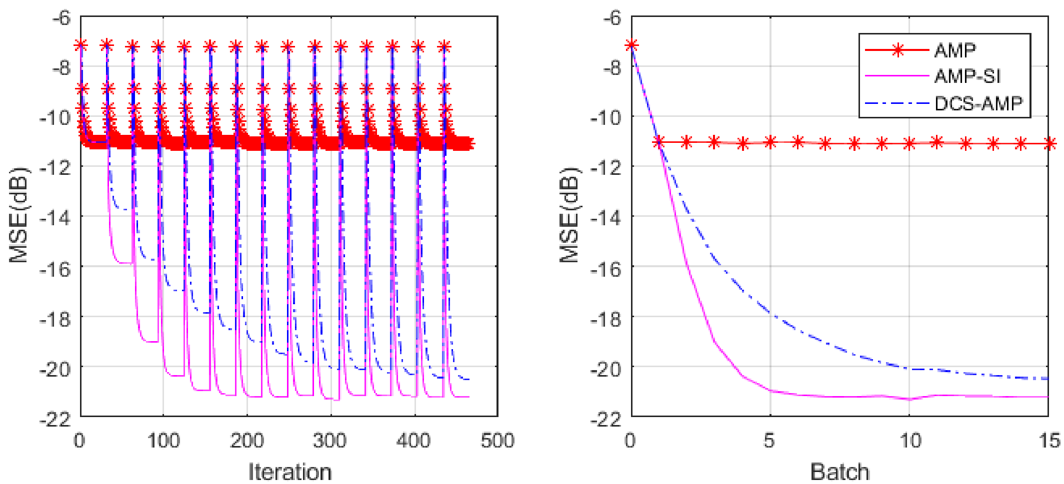



Figure 2.2: Empirical performance of DCS-AMP in filter mode (1), AMP, and AMP-SI as a function of (left) iterations and (right) batches, where the algorithms each spend 30 iterations per batch. (BDD signal, $N = 10000$, $M = 3000$, $\sigma_s = 1$, $\rho = 0.95$, $\sigma_z = 0.077$, $\varepsilon_1 = 0.80$, $\varepsilon_2 = \varepsilon_4 = 0.01$, $\varepsilon_3 = 0.18$. The tuning parameters for DCS-AMP are determined by converting BDD parameters to corresponding Gauss-Markov parameters (1).)

as $\rho$ decreased and $\varepsilon_2$ and $\varepsilon_4$ increased. We also experimented with holding $\varepsilon_2 = \varepsilon_4 = 0.05$ constant and varied $\varepsilon_3$ and $\rho$. Again, as $\rho$ decreased, the gap between AMP-SI and AMP shrank.

That said, for large $\varepsilon_3$, AMP-SI still improved reconstruction quality for small $\rho$.

**Channel estimation with Toeplitz matrices:** So far we used i.i.d. Gaussian matrices, and we now transition to Toeplitz matrices in order to demonstrate that AMP-SI is suitable for channel estimation (details in Sec. 2.1.3). Based on (2.7), the channel estimation problem deviates from the BDD model in three aspects. First, as mentioned, $A$ is Toeplitz rather than i.i.d. Gaussian. It is well known that for non-i.i.d. measurement matrices, the standard AMP prescribed by (2.1), (2.2), and (2.3) often suffers from divergence over iterations. A common approach to improve convergence of iterative algorithms is damping; in AMP, the standard iteration (2.2) is replaced by $x^{t+1} = \lambda x^t + (1-\lambda)\eta_t(x^t + A^T r^t)$. Rangan et al. (39) demonstrate that damping is effective in aiding the convergence of AMP for some non-i.i.d. measurement matrices. It should be noted that supporting theory for AMP-based algorithms is only rigorous for certain classes of random matrices (40; 41), which exclude Toeplitz matrices. Thus we evaluate the performance of AMP-SI for Toeplitz matrices by comparing empirical reconstruction results to standard AMP-SI settings. Lastly, for a pilot sequence $p$, the number of rows of the measurement matrix, $M$, equals $\text{length}(p) + N - 1$, which typically exceeds $N$, the number of columns. This inverse problem is expansive ($M > N$) instead of compressive ($M < N$), where we remind the reader that AMP theory supports arbitrary $\delta > 0$ where $\delta = \frac{M}{N}$.

Our experiment had 5 time batches. We set the length of the channel response $N$ to 4000, the length of the pilot sequence $\text{length}(p) = 1001$, the standard deviation of the steady signal $\sigma_s = 1$, the decay rate of nonzeros $\rho = 0.95$, and the measurement noise standard deviation $\sigma_z \in \{0.01, 0.1, 1\}$. This setting corresponds to SNR= 0dB, 20dB and 40dB, and $\delta = 1.25$. For BDD model parameters, we set $\varepsilon_1 = 0.78$, $\varepsilon_2 = \varepsilon_4 = 0.01$. Thus at each time batch, 21% of the entries of the channel response are nonzero. The individual entries of the pilot $p$ are $\pm 1/\sqrt{\text{length}(p)} = \pm 0.0316$, each with probability 0.5. We performed damping using parameter $\lambda = 0.9$. Table 2.1 demonstrates the empirical channel estimation performance of AMP-SI averaged over 50 realizations. Compared to standard AMP (batch 1 in Table 2.1), AMP-SI consistently achieves lower MSE levels starting from batch 2. One striking observation from Fig. 2.3 is the similar performance of AMP-SI for Toeplitz (channel estimation) and i.i.d. matrices.

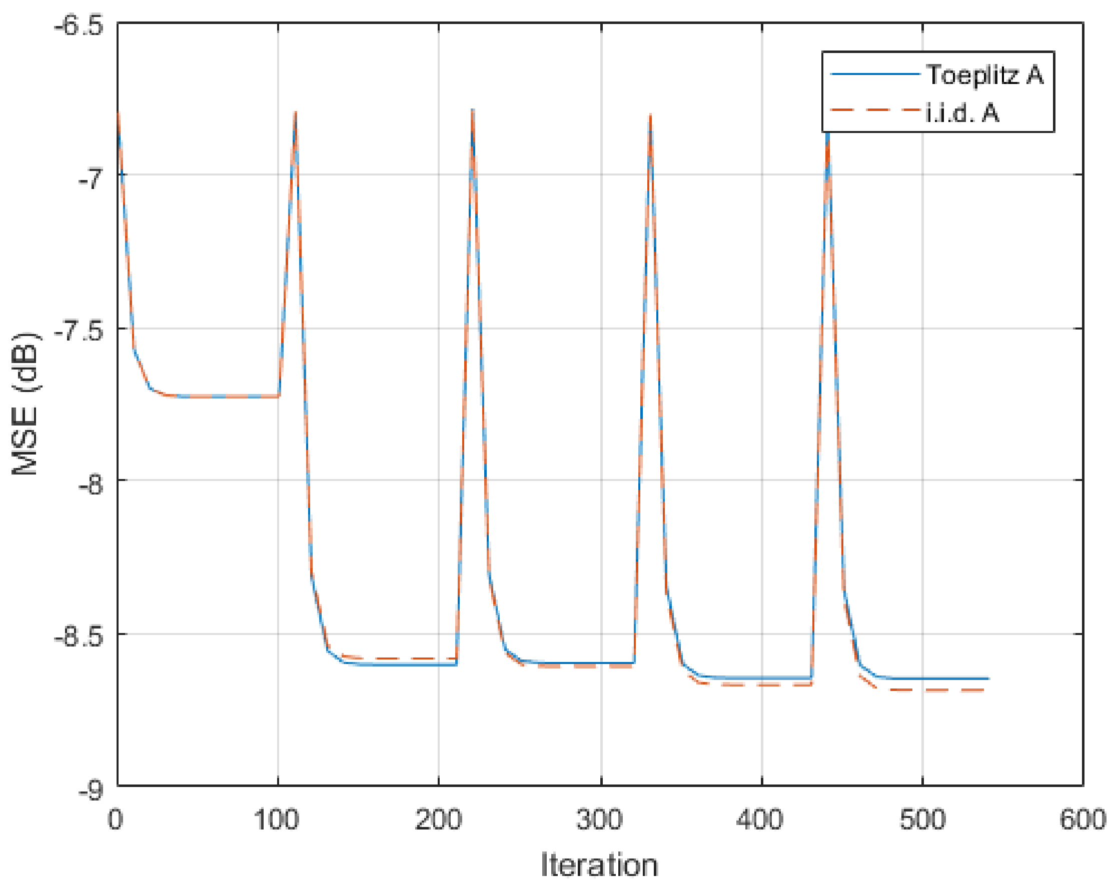



Figure 2.3: Empirical AMP-SI performance with i.i.d. and Toeplitz measurement matrices. (BDD signal, SNR=0dB, averaged over 100 realizations.)

Table 2.1: Empirical AMP-SI performance for channel estimation. (BDD signal, averaged over 50 realizations.)

| | **Channel Estimation MSE(dB)** | | |
|---|---|---|---|
| **SNR** | **Batch 1** | **Batch 2** | **Batch 5** |
| 0dB | -7.71 | -8.61 | -8.68 |
| 20dB | -23.49 | -24.74 | -24.86 |
| 40dB | -45.41 | -45.83 | -45.82 |

Table 2.2: Empirical AMP-SI performance for i.i.d. matrices, SE predictions, and empirical performance for Toeplitz matrices. (BDD signal, Batch 5, averaged over 50 realizations.)

| | **AMP-SI Performance in MSE(dB) at Time Batch 5** | | |
|---|---|---|---|
| **SNR** | **i.i.d. A** | **SE Prediction** | **Toeplitz A** |
| 0dB | -8.68 | -8.64 | -8.65 |
| 20dB | -24.86 | -24.86 | -24.77 |
| 40dB | -45.82 | -45.91 | -45.86 |

### 2.1.5 Summary of Work in AMP-SI Project

This work exposes me to AMP CS algorithms. In this project, I empirically demonstrated AMP-SI's efficacy against a novel AMP-based algorithm for similar problem settings. Furthermore, I implemented simulations for applying AMP-SI to a channel estimation task using the BDD model. Noting that channel estimation is a deconvolution task where the structural measurement matrices do not conform to AMP theory, the simulations empirically extend our knowledge in AMP by illustrating that AMP-SI's performance is not deteriorated by structured measurement matrices for deconvolution.

## 2.2 Nonlinear Function Estimation with Empirical Bayes and Approximate Message Passing

The remaining of the chapter is an adaptation of a conference paper (28). Liu and I contributed equally to the work. My contribution lies in: i). initial empirical testings of the convergence property of the non-standard structured measurement matrix using the vector AMP (VAMP) algorithm (42) and standard AMP with damping (43); ii). the formulation of the empirical Bayesian nonlinear function approximation problem; iii). benchmarking of the empirical Bayesian AMP algorithm against the least absolute shrinkage and selection operator (LASSO) (44; 45) for a nonlinear function approximation task. As an exploration project, the above-mentioned empirical Bayesian component discriminates the work from the standard Bayesian CS problem.

### 2.2.1 Introduction

A pervasive trend in modern society is that ever-larger amounts of data are being collected and analyzed in order to explain various phenomena. In supervised learning, many variables (also referred to as features) that may relate to and thus help explain the phenomena of interest are observed, and the goal is to learn a nonlinear function that relates the explanatory variables to the phenomena of interest. More specifically, we have a multivariate nonlinear function, $\mathbf{f}(\cdot)$, and collect noisy samples of it; our goal is to estimate $\mathbf{f}(\cdot)$. At its core, this is *multivariate nonlinear function estimation*; it could also be interpreted as nonlinear regression or feature selection. Algorithms for solving such problems must be robust to noisy observations and outliers, backed up by fundamental mathematical analysis, support missing data, and have a fast implementation that scales well to large-scale problems. Such algorithms will impact many disciplines, such as health informatics (46), social networks, and finance (47; 48; 49).

**Example applications:** Let us describe how nonlinear function estimation can be used in financial prediction (47; 48; 49). A typical approach to estimate expected returns uses a linear factor model, which is tuned to work well on training data, $y_m = \sum_n \mathbf{X}_{mn}\theta_n + z_m$, where $y_m$ is the price change of asset $m$, $\mathbf{X}_{mn}$ is the exposure of asset $m$ to factor $n$, $\theta_n$ are the returns of factor $n$, and $z_m$ is noise in asset $m$. We can express the linear model in matrix vector form, $\mathbf{y} = \mathbf{X}\theta + \mathbf{z}$, where $\mathbf{X}$ is an input data matrix, by assigning $y_m$ as the $m$-th entry of the vector $\mathbf{y}$, $\theta_n$ as the $n$-th entry of the vector $\theta$, and $\mathbf{X}_{mn}$ as the element of the matrix $\mathbf{X}$ in row $m$ and column $n$ The goal is to estimate $\theta$ from $\mathbf{y}$, $\mathbf{X}$, and possible statistical knowledge about $\theta$ and $\mathbf{z}$. We can see that financial prediction based on linear models relies on solving linear inverse problems. That said, some factors relate to returns in a nonlinear way (50), and financial prediction could be improved using nonlinear schemes.

Nonlinear modeling can also be used in health informatics (46), where $\mathbf{y}$ could measure patients' medical condition, $\mathbf{X}$ contains nonlinear exposure terms, and $\theta$ are explanatory variables that drive the patients' condition. Our goal is to understand the relationships between explanatory variables and patients' medical condition.

**Main idea and contributions:** In this project, as a first step toward learning nonlinear functions, we cast them as linear inverse problems using *polynomial kernels* (51; 52) (Sec. 2.2.3). Incorporating kernels into the matrix maps the nonlinear signal estimation procedure into a linear inverse problem but with an increased feature set, where the features are no longer *independent and identically distributed* (i.i.d.). Unfortunately, the kernels may create poorly conditioned matrices, and many solvers for linear inverse problems struggle with such matrices. Nonetheless, the matrices in our linear inverse problems often contain only mild linear correlations among columns and are reasonably well conditioned.

While polynomial kernels greatly increase the richness of the model class that captures the phenomena of interest, they also significantly increase the dimensionality of the features. For example,

$N$ factors evaluated with quadratic kernels will become approximately $\frac{1}{2}N^2$ new factors. This large scale and well-conditioned linear inverse problem is well suited to *approximate message passing* (AMP) (23; 53), an algorithmic framework for signal reconstruction that is asymptotically optimal for large scale linear inverse problems in the sense that it achieves the best possible reconstruction quality (54; 55). Our AMP-based algorithms improve the reconstruction quality of the coefficient vector $\theta$, leading to better estimation of the nonlinear function.

Two AMP-based approaches are considered. The first follows a *Bayesian framework*, where we assume that the coefficient vector follows some known probabilistic structure. While the Bayesian framework is naive and limited in scope, our past work has shown that universal approaches that adapt to unknown statistical distributions can be integrated within solvers for linear inverse problems (56; 57; 58), thus bypassing the Bayesian limitation. The linear inverse problems resulting from our polynomial kernel expansion are solved using an AMP-based algorithm, whose Bayes optimality ensures that our function estimation procedure can succeed despite using fewer and noisier samples than other methods.

The second AMP-based approach uses *empirical Bayes* (59), where the coefficient vector is assumed to follow some parametric distribution, and in each iteration of AMP we plug maximum likelihood parameter estimates into a parametric Bayesian denoiser.

The resulting algorithms will allow data to better model dependencies between explanatory variables and phenomena of interest. These algorithms could also help reconstruct signals acquired by nonlinear analog systems, allowing hardware designers to exploit nonlinearities rather than avoid them.

**Organization:** The rest of Sec. 2.2 is organized as follows. Sec. 2.2.2 provides background content. Details of our approach for estimating multivariate nonlinear functions appear in Sec. 2.2.3. Numerical results appear in Sec. 2.2.4, and Sec. 2.2.5 concludes the project.

## 2.2.2 Background

### Inverse problems

We present a flexible formulation for nonlinear function estimation in the form of a nonlinear *inverse problem*. We observe $M$ independent samples of the form $\{(\mathbf{x}_m, y_m)\}_{m\in\{1,\ldots,M\}}$, where $(\mathbf{x}_m, y_m) \in \mathbb{R}^N \times \mathbb{R}$, through a nonlinear function $f(\cdot)$ and additive noise,

$$y_m = f(\mathbf{x}_m) + z_m, \tag{2.8}$$

for all $m \in \{1,\ldots,M\}$. In other words, the input data matrix $\mathbf{X} \in \mathbb{R}^{M\times N}$, where $\mathbf{X}$ are locations of samples, will be processed by applying a multivariate operator, $\mathbf{f}(\cdot) : \mathbb{R}^M \times N \to \mathbb{R}^M$, such that $\mathbf{f}$

applies $f$ on each individual row of the data matrix $\mathbf{x}$, with additive noise, $\mathbf{z} \in \mathbb{R}^M$, resulting in noisy measurements,

$$\mathbf{y} = \mathbf{f}(\mathbf{X}) + \mathbf{z} \in \mathbb{R}^M. \tag{2.9}$$

While the reader is likely familiar with linear inverse problems, where the operator $\mathbf{f}$ boils down to multiplication by a coefficient vector $\theta$, i.e., $\mathbf{f}(\mathbf{X}) = \mathbf{X}\theta$, our main interest is in nonlinear inverse problems.

