UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ANALYSIS OF THE NYQUIST FOLDING RECEIVER (NYFR)

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

JAMES CAMDEN MARTIN
Norman, Oklahoma
2018

ANALYSIS OF THE NYQUIST FOLDING RECEIVER (NYFR)

A THESIS APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

Dr. Nathan Goodman, Chair

Dr. Mark Yeary

Dr. Caleb Fulton

Contents

List of Figures

Abstract

Increased utilization of wide swathes of spectrum motivate receivers with wide instantaneous bandwidth. Traditional wideband receivers inherently have high data rates that are difficult to process and store, and receivers that use multiple analog to digital converters to achieve wide bandwidth have high power usage and cost. Compressive sensing (CS) provides a potential low-data-rate and low-power solution in environments where only a small portion of the wide spectrum monitored is in use at one time, through sub-Nyquist sampling at the information rate. The Nyquist Folding Receiver (NYFR), proposed by Fudge et al., is one such promising CS architecture. This thesis investigates the design trade-offs inherent to any NYFR receiver. Basic applications including pulse detection, angle of arrival estimation, and processing of communication signals are simulated. Finally, a prototype receiver was used to experimentally demonstrate the capabilities of a NYFR with an instantaneous bandwidth of 18 GHz while only sampling at 1.5 GSPS.

Chapter 1

Introduction

Increasing utilization of large amounts of spectrum motivates monitoring of wide bandwidths for applications such as cognitive radio where empty channels are identified to more efficiently use the spectrum, or for signal detection and capture for signals intelligence. Traditional methods of accomplishing this are expensive in both price and power usage. Designing analog to digital converters that operate at high frequencies is a difficult affair, and even when the ADC portion is solved, the tremendous amount of data generated requires high throughput to the processor or large quantities of storage.

Current wideband receivers fall largely in two categories. The first have multiple receive channels, each filtered to sample a small slice of spectrum, and an ADC for each. The additional hardware for each channel causes high cost. The second are sweeping receivers. These move between multiple frequency zones, monitoring a wide bandwidth over time with fewer components. The major drawback to the sweeping receiver is that this does not give a wide instantaneous bandwidth; for example, a frequency hopping radar that broadcasts for only a very short time before changing frequencies may get missed as the receiver is sweeping through other frequencies.

The modern field of compressive sensing gives potential for a receiver archi-

tecture that solves both the issues of handling large amounts of data and the cost of wide instantaneous bandwidth. Applications of the theory allow for sampling a wide bandwidth with a single low sample-rate ADC. The Nyquist Folding Receiver (NYFR), proposed by Fudge et al. [1], is one particularly promising design of a compressive sensing system. The NYFR significantly undersamples input signals causing them to alias such that original frequencies would normally be lost, but the NYFR encodes the original frequency as modulation on the input signals so that all of the original information is preserved. Compressive sensing concepts allow for this encoded frequency information to be recovered, with the assumption that the original spectrum was sparse. Fudge developed a model for a prototype NYFR that uses in-phase and quadrature-phase (I/Q) sampling, and demonstrated frequency recovery of continuous wave tones and on-off key (OOK) signals.

This thesis provides a model for a NYFR that uses real-valued samples rather than I/Q data, and adds results from a simulation of the real-valued receiver model used to investigate the trade-offs inherent to any NYFR architecture. The information given about the trade-offs potentially enables the selection of operating parameters that are closer to optimal for specific environments. The thesis shows successful recovery of pulses and communication signals on a prototype receiver, and notably the recovered signals have wider bandwidth than the CW tones and OOK signals demonstrated by Fudge. Finally, this thesis adds experimental verification of basic applications, such as direction finding and demodulation of communication signals.

1.1 Organization

Chapter 2 begins with a short introduction to compressive sensing that is needed to understand the underlying theory that allows the receiver to work. An overview of compressive sensing reconstruction algorithms is given. Finally a survey is given of other compressive sensing receiver architectures. Chapter 3 goes in detail on the specifics of the Nyquist Folding Receiver itself, both the system model and the physical implementation. Chapter 4 describes the effect of various design parameters on the performance of a NYFR. Several basic applications of the receiver are explored through simulation. Chapter 5 gives experimental results from a physical NYFR prototype, confirming that the simulated applications are indeed possible in reality. Chapter 6 gives the final conclusions, overview of the work, and potential future research topics.

Chapter 2

Background

2.1 Compressive Sensing Theory

First, before going into detail about the NYFR, it is helpful to understand the concept that allows the NYFR to sample far below the Nyquist rate, without, under specific circumstances, losing any information. This chapter will give a brief overview of compressive sensing (CS) theory, which is a mathematical framework for sampling a system at the information rate rather than the Nyquist rate. While the work of Nyquist and Shannon established that sampling at a rate twice the highest frequency of a continuous finite-bandwidth signal is a sufficient condition for recovering all the information in the signal, this condition is not necessarily the only sufficient one. Nyquist hinted at the idea that the actual information content may only make up a limited portion of an analog signal in his description of separate intelligence carrying and inert components that form telegraph signals in 1924 [2], insinuating that sampling at the Nyquist-Shannon rate may waste some effort on recovery of the inert component. However, a method for sampling closer to the information rate for a general set of circumstances did not arrive until the modern theory of compressive sensing was developed by Candès, Romberg, Tao, and Donoho over the course of 2004 to 2006 [3]. Compressive sensing added a mathematical framework capable of calculating the lower bound on how many

measurements are required to recover the full information content of a signal. Similarities to modern compressive sensing can be found in the historical use of the ℓ_1-norm in several different fields, ranging from computational statistics to the creation of seismic images in the 1970s from undersampled data [4]. Algorithms applicable to the recovery of sparse solutions have existed at least since matching pursuit was formulated in 1993 [5].

2.1.1 Compressed Signals

As the name implies, signals that are compressible are central to the concept of compressive sensing. If a signal can be represented in some vector space by a small number of linearly independent vectors, it has a sparse representation in some representation basis $\mathbf{\Psi}$. The number of nonzero coefficients required to represent the signal is the sparsity k of the signal. Many real world signals are not strictly sparse, but rather are said to be compressible. A compressible signal has the majority of its energy contained in a small number of coefficients, such that when most of the coefficients are approximated as zero the original signal is still well represented. The level of sparsity of the compressed signal determines the efficiency of the compressive sampling [6].

The mutual coherence μ of two matrices is the absolute value of the maximum cross-correlation between any pair of columns from the matrices, as defined by

$$\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \max_{k,j} |\langle \phi_k, \psi_j \rangle| \quad . \tag{2.1}$$

The value of μ is bounded by $\frac{1}{\sqrt{n}} \leq \mu \leq 1$ for a pair of $N \times N$ matrices [7]. Now consider some sensing basis $\mathbf{\Phi}$ that is incoherent with the representation basis, that is $\mu(\mathbf{\Phi}, \mathbf{\Psi})$ is minimal. The information in the small number of com-

pressible coefficients is spread into many coefficients in the sensing basis. Each measurement in the sensing basis records some information about the full support of the signal in the representation domain. For an illustrative example, a constant frequency sine wave is sparsely represented as a Dirac delta in the frequency domain, and extends infinitely in the time domain. It is known by the Nyquist sampling theorem that in order to recover the frequency content of any signal it is sufficient to sample in the time domain at twice the frequency of the highest frequency component. Additionally, if it is known a priori that there is only a single sinusoid present, then its frequency is known after sampling for one period, not requiring all of the infinite number of time domain coefficients to be sampled. It follows that sampling the entire support of the sensing basis is not a requirement for sampling the total information content of the signal.

For a signal that is sparse in the representation basis, a small number of compressive samples taken in the sensing basis may be sufficient to measure all the needed information. Not every measurement necessarily records new information, consider for the time-frequency domain single sinusoid example if the signal is being uniformly sampled in the time domain at some rate. If, for example, the sinusoid's frequency is exactly a multiple of the sample rate every measurement will be made at the same point on the sine wave and thus the frequency is not measured. If however the sinusoid is sampled in random points in time, each measurement will be at a different point on the sinusoid and thus adds new information. Random sampling is indeed a popular concept used in many compressive sensing systems, as *any* random matrix is likely to be highly incoherent with any representation basis. Random sampling is not the only effective method though, as a structured sampling scheme could be designed to sample at the best points of the sinusoid, or at least a sufficient number of them. The Nyquist Folding

Receiver is such a structured sampling scheme [8]. The question now becomes one of how can the original signal be recovered from these measurements, and how many measurements are truly needed to identify the original sparse signal.

2.1.2 An Underdetermined Problem

Even though the acquired compressed measurements contain most of the information that the original signal had, the measurements are not useful for most applications until the original signal in the sparse representation basis is reconstructed. This applies to compressed information in general, for example, when viewing compressed JPEG2000 computer graphics, the end user does not stare at stored wavelet coefficients, instead the image must first be decompressed to a 2D grid of colours. Recovery of the original signal from compressive measurements is not straightforward, as it requires solving an under-determined linear system. Let $\mathbf{z_\Phi} \in \Re^m$ be a vector of m compressed measurements, $\mathbf{A} \in \Re^{m*n} = \mathbf{\Phi\Psi}$ be the measurement matrix, and $\mathbf{x_\Psi} \in \Re^n$ be the original signal of length n that is sparse in $\mathbf{\Psi}$, then

$$\mathbf{z} = \mathbf{Ax} \tag{2.2}$$

is the linear system that is to be solved for $\mathbf{x}$. This is the recovery of the signal $\mathbf{x}$ from measurements $\mathbf{z}$. As the point of compressive sensing is to take only as many measurements as are necessary to identify the original sparse signal, $m < n$ measurements are taken, thus (2.2) has infinitely many solutions. However, it is known that the original signal is sparse in the representation domain. The ℓ_0 norm is the measure of sparsity, the number of non-zero components of a vector $\mathbf{x}$. By adding the constraint that the original signal is known to be the most sparse, a unique solution can be found.

2.2 Reconstruction Algorithms

If the original signal is know to be sparse, the most straightforward approach for solving the system would be to solve

$$\text{minimize } ||\mathbf{x}||_0 \text{ subject to } \mathbf{Ax} = \mathbf{z} \tag{2.3}$$

by searching through all possible solutions and selecting the one that is the most sparse. This brute-force approach is not computationally feasible though, as it is a NP-hard problem. Instead, there are several major families of algorithms that have been designed to more efficiently solve for the original sparse signal from the compressive measurements [9]. Some may be better suited to certain problems, depending on the structure of the signals and the sensing matrix.

The algorithms can be broadly grouped into five families: greedy iterative, iterative thresholding, combinatorial, convex relaxation, and non-convex minimization. Greedy iterative algorithms solve the linear system in a series of iterative steps. First the residual $\mathbf{r}$ is set equal to $\mathbf{z}$. Now at each step a locally optimal, that is greedy, selection of the column of $\mathbf{A}$ that correlates most with $\mathbf{r}$ is made. The coordinate of the column is recorded as part of the support of $\mathbf{x}$. The contribution of the coordinate to $\mathbf{z}$ is estimated and subtracted from $\mathbf{r}$, then the next iteration is performed on the residual. Removing the contribution of the most highly correlated column decreases the least square error with each iteration. The major changes between different algorithms are the stopping criteria and the details of estimating each column's contribution. In general these algorithms perform best with signals that are very sparse, with the quality of recovery rapidly becoming worse as sparsity decreases. The need for very sparse

signals can cause issues in systems with significant noise, as noise adds additional non-zero coefficients in the representation basis; one algorithm called Regularized Orthogonal Matching Pursuit worked well in the ideal NYFR simulations but had a much lower success rate on recovering the frequency support from actual data with multiple signals and noise present.

Iterative thresholding algorithms limit the solution set for noisy measurements with some form of soft or hard thresholding. There are many variations of these algorithms, with the thresholding function dependent on the specific problem. After thresholding the solution is found through iteration similar to the greedy iterative algorithms. Message passing algorithms are a modification of iterative algorithms that include additional variables with each iteration, messages, that are used to speed convergence.

Combinatorial algorithms use group testing to recover the signal. They are computationally fast, but require the sensing matrix $\mathbf{A}$ to itself be sparse. Thus they are only useful for very specific problems, and the NYFR is not one of them. While it is shown later that the NYFR sensing matrix can be considered block-sparse, consisting of blocks for each Nyquist zone the system operates on, the blocks themselves are large for any practical system with usable baseband bandwidth.

2.2.1 Convex Optimization

The algorithms that give the best reconstruction quality for the NYFR receiver implementation are in the convex optimization family. If instead of attempting to minimize the ℓ_0 norm directly, we relax the problem to minimizing the ℓ_1 norm, then signal recovery becomes a convex problem with the potential for efficient linear solutions. The ℓ_1 norm is the sum of the absolute value of the vector

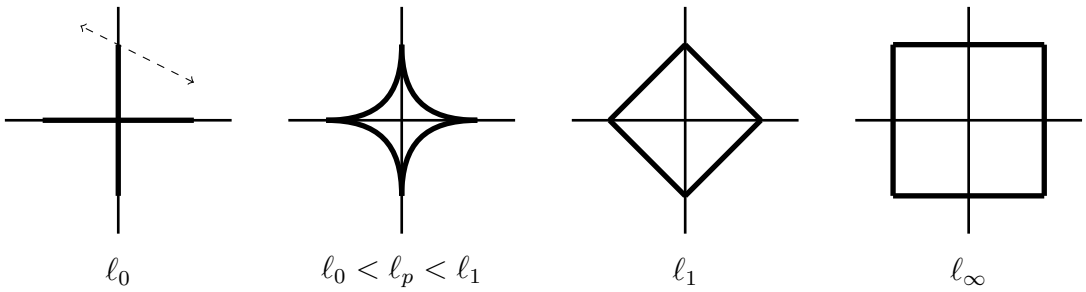

Figure 2.1: Unit sphere with p-norm. The dashed line on ℓ_0 gives an example of a region of feasible solutions.

components as defined by

$$||\mathbf{x}||_1 = \sum_{i=1}^{N} |x_i|. \tag{2.4}$$

In addition to being convex, the ℓ_1 norm is not smooth, so regularization on it also promotes sparsity. Convex optimization has broad application, which has caused it to become a well-studied problem with reliable and efficient solutions [10]. For a strictly convex problem, any local minimum found is guaranteed to also be the unique, global minimum, giving an obvious advantage over a nonconvex problem where one could imagine an algorithm might get stuck in a local minimum between two hills, not knowing that past the next hill is a better solution.

Figure 2.1 illustrates the concept of general ℓ_p norms in two dimensions. In this case the region of feasible solutions to $\mathbf{Ax} = \mathbf{b}$ is a contour. The geometry of the norms are represented by unit spheres in Figure 2.1, that is the plot of all points for which ℓ_p=1. Let a unit sphere grow from the origin until it first intersects the solution region, then that first intersection is the minimum-norm solution. For any solution on the ℓ_0 norm, the support is of size one, as the intersection is either entirely on the x or the y axis. Thus the ℓ_0 norm is the most sparse.

There is active research seeking the potential advantages of using ℓ_p quasi-

norms with $0 < p < 1$, forming the non-convex family of compressive reconstruction algorithms [11]. The $0 < p < 1$ is more sparse than the ℓ_1 norm as many solutions have one large and one small coefficient; it can be seen on the figure that most intersections with the norm will have either a large x or a large y component, but not both. A simple test to see that the norm is not convex is to attempt to draw a line between two points on the norm without intersecting the inside of the norm, this is possible on any pair of points from different axes on the ℓ_0 norm and for many points on the $0 < p < 1$ norms, so these norms are not convex. Thus the $0 < p < 1$ norms are said to further promote sparsity, but are no longer convex causing more difficulty in solving the problem efficiently.

The ℓ_1 norm still promotes sparsity, just to a lesser degree than $p < 1$, and is convex. Since the ℓ_1 norm is convex, when a local minimum is found, it is guaranteed to also the global minimum, giving the optimal solution. Thus the ℓ_1 norm is effective for compressive sensing reconstruction, as we are recovering signals that are sparse, and efficient algorithms for solving convex problems exist.

Any norm with $p > 1$ is also convex, but is also a smooth function with most intersections having both a large x and a y component, so it is not useful for promoting sparsity. The first intersection of the solution region with ℓ_∞ has a support of size two, evenly split between the x and y axes, clearly not sparse. The concept of the two-dimensional unit sphere extends to n many dimensions, with for example the ℓ_0 forming a many dimensional spiky ball for compressive sensing problems with a large representation basis.