We highlight that many "rules of thumb" that the sparse signal processing community has claimed, for example that sparse signals can be reconstructed from a small number of linear measurements, $M < N$, may break down when the measurement noise $\mathbf{z}$ is large or the operator $\mathbf{f}(\cdot)$ contains significant nonlinearities.

### Approximate message passing (AMP)

One approach for solving linear inverse problems is AMP (23; 53), which is an iterative algorithm that successively converts the matrix problem to scalar channel denoising problems with *additive white Gaussian noise* (AWGN). AMP is a fast approximation to precise message passing (cf. Baron et al. (60), Montanari (53), and references therein), and has received considerable attention because of its fast convergence and the *state evolution* (SE) formalism (23; 61; 53), which characterizes how the *mean squared error* (MSE) achieved by the next iteration of AMP can be predicted using the MSE performance of the denoiser being used. AMP solves the following linear inverse problem,

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{z}, \tag{2.10}$$

where the empirical *probability density function* (pdf) of $\theta$ follows $p_\theta(\theta)$, the operator $\mathbf{f}(\mathbf{X})$ multiplies $\mathbf{X}$ by the unknown coefficient vector $\theta$, and $\mathbf{z}$ is AWGN with variance $\sigma_Z^2$. Although the AMP literature mainly considers i.i.d. Gaussian matrices, approaches such as damping (62) and swept AMP (63) have been proposed to deal with more general matrices. After initializing $\theta^0$ and $\mathbf{r}^0$, AMP (23; 53) proceeds iteratively according to

$$\begin{aligned} \theta^{t+1} &= \eta^t(\mathbf{X}^T\mathbf{r}^t + \theta^t), \\ \mathbf{r}^t &= \mathbf{y} - \mathbf{X}\theta^t + \frac{1}{R}\mathbf{r}^{t-1}\langle \eta^{t-1'}(\mathbf{X}^T\mathbf{r}^{t-1} + \theta^{t-1})\rangle, \end{aligned} \tag{2.11}$$

where $(\cdot)^T$ denotes the transpose, $R = M/N$ is the *measurement rate*, $\eta^t(\cdot)$ is a *denoising function*, and $\langle \mathbf{u} \rangle = \frac{1}{N}\sum_{i=1}^N u_i$ for some vector $\mathbf{u} \in \mathbb{R}^N$. The denoising function $\eta^t(\cdot)$ operates in a symbol-by-symbol manner (also known as *separable*) in the original derivation of AMP (23; 53). That is, $\eta^t(\mathbf{u}) = (\eta^t(u_1), \eta^t(u_2), ..., \eta^t(u_N))$ and $\eta^{t'}(\mathbf{u}) = (\eta^{t'}(u_1), \eta^{t'}(u_2), ..., \eta^{t'}(u_N))$, where $\eta^{t'}(\cdot)$ denotes the derivative of $\eta^t(\cdot)$.

A useful property of AMP in the *large system limit* ($N, M \to \infty$ with the measurement rate $R$ constant) is that at each iteration, the vector $\mathbf{X}^T\mathbf{r}^t + \theta^t \in \mathbb{R}^N$ in (2.11) is equivalent to the unknown coefficient vector $\theta$ corrupted by AWGN. This property is based on the decoupling principle (64; 54; 65), which states that the posterior of a linear inverse problem (2.10) is statistically equivalent to a scalar channel. We denote the equivalent scalar channel at iteration $t$ by

$$\mathbf{q}^t = \mathbf{X}^T\mathbf{r}^t + \theta^t = \theta + \mathbf{v}^t, \tag{2.12}$$

where $v_i^t \sim \mathcal{N}(0, \sigma_t^2)$, and $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian pdf with mean $\mu$ and variance $\sigma^2$. AMP with separable denoisers, which are optimal for i.i.d. signals, has been rigorously proved to obey SE (61). However, we will see in Sec. 2.2.2 that non-i.i.d. signals can be denoised better using non-separable denoisers.

Another useful property of AMP in the large system limit involves a Bayesian setting where a prior distribution for the coefficient vector $\theta$ is available. In such Bayesian settings, AMP can use denoiser functions $\eta^t(\cdot)$ that minimize the MSE in each iteration $t$ (61). Using such MSE-optimal denoisers, the MSE performance of AMP (2.11) approaches the *minimum mean squared error* (MMSE) as $t$ increases.

**Non-scalar denoisers**

While i.i.d. signals can be denoised in a scalar separable fashion within AMP, where each signal entry is denoised using the same scalar denoiser, real-world signals often contain dependencies between signal entries. For example, adjacent pixels in images are often similar in value, and scalar separable denoisers ignore these dependencies. Therefore, we apply non-separable denoisers to process non-i.i.d. signals within AMP. For example, if $\theta$ is a time series containing dependencies between adjacent entries, then we can use a sliding window denoiser that processes entry $n$ of $\theta$ using information from its neighbors (66; 58).

We will see in Sec. 2.2.3 that our signal reconstruction problem includes several types of coefficients in $\theta$, and we expect dependencies between coefficients. Therefore, non-scalar denoisers will be used within AMP to process non-i.i.d. coefficients.

### 2.2.3 Learning nonlinear functions

Having reviewed the relevant background material, we now recast nonlinear inverse problems (2.9) as linear inverse problems (2.10) using polynomial kernels (51; 52), which replace our input data matrix $\mathbf{X}$ with transformations of $\mathbf{X}$ (67).

Our nonlinear model (2.9) is motivated by the inadequacy of linear relationships in some applications. One example involves bio-informatics, where genetic factors involve multiplicative interactions among genes (68). Another application involving financial prediction (47; 48; 49), where the research and development expenditures of a firm correlate with future returns in a nonlinear way (50). Similar ideas have been widely used in the machine learning community in the context of polynomial kernel learning (51; 52), and the kernel trick has been introduced to linear inverse problems by Qi and Hughes (69). A related model that learns interactions among variables is the multi-linear model (70), where columns that involve auto-interaction are removed from the polynomial model.

### Basis expansion

Recall that in our inverse problem, $\mathbf{y} = \mathbf{f}(\mathbf{X}) + \mathbf{z}$, we define measurement $m \in \{1, \ldots, M\}$ as $y_m = f(\mathbf{x}_m) + z_m$ (2.8). Linear inverse problems make use of models that are linear in the input factors; they are mathematically and algorithmically tractable and can be interpreted as a first-order Taylor approximation to $f(\mathbf{x})$ (67). However, in many applications, the true function $f(\mathbf{x})$ is far from linear in $\mathbf{x}$.

A basis function expansion replaces $\mathbf{x}$ with transformations of $\mathbf{x}$ (67). For $\ell \in \{1, 2, \ldots, L\}$, $f(\mathbf{x})$ is expressed as in the linear basis expansion of $\mathbf{x}$:

$$f(\mathbf{x}) = \sum_{\ell=1}^{L} \theta_\ell g_\ell(\mathbf{x}).$$

This model is linear in the new variable $g_\ell(\mathbf{x})$, and $\theta_\ell$ are the coefficients. Basis expansions allow us to use a linear model to characterize and analyze nonlinear functions.

### Polynomial regression

We form a polynomial regression problem by applying a Taylor expansion to the multivariate nonlinear function $f(\cdot)$ (68). In polynomial regression, we add to the original columns of the input data matrix $\mathbf{X}_Q$, which represent individual explanatory variables, extra columns that represent interactions among variables.

Let us elaborate on the quadratic case. While we will provide details of a matrix $\mathbf{X}_Q$ that supports a quadratic Taylor expansion (2.14), the reader should be able to employ this concept for cubic expansions and beyond. For each measurement, we use a Taylor expansion of the $N$ factor variables:

$$y_m = \theta_1 + \sum_{n=1}^{N} [\theta_2]_n x_{mn} + \sum_{n=1}^{N} [\theta_3]_n x_{mn}^2 \\ + \sum_{n_1=1}^{N} \sum_{n_2=n_1+1}^{N} [\theta_4]_n x_{mn_1} x_{mn_2}, \tag{2.13}$$

where $\theta_1$ is a constant, $\theta_2, \theta_3 \in \mathbb{R}^N$ are coefficient vectors for linear and quadratic terms, respectively, $\theta_4 \in \mathbb{R}^{\frac{N(N-1)}{2}}$ is a coefficient vector for cross terms, and the subscript $n$ in $[\theta_4]_n$ depends on $n_1$ and $n_2$.

Our quadratic Taylor approximation is a basis expansion, where we have chosen $g(\mathbf{x})$ as follows: (*i*) $g(\mathbf{x}) = 1$ corresponds to a DC constant (*ii*) $N$ linear terms corresponding to the original data, $g(\mathbf{x}) = x_n$, $n \in \{1, \ldots, N\}$; (*iii*) $N$ quadratic terms corresponding to squares of individual linear terms, $g(\mathbf{x}) = (x_n)^2$; and (*iii*) $\frac{N(N-1)}{2}$ cross terms corresponding to products of pairs of linear terms, $g(\mathbf{x}) = x_{n_1} x_{n_2}$, where $n_2 > n_1$, $n_1, n_2 \in \{1, \ldots, N\}$. We assume that the features matrix, $\mathbf{X}$, is i.i.d. zero mean Gaussian for ease of analysis; different types of $\mathbf{X}$ are left for future work.

The polynomial regression model is formulated as a linear inverse problem (2.10) in matrix vector form,

$$\mathbf{y} = \mathbf{X}_Q \theta + \mathbf{z} = \mathbf{X}_Q \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix} + \mathbf{z},$$

where $\theta \in \mathbb{R}^L$ is the coefficient vector, and $L$ is evaluated below (2.15). In our matrix $\mathbf{X}_Q$, each row is an instance or sample, and each column is an attribute or feature, i.e.,

$$\mathbf{X}_Q = \begin{bmatrix} 1 & x_{11} \cdots x_{1N} & x_{11}^2 \ldots x_{1N}^2 & x_{11}x_{12} \ldots x_{1(N-1)}x_{1N} \\ 1 & x_{21} \ldots x_{2N} & x_{21}^2 \ldots x_{2N}^2 & x_{21}x_{22} \ldots x_{2(N-1)}x_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{M1} \ldots x_{MN} & x_{M1}^2 \ldots x_{MN}^2 & x_{M1}x_{M2} \ldots x_{M(N-1)}x_{MN} \end{bmatrix}. \tag{2.14}$$

Our goal is to estimate the regression coefficients in the vector $\theta$ from $\mathbf{X}_Q$ and $\mathbf{y}$. The measurement matrix $\mathbf{X}_Q \in \mathbb{R}^{M \times L}$ will include one DC column, $N$ linear term columns, $N$ quadratics (squared columns), and $\frac{N(N-1)}{2}$ cross terms. This matrix has the form (2.14), and it can be seen that

$$L = 1 + 2N + \frac{N(N-1)}{2}. \tag{2.15}$$

To solve this linear inverse problem using an AMP-based approach, we normalize each column of $\mathbf{X}_Q$, $[\mathbf{X}_Q]_\ell$ to have unit norm, where $\ell \in \{1, 2, \ldots, L\}$, and denote this normalized matrix by $\mathbf{X}'_Q$,

$$\mathbf{y} = \mathbf{X}_Q \theta + \mathbf{z} = \mathbf{X}'_Q \theta' + \mathbf{z}, \tag{2.16}$$

where each entry of $[\mathbf{X}_Q]_\ell$ obeys

$$[\mathbf{X}'_Q]_{\ell m} = \frac{[\mathbf{X}_Q]_{\ell m}}{||[\mathbf{X}_Q]_\ell||_2},$$

and the regression coefficients satisfy

$$\theta'_\ell = \theta_\ell ||[\mathbf{X}_Q]_\ell||_2. \tag{2.17}$$

An SVD analysis was carried out on the structure of the measurement matrix $\theta'_\ell$ by Liu in the original conference paper (28) which is not related to my contribution in the remaining sections, and is thus skipped in the dissertation.

#### AMP-based algorithm

We solve our linear inverse problem (2.16) using AMP, where two points should be highlighted. First, our denoiser can incorporate Bayesian prior information. Specifically, we use conditional expectation denoisers that minimize the MSE (61). Second, owing to the structure of our matrix, various AMP variants that promote convergence can be used (63; 42; 62). That said, these variants all have their shortcomings, and possible divergence of AMP should be tracked carefully.

### 2.2.4 Numerical Results

Our construction of the quadratic polynomial regression model in Sec. 2.2.3 results in a linear inverse problem (2.10) whose solution forms an estimate of a multivariate nonlinear function (2.9) that relates the explanatory variables to the phenomena of interest. This resulting linear inverse problem will now be solved by two AMP-based approaches, Bayesian AMP and empirical Bayes.

#### Bayesian AMP

After I confirmed the convergence of various AMP-based algorithms for the proposed structured input data matrix $\mathbf{X}'_Q$, Liu continued to evaluate AMP's performance for Bayesian nonlinear function estimation against LASSO.

**Non-i.i.d. model for $\theta$:** Our Bayesian approach considers four groups of coefficients (2.13), where $\theta_1 \in \mathbb{R}$, $\theta_2, \theta_3 \in \mathbb{R}^N$, and $\theta_4 \in \mathbb{R}^{\frac{N(N-1)}{2}}$ are the DC, linear, quadratic, and cross-term coefficients, respectively. We modeled each individual entry among these $L$ coefficients as *Bernoulli*

*Gaussian* (BG), where the Bernoulli part is a probability $p$ that the entry is nonzero, in which case its distribution is zero mean Gaussian with some variance. To be specific, (*i*) our DC coefficient obeys $\theta_1 \sim \mathcal{N}(0,10)$, meaning that it is zero mean Gaussian with variance 10; (*ii*) each entry among the $N$ linear coefficient terms satisfies $[\theta_2]_n \sim 0.2\mathcal{N}(0,1) + 0.8\delta_0$, i.e., zero mean unit norm Gaussian with probability 0.2, else zero; (*iii*) the $N$ quadratic coefficient terms obey $[\theta_3]_n \sim 0.2\mathcal{N}(0,0.5) + 0.8\delta_0$; and (*iv*) for the $\frac{1}{2}N(N-1)$ cross terms, $[\theta_4]_n \sim 0.03\mathcal{N}(0,0.1) + 0.97\delta_0$. Although the four groups of coefficients have different distributions, all $L$ entries that follow this model are statistically independent.

**Baseline LASSO algorithm:** The baseline algorithm used to solve (2.16) is the *least absolute shrinkage and selection operator* (LASSO) (45), which minimizes the sum of squared errors subject to a constraint on the $\ell_1$ norm of the coefficients (45). In our polynomial model, the LASSO estimator $\widehat{\theta}$ is calculated in Lagrangian form:

$$\widehat{\theta} = \frac{1}{2}\underset{\theta}{\operatorname{argmin}}||\mathbf{y} - \mathbf{X}_Q\theta||_2^2 + \sum_{j=1}^{4} \lambda_j \|\theta_j\|_1, \tag{2.18}$$

where $\lambda_1,\ldots,\lambda_4$ are tuning parameters. In principle, we could perform grid search over all four parameters, $\lambda_1,\ldots,\lambda_4$, but it is computationally intractable. Therefore, we report the performance obtained by setting all parameters to be equal, which reduces the search space.

**AMP-based approach:** As a proof of concept, we have designed a denoiser specifically for our non-i.i.d. model. Because all $L$ entries that follow this model are statistically independent, we used $L$ scalar denoisers. However, because individual entries among our four groups of coefficients, $\theta_1 \in \mathbb{R}$, $\theta_2, \theta_3 \in \mathbb{R}^N$, and $\theta_4 \in \mathbb{R}^{\frac{N(N-1)}{2}}$ follow different distributions, four different scalar denoisers were used. Details of Bayesian denoisers for BG signals appear in (71).

**Signal generation:** We evaluate the performance of AMP in the Bayesian setting, which is a planted inference problem. The experiment allows us to validate the suitability of AMP for the quadratic basis, e.g., (2.14).