There are three commonly used convex relaxations of the compressive sensing problem. The first is basis pursuit (BP), which is defined as

$$\text{minimize } ||\mathbf{x}||_1 \text{ subject to } \mathbf{Ax} = \mathbf{z} \quad . \tag{2.5}$$

The exact solution sought by basis pursuit does not work well for analog measurements containing noise, so for these cases the basis pursuit denoising (BPDN) formulation is used instead, which is

$$\text{minimize } ||\mathbf{x}||_1 \text{ subject to } ||\mathbf{A}\mathbf{x} - \mathbf{z}||_2 \ \leq \sigma \quad , \tag{2.6}$$

where σ gives a tolerance for how exact the solution must be and is usually derived from an estimate of the noise power. It is worth noting that in most sampling schemes the noise in the original representation basis will be folded into the smaller number of measurements taken, multiplying the noise variance by n/m, so the noise power can be significant in a real system [12]. The least absolute shrinkage and selection operator (LASSO) is the final common relaxation. LASSO pre-dates modern compressive sensing, having originally been created as a regression analysis method for statistics, but is directly equivalent to the BPDN problem [13] as it is defined by

$$\text{minimize } ||\mathbf{A}\mathbf{x} - \mathbf{z}||_2 \text{ subject to } ||\mathbf{x}||_1 \ \leq \tau \quad . \tag{2.7}$$

The problem solved for the NYFR is the BPDN one rather than BP, as the goal is to recover signals on real hardware with noisy measurements. Additionally, in compressive sensing literature, the BPDN formulation is more common than LASSO.

If $\mathbf{x}$ is sufficiently sparse and the sensing matrix $\mathbf{A}$ is appropriately designed, the convex relaxation of the problem will yield the exact solution of $\mathbf{x}$. An s-sparse signal is defined as having a support size of s, that is the vector has s many nonzero elements. Candès introduced the restricted isometry property (RIP) as

a means to characterize the quality of the sensing matrix [14]. For each integer $s = 1, 2, ..., N$ the restricted isometry constant δ_s of the matrix $\mathbf{A}$ is defined as the smallest number such that

$$(1 - \delta_s)||\mathbf{x}||_2^2 \leq ||\mathbf{A}\mathbf{x}||_2^2 \leq (1 + \delta_s)||\mathbf{x}||_2^2 \tag{2.8}$$

for all s-sparse vectors. Thus, a sensing matrix satisfies the (s,δ)-RIP if the above condition holds for every s-sparse vector $\mathbf{x}$, and perfect reconstruction is guaranteed via convex relaxation. Various reconstruction guarantees based on the RIP have also been found for other reconstruction methods [15]. While this is a commonly used property, solving the general problem of whether an arbitrary sensing matrix $\mathbf{A}$ satisfies the (s,δ)-RIP is NP-hard.

As calculating a value for δ is a hard problem for many cases, it is common to turn to the measure of coherence between the sensing and representation bases to predict reconstruction performance. For the case of ℓ_1 optimization with a signal $\mathbf{x} \in \mathbb{R}^n$ and a uniform random sampling basis Candès found that the probability of successful reconstruction exceeds $1 - \delta$, if

$$m \geq C \cdot \mu^2(\mathbf{\Phi}, \mathbf{\Psi}) \cdot s \cdot \log(n/\delta) \tag{2.9}$$

where m is the number of measurements, C is some positive constant, μ is the mutual coherence of the two bases, and s is the sparsity of the signal [16]. This shows that higher incoherence between the bases or greater sparsity lowers the necessary number of compressive measurements for reconstruction.

The spectral projected-gradient SPGL-1 convex reconstruction algorithm in particular was found to give high reconstruction performance with low computa-

tional time for the NYFR system [17]. The specific formulation that solves the BPDN problem is used as there is significant noise in the NYFR chain. Much of the convergence speed of the SPGL-1 algorithm is due to its exploitation of a Pareto frontier [18]. When there are two or more variables being optimized, the Pareto frontier is the curve formed by the set of values where one variable cannot be improved without harming the other variable. As the BPDN problem is constrained by both $||\mathbf{x}||_1$ and $||\mathbf{Ax}-\mathbf{b}||_2$, a Pareto frontier is formed by the two norms. The authors prove that this Pareto curve is continuously differentiable, enabling the use of a Newton root-finding algorithm to solve a non-linear equation describing the curve. This approach gives a fast iterative algorithm that outputs a small search space of optimal solutions. The second innovation of the algorithm is an optimized method of performing least squares minimization with orthogonal projections of vectors on a convex set. This projection algorithm is extended to complex vectors, and thus useful for recovering signals in the frequency domain.

2.3 Sampling Paradigms

So far the mathematical concept of compressive sensing has been discussed, but not how CS systems are realized in hardware. As a random sensing basis gives a low level of coherence with any other basis, random sampling schemes are a popular method of implementing a compressive sensing system. Perhaps one of the most popular examples of such a system is the Rice single-pixel camera. An array of digitally controlled micro mirrors are randomly flipped on or off to reflect random linear combinations of light onto a single CCD sensor. A set of these single-pixel measurements is solved for the random sequence used and the wavelet domain compressed image is reconstructed.

The goal of the NYFR is to perform analog-to-information conversion at low

sample rates. There are several systems that use different sampling schemes to meet the same goal. Most competing systems use a random sensing matrix as opposed to the NYFR's structured sensing matrix.

The Random-Modulation Pre-Integrator architecture was demonstrated with 2 GHz of instantaneous bandwidth and a 320 MS/s sample rate. The RMPI modulates the input signal with a pseudorandom binary sequence (PRBS), integrates the output of the modulator, then samples at a low rate. A potential limitation is that the PRBS generation and mixing must occur at the Nyquist rate. Successful reconstruction of two 450 ns pulses with different carrier frequencies that overlap in time has been shown [19].

Northrop-Grumman's Non-uniform Sampler samples at random time intervals with a nonuniformly clocked sample-and-hold circuit [20]. An 8192-bit PRBS controls the sample times. The instantaneous bandwidth is 2.2 GHz, and a low-sample-rate ADC samples at 230MS/s. Experiments successfully reconstructed GSM signals with a bit error rate under 10^{-5}.

The Modulated Wide-band Converter multiplies an input signal with a bank of periodic waveforms before low pass filtering and sampling at below the Nyquist rate [21]. This architecture is an implementation of the Xampling scheme [22]. The bank of modulators aliases different parts of the spectrum, from the entire Nyquist range, into the baseband. To achieve a certain performance level, this scheme has a trade-off between the sample rate of each channel and the number of channels.

The above sampling paradigms for analog signals are all limited in that either some system components still have to operate at the Nyquist sampling frequency, specifically to generate the PRBS, or multiple receive channels are required, increasing power requirements and cost. The NYFR does not have these con-

straints; the highest frequency that must be generated is only slightly higher than the sampling clock, and only a single receive chain is required.

Chapter 3

Nyquist Folding Receiver

Now that a brief overview of compressive sensing has been given, this chapter will cover the hardware of the Nyquist Folding Reciever (NYFR) and the CS model that links the NYFR's sampling and representation bases. In 2008 Fudge et al. considered a compressive sampling scheme that would use structured non-uniform sampling rather than random sampling for recovery of signals that are sparse in the frequency domain [1]. Random sampling schemes require some components to operate at the high frequency Nyquist rate of the signals, whereas a non-uniform sampling scheme could potentially be realized with only components at the Nyquist rate of the baseband. If a signal is undersampled, below the Nyquist rate, it will alias to some lower frequency and thus information on the original frequency will be lost. However, the original structure of the signal will be largely preserved, as long as the sampling rate is higher than the bandwidth of the signal. The NYFR encodes the original frequency information onto the aliases of input signals, in such a way that compressive sensing concepts can be used to reconstruct the original inputs completely.

A prototype system was built with experimental results and a matrix model for the compressive sensing equation published in 2012 [8]. In depth information about Fudge's prototype hardware is not provided however, and the model is not

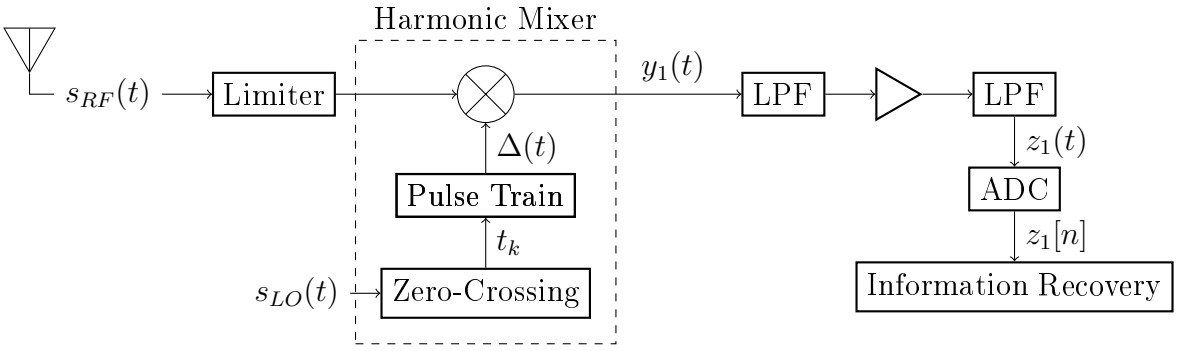

Figure 3.1: Basic block diagram for one NYFR channel.

directly usable for the system used in this thesis. In particular Fudge's model uses a complex in-phase and quadrature-phase representation with an ADC that provides I/Q data. How the NYFR hardware manipulates input signals naturally leads to the model describing the system, so first we give a detailed description of the prototype receiver hardware.

3.1 Hardware Implementation

A high-level block diagram of the prototype NYFR receiver is shown in Figure 3.1. A Tektronix arbitrary waveform generator (AWG) is used along to generate the reference local oscillator signal $s_{LO}(t)$. Use of the AWG gave flexibility in experimenting with additional modulation patterns and parameters. In a production system, a cheaper device that only requires components operating up to the desired baseband sampling rate could be used to generate the modulated LO signal.

Inside the harmonic mixer, a step recovery diode is used to create a very narrow, high-bandwidth pulse on each positive zero crossing of $s_{LO}(t)$. Figure 3.2 illustrates the pulse generation for a constant-frequency LO and a frequency-modulated LO. Let $s_{LO}(t)$ be a real sinusoid $\sin(\omega_{LO}t)$, then the ideal approxi-

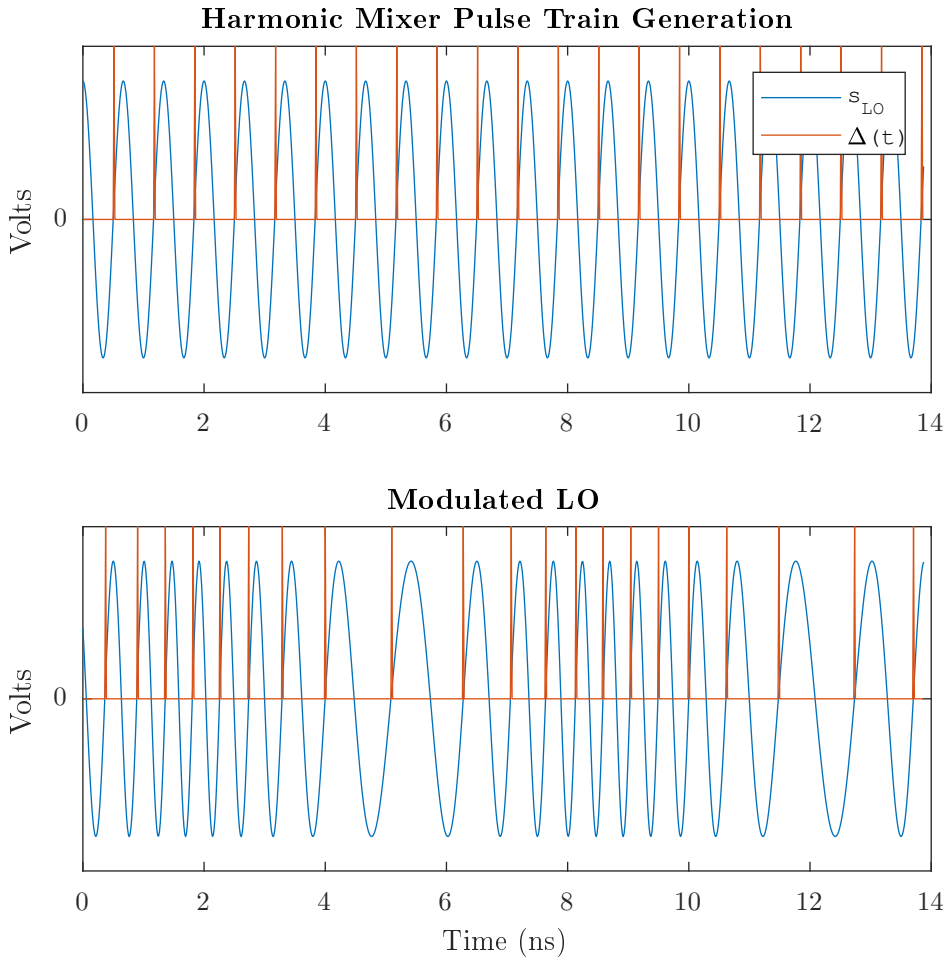

Figure 3.2: Narrow pulses are generated on the zero crossings of the LO port. The lower plot shows how the frequency of pulse generation varies along with a frequency-modulated LO.

mation of the diode output is by a train of Dirac delta functions

$$\Delta(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT) \tag{3.1}$$

with T being the period of $s_{LO}(t)$. The Fourier transform of a pulse train is a

comb function with a spike at the harmonics of ω_{LO} such that

$$\Delta(\omega) = \sum_{k=-\infty}^{\infty} \delta(\omega - k\omega_{LO}) \quad . \tag{3.2}$$

In real hardware the amplitudes of the higher harmonics roll off, as the pulses generated by the mixer are not ideal impulse functions, limiting the useful bandwidth of the device. The HL9313 mixer used for the receiver has a nominal 3 dB bandwidth of 15 GHz. Thus the nonideal mixer creates pulses with some shape $p(t)$, so that the pulse train takes the form

$$\Delta(t) = (\sum_{k=-\infty}^{\infty} \delta(t - kT)) * p(t) \quad , \tag{3.3}$$

and has a spectrum defined by

$$\Delta(\omega) = P(\omega) \sum_{k=-\infty}^{\infty} \delta(\omega - k\omega_{LO}) \tag{3.4}$$

where $P(\omega)$ drops by 3 dB at approximately 15 GHz.

The pulse train is then mixed with the RF input port. As the RF input $s_{RF}(t)$ mixes with each component of $\Delta(t)$, copies of the input are created at $\omega_{RF} \pm k\omega_{LO}$. Thus, the harmonic mixer is effectively sampling s_{RF} at each zero crossing which causes aliasing harmonics of the of the mixer's sample rate, ω_{LO}. The copy at baseband with a frequency less than $\frac{s_{LO}}{2}$ is the copy produced by the harmonic closest to the original frequency.

Now let the $s_{LO}(t)$ be a real frequency modulated sinusoid centered at ω_{LO} with phase modulation $\theta(t)$

$$s_{LO}(t) = \sin(\omega_{LO}t + \theta(t)) \tag{3.5}$$

and $\theta(t)$ a sinusoidal modulation defined as

$$\theta(t) = \frac{F_\Delta}{F_{mod}} \sin(2\pi F_{mod} t) \quad . \tag{3.6}$$

For the hardware implementation used in this thesis, ω_{LO} is 1.5 GHz. According to the Dirac scaling property [8], the non-uniform pulse train generated in the harmonic mixer takes the form

$$\widetilde{\Delta}(t) = \varphi'(t) \sum_k 2\pi\delta(\varphi(t) - 2\pi k) \tag{3.7}$$

where $\varphi = \omega_{LO} t + \theta(t)$. The identity $2\pi \sum_k \delta(t - 2\pi k) = \sum_k e^{jkt}$ is then used to rewrite (3.7) as

$$\widetilde{\Delta}(t) = (\omega_{LO} + \theta'(t)) \sum_k e^{jk(\omega_{LO} t + \theta(t))} \quad . \tag{3.8}$$

Considering that the modulation is narrowband such that $|\theta'(t)| << \omega_{LO}$, (3.8) can be approximated by

$$\widetilde{\Delta}(t) \approx \omega_{LO} \sum_k e^{jk(\omega_{LO} t + \theta(t))} = \omega_{LO} \sum_k e^{jk\omega_{LO} t} e^{jk\theta(t)} \quad . \tag{3.9}$$

The $e^{jk\theta(t)}$ term shows that modulating the frequency of $s_{LO}(t)$ spreads the spectrum of the pulse train harmonics, with the amount of spread increasing as the magnitude of the harmonic number k increases. Thus each harmonic has a unique pattern encoded on it. Figure 3.3 shows the pulse train spectrum for such a modulated pulse train with a peak frequency deviation F_Δ of 4 MHz. Figure 3.4 shows the spectrum of a pulse train for a higher frequency deviation, F_Δ=15 MHz. The energy content of each harmonic is spread out over a wider range of frequencies.

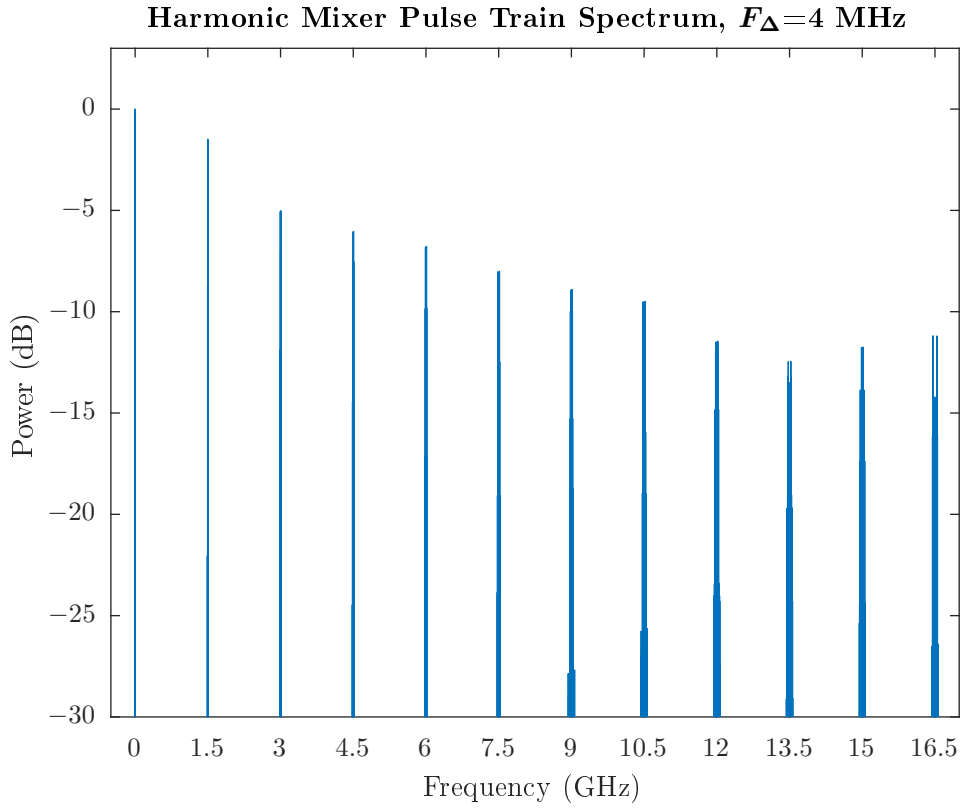

Figure 3.3: Simulated spectrum of the pulse train inside the harmonic mixer for a sinusoidal $s_{LO}(t)$ with a F_Δ of 4 MHz. As the frequency increases the amplitude of the harmonics drop with $P(\omega)$, and the energy of the higher harmonics is spread out over a wider frequency range.

The selection of F_Δ requires a performance trade off that will be examined in depth later in the thesis. Note that the typical peak frequency deviation F_Δ used throughout this thesis is 4 MHz unless otherwise stated.

Next, let there be a real, narrowband RF input signal with a center frequency at ω_{RF} and phase ϕ

$$x(t) = \cos(\omega_{RF}t + \phi) \quad . \tag{3.10}$$

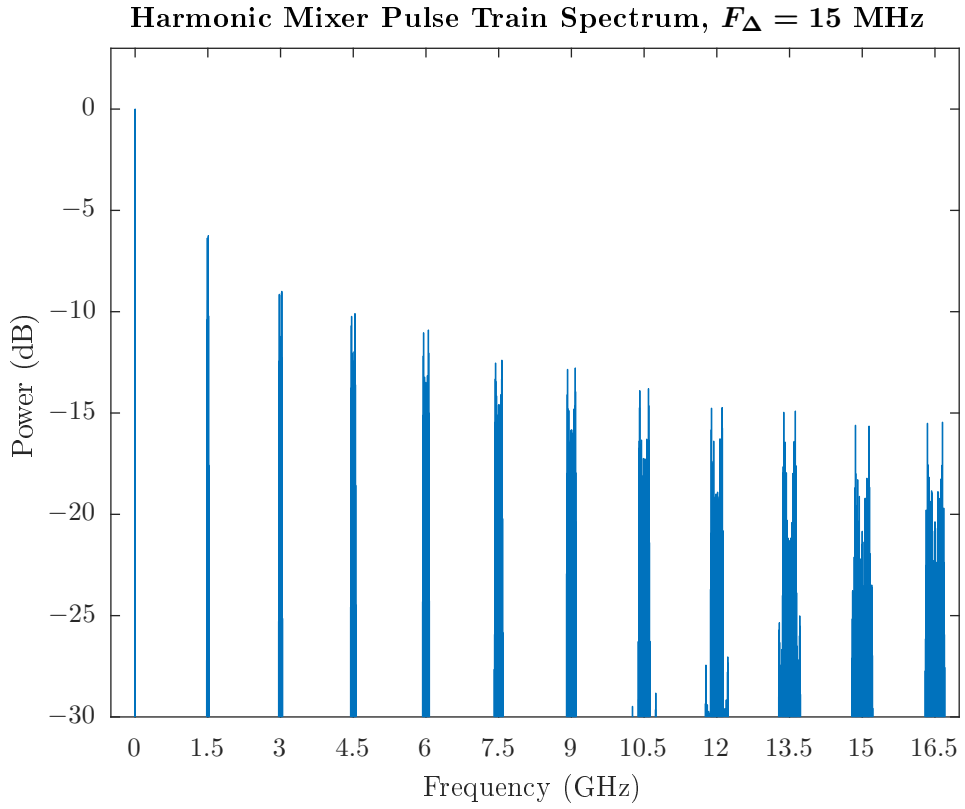

Figure 3.4: Spectrum of the pulse train inside the harmonic mixer for a sinusoidal $s_{LO}(t)$ with a F_Δ of 15 MHz. Each harmonic has its energy spread over a wider range of frequencies than the equivalent harmonics in Figure 3.3.

The harmonic mixer output then becomes

$$y(t) = x(t)\Delta(t) \approx \sum_k \cos(\omega_{RF}t + \phi)e^{jk\omega_{LO}t}e^{jk\theta(t)} \quad . \tag{3.11}$$

The output of the harmonic mixer is followed by an anti-aliasing filter. Let this low pass anti-aliasing filter (LPF) have a cutoff frequency of $\frac{1}{2}\omega_{LO}$, and let k_H be the harmonic in the Fourier series of the impulse train that satisfies $0 \leq |\omega_{RF} - k_H\omega_{LO}| \leq \frac{1}{2}\omega_{LO}$, in other words $k_H\omega_{LO}$ is the harmonic closest to the

input RF frequency. The output of the LPF is then

$$z(t) = \cos((\omega_{RF} - k_H\omega_{LO})t + \phi)e^{jk_H\theta(t)} \quad . \tag{3.12}$$

It is clear that the copy of the RF input passed by the filter is the product of only the harmonic closest to ω_{RF}. Recall that each harmonic has a unique frequency spread from the $s_{LO}(t)$ modulation. This means the resulting mixer and filter output for two copies of an input signal, which are identical except for having carrier frequencies from different Nyquist zones, are unique even if the inputs fold to the same baseband frequency. Since information on the original frequency support is preserved even though the input signal is aliased, it is possible to reconstruct the original signal. Figure 3.5 illustrates this concept. The figure plots the simulated output spectrum for a 9.2 GHz continuous-wave (CW) RF input and a sinusoidally modulated $s_{LO}(t)$. Examination of the plotted spectrum shows that no two harmonic products are identical, and only a single product is to the left of the LPF cutoff frequency, thus the original zone is uniquely identified. The spectrogram in Figure 3.6 shows a simulation of the mixer and filter output for a 3.5 GHz and 9.2 GHz tone with a sinusoidal modulated $s_{LO}(t)$ with F_Δ=15 MHz. The 9.2 GHz tone is clearly modulated over a wider frequency range.

For a non-ideal system the careful consideration of the frequency response of the LPF must be made. The ideal cut-off in Figure 3.5 passes the desired harmonic product without any attenuation, while perfectly blocking the harmonic products from the wrong Nyquist zones. The actual filter used in this NYFR implementation has a cutoff frequency f_c of 750 MHz, half the sampling rate of the system. If there is non-negligible attenuation at half the $s_{LO}(t)$ center frequency, as is the case with the receiver prototype, then the receiver will have

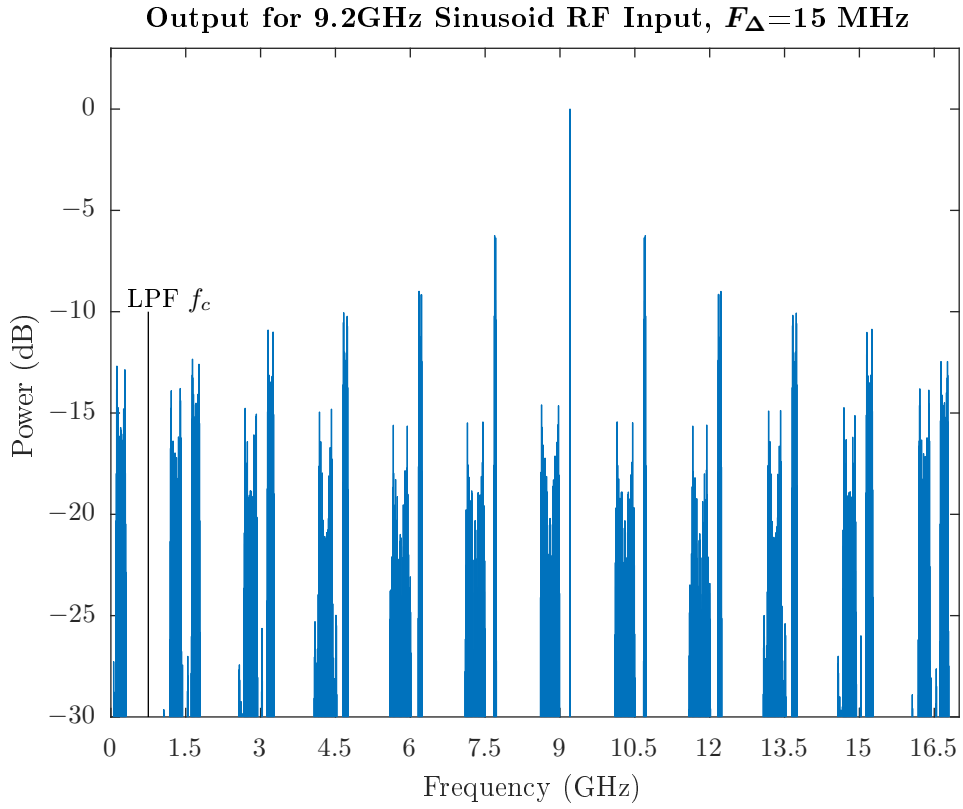

Figure 3.5: Output spectrum of the harmonic mixer for a 9.2 GHz tone RF input

blind-spots at all multiples of $\frac{f_{LO}}{2}$. Any signal that folds to a frequency close to $\frac{f_{LO}}{2}$ will be attenuated below the noise floor. The blind zone issue can be mitigated by using two channels with different sampling frequencies chosen such that the two resulting blind zones never occur at the same frequencies throughout the desired bandwidth of the receiver, but this obviously leads to higher system cost.

On the other hand, if the LPF cutoff is set to a higher frequency, or the filter order is low causing the frequency roll-off of the filter to be less steep, such that frequencies higher than $\frac{f_{LO}}{2}$ are not attenuated enough, unwanted aliases from harmonics outside of the RF signal's original Nyquist zone will be passed by the

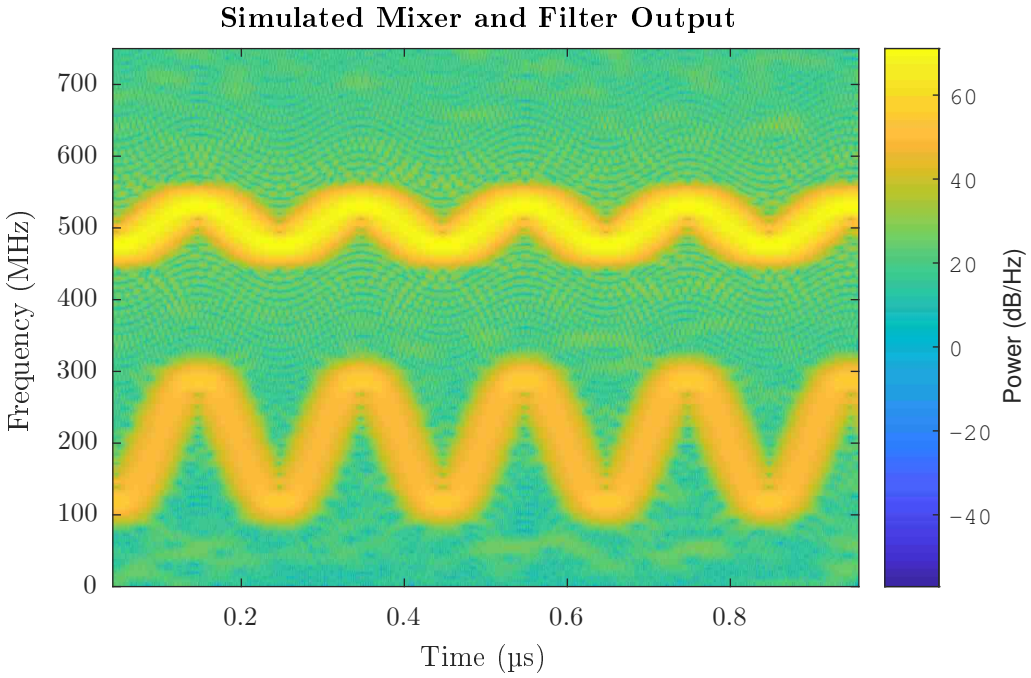

Figure 3.6: Spectrogram of the simulated harmonic mixer and LPF output for a 3.5 GHz and 9.2 GHz tone RF input. For simple RF input cases the spectrogram may be used to visually estimate the original input frequencies. The $s_{LO}(t)$ used is sinusoidal modulation with F_{Δ} =15 MHz and F_{mod}=5 MHz. As the center frequency of $s_{LO}(t)$ is f_{LO} =1.5 GHz, the Nyquist rate baseband has a bandwidth of 750 MHz. The 9.2 GHz tone folds to 3.5 GHz - 6·1.5 GHz=200 MHz, and the 3.5 GHz tone folds to 500 MHz. The compressed 9.2 GHz signal deviates $6 \cdot F_{\Delta}$=90 MHz from the 200 MHz center. The period of the modulation is 0.2 µs due to the F_{mod} of 5 MHz.

filter. As the unwanted harmonic products in the output are indistinguishable from a true input signal, they may lead to false positives during reconstruction of the input. General reconstruction performance is also lowered by the false positives, as now the signal being reconstructed is less sparse.