We generated the feature matrix, $\mathbf{X}$, as i.i.d. Gaussian with dimension $N = 100$. These linear terms were then transformed into quadratic form $\mathbf{X}'_Q$ with normalized columns (2.16). The number of columns in the normalized matrix was $L = 5151$ (2.15), and the number of rows $M = 5400$. Next, we created quadratic multivariate functions by generating $\theta$ vectors following our non-i.i.d. model. The expected energy of each group of coefficients satisfies $E_{DC} = 10$, $E_{linear} = 0.2 \times N = 20$, $E_{quadratic} = 0.2 \times N \times 0.5 = 10$, and $E_{cross} = 0.03 \times \frac{N^2-N}{2} \times 0.1 = 14.85$. Finally, the measurement noise $\mathbf{z}$ was AWGN with variance $\sigma_Z^2 = 0.004$.

**MSE performance:** Fig. 2.4 shows the MSE performance for estimated coefficients, $\theta$. We estimated the coefficients using LASSO, *swept AMP* (SwAMP) (63) and *vector AMP* (VAMP) (42). The left panel of the figure shows the MSE obtained when estimating the original coefficients $\theta$, where the estimator $\widehat{\theta}$ can be calculated using (2.19),

$$\widehat{\theta}_\ell = \frac{\widehat{\theta'}_\ell}{||[\mathbf{X}_Q]_\ell||_2}, \tag{2.19}$$

$l \in \{1,\ldots,L\}$, and $\widehat{\theta'}$ are estimated coefficients of $\theta'$. SwAMP and VAMP both converge well for normalized quadratic matrices. However, it can be seen in Fig. 2.4 that VAMP requires fewer than 100 iterations to converge; SwAMP requires a few hundred, and its individual iteration requires more computation than that of VAMP; our specific implementation of LASSO requires thousands of iterations. Because our AMP-based approaches are expected to be Bayes optimal while LASSO does not share these optimality properties, there is no surprise that AMP-based approaches obtain lower MSE.

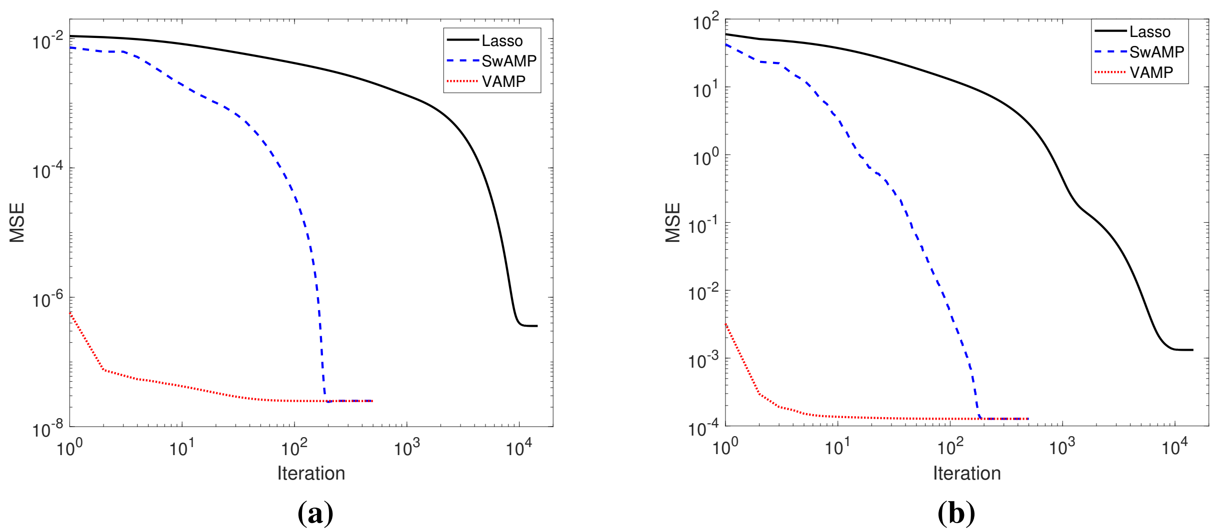



Figure 2.4: Performance of LASSO, SwAMP, and VAMP. The MSE is shown in the vertical axis, while the horizontal axis reflects the iteration number, $t$. Left panel **(a)**: MSE performance in recovering unknown coefficients, $\theta$. Right panel **(b)**: MSE performance in predicting the test data.

To ensure that our function reflects the nonlinear function well, the right panel of Fig. 2.4 shows the MSE obtained when applying our estimated polynomial function to predict test data,

$$\frac{||\mathbf{y}_{test} - \mathbf{X}_{test}\widehat{\theta}||_2^2}{K} = \frac{||\mathbf{X}_{test}(\theta - \widehat{\theta})||_2^2}{K},$$

where we held back $K = 600$ test measurements (recall that $M = 5400$), $\mathbf{X}_{test} \in \mathbb{R}^{K \times L}$ has the same format as $\mathbf{X}_Q$ and $\mathbf{y}_{test} \in \mathbb{R}^K$. Note that the MSE for coefficients, $\theta$, is inapplicable to real-world problems because the true coefficients do not exist, and we are merely modeling some nonlinear dependence as a low-order Taylor series. In our synthetic experiment, we are using the MSE over the test data as a metric of interest.

### Empirical Bayes

**Nonlinear function:** Nonlinear function learning is now performed using empirical Bayes within AMP (59). We employ the quadratic formulation (2.16) and learn the coefficient vector $\theta$ to approximate a family of (mildly) nonlinear functions,

$$y = \sum_{i=1}^{3} w_i \sin\left(X\rho_i + \phi_i\right) + \mathbf{z}, \tag{2.20}$$

where $w_1 = 0.1$, $w_2 = 0.3$, and $w_3 = 0.6$ are weights of the sinusoids, $\rho_i \in \mathbb{R}^N$ is a BG vector, $X\rho_i \in \mathbb{R}^M$, $\phi \in \mathbb{R}^M$ are phase shifts uniformly distributed between 0 and $2\pi$, the sine is applied element-by-element, and the noise $\mathbf{z} \in \mathbb{R}^M$ is AWGN with variance $10^{-4}$. Note that the vectors $\rho_i$ are chosen to be sparse BG, in order for the coefficient vector $\theta$ fit by AMP to the quadratic expansion to also be sparse, which allows us to utilize the rich knowledge available in CS literature on sparse signal recovery.

**AMP-based empirical Bayes:** In contrast to the Bayesian case, we assume that $\theta_2$, $\theta_3$, and $\theta_4$ are BG. The DC coefficient $\theta_1$ is assumed to be Gaussian. Formally,

$$\theta_1 \sim \mathcal{N}\left(0, \sigma_1^2\right), \theta_i \sim (1-\varepsilon)\,\delta_0 + \varepsilon(N)\left(0, \sigma_i^2 \mathbf{I}_{N_i}\right),$$

where $i \in \{2,3,4\}$, $N_2 = N_3 = N$, $N_4 = \frac{N(N-1)}{2}$. The parameters of these distributions are estimated using *maximum likelihood* (ML) and pseudo-data, $V^t \sim \theta' + \mathcal{N}\left(0, \sigma_t^2\right)$, at each iteration $t$. We note that $\theta$ and $\theta'$ satisfy (2.19), thus $V^t \sim \mathbf{a}^T\theta + \mathcal{N}\left(0, \sigma_t^2\right)$, where $\mathbf{a} \in \mathbb{R}^L$ and its $\ell_{\text{th}}$ element $a_\ell = ||[\mathbf{X}_Q]_\ell||_2$ for $l \in \{1, \ldots, L\}$. Then parameters for the distributions of $\theta$ may be estimated at each AMP iteration by maximizing the likelihood

$$L(\theta|\mathbf{v}^t) = \mathcal{N}\left(\mathbf{v}_1^t; 0, a_1^{-2}\sigma_1^2 + \sigma_{Z_1}^2\right) + \sum_{i=2}^{1+2N+\frac{(N-1)N}{2}} \varepsilon\mathcal{N}\left(\mathbf{v}_i^t; 0, a_i^{-2}\sigma_i^2 + \sigma_{Z_i}^2\right) + (1-\varepsilon)\,\mathcal{N}\left(\mathbf{v}_i^t; 0, \sigma_{Z_i}^2\right) \tag{2.21}$$

such that

Table 2.3: Empirical MSE on test data for nonlinear function estimation.

| **Measurement Rate** $R = \frac{M}{L}$ | **Median MSE over 20 Realizations** | | |
|---|---|---|---|
| | **LASSO** | **AMP** | **Pseudoinverse** |
| 0.14 | 0.0382 | 0.0293 | 0.041 |
| 0.28 | 0.0298 | 0.0228 | 0.033 |
| 0.56 | 0.0063 | 0.0036 | 0.01 |

$$\sigma_i := \begin{cases} \sigma_a, \text{ for} i = 1; \\ \sigma_b, \text{ for} i = [2, N+1]; \\ \sigma_c, \text{ for} i = [2+N, 2N+1]; \\ \sigma_d, \text{ for} i = \left[2+2N, \frac{(N-1)N}{2}\right]. \end{cases} \tag{2.22}$$

Then by substituting the maximum likelihood estimator

$$\widehat{\sigma}^t = \underset{\sigma_i,\ i \in \left[1, \frac{(N-1)N}{2}\right]}{\operatorname{argmax}} L\left(\theta | V^t\right) \tag{2.23}$$

back into Bayesian denoiser $\eta^t$ defined by (2.4) at every AMP iteration $t$, we obtain the empirical Bayesian AMP algorithm for approximating the specific nonlinear function. As an exploration project, the maximum likelihood problem may be solved by any efficient nonlinear solver packages.

**MSE performance:** We generated nonlinear functions and ran our empirical Bayes algorithm, LASSO, and a pseudoinverse approach (least squares). Each run of LASSO requires many iterations, and we use cross validation to regularize the parameter selection procedure. AMP with damping requires fewer iterations than LASSO. Empirical results for different measurement rates, $R = M/L$, appear in Table. 2.3. AMP obtains lower MSE than LASSO, which in turn obtains lower MSE than pseudoinverse.

### 2.2.5 Discussion

In this work, we study nonlinear function estimation, where a nonlinear function of interest is regressed on a set of features. We linearized the problem by considering low-order polynomial kernel expansion, and solved the resulting linear inverse problem using *approximate message passing* (AMP). Numerical results confirm that our AMP-based approaches learn the function better than the widely used *least absolute shrinkage and selection operator* (LASSO) (45), offering markedly lower error in predicting test data for both Bayesian and non-Bayesian settings.

While we have presented a first step toward approximating nonlinear functions by applying AMP to polynomial regression, many open problems remain for future work.

## 2.3 Summary

This chapter covers two projects that successfully extend AMP algorithmically for inverse problems. AMP has demonstrated superior efficacy in solving various generative problems, of which compressive sensing (CS) is a particular example. However, it is worth highlighting that AMP can be considered as a specific example of Bayesian single-layer neural networks, the drawback of which has been widely known in the machine learning community (72). When expanding AMP for solving general data science problems in high dimension, it cannot be emphasized enough that one should be aware of the generative nature assumed by AMP and the limitations inherited by such assumption.

# Chapter 3

# Diffraction in Millimeter Wave Single Pixel Compressive Imaging

The second half of the dissertation provides a complete analysis of the pixel diffraction limit for compressive imaging. This chapter provides background information on the problem, proposes a hypothetical slit diffraction example for studying pixel diffraction, and utilizes the Fraunhofer equation to identify a diffraction limit on the minimum coded aperture pixel size for the proposed compressive imaging hardware arrangement.

## 3.1 INTRODUCTION

Millimeter wave imaging has gained popularity in applications where optical imaging systems are ineffective due to occlusions, concealment, or poor lighting conditions (73; 74; 75; 76; 77; 78). In contrast to conventional X-ray technology (79; 80), the non-ionizing property of millimeter wave makes it particularly effective in applications that involve body scanning, for example in security screening (76; 77; 78; 73). Although state-of-the-art commercial millimeter wave imaging systems enjoy high image resolution and 3D focusing flexibility, these nice features are realized through holographic imaging theory (76; 77), whose sophisticated signal modulation process results in bulky and expensive final products. Unlike imaging systems that actively utilize phase and frequency information from sources (active millimeter wave imaging systems), the imaging approach relying solely on radiation intensities, known as passive millimeter wave imaging systems, offers an alternative solution due to its simplicity (78; 73). To further address the lack of high density millimeter wave sensor arrays, the single pixel camera (SPC) method, originated from compressive sensing theory and intended for the visible spectrum (81), is often adopted(82; 83; 84; 85; 86). This work focuses on the implementation aspects of SPC on millimeter wave imaging, in an effort to abridge the the knowledge gap between SPC literature and diffraction properties of millimeter wave signals. First, we briefly cover the background in compressive image sensing in order for readers to better appreciate the topic.

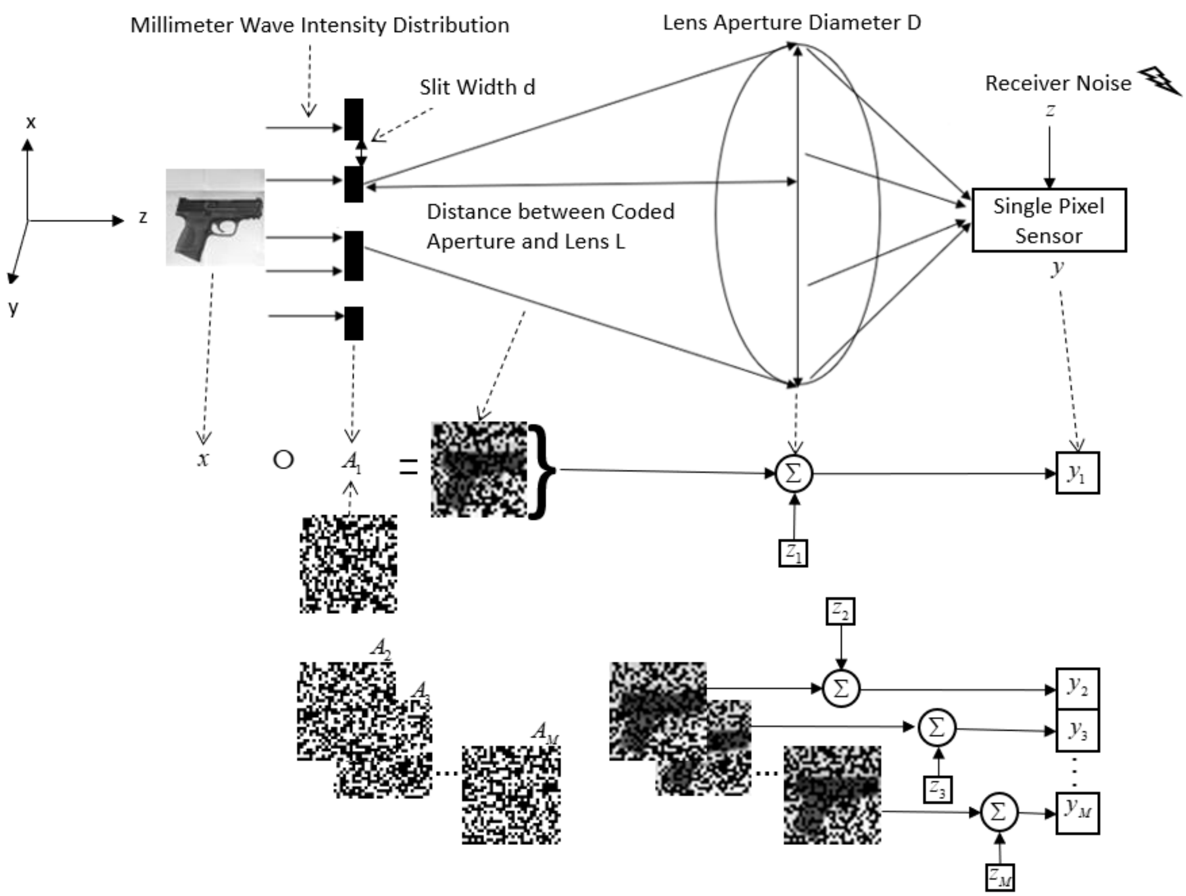



Figure 3.1: SPC data acquisition process and hardware schematic diagram. The symbol ∘ represents Hadamard products. The coded aperture for the $m_{th}$ measurement $y_m$ is denoted by $A_m$, which also corresponds to the $m_{th}$ row in the measurement matrix $A$ in (3.1). The summations over both the horizontal and vertical axes are denoted by $\Sigma$.

### 3.1.1 COMPRESSIVE IMAGE SENSING

Compressive image sensing refers to a class of image signal acquisition systems that recover unknown images from far fewer measurements than the number of pixels in the image (81). Fig. 3.1 illustrates the basic concept of compressive image sensing, as well as the single pixel camera (SPC) hardware arrangement which is commonly used to realize the model. The signal, either from the visible spectrum or from a millimeter wave source, is reflected or obscured by the object. The energy is then modulated through a pseudorandom mask, which contains $N$ locations, or pixels, which can block or transmit electromagnetic energy. The transmitted energy is then collected by a lens of diameter $D$ and measured by the power sensor (Single Pixel Sensor in the figure) to generate one single pixel measurement. For non-compressive single pixel imaging, one could take $N$ masks $[A_1..A_N]$, each of which contains only one transparent pixel, and $N$ power readings $[y_1..y_N]$ to straightforwardly reconstruct the full image. This is denominated as the raster scan method in the following and is used as the ground truth for system performance evaluations. The benefit of the single pixel method over a standard multi-pixel camera is the need for only a single power sensing device. The downside is that one needs $N$ masks and measurements. Alternatively, we leverage the sparsity of the image data (after appropriate transformation) and use compressive sensing (CS) algorithms to reduce the number of measurements needed by a factor of 2-10.

**Compressive sensing:** Mathematically, the CS algorithm resolves the problem of

$$y = Ax + z, \tag{3.1}$$

where the noisy measurement vector $y \in \mathbb{R}^M$ is generated by the matrix-vector product of a known sensing matrix $A \in \mathbb{R}^{M \times N}$, $M \ll N$, and the unknown sparse input signal $x \in \mathbb{R}^N$; interference and measurement noise are denoted by $z \in \mathbb{R}^M$. Compressive sensing algorithms are particularly effective when the energy of the unknown signal $x$, after certain transforms is dominated by a few large coefficients while the majority of the signal's coefficients remain small, which is known as the approximately sparse condition (87). The wavelet transform (88) has been shown to produce sparse representations for image signals. For the detection matrices $A$, their entries are sampled independently of $\{0,1\}$ with equal probability. This choice of entry values corresponds to coded apertures consisting of completely transparent and opaque pixels; 0-1 random code aperture patterns not only reduce fabrication complexity, but also guarantee competitive performance from CS theoretical perspectives (87; 24). For imaging applications, $x$ in (3.1) and Fig. 3.1 corresponds to the 2D image intensity distribution vectorized into a single column; and the $m_{th}$ measurement, $y_m$, where $m \in \{1,...,M\}$, is obtained by the inner product of $x$ and the corresponding coded aperture pattern, also vectorized into a column vector. And hence the coded aperture pattern used for measurement $y_m$ corresponds to the $m_{th}$ row of the sensing matrix $A$.

In this work, we investigate the mechanism of how and to what extent diffraction degrades millimeter wave compressive imaging performance. In this Chapter, Sec. 3.2 provides the literature review. Sec. 3.3 identifies the mechanism of how pixel diffraction interferes with the linear signal model (3.1) and applies the Fraunhofer equation (25) to identify the compressive imaging resolution limit due to pixel diffraction. In Chapter 4, the Fraunhofer analysis is replaced by rigorous diffraction theory (26), which concludes similar pixel diffraction limit as Fraunhofer analysis. In

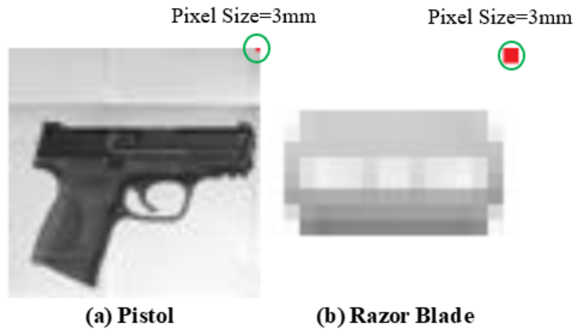



Figure 3.2: (a) 64-by-64 pistol image accurate to scale where each pixel dimension is 3mm; (b) 16-by-16 blade image accurate to scale where each pixel dimension is 3mm.

Chapter 5, the EM simulation (HFSS) demonstrates the deterioration due to pixel diffraction on the actual CS imaging reconstruction results. Last but not least, a method for rapid evaluation of the impact of pixel diffraction on CS imaging quality without full-blown EM simulation is proposed.

## 3.2 DIFFRACTION IN MILLIMETER WAVE IMAGING

Previous attempts in applying SPC-style approach to millimeter wave imaging (82; 83; 84) often adopt the linear signal model (3.1) intended for visible-spectrum SPC systems. Although such approaches produced promising results (82; 83; 84), directly applying visible-spectrum signal models to the millimeter wave spectrum may not be appropriate. Diffraction is one particular physical phenomenon that is secondary in optical compressive sensing, yet becomes important in the millimeter wave spectrum. More specifically, at the heart of any CS imaging system is the signal modulation and summation stage. In the SPC arrangement depicted in Fig. 3.1, these two stages are realized by coded apertures (modulating masks with pseudo-random patterns) and a lens, respectively. In this setup, the dimension of the reconstructed image is determined by the dimension of coded apertures, and the resolution is determined by the pixel size of coded apertures. For security scanning applications, the details required in reconstructed images for reliable object detection bound the system's imaging resolution, and hence coded apertures' pixel size. In the example of Fig. 3.2, each pixel (red square on the top right) is scaled to $3mm$ ($\lambda$ at 94GHz). We observe that such choice of coded aperture pixel size/image resolution is adequate for pistol detection, but would struggle with detecting razor blades. To reliably detect a razor blade object, we require coded aperture pixel size at fraction of a wavelength for a 94GHz source, which would inevitably result in severe diffraction.

The diffraction caused by coded aperture pixels is called pixel diffraction (89). Pixel diffraction is a distinctive phenomenon in millimeter wave SPC where coded apertures are introduced for power modulation and is covered neither by the visible-spectrum SPC nor by the non-compressive millimeter wave imaging literature. In existing millimeter wave compressive imaging works, pixel diffraction is often absent from system design considerations (82; 83). In (84), the impact of pixel diffraction is empirically approximated by reducing the signal-to-noise ratio (SNR). Although this approximation is plausible for its specific hardware arrangement, a heuristic single SNR reduction figure does not provide the complete picture of how different diffraction severity levels interfere with the image reconstruction process. Alternatively, Schmitt et al. (89) studied pixel diffraction by simulating the reconstruction of 1D patterns. This simplified 1D EM simulation does not correspond to realistic imaging tasks and cannot provide adequate intuitions about the impact of pixel diffraction. In the terahertz band, the significantly smaller wavelength allows for a much smaller coded aperture pixel size without needing to consider pixel diffraction. In other words, the compressive image resolution is less constrained by pixel diffraction and hence is not within the scope of this study. In the following sections, we will conduct an illustrative yet insightful study of the pixel diffraction limits on passive millimeter wave compressive imaging based on SPC, so as to eliminate the guess work from any future projects.

## 3.3 Diffraction Analysis Based on Fraunhofer Analysis

To accurately evaluate the impact of pixel diffraction on compressive image sensing whose outcome relies on a specific linear signal model (3.1) and sophisticated CS algorithms (3; 17), it is essential to isolate other potential uncertainty sources that would cause the physical system to deviate from the theoretical model (3.1). In practical experiments, our experience shows that it is extremely challenging to quantitatively evaluate the impact of pixel diffraction independently from other sources of distortions, e.g. fabrication imprecision, experiment implementation variations, and nonlinearity in electronic and optical components at the millimeter wave band. Instead, we first explore the possibility of applying first principles in physics. For far-field diffraction (90), the spatial intensity distribution can be formulated by the Fraunhofer equation for rectangle slit. In polar coordinates, the spatial intensity distribution follows,

$$I_{2D}(\theta,\phi) \propto \mathrm{sinc}^2\left(\frac{\pi d \sin\theta}{\lambda}\right)\mathrm{sinc}^2\left(\frac{\pi d \sin\phi}{\lambda}\right), \tag{3.2}$$

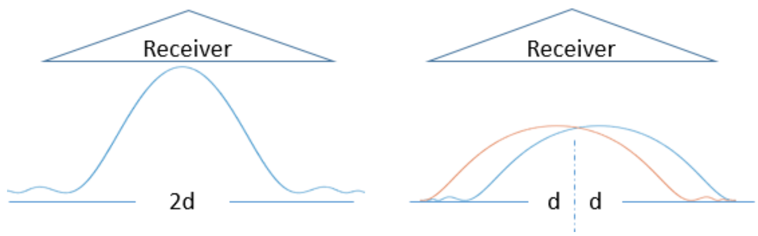
with parameters $\theta = \mathrm{atan}\left(\frac{x}{z}\right)$, and $\phi = \mathrm{atan}\left(\frac{y}{z}\right)$. In Cartesian coordinates, $\theta$ represents the angle of incidence within the x-z plane at corresponding $x$ and $z$ coordinates, assuming the diffraction source is at the origin; $\phi$ is the angle of incidence within the y-z plane at corresponding $y$ and $z$ coordinates assuming the diffraction source is at the origin. If each slit diffracts independently, the EM intensity distribution by the coded aperture can be considered as the spatial superposition of all diffracting slits. The diffraction of each slit can be modelled as a convolution of the millimeter wave intensity level before the slit, $G(x,y)$, and the diffracted point spread function (3.2) for the slit, denoted by $I_{2D,L}(x,y)$. Using $D(x,y)$ to indicate the location of slits in the coded apertures, each single pixel measurement can be formulated as

$$Y = \frac{\sum_{(x,y)\in \text{lens area}} I_{2D,L}(x,y) * (G(x,y) \times D(x,y))}{Z}, \tag{3.3}$$

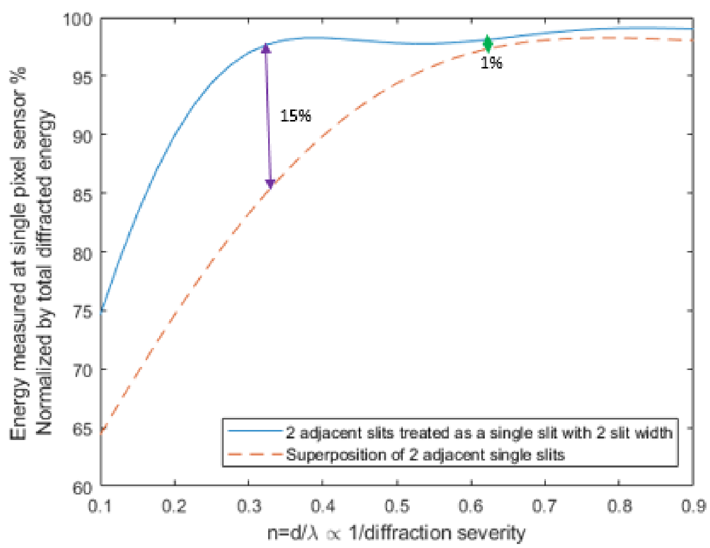
where $*$ denotes convolution. Because (3.2) does not sum to one after discretization, in (3.3) a normalizing constant $Z$ defined in (3.4) ensures that the energy is conserved:

$$Z = \sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} I_{2D,L}(x,y). \tag{3.4}$$

Model 3.3 varies from the standard CS imaging model 3.1 in the extra convolution with the point spread function 3.2. If the point spread function for each slit pixel is approximately identical, or in other words, if each transparent pixel is diffracting approximately identically, we may factor the convolution, a linear operator, out of the summation 3.3 and its effect can be removed from the reconstructed image. To investigate how variations in point spread function due to different slit sizes influence single pixel measurements, we studied the diffraction of a 1D slit (gap) at different slit widths with respect to wavelength $\lambda$. To be consistent with Chapter 5, we fixed the hardware dimensions and positioning and varied the wavelength, as opposed to the typical practice in physical experiments where $\lambda$ would be fixed. The ideal receiver lens was set as 20mm in length and 9mm away from the diffraction slit. We then fixed the single pixel slit $d = 0.5mm$ and varied the wavelength $\lambda$ so that $n = \frac{d}{\lambda} \in [0.1, 1]$, in order to investigate the impact of different diffraction levels. Explicitly, $n = 0.2$ refers to slit width $d = 0.2\lambda$, i.e., severe diffraction; and $n = 1$, i.e., $d = \lambda$ refers to negligible diffraction. The proposed spatial arrangement ensures that the criteria for applying the Fraunhofer equation in the classical optics theory is satisfied (90). The experiment compares the single pixel measurement by directly employing a wide slit at two-pixel width $2d$ (left panel Fig. 3.3(a)), to the single pixel measurement obtained from the superposition of two adjacent single pixel slits at $d = 0.5mm$ diffracting independently (right panel Fig. 3.3(a)). The single pixel power measurements are obtained by discretizing 3.3 within the receiver lens aperture plane. According to Fig. 3.3(b), the discrepancies between the measurements from a single wide slit and those from superposition of two narrow slits are within 1% for $n = 0.65$ and above, 5% for $n = 0.5$, and remain $> 10\%$ for more severe diffraction scenarios. Noting the two curves in Fig. 3.3(b) represent the portion of power that arrives at the single pixel receiver over the total power transmitted per single pixel slit, we designate the values on the two curves as the percent transmittance ($T\%$) of a pixel. In addition to diffraction severity, $T\%$ also depends on the positioning and dimension of the receiver lens, hence in the following chapters it is ensured that the spatial




(a) **Left:** intensity diffraction pattern for a two-pixel slit as one wide slit (solid curve in the lower panel)
**Right:** intensity diffraction pattern for a 2-pixel slit as superposition of two diffracting pixels (dashed curve in the lower panel)




(b) Effective percent transmittance vs 1/diffraction severity

Figure 3.3: 1D slit experiment, one wide slit diffraction vs superposition of two adjacent narrow slits

configurations of the receiver lens are consistent with the hypothetical 1D slit experiment. In the visible-spectrum (diffraction-free) model (3.1), $T\%$ for a pixel corresponds to the value of an entry in the sensing matrix A $\in \{0, 100\%\}$. In the case of mild diffraction where $n > 0.6$ in Fig. 3.3(b), regardless whether the 1D slit is of width $d$ or $2d$, $T\%$ remains $\simeq 95\%$. The diffraction-free model is approximately valid, only with 5% lower signal-to-noise ratio (SNR). However, when severe diffraction presents ($n = 0.33$ in the figure for example), the $T\%$ of a pixel is $\in \{0\%, 78\%, 93\%\}$. Depending on the random patterns, the entries in the sensing matrix can take values in a significantly wider range. We expect this deviation from visual-spectrum model to significantly degrade the performance of CS imaging. By now, we have successfully observed the mechanism of how diffraction interferes with the CS imaging model.

# Chapter 4

# Rigorous Analysis of Severe Diffraction

In the previous chapter, the hypothetical experiment identified the mechanism of how pixel diffraction deteriorates on the CS imaging model; the pixel diffraction limit on coded aperture pixel size was also derived. The Fraunhofer diffraction equation was used to model the diffraction effects due to its simplicity and consistency with empirical results in the far field (25; 90). However, in the classical optics literature, it is either explicitly specified or implicitly assumed that the dimensions of the diffracting object are $> \lambda$ (wavelength) (25). Then by neglecting the interactions between the scattered E and H fields, three-dimensional (3D) Maxwell's equations may be decoupled into scalar wave equations with appropriate boundary conditions at the diffraction plane, and result in the scalar diffraction theory, i.e., the popular Kirchhoff and Rayleigh-Sommerfeld solutions (25). In the far field, both Kirchhoff and Rayleigh-Sommerfeld diffraction solutions converge to the Fraunhofer equation (25). While these approximations are naturally satisfied in conventional optical systems, the validity of these approximations becomes questionable for the proposed millimeter wave single pixel compressive imaging system, as the scalar diffraction theory from classical optics is obviously not intended for general millimeter wave problem settings (25). To accommodate arbitrary hardware spatial arrangement for millimeter wave imaging, we resort directly to Maxwell's equations, the first principle of electromagnetic wave to better understand the impact of the pixel diffraction on compressed imaging. Furthermore, the rigorous diffraction results derived in this chapter provide invaluable empirical support for the EM simulation study in the following chapter.