The next component in the receive chain is an amplifier. As the harmonic mixer spreads the energy of signals from higher zones across the spectrum, amplification is required to match to the dynamic range of the analog to digital converter (ADC). Then another low pass filter is used to remove spurious prod-

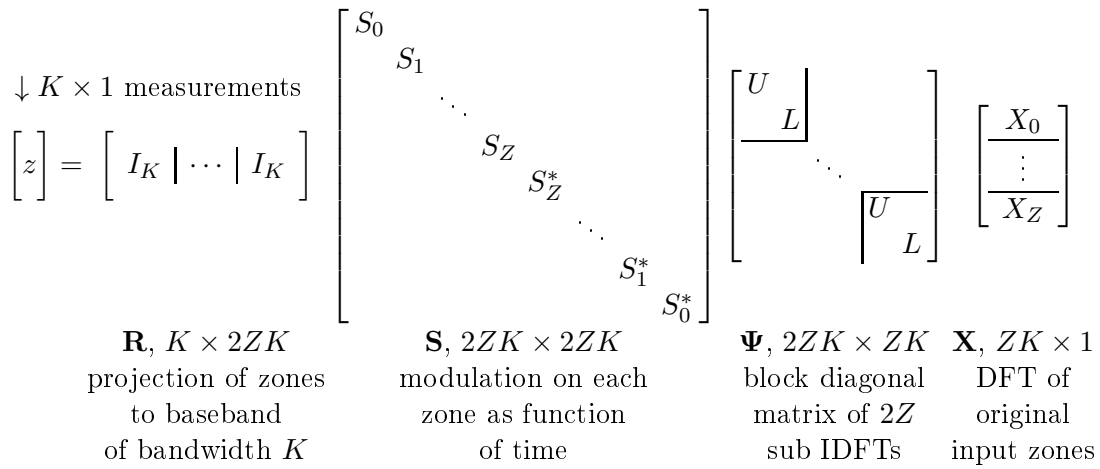

Figure 3.7: Real valued compressive sensing matrix model for the NYFR implementation.

ucts from the amplifier. Finally the receiver output is sampled at 1.5 GS/s by a Red Rapids Model 276 ADC with 12-bit resolution. Three of these PCIe cards are hosted in a Linux PC with MATLAB used to process the output of the prototype three channel NYFR.

In order to match the compressive sensing model with the physical hardware, the phase of the modulation induced on the RF signal, and thus the starting phase of $s_{LO}(t)$ must be known. The phase offset between the start of an ADC capture and the $s_{LO}(t)$ generated by the AWG is controlled by triggering both the ADCs and the AWG with a multichannel digital pulse generator. The delay between the ADC trigger and AWG trigger was experimentally adjusted to account for the delay of the whole system, including the delays between the triggers and actual action and the phase response of the receive chain. In a production system, this timing and phase must also be accurately known.

3.2 Compressive Sensing Model

Now a CS model of the measurement process performed by the hardware is needed to enable reconstruction of the original RF input from the compressed receiver

output. The real-valued signal model for the hardware used by this thesis is developed in [23]. Figure 3.7 shows the full model. The compressed measurements $\mathbf{z}$ are captured by the ADC in the time domain. $\mathbf{R}$ is a set of identity matrices that project Z number of Nyquist zones onto a baseband of K samples.

The conjugate symmetric diagonal matrix $\mathbf{S}$ contains the modulation induced on each zone as a function of time. As shown in Section 3.1 when given a frequency-modulated $s_{LO}(t)$ the harmonic mixer spreads the input signal over a range of frequencies, with the spread increasing for higher harmonics of $s_{LO}(t)$ and thus higher frequency input signals. The modulation index $M = 0, -1, 1, 2, -2, ..., Z$ is the parameter that uniquely defines the original frequency of the compressed signals giving an output with a peak frequency deviation MF_{Δ}. The sign of the modulation is set by whether the signal is from the upper or lower sideband, and negative modulation is shifted 90 degrees from the positive modulation. The modulation index depends on the Nyquist zone of the RF input frequency f_{RF}, giving the pattern

$$M(f_{RF}) = \begin{cases} 0, & \text{for} \quad & & f_{RF} < & \frac{f_{LO}}{2} \\ -1, & \text{for} \quad & \frac{f_{LO}}{2} & < f_{RF} < & f_{LO} \\ 1, & \text{for} \quad & f_{LO} & < f_{RF} < & \frac{3f_{LO}}{2} \\ -2, & \text{for} \quad & \frac{3f_{LO}}{2} & < f_{RF} < & 2f_{LO} \\ \vdots & & & \vdots & \\ -Z, & \text{for} \quad & \frac{(2Z-1)f_{LO}}{2} & < f_{RF} < & Zf_{LO} \\ Z, & \text{for} \quad & Zf_{LO} & < f_{RF} < & \frac{(2Z+1)f_{LO}}{2} \end{cases} . \tag{3.13}$$

Thus $\mathbf{S}$ is also said to be split into two conjugate symmetric blocks for each zone, consisting of the time modulation pattern common to all zones modified for that

particular zone's value of M.

$\mathbf{\Psi}$ is a block diagonal matrix with each block consisting of the inverse discrete Fourier transform (IDFT) matrix for one sub-band. The IDFT blocks are split into upper and lower side-bands in order to match the sign of the induced modulation. Finally $\mathbf{X}$ is the complex-valued discrete Fourier transform of the full-bandwidth signal. The full model is an under-determined equation, as only $\mathbf{R}$, $\mathbf{S}$, $\mathbf{\Phi}$, and $\mathbf{z}$ which is shorter than $\mathbf{X}$ are known. The system samples in the time domain and represents the signal in the frequency domain. The two bases are incoherent, which is optimal for compressive sensing. The equation may be solved for $\mathbf{X}$ with CS algorithms if $\mathbf{X}$ is sparse and $\mathbf{z}$ consists of a sufficient number of measurements that have low coherence. The assumption that the original frequency spectrum is sparse limits recovery of signals with very wide bandwidth, especially those wide enough to fill the whole baseband.

Figure 3.8 shows the results of using the compressive sensing model with a convex recovery algorithm to recover the original frequencies for the inputs show in Figure 3.6. The original frequencies of the two CW tones are correctly recovered. Note that these inputs fit the assumption of a sparse frequency spectrum. For cases with wide-bandwidth signals, especially ones wide enough to fill the baseband defined by ADC's low sample rate, recovery of the spectrum will fail as the sparsity assumption is no longer true. The left panel Figure 3.9 shows an LFM chirp that starts at 2.3 GHz and ends at 3 GHz, having significant bandwidth compared to the sample rate. The right panel shows the failed spectrum recovery. It is worth noting, however, that the chirp is still sparse in frequency over a small period of time, thus for short enough time windows, recovery of some kinds of LFM signals is still possible. Additionally, a CS model using a different basis could potentially be found, for example a dictionary of various expected

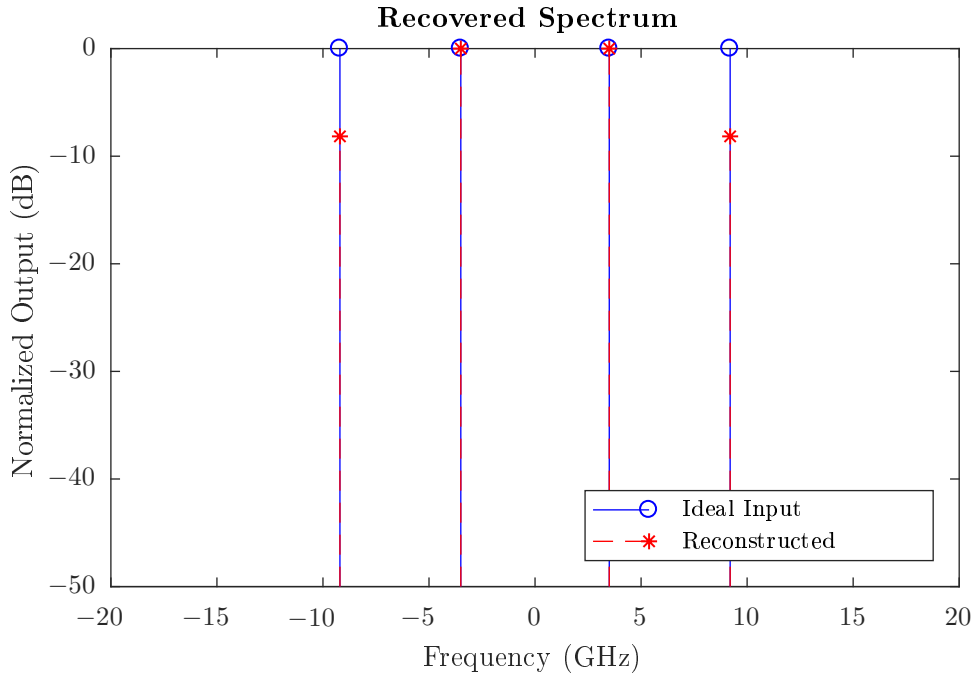

Figure 3.8: Output of the SPGL-1 convex recovery algorithm ran on the simulated inputs from Figure 3.6 for the CS model given in this chapter. Both of the input frequencies, 3.5 GHz and 9.2 GHz, are correctly recovered. Thus the CS model is shown to work for a sparse case with $s = 4$.

chirp signals, then this example input would have a sparse representation in that basis. Now that it has been shown that the sparsity of the spectrum can cause reconstruction to fail, the RIP will be used in order to get an idea of what number of measurements or level of sparsity is sufficient for reconstruction.

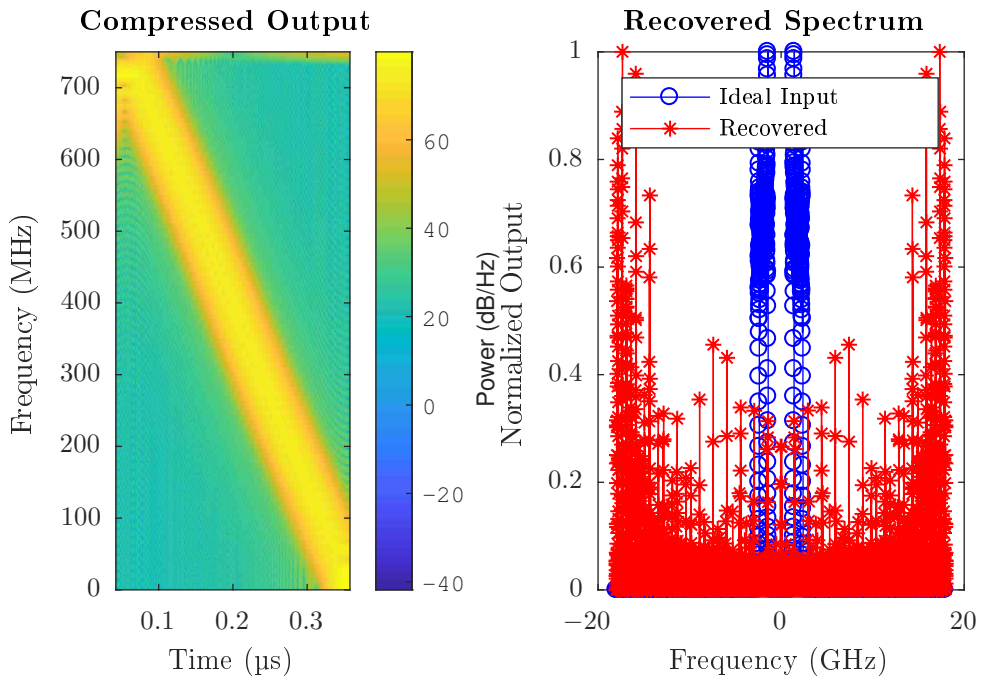

Figure 3.9: The simulated receiver output for wideband LFM chirp starting at 2.3 GHz and ending at 3 GHz is plotted in the left panel. The right panel shows the ideal spectrum of the chirp, and the attempted recovery of the spectrum through convex optimization. Recovery is not successful as the input chirp is not sparse in the basis used for the CS model.

3.2.1 RIP Analysis

As mentioned in Chapter 2, the restricted isometry property can be used to establish whether a sensing matrix is guaranteed to exactly recover $\mathbf{x}$ for a sparsity level s. Of course, the difficult computation is determining δ_s for a matrix. By the definition of the RIP, the problem of finding the RIP constant is equivalent to finding the signal $\mathbf{x}$ that is s-sparse and gives the largest value of δ_s

$$||\mathbf{Ax}||_2^2 = (1 \pm \delta_s)||\mathbf{x}||_2^2. \tag{3.14}$$

Additionally the spectral norm-deviation of $\mathbf{x}$ is defined as

$$\delta_s(x) = \left| \frac{||\mathbf{A}^*_{\Lambda(\mathbf{x})}\mathbf{Ax}||_2}{||\mathbf{x}||_2} - 1 \right| \tag{3.15}$$

where $\Lambda(\mathbf{x})$ is the support of $\mathbf{x}$ [24]. The spectral norm of $\mathbf{x}$ is the maximum singular value of $\mathbf{x}$. Finding δ_s is then equivalent to finding the maximum value spectral norm-deviation of all possible combinations of $\mathbf{x}$.

Determining δ_s exactly for a given s would require examining all possible values $\mathbf{x}$ can take with a support size of s, which is clearly not feasible. Instead Fudge et al. consider what should be the worst case for the NYFR system, frequency-bin-centered tones at frequencies that fold to the same baseband frequency [24]. First, the system is considered as if there no zone-dependent modulation induced. Without modulation, it is impossible to distinguish between two signals that originate from different Nyquist zones but fold to the same frequency, which is the worst case. By definition the worst case gives the maximum value of $\delta_s(\mathbf{x})$, which is then also the restricted isometry constant of $\mathbf{A}$ for any s-sparse signal. Fudge shows that for cases where aliased, bin-centered frequencies do not fold to the

same compressed bin, δ_s is identically zero, and establishes a lower bound of one for folds to the same bin.

A RIP constant of one or greater does not guarantee successful reconstruction for any algorithm. This leads to an alternative consideration of how often these worst cases will occur. If the original signal $\mathbf{x}$ consists of s continuous-wave (CW) tones that have a uniform probability of originating from any particular frequency bin in the receiver's bandwidth, then it follows that the probability of $\delta_s(x)$ being identically zero, guaranteeing reconstruction, is the probability of any two of the tones aliasing to the same baseband frequency-bin, which is found through a short derivation to be

$$P(\delta_s(x)) = \left(\frac{f_{LO}T - 1}{f_{LO}T}\right)^{s!/(2(s-2)!)} \tag{3.16}$$

where T is length in seconds of the reconstruction window. For the values most commonly used for the implementation of the NYFR in this thesis, f_{LO}=1.5 GHz and T=1 µs, which gives a probability of 0.9993 of successful recovery of any spectrum containing two signals with a high-signal-to-noise ratio (SNR). Of course, in a real environment the distribution of signals across the spectrum is far from uniform, and rarely will a signal be a bin-centered CW tones, so this result is of limited utility.

Next, we extend this consideration of the worst case to the NYFR system that induces a zone-dependent modulation. Consider two signals that fold to the same compressed center frequency. The further away the original Nyquist zones are, the more their modulation differs and thus the less correlated they are. The worst case will be when the two signals are highly correlated; that is, when they

are from adjacent zones. An upper bound for δ_s was then calculated in [24] to be

$$\delta_s \leq sC\sqrt{\frac{1}{F_\Delta T}} \tag{3.17}$$

where C is some constant that must be empirically found for a particular modulation pattern. In general it can be concluded that for a high-SNR case, increasing the peak frequency deviation of $s_{LO}(t)$ or increasing the number of measurements will improve the level of sparsity that is guaranteed to be reconstructed. As the exact performance bound of the NYFR is difficult to derive mathematically, such that the current literature requires C be empirically found, a useful step for designing the system was the creation of a simulation based on the hardware. The simulation was then used to quickly find values for adjustable parameters such as the $s_{LO}(t)$ modulation pattern that should later give reasonable performance for the hardware prototype.

Chapter 4

Simulation

This chapter will detail results found through the use of a MATLAB simulation of the NYFR system. The simulation enabled analysis of some of the design constraints inherent with the NYFR architecture, eventually leading to a reasonable set of parameters that were used with the receiver prototype. The simulation was designed to match the operation of the harmonic mixing hardware. Short pulses are generated on the zero crossings of the simulated $s_{LO}(t)$, and mixed with simulated input signals. The bandwidth of the mixer is modeled as having the -3 dB point at the 15 GHz harmonic. The LO center frequency is 1.5 GHz. An accurate simulation of the harmonic mixer behavior is sufficient to model the compressive sensing performance. The other components of the system are only simulated for the ideal case: the digital LPF used to simulate the analog LPF after the harmonic mixer output has a significantly sharper cutoff at 750 MHz than the filter used in the hardware implementation, and other loss in the system components such as the limiters and ADC quantization loss are not considered.

4.1 Frequency Recovery

The effects of several system parameters on the ability to recover the frequency support of narrowband input signals are analyzed in order to obtain optimal re-

construction performance. The design of the LO sampling waveform is of particular interest. For the case of inducing the sinusoidal frequency modulation defined in (3.6) on the RF input by modulating the phase of the LO waveform, the peak frequency deviation F_Δ and the modulation rate F_{mod} must be selected. F_Δ involves a trade-off between recovering low-power signals in higher Nyquist zones and the ability to determine the correct origination zone. Increasing F_{mod} allows for recovery of shorter pulses by increasing the incoherence of the small number of measurements taken during the short pulse, but the maximum modulation rate is limited by phase noise. Both parameters limit the dynamic reconstruction range of the system.

4.1.1 Selection of Peak Frequency Deviation

Increasing F_Δ decreases the correlation between compressed signals in adjacent Nyquist zones, thus making it easier for compressive reconstruction algorithms to identify the correct zone. However, increasing F_Δ also spreads the energy content of the input signal over a wider frequency range, eventually causing the signals at higher zones to fall below the noise floor. Thus a trade-off must be made between the probability of selecting the correct Nyquist zone and the signal-to-noise ratio (SNR) requirements of a practical system.

Figure 4.1 shows the results of a Monte Carlo simulation for the probability of successful reconstruction versus SNR for signals in the first seven Nyquist zones. F_Δ is set to 4 MHz. Each trial consists of randomly generating a single input frequency with added white noise, attempting to recover the frequency support with the SPGL-1 solver, and finally classifying the reconstruction as successful if the energy in the bins containing the original frequency is greater than 80% of the maximum reconstructed frequency bin. The two frequency bins surrounding the

random frequency are summed in order to handle cases where the frequency does not exactly fit on the simulation grid. This approach does allow a reconstruction with a limited number of false positives to still be classified as successful. Forty-eight trials were performed for each zone and SNR step. The reconstruction performance decreases rapidly as the spectrum of the modulated input signal falls into the noise floor. The SNR requirements for reconstruction grow at an increasing rate as the modulation index increases.

Now consider that signals that fold to the same baseband frequency, with all things identical other than the original zone, are still somewhat similar as the same modulation pattern is applied to each. The coherence between identical signals from different zones is not zero. In the most extreme case of F_Δ=0 Hz, the mutual coherence is one, and which zone the signal originally came from may not be determined. An important effect caused by the modulation patterns for each zone not being completely incoherent is that when convex optimization is performed to recover a signal at a frequency f_{RF}, spurious values appear in the reconstruction at $f_{RF} \pm f_{LO}$. The highest magnitude spurs are always in the zones to either side of the correct one, as these zones have the least difference in modulation pattern and thus have the highest coherence with the true zone. Algorithm families such as iterative thresholding make the coherence issue less clear, as they stop iterating after reconstructing a selected number of the highest magnitude coefficients rather than optimizing the whole signal basis, and the weaker $f_{RF} \pm f_{LO}$ will often fall below the threshold for reconstruction, but there is still a nonzero probability that an incorrect Nyquist zone will be selected.

A second Monte Carlo simulation was performed to find the relative levels of the adjacent Nyquist zones versus F_Δ. Thirty-two simulations per F_Δ step were performed for a constant RF input with frequency f_{RF} on the simulation

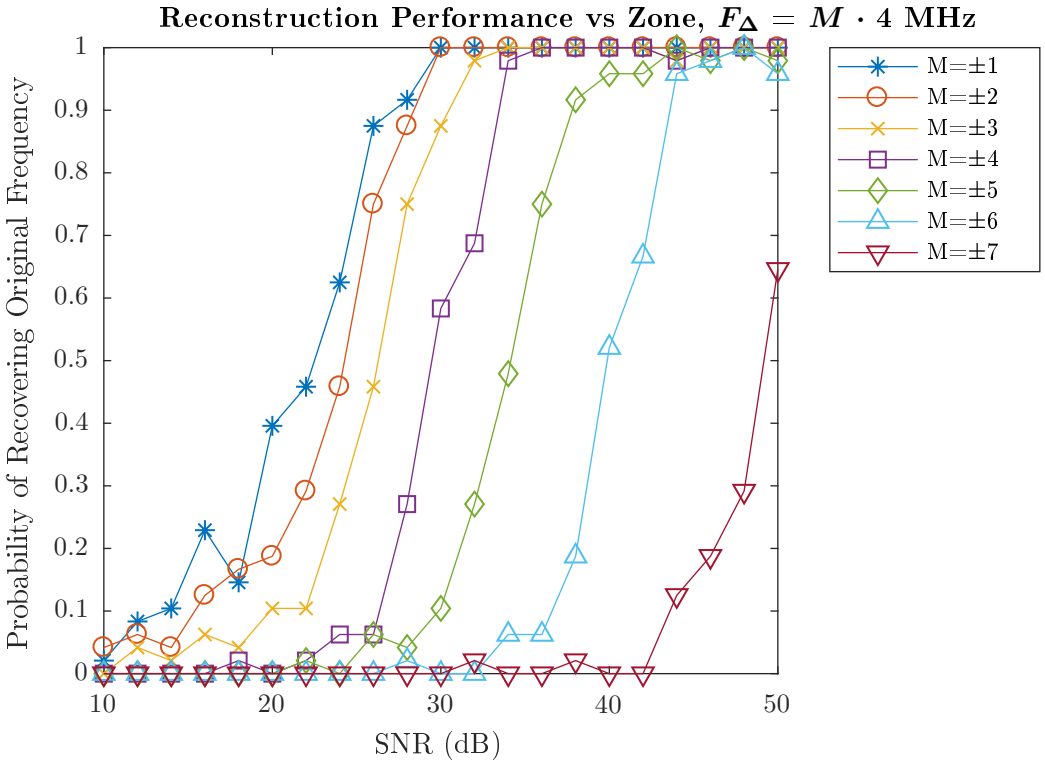

Figure 4.1: Results of Monte Carlo Simulation for the reconstruction of random frequencies in each Nyquist zone versus SNR. Trials are ran for seven Nyquist zones each with a corresponding modulation index $M = \pm 1, \pm 2, ..., \pm 7$. For each trial, a CW tone is randomly selected with uniform probability from the frequency range of the current zone. AWG is added to the generated input to give a specific SNR. Reconstruction is then performed for each trial. The number of reconstructions that correctly identify the random tone is used to estimate the probability of successful reconstruction at various SNRs for each zone. As the SNR decreases, the modulated harmonic mixer output falls below the noise floor, causing reconstruction to fail. As the energy of input signals from higher Nyquist zones is spread out over a wider frequency range according to MF_Δ, high frequency signals require a higher SNR for the compressed output to remain above the noise floor.

frequency grid with an SNR of 40 dB. Peak frequency deviations of $s_{LO}(t)$ were stepped through from 0.5 MHz to 6 MHz with a constant F_{mod} of 5 MHz. The mean relative level of the peak side zones is plotted in Figure 4.2. It can be seen that the side-zone levels are decreasing as the peak frequency deviation increases. This intuitively makes sense, as the higher the peak frequency deviation, the less correlated different zones are.

To illustrate the relation between F_{mod} and F_{Δ}, the simulation is performed again with an F_{mod} of 10 MHz and plotted in Figure 4.3. Increasing the modulation rate without also increasing the underlying sampling rate of the system causes there to be fewer measurements at each intermediate sampling frequency. Fewer measurements lowers the reconstruction performance of the system, raising the side-zone levels. Thus the selection of F_{Δ} is a trade-off that depends on the required attenuation of the side-zones, the modulation rate, and the needed sensitivity of the system to low-power signals in the higher Nyquist zones.

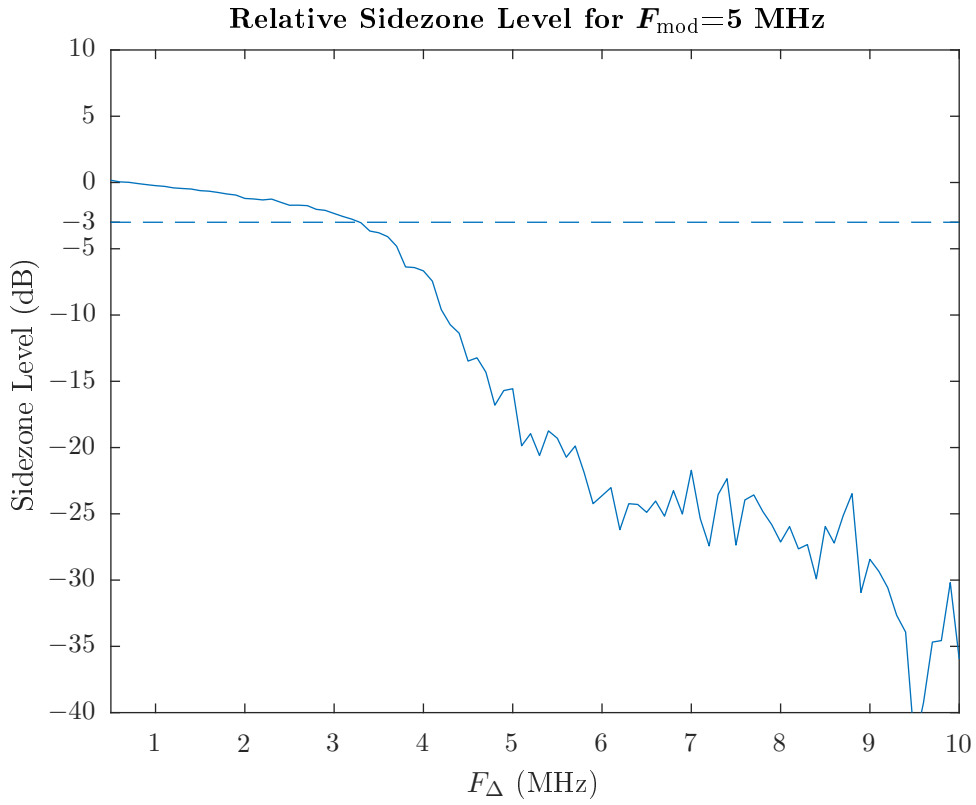

Figure 4.2: Mean relative level of the spurs from adjacent zones for different peak frequency deviation values. As the same signal, except for originating in different Nyquist zones, is not perfectly incoherent, the reconstruction of the original signal $\mathbf{X}$ appears to have nonzero components at the $f_{RF} \pm n f_{LO}$ where n is integer multiples of f_{LO}. Increasing F_Δ decreases the coherence of the zones with each other. The zones to either side of the zone the signal truly originates from, the "side-zones", have the least difference in modulation pattern and thus are the most likely to be reconstructed with significant false magnitude. A CW input of 9.2 GHz with added AWGN for an SNR of 60 dB is used for each trial. Trials are ran for each value of F_Δ. The mean of all side-zone power levels, including both adjacent zones, are plotted.

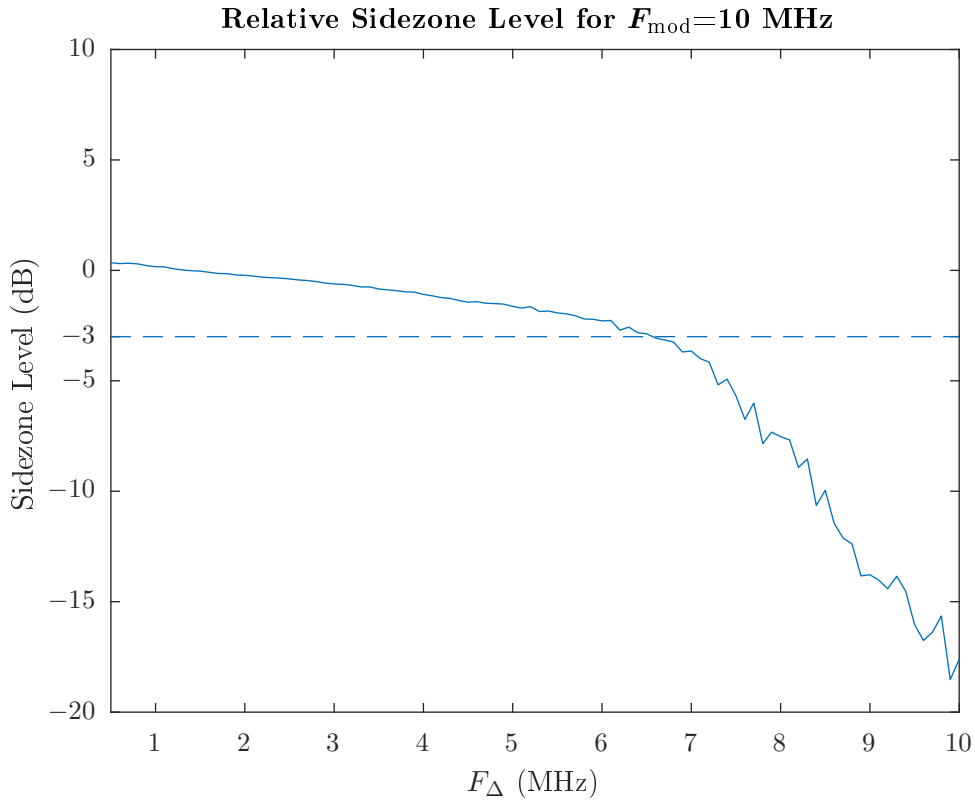

Figure 4.3: Relative level of the adjacent zones for different modulation frequency deviation values, with the modulation rate set to 10 MHz. A CW input of 9.2 GHz with added AWGN for an SNR of 60 dB is used for each trial. Trials are ran for each value of F_Δ. The mean of all side-zone power levels, including both adjacent zones, are plotted.

4.1.2 Reconstruction Dynamic Range

For a practical case where there is not a priori knowledge of the number of signals or their power levels, some discussion must be made of how to select a threshold to determine whether there is an actual signal located at a frequency in the reconstructed spectrum, or just a spurious reconstruction or noise. One simple method is to set a minimum amplitude threshold for a reconstructed signal to be considered a true signal. However when the original input had multiple signals of varying power levels, selecting which signals are truly present becomes more difficult.

As the magnitudes of side-zone harmonics, dependent on the selection of F_{mod} and F_{Δ}, can be significant, the side-zones limit the dynamic range of the system when performing reconstruction. As the convex optimization is not guaranteed to determine exactly the original zone of a signal, shown in Section 4.1.1, harmonics of the signal will appear at $f_{reconstructed} \pm f_{LO}$ with a power level relative to the original signal power. If one signal source has a significantly higher power than another source, the reconstructed harmonics of the first signal may also have a higher power than the second source. If the threshold for signal detection is set relative to the strongest reconstruction, in order to ignore the harmonics, detection of the second source will fail. An example for three CW tones of unequal power at 3.5, 3.65, and 3.8 GHz is shown in Figure 4.4. The signals are all located in the same zone, in order to ignore any affects from the frequency spread increasing with higher zones. The signal at 3.8 GHz has the highest power, at 80 dB above the noise floor, 3.65 GHz is 20 dB above, and 3.5 GHz is 30 dB above. The side harmonics of the 3.8 GHz tone are higher for the 10 MHz F_{mod} case than the 5 MHz case, as expected. Indeed they are higher than the recovered