## 4.1 INTRODUCTION

The vectorial diffraction problem involves 3D Maxwell's equations. But for a slit or a strip, which is a one-dimensional pattern, the problem can be simplified while still being rigorous due to symmetry (26). The vectorial solutions derived in the sequel depend on the polarization of the EM wave, as opposed to the case of scalar diffraction theory such as the Fraunhofer model, where the

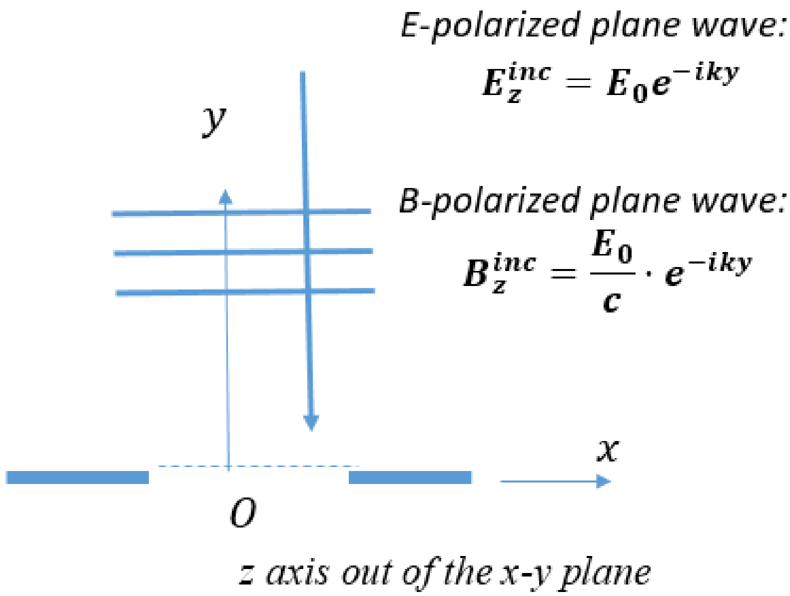



Figure 4.1: Schematic of the slit diffraction problem

solutions are insensitive to polarization. We simplify the numerical approach developed in (26) to match the slit diffraction problem described in the previous chapter. The derivation does neither depend on assumptions in the distance between the point of observation and the point of diffraction (far, Fresnel, or near field), nor on the size of the slit with respect to wavelength ($\lambda$). The universality of the approach is particularly valuable for our unconventional imaging hardware. For completeness, we include the derivations of the explicit expressions following (26) to remove any ambiguities in coding.

As demonstrated in Fig. 4.1, we consider a perfect electric conductor (PEC) screen in a vacuum 3D space that spreads the complete $x-z$ plane with infinitesimal thickness at $y = 0$; a strip of infinite length in $z$ direction and width $a$ in $x$ direction is cut out of the screen. We are interested in finding the diffracted field from the remaining slit when a monochromatic incident plane wave propagates in the $-y$ direction from the upper half plane of $y > 0$. To avoid integrating over PEC surface in $x \in (-\infty, -a) \cup (a, \infty)$, Nye (26) proposes to first solve the strip problem which only requires integrating across the strip of $x \in [-a/2, a/2]$; and then the solution for the complementary slit can be derived from Babinet's principle.

To implement the rigorous Babinet's principle (26), we consider two particular polarization orientations, namely the E-polarized problem where monochromatic plane wave travels in $-y$ direction as $E^{inc} = E_z^{inc} = E_0 e^{-iky}$ with the E field polarized in $z$ direction, and the B-polarized problem where the B field is polarized in $z$ direction as $B^{inc} = B_z^{inc} = \frac{E_0}{c} e^{-iky}$. $c$ is speed of light in the vacuum free space. Without loss of generality, we let $E_0 := 1$ in the following sections. We note that the scattered field for planar waves with other polarization could be considered as a linear combination of these two results (26).

## 4.2 E-polarized Strip Problem

In slit and strip diffraction problems, any terms associated with $\frac{\mathrm{d}\cdot}{\mathrm{d}z}$ remain 0 as the geometry maintains unchanged in $z$ direction. Hence from Maxwell's equations we have for the incident fields

$$E_z = e^{-iky},\ E_x = E_y = 0,\ B_x = \frac{1}{ikc}\frac{\partial e^{-iky}}{\partial y},\ B_y = -\frac{1}{ikc}\frac{\partial e^{-iky}}{\partial x} = 0,\ B_z = 0, \tag{4.1}$$

where $k = \frac{2\pi}{\lambda}$ is the angular wave number. To avoid infinite current density on the PEC screen, the total E field is 0 on the PEC strip surface, which is the boundary condition, i.e., at $y = 0$

$$E_z^{total} = E_z^{inc} + E_z^{sc} = 0. \tag{4.2}$$

Using the tools of scalar and vector potentials (26; 25), we arrive at the integral equation for the scattered field due to surface current density $j_z$,

$$E_z^{sc} = -\frac{kZ_0}{4}\int \mathrm{d}\xi\, j_z(\xi) H_0^{(1)}(k\rho), \tag{4.3}$$

where $\rho = \sqrt{(x-\xi)+y^2}$. The surface current density $j_z$ is solved by realizing $E_z^{sc} = -E_z^{inc}$ as $y \uparrow 0$ due to (4.2). Thus (4.3) becomes

$$E_z^{inc}(y=0) = \frac{kZ_0}{4}\int \mathrm{d}\xi\, j_z(\xi) H_0^{(1)}(k|x-\xi|). \tag{4.4}$$

To solve $j_z$ numerically, the Riemann sum is used to discretize the integral by introducing N samples from $\xi \in [-\frac{a}{2}, \frac{a}{2}]$ on the strip surface and N samples just off the strip surface. The interval between adjacent samples is $\Delta = \frac{a}{N-1}$. Define $L_{ik} = \frac{kZ_0 H_{ik}\Delta}{4}$, where $H_{ik} = H_0^{(1)}(k|x_i - \xi_k|)$. Then $\sum_{k=0}^{N} L_{ik} j_k = R_i$, for $i,k \in [1,N]$. In the E-polarized problem, $R_i = E^{inc}(y=0) = e^{-ik\cdot 0} = 1$ and $j_k = j_z(\xi_k)$ on the slit surface. There is a singularity at $H_0^{(1)}(0)$. A standard technique to integrate the logarithmic singularity of the Hankel function is to replace the singularity $H_0^{(1)}(0)$ by $1 + i\frac{2}{\pi}[\ln(\Delta\gamma_s/2) - 2]$, where $\gamma_s = 1.7807$ (26). Then $j_z = L^{-1}\mathbf{1}$, where $\mathbf{1} = [1,\ldots,1]^T$, $L \in \mathbb{C}^{N\times N}$

and

$$L(i,k) := \begin{cases} \frac{kZ_0\Delta}{4} H_0^{(1)}(k|x_i - \xi_k|), \text{ if } i \neq k; \\ \frac{kZ_0\Delta}{4} \left\{ 1 + i\frac{2}{\pi}[\ln(\Delta\gamma_s/2) - 2] \right\}, \text{ if } i = k. \end{cases} \tag{4.5}$$

After solving $j_z$, for arbitrary position in the lower half plane ($y < 0$), the E field can be calculated by (4.3) and the B field by Maxwell's equation, similar to (4.1), i.e.

$$\begin{cases} B_x^{sc} = \frac{\mu_0 k}{4i} \int_{\frac{-a}{2}}^{\frac{a}{2}} \mathrm{d}\xi\, j_z(\xi) H_1^{(1)}(k\rho)\frac{y}{\rho}; \\ B_y^{sc} = -\frac{\mu_0 k}{4i} \int_{\frac{-a}{2}}^{\frac{a}{2}} \mathrm{d}\xi\, j_z(\xi) H_1^{(1)}(k\rho)\frac{x-\xi}{\rho}; \\ B_x^{inc} = -\frac{e^{-iky}}{c}. \end{cases} \tag{4.6}$$

With the diffraction problem for E-polarized strip solved, according to Babinet's principle (26), the B-polarized slit problem is resolved for $y < 0$ by

$$E^{strip} + cB^{slit} = E^{inc},\ cB^{strip} - E^{slit} = cB^{inc}. \tag{4.7}$$

Note superscripts $\cdot^{strip}$ corresponds to the total field of the E-polarized strip problem and $\cdot^{slit}$ to the total field for the B-polarized slit problem; $E^{inc}$ and $B^{inc}$ correspond to the incident the E and B fields for E-polarized problem respectively (4.1). Thus for B-polarized slit problem for $y < 0$,

$$\begin{cases} B_z^{slit} = \frac{1}{c}\left(e^{-iky} - E_z^{strip}\right); \\ E_x^{slit} = e^{-iky} + cB_x^{strip}; \\ E_y^{slit} = cB_y^{strip}. \end{cases} \tag{4.8}$$

All other fields for the B-polarized slit problem are 0 following (4.1).

## 4.3 B-polarized Strip Problem

Recall that the B-polarized problem refers to the case of the source monochromatic plane wave traveling in $-y$ direction, with the B field polarized in $z$ direction. The incident field for $y \leq 0$ is given by

$$B_z = \frac{e^{-iky}}{c},\ B_x = B_y = 0,\ E_x = -\frac{c}{ik}\frac{\partial B_z}{\partial y},\ E_y = \frac{c}{ik}\frac{\partial B_z}{\partial x} = 0,\ E_z = 0. \tag{4.9}$$

It is known that PEC boundary condition causes discontinuities in B field. We abuse the notation for $\frac{\partial \cdot}{\partial x}$ and $\frac{\partial \cdot}{\partial y}$ at $y = 0$ which refers to the one-sided operators $\frac{\partial^- \cdot}{\partial x}$ and $\frac{\partial^- \cdot}{\partial y}$ as we are only interested in the half plane of $y<0$. From (4.9), the only nonzero component of surface current density is $j_x$ and so is the vector magnetic potential $A_x$, i.e.,

$$A_x(x,y)=\frac{i\mu_0}{4}\int \mathrm{d}\xi\, j_x(\xi)\, H_0^{(1)}(k\rho)\,. \tag{4.10}$$

At $y=0$, the Helmholtz wave equation becomes

$$\left(\frac{\partial^2}{\partial x^2}+k^2\right)A_x=-\frac{\partial^2 A_x}{\partial y^2}=\frac{\partial B_z^{sc}}{\partial y}=-\frac{\partial B_z^{inc}}{\partial y}=i\frac{k}{c}, \tag{4.11}$$

where the second equality from the left is from the definition of $E=\nabla\times A$ and (4.9), and the third equality is due to PEC boundary condition, i.e. $E_x(x,0)=-\frac{c}{ik}\frac{\partial B_z}{\partial y}=-\frac{c}{ik}\frac{\partial(B_z^{inc}+B_z^{sc})}{\partial y}=0$.

The general solution for (4.11) is

$$A_x=A+Bcos(kx)+Csin(kx). \tag{4.12}$$

Substitute the general solution into (4.11), and use the property of sinusoidal functions,we get

$$A=\frac{i}{ck},\ B=\frac{\left(A_x(\frac{a}{2})+A_x(-\frac{a}{2})\right)}{2}-A,\ C=\frac{\left(A_x(\frac{a}{2})-A_x(-\frac{a}{2})\right)}{2}. \tag{4.13}$$

Then substitute (4.10) into (4.13), then (4.13) into both sides of (4.12) and lastly collect all the terms with $j_x$,

$$\begin{aligned}\int \mathrm{d}\xi\, j_x\Bigg[H_0^{(1)}(k|x-\xi|)-\frac{1}{2}H_0^{(1)}(k|\frac{a}{2}-\xi|)\left(\frac{sin(kx)}{sin(\frac{1}{2}ka)}+\frac{cos(kx)}{cos(\frac{1}{2}ka)}\right)+&\\ \frac{1}{2}H_0^{(1)}(k|\frac{a}{2}+\xi|)\left(\frac{sin(kx)}{sin(\frac{1}{2}ka)}-\frac{cos(kx)}{cos(\frac{1}{2}ka)}\right)\Bigg]&=\frac{4E_0}{Z_0k}\left[1-\frac{cos(kx)}{cos(\frac{1}{2}ka)}\right]\end{aligned} \tag{4.14}$$

Riemann sum is used to discretize the integral by introducing N samples of $\xi\in[-\frac{a}{2},\frac{a}{2}]$ within the strip and N samples of $x$ adjacent to the strip surface. Again, the interval between adjacent samples is $\Delta=\frac{a}{N-1}$. Define $L\in\mathbb{C}^{N\times N}$, where $L_{ik}=\frac{kZ_0\Delta}{4}H_{ik}$, and

$$\begin{aligned}H_{ik}=&H_0^{(1)}(k|x_i-\xi_k|)-\frac{1}{2}H_0^{(1)}(k|\frac{a}{2}-\xi_k|)\left(\frac{sin(kx_i)}{sin\left(\frac{1}{2}ka\right)}+\frac{cos(kx_i)}{cos\left(\frac{1}{2}ka\right)}\right)+\\ &\frac{1}{2}H_0^{(1)}(k|\frac{a}{2}+\xi_k|)\left(\frac{sin(kx_i)}{sin((\frac{1}{2}ka)}-\frac{cos(kx_i)}{cos\left(\frac{1}{2}ka\right)}\right).\end{aligned} \tag{4.15}$$

As in the previous section, we also replace the diagonal of $H_0^{(1)}(k|x_i - \xi_i|) = H_0^{(1)}(0)$ by $1 + i\frac{2}{\pi}[\ln(\Delta\gamma_s/2) - 2]$. In addition, at the edges of the strip $x = \pm\frac{a}{2}$, the right hand side of (4.14) is always zero, and so is the value encapsulated in $[\cdot]$ on the left hand side. We remove the contribution from the edges in the discretization process to avoid rank deficiency. Hence, we have $L \in \mathbb{C}^{(N-2)\times(N-2)}$ for $i,\ j = [2, N-1]$ such that

$$L(i-1,k-1) := \begin{cases} \frac{kZ_0\Delta}{4}\Big[H_0^{(1)}(k|x_i-\xi_k|) - \frac{1}{2}H_0^{(1)}(k|\frac{a}{2}-\xi_k|)\Big(\frac{sin(kx_i)}{sin(\frac{1}{2}ka)} + \frac{cos(kx_i)}{cos(\frac{1}{2}ka)}\Big) + \frac{1}{2}H_0^{(1)}(k|\frac{a}{2}+\xi_k|)\Big(\frac{sin(kx_i)}{sin(\frac{1}{2}ka)} - \frac{cos(kx_i)}{cos(\frac{1}{2}ka)}\Big)\Big], \text{ if } i \neq k; \\ \frac{kZ_0\Delta}{4}\Big\{1 + i\frac{2}{\pi}[\ln(\Delta\gamma_s/2) - 2] - \frac{1}{2}H_0^{(1)}(k|\frac{a}{2}-\xi_k|)\left(\frac{sin\left(kx_i\right)}{sin\left(\frac{1}{2}ka\right)} + \frac{cos\left(kx_i\right)}{cos\left(\frac{1}{2}ka\right)}\right) + \frac{1}{2}H_0^{(1)}(k|\frac{a}{2}+\xi_k|)\left(\frac{sin\left(kx_i\right)}{sin\left(\frac{1}{2}ka\right)} - \frac{cos\left(kx_i\right)}{cos\left(\frac{1}{2}ka\right)}\right)\Big\}, \text{ if } i = k. \end{cases} \tag{4.16}$$

Lastly, careful investigation into the numerical experiments reveals that when strip width $a$ is a rational fractions or multiples of $\lambda$, the denominators in $\frac{sin(kx)}{sin(ka)}$ and $\frac{cos(kx)}{cos(ka)}$ may cause numerical instability. We follow the suggestion in (26) to add a small $\varepsilon \sim 10^{-6}\lambda$ to strip width $a$. Then $j_x = L^{-1}R$ where $R \in \mathbb{R}^{N-2}$ and is defined by $\left[1 - \frac{cos(kx_2)}{cos(\frac{1}{2}ka)}, \cdots, 1 - \frac{cos(kx_{N-1})}{cos(\frac{1}{2}ka)}\right]^T$. After obtaining $j_x$, we solve $A_x$ from 4.10, then $B_z^{sc}$ from (4.11), and lastly $E_x^{sc}$ and $E_y^{sc}$ from (4.9) for $y < 0$.

$$\begin{cases} B_z^{sc} = -\dfrac{\partial A_x}{\partial y} = \dfrac{\mu_0 k}{4i}\displaystyle\int_{-\frac{a}{2}}^{\frac{a}{2}} \mathrm{d}\xi\, j_x(\xi) H_1^{(1)}(k\rho)\frac{y}{\rho}; \\ E_x^{sc} = -\dfrac{c}{ik}\dfrac{\partial B_z^{sc}}{\partial y} = -\dfrac{c\mu_0}{4}\displaystyle\int_{\frac{-a}{2}}^{\frac{a}{2}} \mathrm{d}\xi\, j_x(\xi)\left[kH_0^{(1)}(k\rho)\frac{y^2}{\rho^2} - \frac{1}{\rho}H_1^{(1)}(k\rho)\left(2\frac{y^2}{\rho^2} - 1\right)\right]; \\ E_y^{sc} = \dfrac{c}{ik}\dfrac{\partial B_z^{sc}}{\partial x} = -\dfrac{c\mu_0}{4}\displaystyle\int_{\frac{-a}{2}}^{\frac{a}{2}} \mathrm{d}\xi\, j_x(\xi)\left[kH_0^{(1)}(k\rho) - \frac{2}{\rho}H_1^{(1)}(k\rho)\right]\frac{(-y)(x-\xi)}{\rho^2}. \end{cases} \tag{4.17}$$

With the B-polarized scattered field resolved for the strip, Babinet's principle is applied to resolve the E-polarized strip diffraction for $y < 0$, i.e.

$$\begin{cases} E_z^{slit} = e^{-iky} - cB_z^{strip}; \\ cB_x^{slit} = -e^{-iky} + E_x^{strip}; \\ cB_y^{slit} = E_y^{strip}. \end{cases} \tag{4.18}$$

## 4.4 Diffraction Pixel Resolution Limit

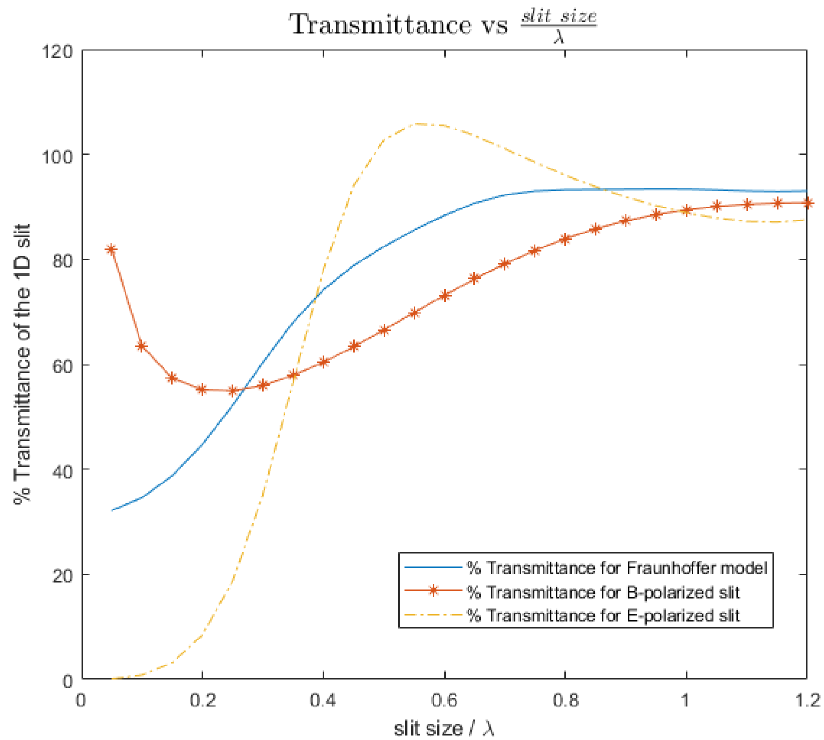
In Chapter 3, we implemented a hypothetical experiment to identify a pixel diffraction limit on reconstructed image pixel dimension. When the pixel dimension falls below this limit, the transmittance of each pixel varies significantly according to neighboring pixels' values. This violates the single pixel CS model where the pixel transmittance only admits values from $\{0, 1\}$. The hypothetical 1D experiment clearly demonstrates the mechanism of how the diffraction phenomenon deteriorates the performance of CS imaging systems. The Fraunhofer diffraction model can now be further replaced by the more general diffraction model derived above. A key feature the Fraunhofer model misses out is the impact of polarization. We use the surface integral over the lens area (20mm, as in the previous chapter) of the electromagnetic power flux (poynting vector) as the ideal receiver reading, i.e.,

$$\begin{aligned} P_{rec} &= \int_{S_{lens}} \mathrm{d}s(-\hat{j}) \cdot \frac{1}{2}\Re\left(E \times H^*\right) = \int_{S_{lens}} \mathrm{d}s(-\hat{j}) \cdot \frac{1}{2\mu_0}\Re\left(E \times B^*\right) \\ &= \begin{cases} \dfrac{1}{2\mu_0}\displaystyle\int_{S_{lens}} \mathrm{d}s E_x B_z^*, \text{ if B-polarized on the slit ;} \\ -\dfrac{1}{2\mu_0}\displaystyle\int_{S_{lens}} \mathrm{d}s E_z B_x^*, \text{ if E-polarized on the slit.} \end{cases} \end{aligned} \tag{4.19}$$

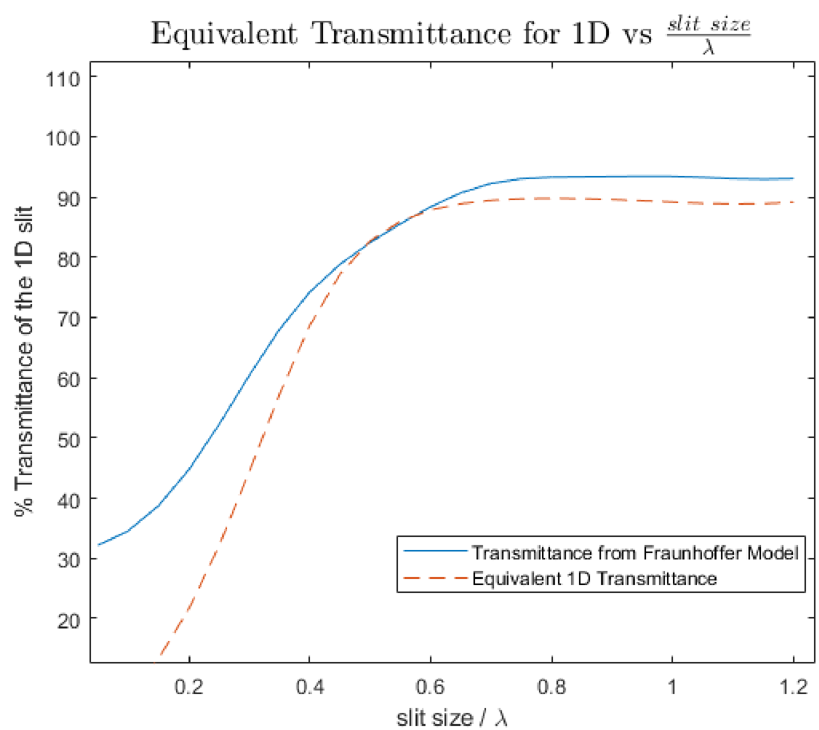
Since the E and B fields have been resolved in the previous section for both polarization orientations, we can readily apply Riemann sum to obtain $P_{rec}$. In the case when diffraction does not exist, then for both E- and B- polarization the power flux through the slit is $\frac{1}{2\mu_0 c}$. Thus the transmittance $T\% = P_{rec}/(\frac{a}{2\mu_0 c}) \cdot 100\%$, where $a$ is the slit width.

For the Fraunhofer model, we identify the knee point (slit size $\sim 0.6$, refer to the solid line in Fig. 3.3(b)) as the pixel diffraction limit. In this more general vector diffraction model, Fig. 4.2 indicates an overshoot of transmittance $> 100\%$ as we vary the slit size for the E-polarized case. Similar overshoot in transmittance is observed in (91) when the author attempted to address the inconsistency in the scalar diffraction theory. Practically, a pixel is a "2D slit" where, in one direction, it is a B-polarized slit diffraction problem and, in the orthogonal direction it is an E-polarized slit diffraction problem. Noting that for other orientations other than E- and B-polarized plane waves, the diffraction field from a slit is the linear combination of the E- and B-polarized problems (25; 26), to take both polarization scenarios into account in a 1D model, we propose heuristically $T_{eqv} = \sqrt{T_E \cdot T_B}$ as the equivalent transmittance for the 1D slit. From Fig. 4.2, we observe that the diffraction limit (knee point of the equivalent transmittance curve) is close to the Fraunhofer analysis at $0.6\lambda$.

The introduction of rigorous diffraction theory in this chapter not only provides us with more confidence in the pixel diffraction limit proposed in Chapter 3, but also paves the way for the EM simulation study in the following chapter. Although the severe diffraction scenario violates the fundamental assumptions of scalar diffraction theory, the Fraunhofer result on the pixel diffraction limit is impressively close to the result from the significantly more sophisticated approach presented in this chapter. This is consistent with the empirical finding that the scalar diffraction theory can perform satisfactorily accurately, quite often well beyond the limitations imposed by the physical assumptions of these models (92). And this is probably the reason why the scalar diffraction theory, e.g., the Kirchhoff solution, the Fraunhofer equation, and the Fresnel equation have been so well received in studying diffraction.




(a) % Transmittance for different diffraction models




(b) Geometric mean of % transmittance under two polarization orientations

Figure 4.2: % Transmittance for 1D slit width $a \in [0.05\lambda, 1.2\lambda]$

# Chapter 5

# Electromagnetic Simulations of Diffraction in Millimeter Wave Single Pixel Compressive Imaging

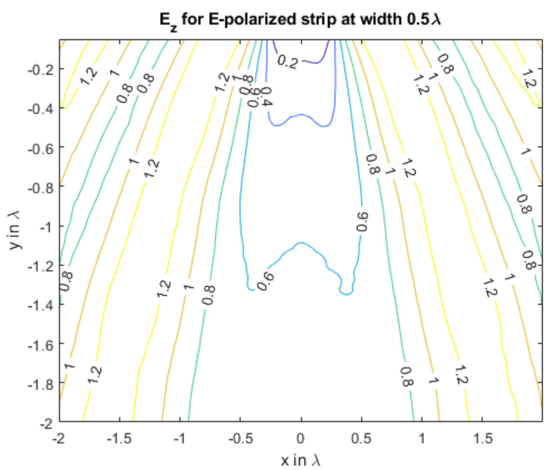
## 5.1 Introduction

In Chapter 4, the Fraunhofer diffraction model is replaced by a general slit diffraction model, which eliminates any approximations encountered in the classical diffraction theory. This is particularly valuable when developing a general approach for studying the pixel diffraction limit on the non-conventional millimeter wave CS imaging systems. For the nominated hardware spatial arrangement, the pixel diffraction limit identified in Chapter 4 closely matches the Fraunhofer result in Chapter 3. In an attempt to validate the proposed pixel diffraction limit $\sim 0.6\lambda$ via physical experiments, it becomes apparent that isolating image reconstruction performance degradation due to pixel diffraction from other sources of uncertainties, e.g., the inherent imprecision, nonlinearities, thermal noise and interference embedded in the electronic components, is rather infeasible. We resort to EM simulation software (HFSS) to facilitate an ideal single pixel CS imaging system. EM simulation excludes inevitable distortions in physical systems and allows us to focus the study on pixel diffraction phenomenon.

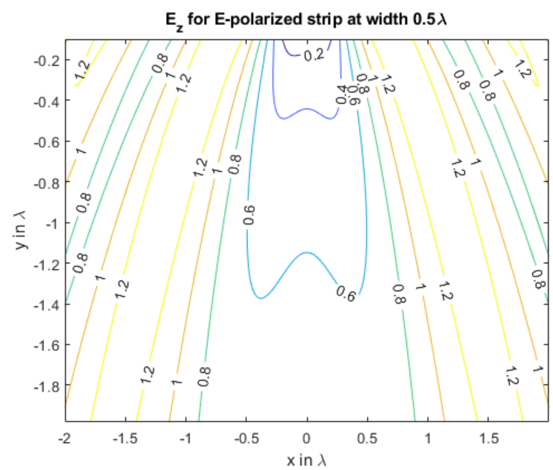
## 5.2 EM Simulations of Slit and Strip

Before performing CS imaging EM simulation, the performance of EM simulation is validated by comparing the simulated strip and slit diffraction results with the results in Chapter 4, as illustrated in Figs. 5.1 and 5.2. We notice that there are some ripples in $E_z$ contours for slit EM simulation.

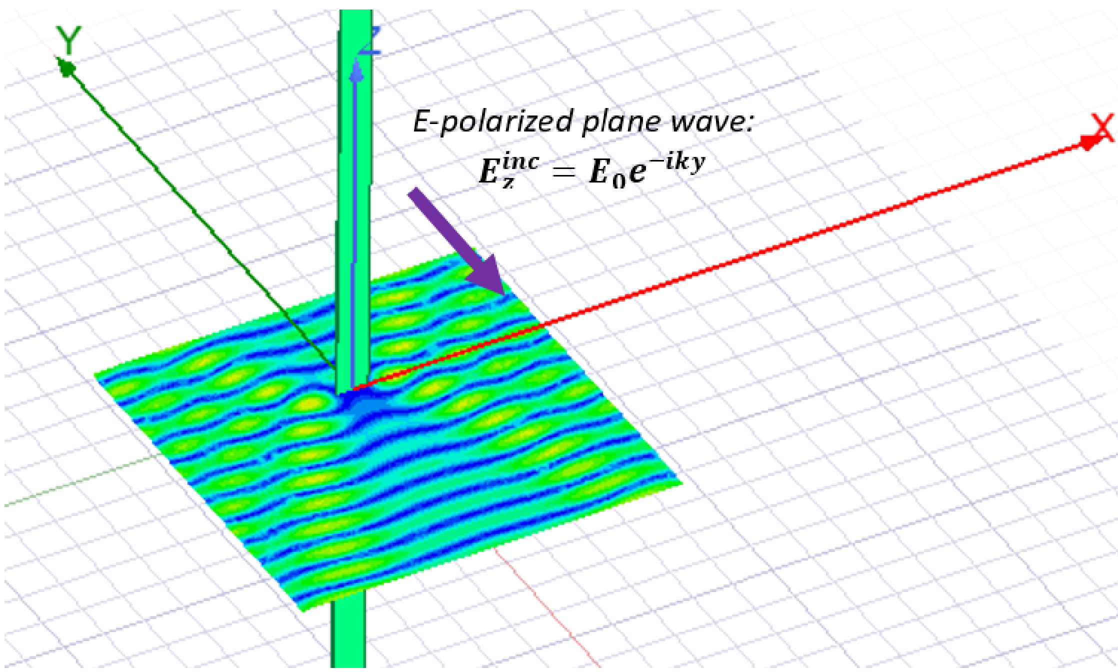
We attribute these ripples to the finite thickness of the PEC screen. Similar ripples can be observed in physical measurements when Nye validated the proposed numerical method (26).