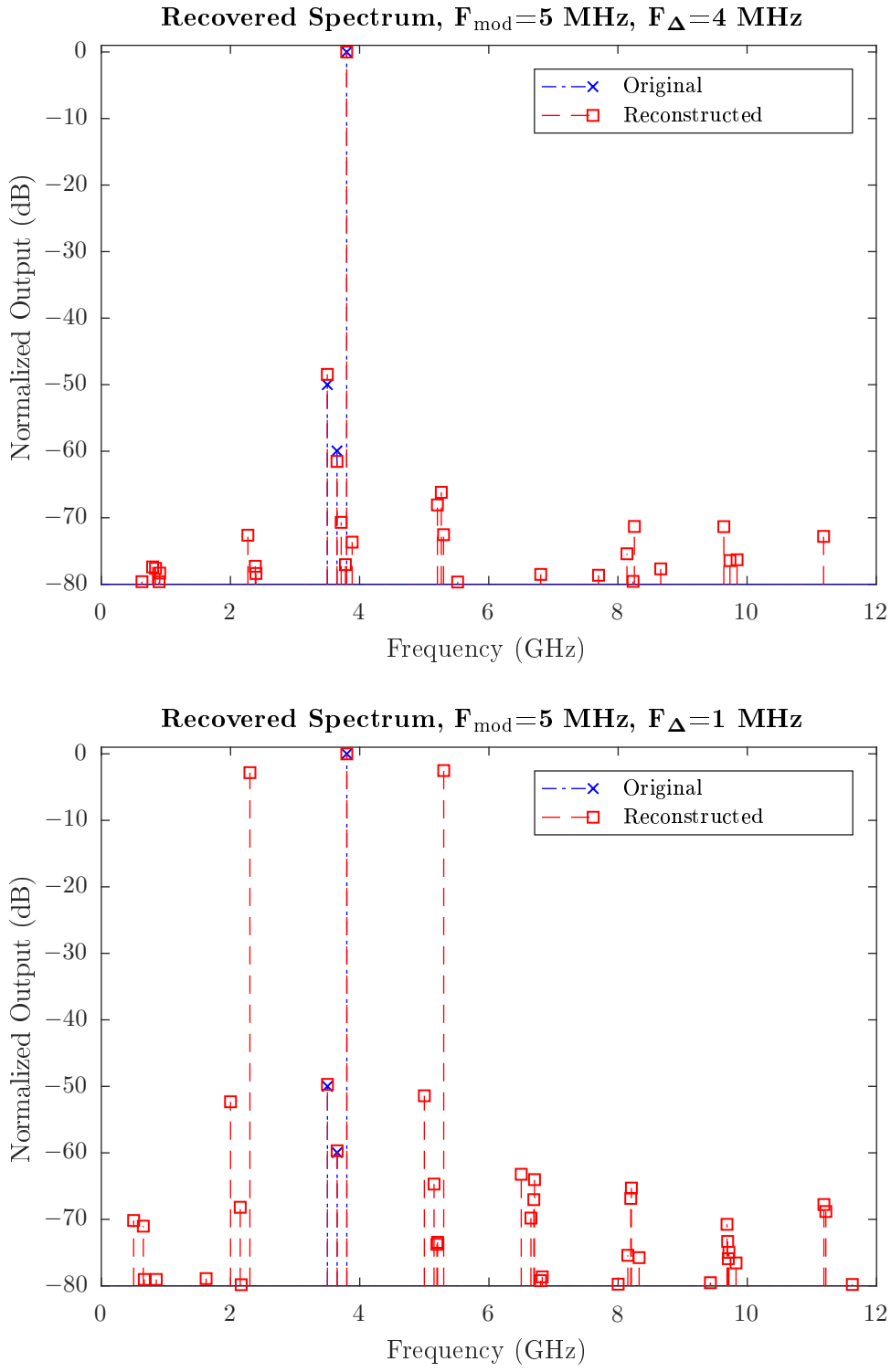

Figure 4.4: The reconstruction dynamic range is limited by the harmonics located in the zones to either side of the strongest signal. In the top figure the higher value of F_Δ causes the average magnitude of the side-zone harmonics to be lower, such that in this example all original signals are reconstructed above the harmonics. In the bottom figure F_Δ is decreased, raising the side-zones of the strongest signal above the correct frequencies of the other two signals, complicating identification of the signals present.

tones at 3.65 GHz and 3.5 GHz tones, so a simple thresholding method would either give false positives or false negatives.

There is a potential for more intelligent detection methods to alleviate this effect. An iterative method could select the strongest reconstructed signal, subtract that signal and the predicted harmonics from the reconstructed spectrum, then repeat until the strongest detection is below some threshold that is selected from the receiver noise floor. This method would have a few drawbacks, however, as this could cause the presence of a real, but weak, signal that is located at a harmonic of the one strong signals to not be detected. This method also becomes more fragile in the presence of signals with significant bandwidth, as the bandwidth must be estimated before subtracting the harmonics. The performance of this method will be shown in Chapter 5.

4.1.3 Phase Accuracy Requirement

The question of how close the phase of the sensing matrix used for reconstruction must match the phase of the measurements should be considered, in order to be informed of the timing requirements for a real system. A single simulated CW tone in the 16-17.5 GHz Nyquist zone matching the highest zone of the physical prototype, is input into the harmonic mixer along with a sinusoidal $s_{LO}(t)$ with the modulation $\theta(t,\tau) = \frac{F_\Delta}{F_{mod}} \sin(2\pi F_{mod} t + \tau)$. The mixer output for various values of phase drift τ is then reconstructed via a sensing matrix assuming $\tau = 0$. This procedure simulates the unknown alignment between the $s_{LO}(t)$ generator and the start of an ADC capture in a the real system. The results for F_Δ=4 MHz and F_{mod}=5 MHz are plotted in Figure 4.5.

There is a sharp transition between successful and failed recovery at the same delay for both the ROMP greedy iterative thresholding algorithm and the SPGL-

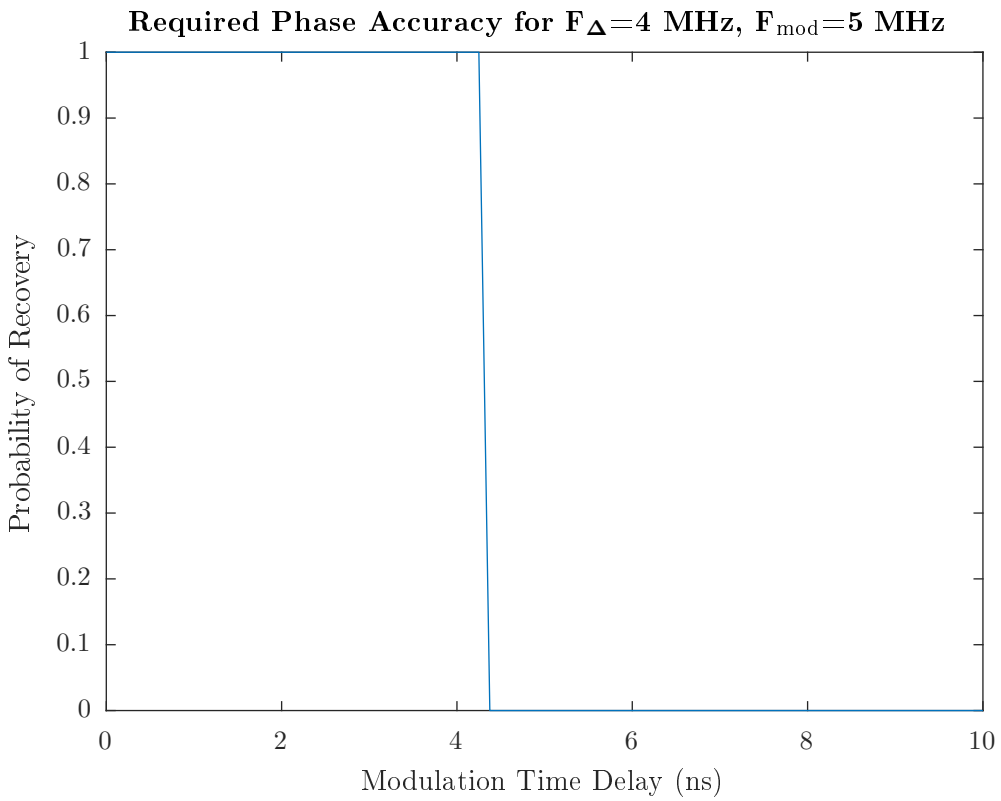

Figure 4.5: Simulation performed to test how closely the phase of the CS model and the compressive measurements must match in order to get successful recovery. One hundred trials were performed for each time delay, generating a CW tone in the highest Nyquist zone with added AWGN. The estimated probability of recovery is plotted. For the tested LO parameters recovery fails when there is a difference of approximately 4.25 ns between the model and measurements.

1 convex optimization. Increasing the slope of the modulation by increasing the modulation rate or peak deviation decreases the misalignment at which reconstruction fails. Depending on the selected LO parameters, this constraint is not too difficult to meet; for the values used for the majority of the experiments performed with the hardware prototype, the timing must remain within only 4.25 ns compared to the modulation period of 0.2 µs.

4.2 Recovery of Pulses

The recovery of pulses is of interest as few signal sources in the real world are constant-frequency tones. Digital communication systems transmit in bursts as needed, and in several schemes only transmit part of the time to allow other devices to share the same frequency. Frequency hopping radar is another interesting signal source, especially since existing swept frequency receivers will miss some of the pulses. If the duration of a pulse is long enough for the system to capture a number of incoherent measurements to satisfy the RIP, reconstruction performance is the same as that of a CW tone. A pulse length requirement of greater than one LO modulation period was discovered to be a general rule of thumb for there to be measurements taken with low enough coherence for successful recovery. Increasing F_{mod} decreases the coherence of each measurement in the pulse, allowing reconstruction of a shorter pulse where only a small number of measurements may be taken, but as noted in Section 4.1.1 there is also a negative effect in that increasing F_{mod} raises the side-zone magnitudes. Figure 4.6 shows an example of two simulations for a pulse length three microseconds shorter than half a modulation cycle and for a pulse length equal to half the modulation cycle. The underlying sampling rate of the system defined by f_{LO} and the ADC sampling rate limits the shortest pulse than can be reconstructed as well, since the sample rate obviously limits the number of measurements of the pulse that are taken. Figure 4.7 shows an example of reconstruction of three pulses, each lasting more than one modulation period. With three pulses in the reconstruction window, intermodulation products also appear in the reconstruction.

As continuous reconstruction is computationally expensive, to the point that with current technology it is difficult to scale up to real-time processing, it is

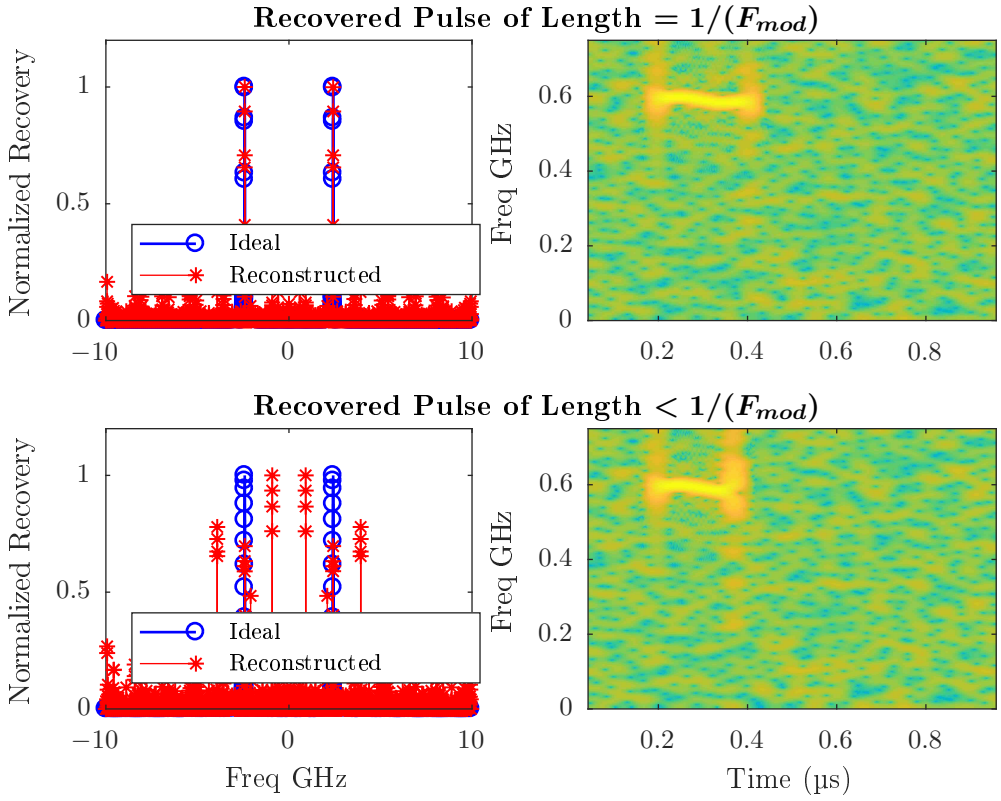

Figure 4.6: Simulated reconstructions to show an example of the pulse length requirements for success. The top figure has a pulse length equal to one LO modulation period, giving it a sufficient number of measurements with low coherence for successful reconstruction. The bottom figure has the same pulse except now it is three microseconds shorter, giving it fewer incoherent measurements. The smaller number of measurements causes the reconstruction to be noticeably less clean.

useful to examine alternatives. Rather than intensive processing of either large time windows or many windows, it is faster to only perform reconstruction on small time windows that may contain new signals of interest. Classical pulse detection techniques, such as those used for radar, are useful for time localization, allowing a fast reconstruction to be performed to identify the original frequency support only when the signals present change. It is later shown in Section 4.5 that knowledge of the frequency support can be used to demodulate the compressed data, arriving at the original signals of interest. This three-step scheme can reduce

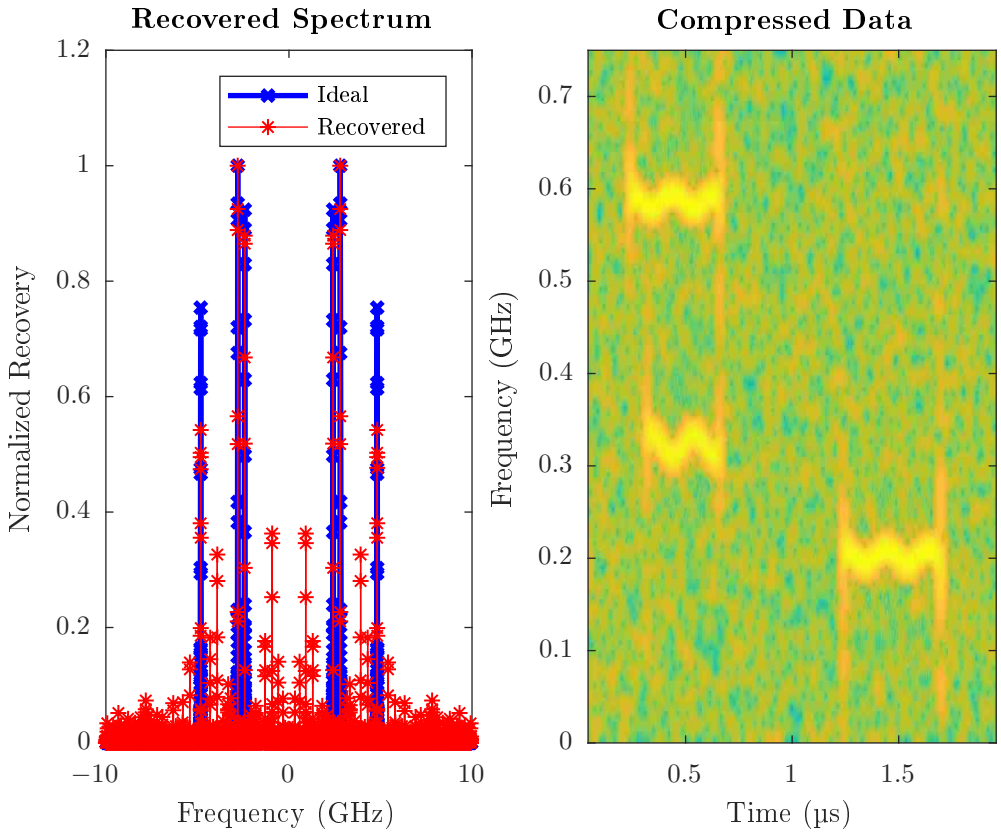

Figure 4.7: Simulation and recovery of three pulses of varying duration at 2.415, 2.8, and 4.82 GHz with an SNR of 20 dB. The left pane shows the recovered spectrum, with the frequency support of all three pulses correctly identified. The right pane shows a spectrogram of the simulated compressed pulse data.

computational cost for recovery of pulses or bursty communication signals. This method is limited by how effective the pulse detection method is, and the shortest duration that is reconstructible is set by the selection of F_{mod}. As high values of F_{mod} are both limited by phase noise and the required dynamic range of the reconstruction, there is a practical limit on the minimum pulse length.

Constant False Alarm Rate (CFAR) is used for such an example of pulse detection preprocessing. CFAR is used in radar to determine whether range cells from time series data contain targets, with the goal of providing a constant rate

of false alarm [25]. The time data is split into several cells. Then the cells are iterated over, with the current cell being the cell under test (CUT). Reference cells on either side of the CUT are used to estimate the power level of noise and interferers in order to adaptively set a threshold for detection. For radar, often guard cells are a small number of ignored cells in between the CUT and reference cells, used to ensure returns from the target in the CUT does not effect the noise power estimation, but for the NYFR we're using the CFAR technique to detect the presence of a new pulse and the guard cells are not needed. The threshold is set above the noise power estimated from the reference cells, so that there is a constant rate of false alarm calculated from the probability of noise in one sample exceeding the threshold. The adaptive threshold causes continuous signals and short pulses to only be detected once, as after the detection the threshold raises above the new power level, and long pulses are detected once near the beginning and ends of the pulse as the power level of the NYFR output changes. The system must wait until the reference cells after the CUT arrives before the threshold can be calculated, so CFAR will add some latency to the system.