(a) mag $(E_z)$ in the $x-y$ plane from HFSS Simulation

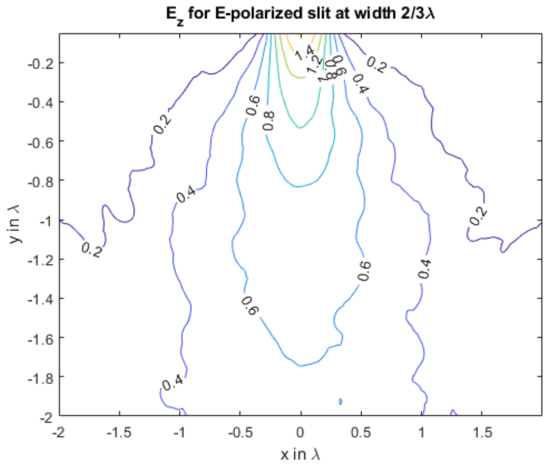



(b) mag $(E_z)$ in the $x-y$ plane based on rigorous analysis

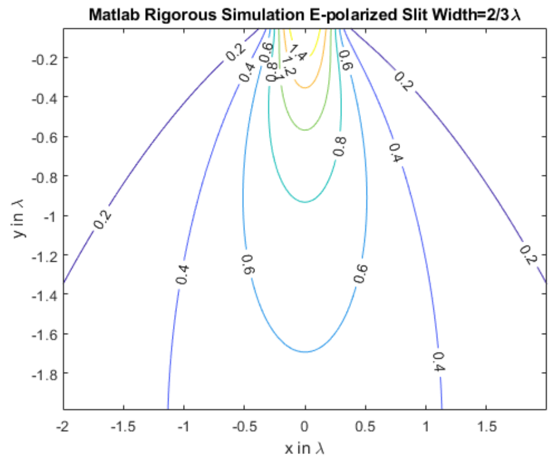



(c) Demonstrative HFSS 3D Model for Strip, $E_k^{inc} = e^{-iky}\tilde{k}$

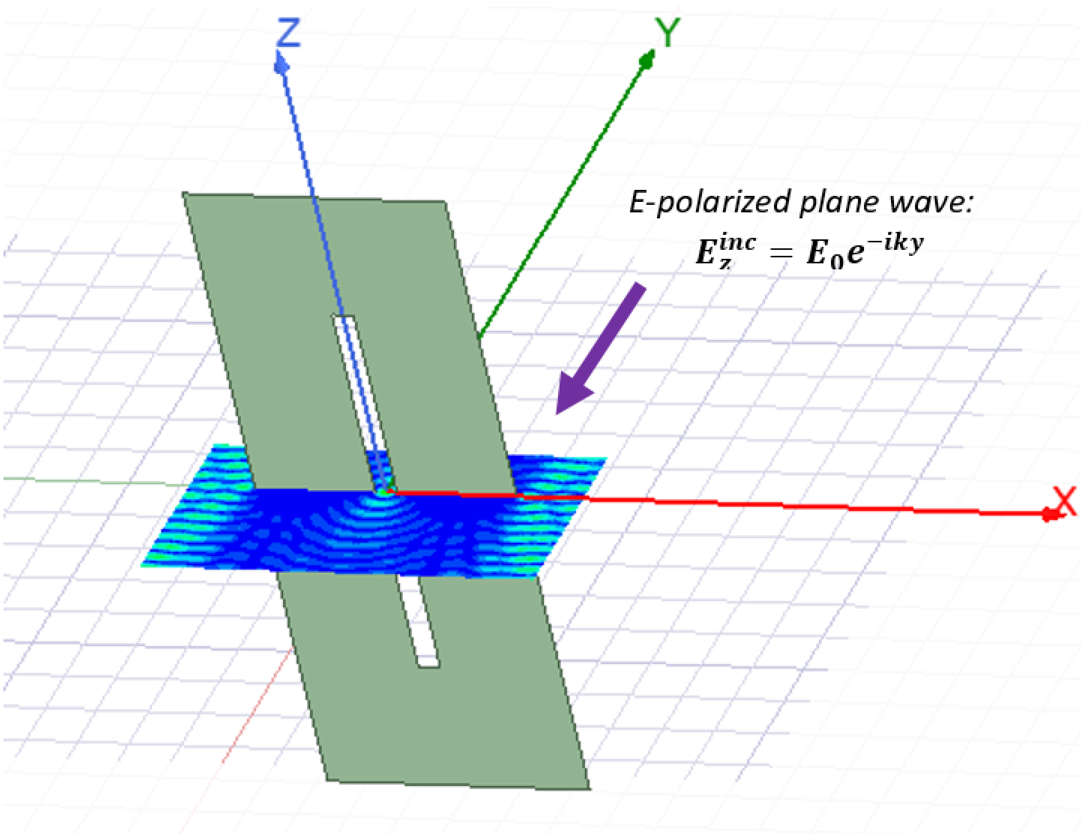
Figure 5.1: Validation of EM simulation




(a) mag $(E_z)$ in the $x-y$ plane from HFSS Simulation




(b) mag $(E_z)$ in the $x-y$ plane based on rigorous analysis




(c) Demonstrative HFSS 3D Model for Slit, $E_k^{inc} = e^{-iky}\tilde{k}$

Figure 5.2: Validation of EM simulation

Table 5.1: SPC compressive imaging EM simulation experiment parameters

| **EM Simulations Experimental Parameters** | |
|---|---|
| Illuminating EM Frequency | 94GHz/188GHz |
| Coded Aperture Pixel Slit Width $d$: | 1mm |
| Character Size | $12 \times 12$ mm |
| Coded Aperture Dimension | $32 \times 32$ pixels |
| Distance L between Coded Aperture and Receiver Lens | 9 mm |
| Lens size | $48 \times 48$ pixels |
| Measurement Rate $\delta = M/N$ | 0.3 |

## 5.3 Diffraction Analysis Based on EM Simulation

Having demonstrated the efficacy of EM simulation software (HFSS), we emphasize that EM simulation, though highly flexible, is limited by computation resources and may not save time over physical experiments. Each of the following EM simulated experiment (each image reconstruction) consumed 10 to 30 days of computational time on a dedicated 24-core Xeon server with 384Gb of memory.

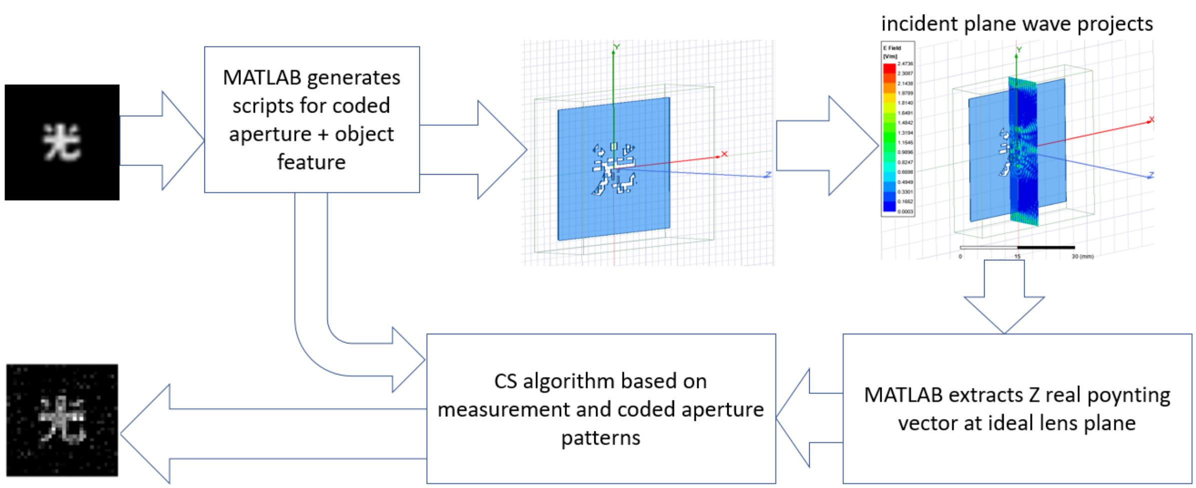
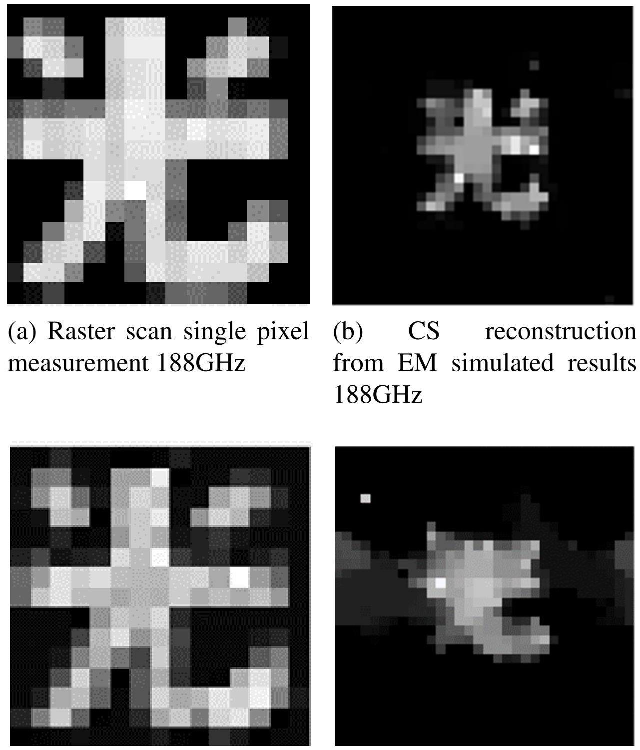



Figure 5.3: EM simulation study of the impact of diffraction on compressive imaging performance

The object feature under investigation is adopted from (82) as it has demonstrated superior reconstruction quality using TVAL (17) CS algorithm. The experiment arrangement is listed in



(a) Raster scan single pixel measurement 188GHz

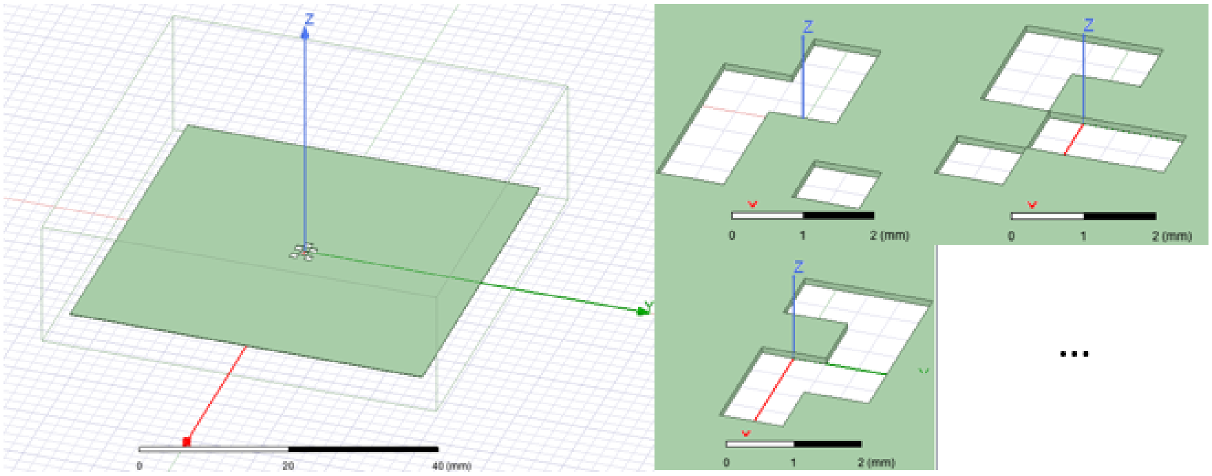
(b) CS reconstruction from EM simulated results 188GHz

(c) Raster scan single pixel measurement 94GHz

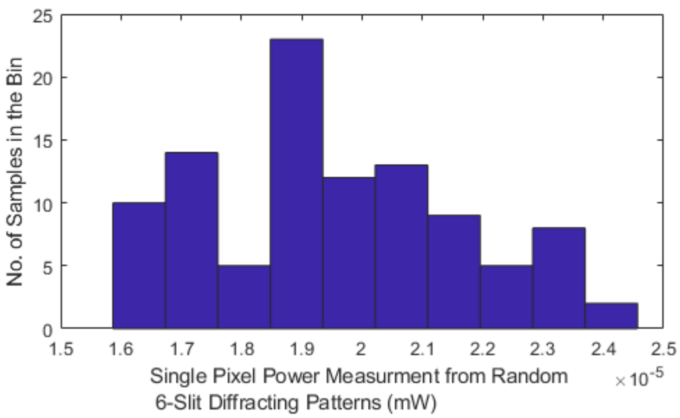
(d) CS reconstruction from EM simulated results 94GHz

Figure 5.4: Compressive imaging reconstruction based on EM simulation measurement

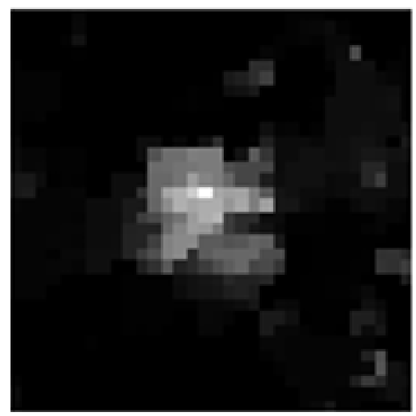
TABLE 5.1 and the complete simulation experiment workflow is demonstrated in Fig. 5.3. MATLAB is used to generate $M = \delta \cdot N = 0.3 \cdot 32^2 = 307$ coded aperture patterns; then in HFSS the coded aperture plates are merged with the planar object feature and embedded in a vacuum environment. In contrast to physical experiments where the radiation source is fixed and the dimensions of the hardware are altered, we fix the hardware dimensions and vary the radiation source between 94 GHz (pixel size $d = \frac{\lambda}{3}$) and 188 GHz (pixel size $d = \frac{2\lambda}{3}$) to minimize reconfigurations in HFSS. The radiation source is configured as a planar incident wave propagating orthogonally to the coded aperture plane. The real poynting vector in the propagating direction $L = 9mm$ away from the coded aperture plane is integrated/summed within the lens aperture area to represent an ideal receiver. The background field power obtained by similar method but with coded apertures replaced by a PEC plate with no slit holes, is lastly subtracted to form the final single pixel measurements. Table 5.1 summarizes the spatial arrangement for the simulated experiments. For each experiment, $M$ HFSS simulation projects are executed for the $M$ coded apertures. After the simulations are completed, MATLAB loads the $M$ single pixel measurements from HFSS and their corresponding coded aperture patterns to implement TVAL (17) CS algorithm.




(a) Monte Carlo random samples of 2D random patterns formed by 6 slits in a 3-by-4-pixel area




(b) Mean= $1.95 \cdot 10^{-5}$mW, Std= $0.211 \cdot 10^{-5}$, Equivalent SINR≃10dB



(c) MATLAB CS reconstruction, without diffraction with added Gaussian noise so that SINR=10dB

Figure 5.5: Modelling diffraction variations as AWGN

Raster scan results are included in Fig. 5.4 for moderate (188GHz) and severe (94GHz) diffraction as the ground truth for CS performance evaluation. Since raster scans only travel across the character feature area, images reconstructed by raster scans are 12-by-12 pixels based on TABLE 5.1; the CS reconstructed images follow the dimension of coded apertures at 32-by-32 pixels. Firstly, the reconstruction quality matches closely for the two raster scan results in Figs. 5.4(a) and 5.4(c), which implies that for an ideal hardware system the raster scan scheme does not have similar pixel diffraction issue. We observe from Fig. 5.4(b) that moderate pixel diffraction (pixel size $\sim \frac{2\lambda}{3}$) causes mild noise in the CS reconstructed image, yet the reconstructed image contains adequate details for observing the object feature; in the case of severe diffraction (Fig. 5.4(d), pixel size $\sim \frac{\lambda}{3}$), the fine details of the CS reconstructed image are severely distorted and illegible. The results are consistent with our expectation based on 1D first principle analysis. *In short, when pixel diffraction becomes a prominent effect for the system configuration, the variations in diffraction severity cause the transmittance of pixels to deviate significantly from the linear model. This refers to entries* $\notin \{0,1\}$ *for sensing matrix A in* (3.1). *The deviation from the linear model assumption becomes the limiting factor for CS imaging performance.* If under physical CS imaging experiments with inadequate SNR, it would have been impossible to discriminate the impact from pixel diffraction and from the changing SNR.

## 5.4 Equivalence of Diffraction and AWGN

EM simulation result is consistent with the diffraction limit derived from the hypothetical 1D experiment. However, executing a full-blown EM simulation is extremely computation intensive. Instead, in this section we propose a method that enables researchers to visually inspect the image reconstruction quality degradation due to pixel diffraction using a desktop PC instead of a dedicated server, and within a time frame of a few days instead of weeks. It has been proven that as the dimension of the CS problem increases, the effect of uniform measurement matrix uncertainty can be modelled by additive white Gaussian measurement noise (AWGN), with rescaled variance (93). In our application, the measurement matrix uncertainty is correlated across slit pixels and hence not uniform. Still borrowing the idea, we empirically obtain the equivalent signal to interference and noise ratio (SINR) corresponding to the pixel diffraction for CS imaging experiments in Sec. 5.3 and investigate if the interference due to diffraction can be approximated by additive white noise at the equivalent SINR. To obtain the equivalent SINR due to diffraction, 6-slit patterns are randomly sampled in a 3-by-4-pixel rectangular area at the center of a PEC plate (Fig. 5.5(a)); then we observe the single pixel measurement variance via the measurement histogram obtained from 100 Monte Carlo 6-slit random patterns (Fig. 5.5(b)). The EM frequency is fixed at 94GHz, the pixel size remains at $1mm$ or $\frac{\lambda}{3}$, and the distance $L$ between coded apertures and the receiver plane remains $9mm$ as in previous experiments. The receiver lens dimension is set to 40-by-40 pixels, so that the distance from the 3-by-4-pixel random patterns to the boundary of the receiver lens is $\frac{40-4}{2} = 18$pixels$\approx 18mm$, which is identical to the distance in the previous experiment settings, i.e., $\frac{48-12}{2} = 18$pixels$\approx 18mm$. From Fig. 5.5(b), we observe $SINR = 10log_{10}\frac{Mean}{Std} \approx 10dB$ (note that the single pixel measurement is a power term, not an amplitude). Lastly, a MATLAB compressive image reconstruction experiment is implemented on the original input feature image, without considering diffraction but with additive white Gaussian noise in the measurements such that $SINR = 10dB$. CS algorithm produces similar reconstruction quality when diffraction interference (Fig. 5.4(d)) is replaced by AWGN at the equivalent SINR (Fig. 5.5(c)). The result empirically supports the equivalence of pixel diffraction and measurement AWGN. Compared to the full-blown EM simulation, the Monte Carlo simulation on the checkerboard patterns demands significantly less computation resources and can be completed in a desktop PC.