A simulation is created that has one CW tone present at 11.5 GHz, and a pulse-gated tone at 7.4 GHz with a pulse duration of 1 µs. The probability of false alarm is set to 10^{-6}. The CFAR output detects a time close to the center of each pulse and is plotted in Figure 4.8. The plot of the threshold increases as each pulse begins and decreases as the pulses end. For each continual set of cells that all exceed the threshold, the center cell marks the beginning or end of a pulse and is plotted on the figure as detections. A short 0.4 µs reconstruction window is created for each detection, aligned with the start of the modulation period prior to the pulse center. The reconstruction output for one of the detected pulses is shown in Figure 4.9. This method works well for any case where the total power

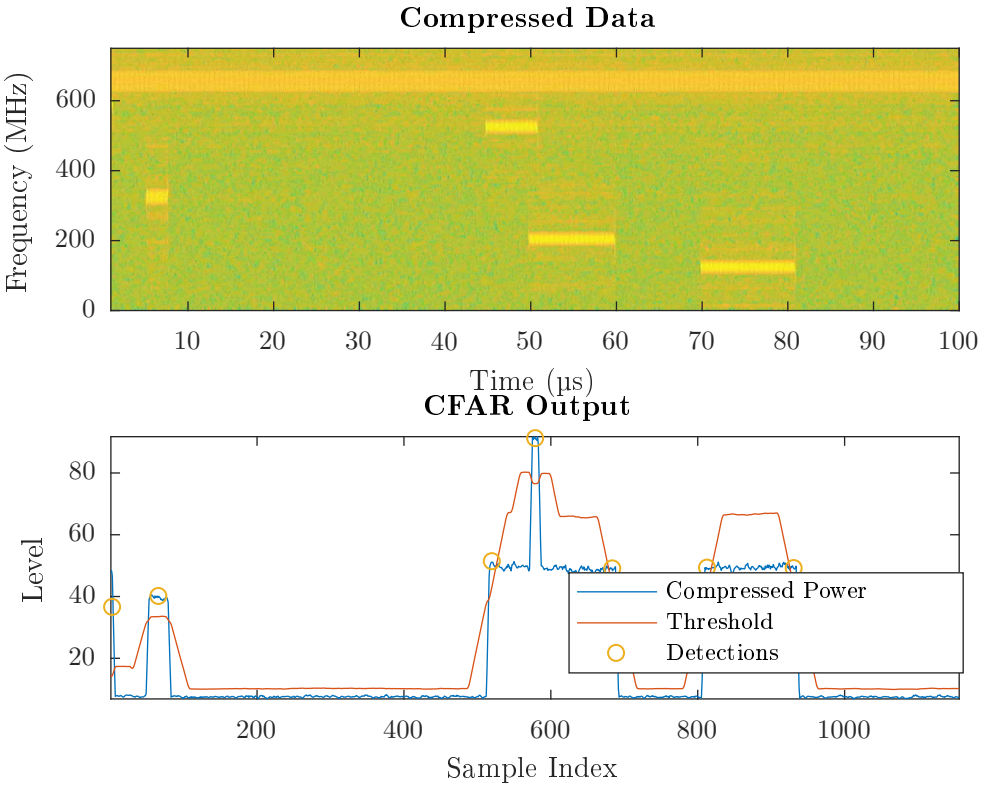

Figure 4.8: Constant False Alarm Rate detection is used to detect pulses. The top figure pane displays the spectrogram of the simulated input. A CW tone is present at 8.35 GHz in order to show how CFAR sets the threshold when a signal is present for the whole capture, and four pulses of different frequencies and duration are simulated. The bottom panel shows the CFAR output. The blue line tracks the total power of the NYFR output at that time. The other line plots the threshold adaptively set by CFAR. The circles mark the pulse detections that are used to select windows of the compressed data for reconstruction.

in the RF spectrum changes when any signals of interest start or stop; the one weakness is that if the next pulse starts immediately when the previous one ends, and has nearly the same power level, the change will not be detected. A more complex detection method that also takes into account the baseband frequency as well, possibly a wavelet based method, would handle such a case.

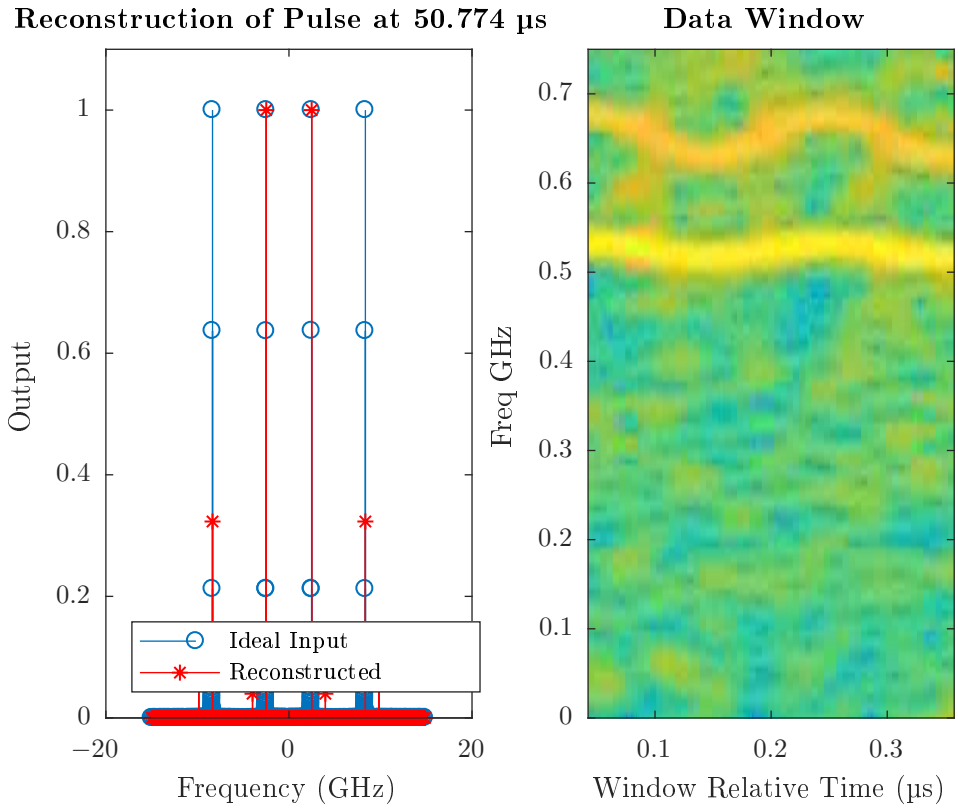

Figure 4.9: One of the pulses is detected by CFAR then the frequency support is recovered with a short reconstruction window. Both the CW tone that is present throughout the whole capture and the detected pulse's frequency are recovered. The pulse frequency was recovered without performing reconstruction for the entire 100 µs simulation, significantly lowering the number of computations performed.

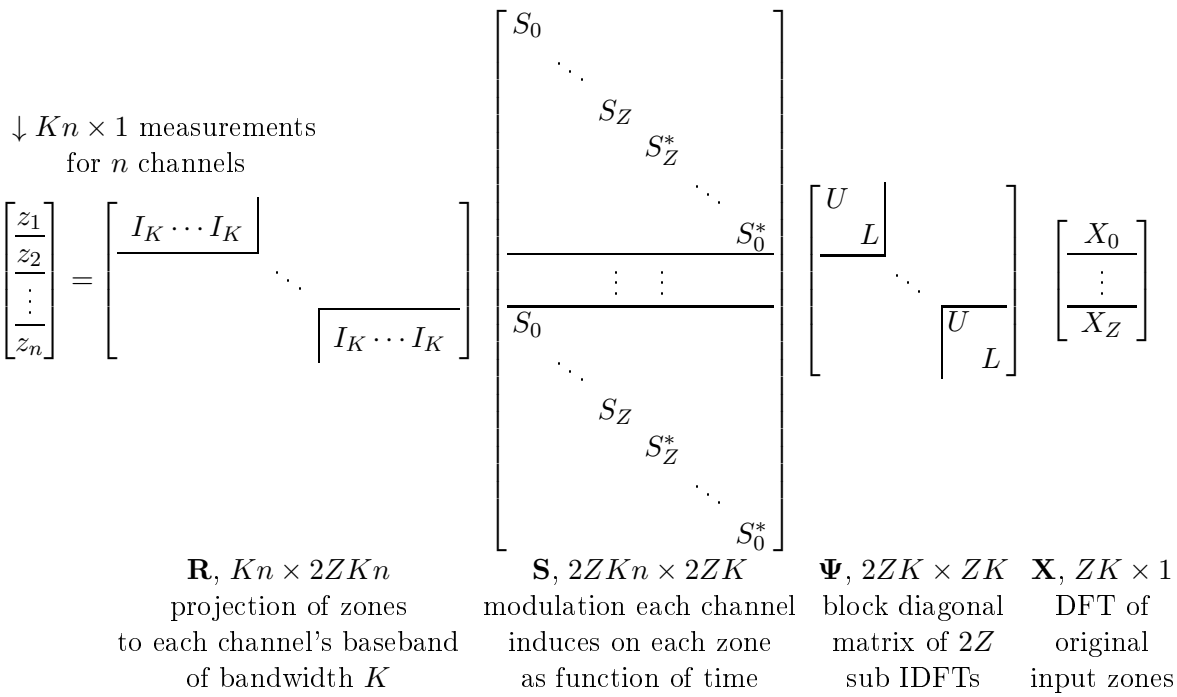

Figure 4.10: Real valued compressive sensing matrix model for combining measurements from multiple channels.

4.3 Multichannel Recovery

Measurements from multiple receiver channels may be combined in the recovery step, capturing more samples in the same time period and thus improving reconstruction performance. Expanding the compressive sensing model to include this is simple; the data from each channel may simply be appended together as shown in Figure 4.10. The matrix $\mathbf{R}$ becomes a block diagonal matrix, with one block for each channel. The time modulation matrix $\mathbf{S}$ is expanded vertically for each channel, allowing for different modulation patterns for each channel. The remainder of the model is the same as for the single-channel case. If identical modulation is used for two ideal channels with perfectly synchronous timing then the measurements will be the same, adding no information to the system. In reality the timing will never be perfect, so the measurements will be different, but as shown earlier in the single-channel case, if significant phase differences are not accounted for in the sensing model, reconstruction will fail. Instead, different

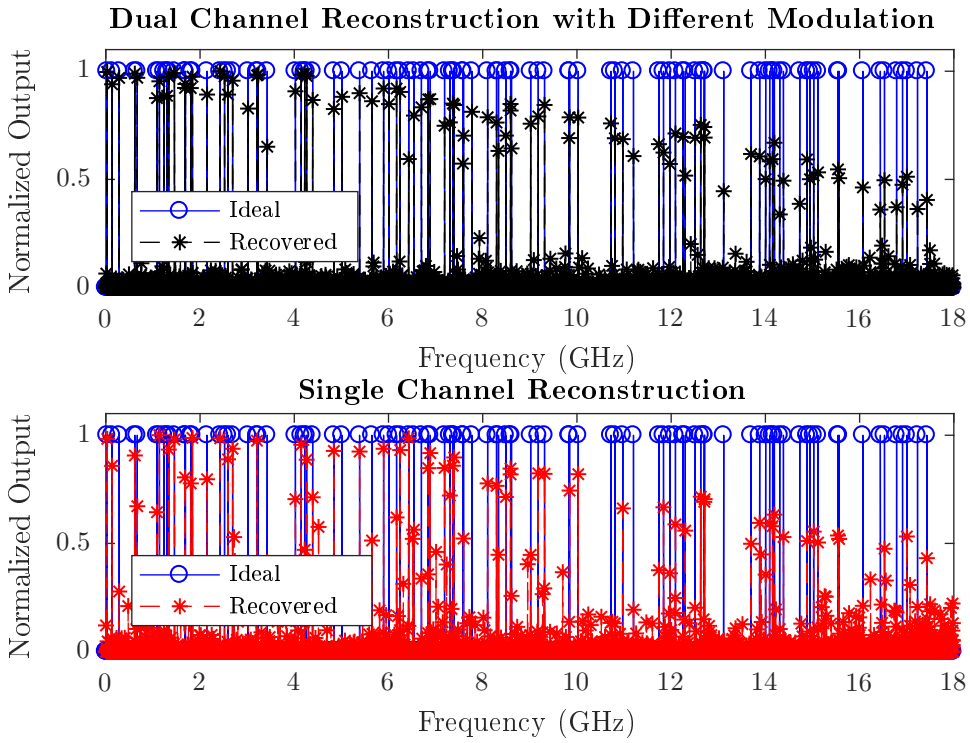

Figure 4.11: Results for recovery using only the channel with sinusoidal modulation and for recovery using also the measurements from the chirp modulated channel. The dual channel case recovers all of the input frequencies, while the single-channel misses eight of them. As predicted the RIP, increasing the number of low coherence measurements used for recovery improves the recovery performance, allowing for full recovery in less sparse environments.

modulation schemes can be applied to each $s_{LO}(t)$ to guarantee that the channels add unique information.

By the RIP it is known that additional incoherent measurements should allow reconstruction of less sparse signals. A simulation of two channels is created. One channel has sinusoidal modulation with F_{mod}= 5 MHz and F_{Δ}= 4 MHz. The second channel is modulated with a linear upchirp with a rate of 6 MHz and 0.2 µs period. One hundred frequency-bin-centered CW tones of equal power are simulated as input. SPGL-1 is used to reconstruct the spectrum over a 0.5 µs window. The results for recovery on only the sinusoidal channel and on the

combined channels are plotted in Figure 4.11.

If for this case successful recovery is defined optimistically as the original frequency bin being recovered with greater than one percent of the original amplitude, the single-channel case recovers 92% of the tones. The dual channel recovery for different modulations, giving twice the effective number of measurements, recovers 100% of the tones. Qualitatively the top panel is a cleaner reconstruction as well. Thus, it is shown that for an ideal multichannel system, if the channels induce different modulation on $s_{LO}(t)$, the reconstruction performance can be improved.

Additionally, the same inputs are recovered with measurements from two channels that both use the sinusoidal modulation from earlier. One channel is delayed by 1.3 ns to simulate the imperfect timing match a real system would have, making the measurements from each channel slightly different while still being within the phase accuracy requirements established in Chapter 3. Figure 4.12 displays an example of the recovery. This case of combining two identically modulated channels also recovers 92% of the tones, just as if only a single-channel were used, showing that combining measurements from multiple channels using an identical $s_{LO}(t)$ makes little difference on reconstruction performance.

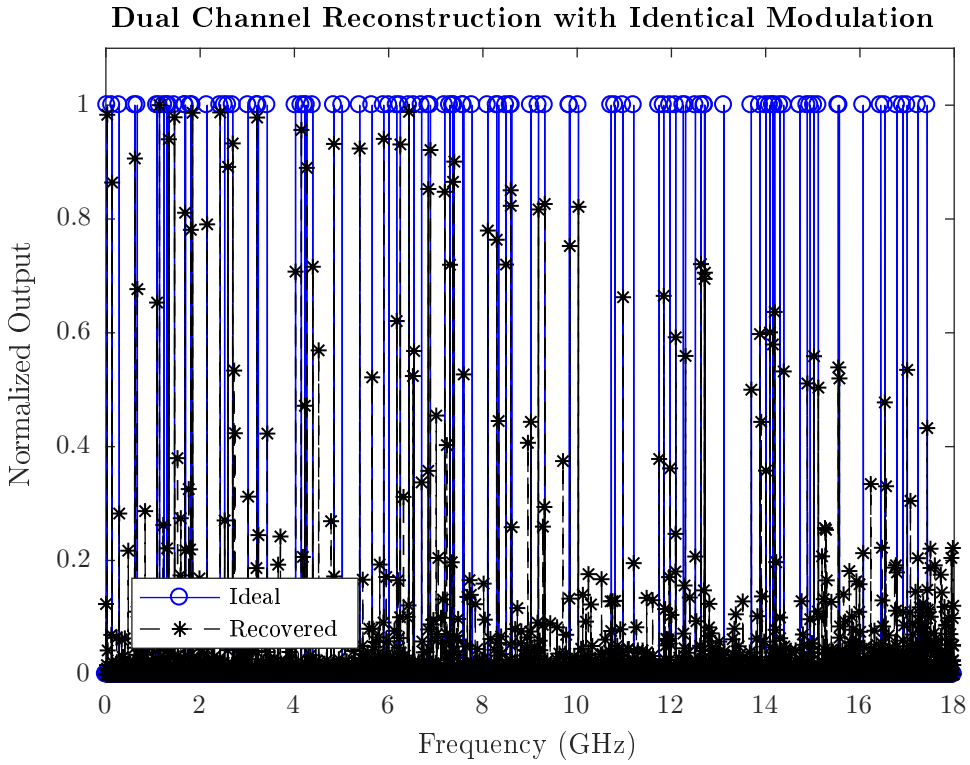

Figure 4.12: Results for recovery using combined measurements from two channels that both have the same sinusoidal modulation. Eight input signals are missed, just as in the single-channel case, demonstrating that combining measurements from multiple channels does not have a significant effect on recovery performance if the measurements from the two channels have high coherence. Taking nearly identical measurements in multiple channels adds no new information to the system.