# Chapter 6

# Conclusions and Future Directions

This dissertation investigates the applications of compressive sensing in both the algorithmic and system application fronts. In the algorithmic direction, the AMP CS algorithm is extended in an optimal manner to address side information; furthermore, the AMP CS algorithm is applied in an empirical Bayesian scheme for approximating a nonlinear function, which is a first published study of its kind.

In the system application direction, we investigate the pixel diffraction limit on passive millimeter wave CS imaging. In compressive imaging systems where coded apertures are present, such as SPC, it is not the diffraction severity, but the *variations* in diffraction severity levels that limit the CS reconstruction performance. The concept of transmittance is used to summarize such variations. For a visible-spectrum system, we may consider the transmittance to vary between 0% for opaque pixels and 100% for slits; in the case of millimeter wave applications, if the pixel size is greater than $0.6\lambda$, the transmittance remains close to 100% and only fluctuates within 1%. The linear model (3.1), intended for the visible spectrum is approximately applicable. However, if the pixel size is less than $0.6\lambda$, the transmittance of a slit becomes highly sensitive to the patterns of adjacent pixels. Conceptually, the entries in the sensing matrix $A$ in (3.1) deviate from $\{0,1\}$, invalidating the standard linear inverse model. The $0.6\lambda$ pixel diffraction limit is supported by both the Fraunhofer equation and a more general numerical diffraction model (26). Extensive EM simulations are then performed to empirically demonstrate the interference of pixel diffraction with CS imaging performance. The results support our fast 1D first principle analysis. Last but not least, a method to empirically inspect the CS imaging performance under pixel diffraction is proposed, so that researchers can readily evaluate the performance of a specific hardware arrangement using a desktop PC without running the full-blown EM simulation on a dedicated server for multiple weeks.

As for future directions, a common rule of thumb in the classical antenna design literature that the antenna resolution can be improved by enlarging the effective antenna aperture applies similarly to CS imaging systems. In other words, with an adequately large receiver lens, we can always ensure that most power diffracted by coded apertures is captured by the receiver regardless of the diffraction patterns, so that the transmittance for a slit pixel is always close to 100%. While this is not feasible for standard millimeter wave imaging application due to spatial dimension constraints, the insights may be applied to other related fields, such as near-field computational imaging. Furthermore, though violating the assumptions for scalar diffraction theory, the results from the Fraunhofer diffraction model and the vector diffraction model match well. Hence, methods developed for suppressing diffraction in Fresnel or Fraunhofer scalar models, though developed upon different theoretical assumptions, may still be effective in improving the millimeter wave CS imaging result, which will be studied in the future.

# REFERENCES


[1] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, pp. 5270–5284, Nov. 2013.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[3] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, pp. 5117–5144, Sep. 2016.

[4] A. Petrosian and F. Meyer, *Wavelets in signal and image analysis: From theory to practice*. Springer Science & Business Media, 2001.

[5] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Communications*, vol. 50, no. 3, pp. 374–377, 2002.

[6] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[7] M. Davenport and M. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.

[8] D. Donoho, Y. Tsaig, I. Drori, and J.-L. Stark, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[9] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials II - Orthogonal matching pursuit versus basis pursuit," *Found. Comput. Math.*, vol. 8, no. 6, pp. 737–763, 2008.

[10] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, pp. 33–61, Aug. 1998.

[11] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, Aug. 1998.

[12] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Computing*, vol. 31, pp. 890–912, Nov. 2008.

[13] A. Talari and N. Rahnavard, "GBCS: A two-step compressive sensing reconstruction based on group testing and basis pursuit," in *Military Comm. Conf.*, pp. 157–162, Nov. 2011.

[14] J. Mota, J. Xavier, and P. Aguiar, "Distributed basis pursuit," *IEEE Trans. Signal Process.*, vol. 60, pp. 1942–1956, Apr. 2012.

[15] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

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

[17] C. Li, *Compressive Sensing for 3D Data Processing Tasks: Applications, Models and Algorithms*. PhD thesis, Rice University, Houston, TX, Apr. 2011.

[18] C. Li, *An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing*. BiblioBazaar, 2011.

[19] J. Zhang, S. Liu, R. Xiong, S. Ma, and D. Zhao, "Improved total variation based image compressive sensing recovery by nonlocal regularization," in *Proc. IEEE Int. Symp. Circuits Systems (ISCAS)*, (Beijing, China), pp. 2836–2839, May 2013.

[20] A. Maleki and A. Montanari, "Analysis of approximate message passing algorithm," in *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–7, Mar. 2010.

[21] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices," *J. Stat. Mech. - Theory E.*, vol. 2012, p. P08009, Aug. 2012.

[22] S. Rangan, A. Fletcher, P. Schniter, and U. Kamilov, "Inference for generalized linear models via alternating directions and Bethe free energy minimization," in *Proc. Int. Symp. Inf. Theory (ISIT)*, pp. 1640–1644, June 2015.

[23] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Academy Sci.*, vol. 106, pp. 18914–18919, Nov. 2009.

[24] A. Maleki, *Approximate message passing algorithms for compressed sensing*. Stanford University, Nov. 2010.

[25] J. W. Goodman, *Introduction to Fourier optics / Joseph W. Goodman.* Englewood, Colo: Roberts  Co., 3rd ed. ed., 2005.

[26] J. Nye, "Numerical solution for diffraction of an electromagnetic wave by slits in a perfectly conducting screen," *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 458, Feb. 2002.

[27] A. Ma, You, Zhou, C. Rush, D. Baron, and D. Needell, "An approximate message passing framework for side information," 2018.

[28] H. Liu, You, Zhou, A. Beirami, and D. Baron, "Nonlinear function estimation with empirical bayes and approximate message passing," 2019.

[29] D. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, pp. 18914–18919, Nov. 2009.

[30] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. Arce, "Compressive hyperspectral imaging via approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, pp. 389–401, Mar. 2016.

[31] J. Vila, P. Schniter, S. Rangan, F. Krzakala, and L. Zdeborová, "Adaptive damping and mean removal for the generalized approximate message passing algorithm," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2021–2025, Apr. 2015.

[32] A. Ma, Y. Zhou, C. Rush, D. Baron, and D. Needell, "An approximate message passing framework for side information," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1875–1888, 2019.

[33] X. Wang and J. Liang, "Approximate message passing-based compressed sensing reconstruction with generalized elastic net prior," *Signal Process. Image*, vol. 37, pp. 19–33, Sept. 2015.

[34] H. V. Luong, J. Seiler, A. Kaup, S. Forchhammer, and N. Deligiannis, "Measurement bounds for sparse signal reconstruction with multiple side information," *Arxiv preprint arXiv:1605.03234*, Jan. 2017.

[35] F. Renna, L. Wang, X. Yuan, J. Yang, G. Reeves, A. Calderbank, L.Carin, and M. Rodrigues, "Classification and reconstruction of high-dimensional signals from low-dimensional features in the presence of side information," *IEEE Trans. Inf. Theory*, vol. 62, pp. 6459–6492, Sept. 2016.

[36] J. Mota, N. Deligiannis, and M. Rodrigues, "Compressed sensing with prior information: Strategies, geometry, and bounds," *IEEE Trans. Inf. Theory*, vol. 63, pp. 4472–4496, July 2017.

[37] M. Chen, F. Renna, and M. Rodrigues, "On the design of linear projections for compressive sensing with side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Barcelona, Spain), pp. 670–674, July 2016.

[38] A. Manoel, F. Krzakala, E. W. Tramel, and L. Zdeborová, "Streaming bayesian inference: theoretical limits and mini-batch approximate message-passing," in *Communication, Control, and Computing (Allerton), 2017 55th Annual Allerton Conference on*, pp. 1048–1055, IEEE, 2017.

[39] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 236–240, Feb. 2014.

[40] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, pp. 764–785, Feb. 2011.

[41] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *Arxiv preprint arXiv:1610.03082v1*, Oct. 2016.

[42] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *Proc. Int. Symp. Inf. Theory (ISIT)*, pp. 1588–1592, July 2017.

[43] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. Arce, "Compressive hyperspectral imaging via approximate message passing," *arXiv preprint arXiv:1507.01248*, 2015. submitted.

[44] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, England: Cambridge Univ. Press, 2004.

[45] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Royal Stat. Soc. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[46] L. A. Dalton and E. R. Dougherty, "Optimal classifier with minimum expected error within a Bayesian framework - Part 1: Discrete and Gaussian model," *Pattern Recognition*, vol. 46, pp. 1301–1314, Nov. 2012.

[47] R. C. Grinold and R. N. Kahn, *Active portfolio management: a quantitative approach for providing superior returns and controlling risk*. McGraw-Hill Companies, 2000.

[48] E. Fama and K. French, "Common risk factors in the returns on stocks and bonds," *J. Finan. Econ.*, vol. 33, no. 1, pp. 3–56, 1993.

[49] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *J. Finance*, vol. 48, no. 1, pp. 65–91, 1993.

[50] L. K. Chan, J. Lakonishok, and T. Sougiannis, "The stock market valuation of research and development expenditures," tech. rep., National Bureau of Economic Research, 1999.

[51] J. Fan, N. E. Heckman, and M. P. Wand, "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions," *J. Amer. Stat. Assoc.*, vol. 90, no. 429, pp. 141–150, 1995.

[52] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 40–42, 2002.

[53] A. Montanari, "Graphical models concepts in compressed sensing," *Compressed Sensing: Theory and Applications*, pp. 394–438, 2012.

[54] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. Allerton Conf. Commun., Control, and Comput.*, pp. 52–59, Sept. 2009.

[55] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *CoRR*, vol. abs/0906.3234, June 2009.

[56] D. Baron and M. F. Duarte, "Universal MAP estimation in compressed sensing," in *Proc. Allerton Conf. Commun., Control, and Comput.*, pp. 768–775, Sept. 2011.

[57] J. Zhu, D. Baron, and M. F. Duarte, "Recovery from linear measurements with complexity-matching universal signal estimation," *IEEE Trans. Signal Process.*, vol. 63, pp. 1512–1527, Mar. 2015.

[58] Y. Ma, J. Zhu, and D. Baron, "Approximate message passing algorithm with universal denoising and Gaussian mixture learning," *IEEE Trans. Signal Process.*, vol. 65, pp. 5611–5622, Nov. 2016.

[59] Y. Ma, J. Tan, N. Krishnan, and D. Baron, "Empirical Bayes and full Bayes for signal estimation," *Arxiv preprint arxiv:1405.2113v1*, May 2014.

[60] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, pp. 269–280, Jan. 2010.

[61] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, pp. 764–785, Feb. 2011.

[62] S. Rangan, P. Schniter, and A. Fletcher, "On the convergence of approximate message passing with arbitrary matrices," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 236–240, July 2014.

[63] A. Manoel, F. Krzakala, E. W. Tramel, and L. Zdeborová, "Swept approximate message passing for sparse estimation," in *Proc. Int. Conf. Machine Learning*, vol. 37, pp. 1123–1132, July 2015.

[64] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1983–2010, June 2005.

[65] D. Guo and C. C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 421–431, Apr. 2008.

[66] K. Sivaramakrishnan and T. Weissman, "A context quantization approach to universal denoising," *IEEE Trans. Signal Process.*, vol. 57, pp. 2110–2129, June 2009.

[67] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, Aug. 2001.

[68] V. Kekatos and G. Giannakis, "Sparse Volterra and polynomial regression models: Recoverability and estimation," *IEEE Trans. Signal Process.*, vol. 59, pp. 5907–5920, Dec. 2011.

[69] H. Qi and S. Hughes, "Using the kernel trick in compressive sensing: Accurate signal recovery from fewer measurements," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 3940–3943, IEEE, 2011.

[70] B. Nazer and R. Nowak, "Sparse interactions: Identifying high-dimensional multilinear systems via compressed sensing," in *Proc. Allerton Conference Commun., Control, and Comput.*, pp. 1589–1596, Sept. 2010.

[71] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proc. IEEE 45th Asilomar Conf. Signals, Syst., and Comput.*, pp. 799–803, Nov. 2011.

[72] S. S. Haykin, *Neural networks and learning machines*. Upper Saddle River, NJ: Pearson Education, third ed., 2009.

[73] R. Appleby and R. N. Anderton, "Millimeter-wave and submillimeter-wave imaging for security and surveillance," in *Proc. IEEE*, vol. 95, pp. 1683–1690, Aug. 2007.

[74] R. Appleby, D. A. Robertson, and D. Wikner, "Millimeter wave imaging: a historical review," in *Proc. SPIE*, vol. 10189, pp. 1018902–1018902–13, May 2017.

[75] D. McMakin, D. SHEEN, T. HALL, J. Tedeschi, and A. Mark JONES, "New improvements to millimeter-wave body scanners," in *Proc. 3DBODY.TECH*, (Montreal, Canada), pp. 263–271, Oct. 2017.

[76] D. M. Sheen, D. L. McMakin, and T. E. Hall, "Three-dimensional millimeter-wave imaging for concealed weapon detection," *IEEE Trans. Microw. Theory Tech.*, vol. 49, pp. 1581–1592, Sep. 2001.

[77] J. L. Fernandes, C. M. Rappaport, and D. M. Sheen, "Improved reconstruction and sensing techniques for personnel screening in three-dimensional cylindrical millimeter-wave portal scanning," in *Proc. SPIE*, vol. 8022, pp. 34–41, May 2011.

[78] L. Yujiri, M. Shoucri, and P. Moffa, "Passive millimeter wave imaging," *IEEE Microw. Mag.*, vol. 4, pp. 39–50, Sep. 2003.

[79] D. Hardmeier, F. Hofer, and A. Schwaninger, "The X-ray object recognition test (X-ray ort) - a reliable and valid instrument for measuring visual abilities needed in X-ray screening," in *Proc. Int. Carnahan Conf. Security Tech. (ICCST)*, (Las Palmas de Gran Canarias, Spain), pp. 189–192, Oct. 2005.

[80] J. P. O. Evans, M. Robinson, and D. Lacey, "Testing and evaluation of an advanced (3-d) X-ray screening system," in *Proc. Int. Carnahan Conf. Security Tech. (ICCST)*, (Alexandria, VA), pp. 50–54, Oct. 1998.

[81] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 83–91, Mar. 2008.

[82] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz imaging system based on compressed sensing," *Appl. Phys. Lett.*, vol. 93, p. 121105, Sep. 2008.

[83] S. D. Babacan, M. Luessi, L. Spinoulas, A. K. Katsaggelos, N. Gopalsami, T. Elmer, R. Ahern, S. Liao, and A. Raptis, "Compressive passive millimeter-wave imaging," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, (Brussels, Belguim), pp. 2705–2708, Sep. 2011.

[84] N. Gopalsami, S. Liao, T. W. Elmer, E. R. Koehl, A. Heifetz, A. P. C. Raptis, L. Spinoulas, and A. Katsaggelos, "Passive millimeter-wave imaging with compressive sensing," *Opt. Eng.*, vol. 51, pp. 091614 – 091614 – 9, Sep. 2012.

[85] V. M. Patel and J. N. Mait, "Passive millimeter-wave imaging with extended depth of field and sparse data," in *2012 Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 2521–2524, Mar. 2012.

[86] H. Shen, *Compressive Sensing on Terahertz Imaging*. PhD thesis, University of Liverpool, Liverpool, U.K., Jan. 2017.

[87] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, Mar. 2008.

[88] S. G. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Orlando, FL: Academic Press, Inc., 3rd ed., 2008.

[89] K. M. Schmitt and M. Rahm, "Evaluation of the impact of diffraction on image reconstruction in single-pixel imaging systems," *Opt. Express*, vol. 24, pp. 23863–23871, Oct. 2016.

[90] B. Abbey, "From Grain Boundaries to Single Defects: A Review of Coherent Methods for Materials Imaging in the X-ray Sciences," *JOM - Journal of the Minerals, Metals and Materials Society*, vol. 65, pp. 1183–1201, Sep. 2013.

[91] K. D. Mielenz, "Issues in optical diffraction theory," *J Res Natl Inst Stand Technol.*, vol. 114, pp. 83–97, Apr. 2009.

[92] J. Z. Buchwald and C.-P. Yeang, "Kirchhoff's theory for optical diffraction, its predecessor and subsequent development: the resilience of an inconsistent theory," *Archive for History of Exact Sciences*, vol. 70, pp. 463–511, Jan. 2016.

[93] J. T. Parker, V. Cevher, and P. Schniter, "Compressive sensing under matrix uncertainties: An approximate message passing approach," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 804–808, 2011.