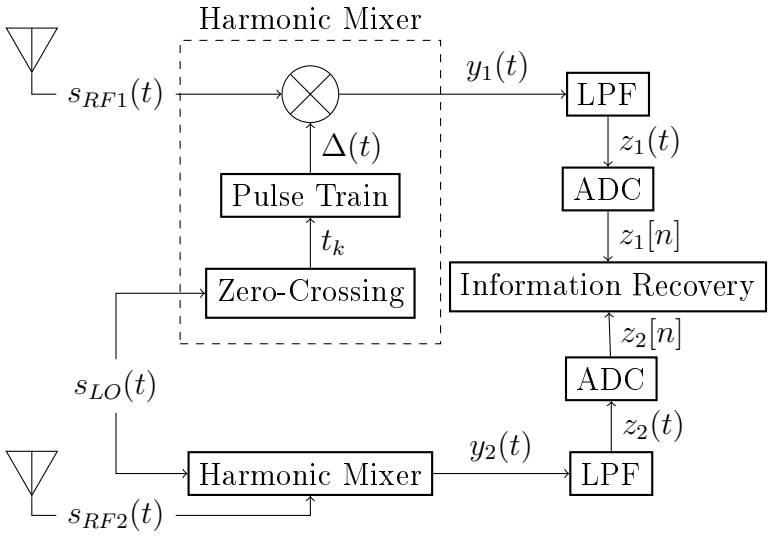

Figure 4.13: Block diagram of the two element NYFR array used for direction of arrival estimation.

4.3.1 Direction of Arrival Estimation

One advantage of the structured sampling approach of NYFR is that it allows for simple implementation of existing direction of arrival (DOA) techniques, as long as the same modulation is applied to two or more channels. It should also be noted that joint estimation of both frequency and direction of arrival without prior information about either is an active area of interest [26][27]. The basic DOA estimation for multiple sources problem assumes that the frequencies of the signals arriving at the antenna array are known [28]. Common solutions include the MUltiple SIgnal Classification (MUSIC) technique and the ESPRIT algorithm for uniform linear arrays. MUSIC is applied to frequency estimation in [29]. The narrowband MUSIC technique has been extended to wideband cases [30]. The NYFR provides a promising solution for sparse RF environments. Figure 4.13 gives the block diagram of a two-channel NYFR system which is a linear array.

Let $s_{RF}(t)$ be a narrowband real signal $cos(\omega_c t+\phi)$. The $s_{LO}(t)$ with sinusoidal

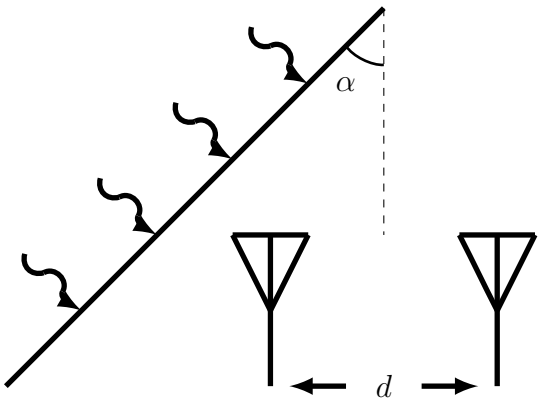

Figure 4.14: Geometry of antenna array for DOA estimation.

modulation that is described in Chapter 3 is used for both channels. The output of the first harmonic mixer and LPF is given in (3.12).

Now the geometry for a uniform linear array with elements separated by a distance d meters is shown in Figure 4.14. The narrowband plane wave $s_{RF1}(t)$ is incident on the array at an angle α. Let $s_{RF1}(t)$ arrive at element 1 at time t, then the signal incident at element 2 is the time delayed $s_{RF1}(t-\tau) = s_{RF2}(t)$. From the geometry it is derived that the time delay τ between the signal arriving at the two elements is

$$\tau = \frac{d\sin(\alpha)}{c} \quad . \tag{4.1}$$

The second harmonic mixer samples the time delayed $s_{RF2}(t)$, giving

$$y_2(t) = y_1(t-\tau) \approx \sum_k s_{RF2}(t) e^{jn_H(\omega_{LO}t+\theta(t))} e^{-j\tau} \quad . \tag{4.2}$$

The modulation induced by the harmonic mixers is independent of the time delay at each element. This means that CS reconstruction can be applied using the same measurement matrix for either digitized channel to recover the frequency support. The recovered center frequency also gives the wavenumber $\frac{2\pi}{\lambda}$. Since

equal modulation has been induced on each signal, they are simply delayed copies of each other allowing traditional direction of arrival estimation techniques to be applied without demodulating the channels.

Capon Minimum Variance Distortionless Response (MVDR) beamforming is implemented for a reasonable tradeoff between complexity and resolving power. The MVDR power spectrum is defined as

$$P(\phi) = \frac{1}{a^H(\phi)\hat{R}^{-1}a(\phi)} \tag{4.3}$$

where a is the array manifold and $\hat{R}$ is the sample covariance matrix. It must be considered that the original signal frequency may not exactly align with the reconstruction frequency grid. Error in estimating the original frequency also introduces error to the angle of arrival estimation. To improve the angle estimation performance, an angle-frequency spectrum is created by calculating the MVDR power for the recovered frequency ± 10 KHz. The peak power from the spectrum is then used as a refined angle estimate.

A simulation for a three element ULA spaced by 1.5 cm is created with three CW sources: 6.02 GHz at $-65°$, 4.44 GHz at $0°$, and 8.52 GHz at $42°$. The source frequencies are correctly identified by reconstruction with one element. The output of the MVDR-based DOA estimation is shown in Figure 4.15. The signals are estimated to be at $-69.5°$, $-0.516°$, and $+44.0°$, all close to the correct angles.

It was shown earlier that using two or more channels with different modulation patterns adds additional incoherent measurements, allowing for faster convergence to the original sparse solution. However, multichannel reconstruction is not fully compatible with using multiple channels for direction finding. Specifi-

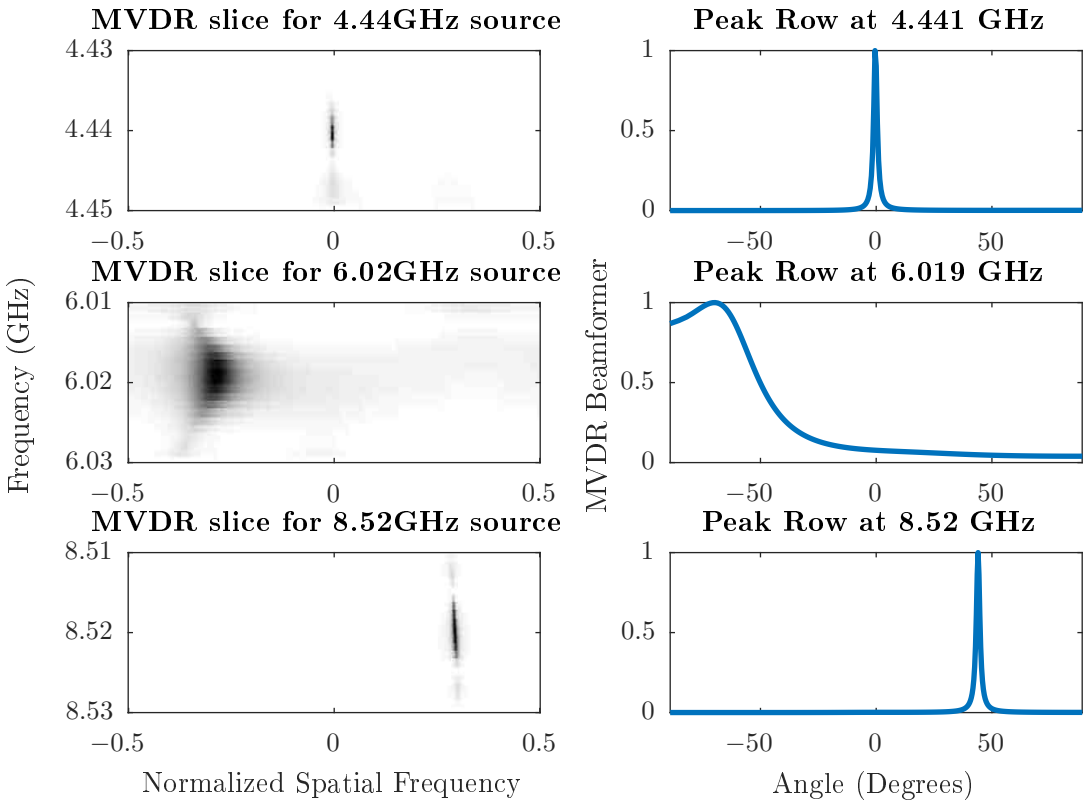

Figure 4.15: Output of direction of arrival simulation. The jointly estimated frequencies and angles are all close to the ideal ones.

cally, the phase difference between the channels connected to different antennas will cause the current system model and compressive sensing solvers to fail when the time delay between the signal arriving at each antenna becomes significant. The NYFR CS model solves for the complex-valued DTFT of the original signal. The system is not just solving for the magnitude of the frequency spectrum, but also the phase. When an RF signal impinges on an array, there is a phase difference between the copy at each channel that is exploitable for direction finding. This phase difference means that the correct solution for a signal that is to be reconstructed for one particular channel is an entirely different solution from the correct one for a separate channel. Now if the measurements from the two different channels were combined, based on the CS model, they give conflicting

information on what the original signal was, lowering reconstruction performance or causing it to fail. There is an additional complication in that when different modulations are applied to two channels they must be demodulated before being used for angle estimation, as the channels are no simply time delayed copies.

The most straightforward to implement solution to have both direction finding capabilities and improved reconstruction performance from multiple channels would be to use a three channel system, with two phase matched channels with different modulation patterns processing the input of one antenna, and the third providing the angle of arrival estimation with a second antenna. However, this obviously has the downside of higher cost. There is some potential to apply algorithms that promote group sparsity or block sparsity to the channels, in which case the solution that separately is the most sparse for each channel is found. Another alternative would be the creation of a phase independent representation basis.

4.4 Communication Signal Recovery

Another advantage of the structured sampling approach of NYFR has over systems with a random sampling matrix is the capability to recover the original signals with a simple demodulation operation, which is more specifically termed as decompression. When the original carrier frequency of a signal is identified, the modulation applied by the harmonic mixer is also known. Demodulating the compressed data using the known Nyquist zone gives the original, decompressed, signals that were present in that zone.

Another nice property is that signals from other zones that fold onto the same compressed frequency remain frequency modulated. Consider that there is one particular signal of interest, and we want to ignore all other signals, so

the other signals are unwanted interferes. The frequency deviation becomes $(M_i - M_{demod})F_\Delta$ where M_i is the original modulation index of an unwanted interferer and M_{demod} is the modulation index for which the data is being decompressed. Thus for an interfering signal with the same modulation sign as the signal of interest, that is an interferer that originates from the same of either the upper or lower zone sidebands, the interferer's frequency deviation increases for each zone between it and the signal of interest. For signals from the other sideband, with the opposite modulation sign, the closer the zones are, the further energy is spread. The important takeaway is that any interfering signals that alias to the same frequency still have their spectrum spread, effectively whitening the interferers and thus improving the SNR of the particular signal of interest.

A simulation is performed for a random QPSK bitstream. For this exercise, 10 symbols are generated at a modulation rate of 5 Msym/s, and AWGN is added to simulated RF input to have an SNR of 30 dB. To simulate interference from other signal sources, two other RF inputs of equal power, one a CW tone at 8.4 GHz, $M = -6$, and the other a FM modulated signal with a bandwidth of 44.1 KHz at 9.6 GHz, $M = 6$, are generated. The QPSK signal has a carrier frequency of 2.4 GHz, $M = -2$, which aliases to 600 MHz. The two interfering signals fold down to the same alias at 600 MHz. The left panel of Figure 4.16 shows the spectrogram of the compressed and digitized output of the receiver. All ten symbols are present in the plot. As the QPSK signal is in the lower Nyquist zone, it can be visually identified as the signal with the lowest peak frequency deviation out of the three signals. Additionally, the one-sided power spectral density of the compressed and decompressed NYFR output is plotted in Figure 4.17.

Reconstruction is performed with SPGL-1 to identify the signals present and their carrier frequencies, with the result shown in Figure 4.18. The reconstruction

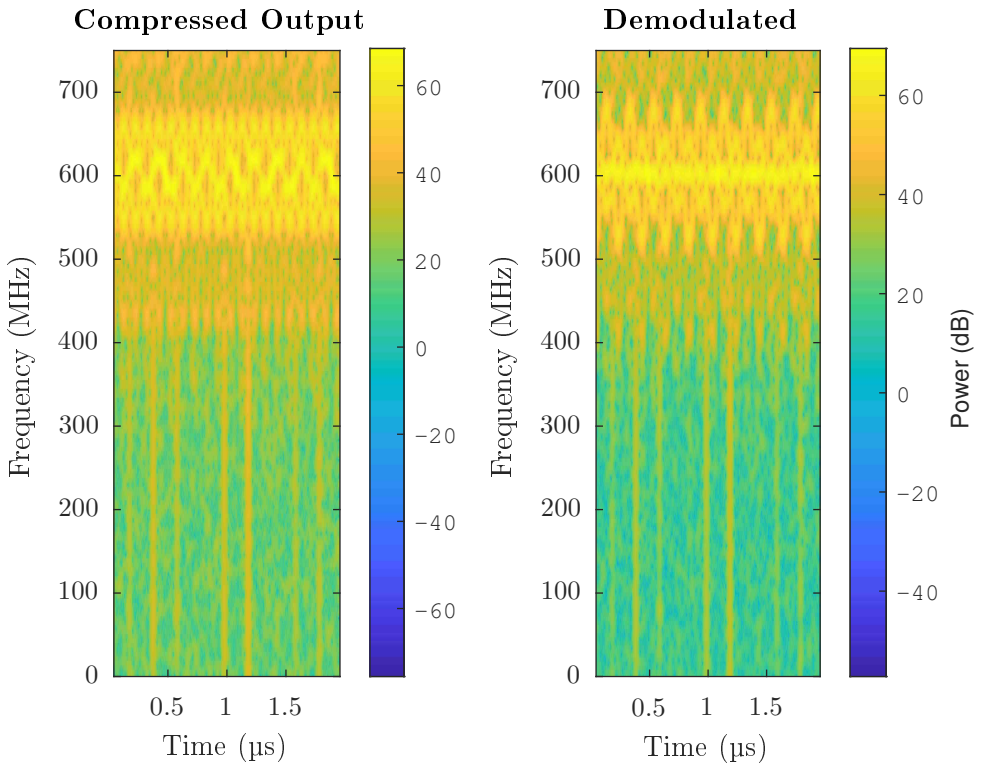

Figure 4.16: The left panel contains the spectrogram of the compressed QPSK signal with interferers. The QPSK signal is from the lowest zone, and as such appears as the signal around 600 MHz with the lowest peak frequency deviation. The right panel shows the spectrogram of the data that has been decompressed for the Nyquist zone that the compressed QPSK signal originates from. The QPSK signal now appears as a straight line at 600 MHz, as the zone dependent frequency modulation has been removed from it.

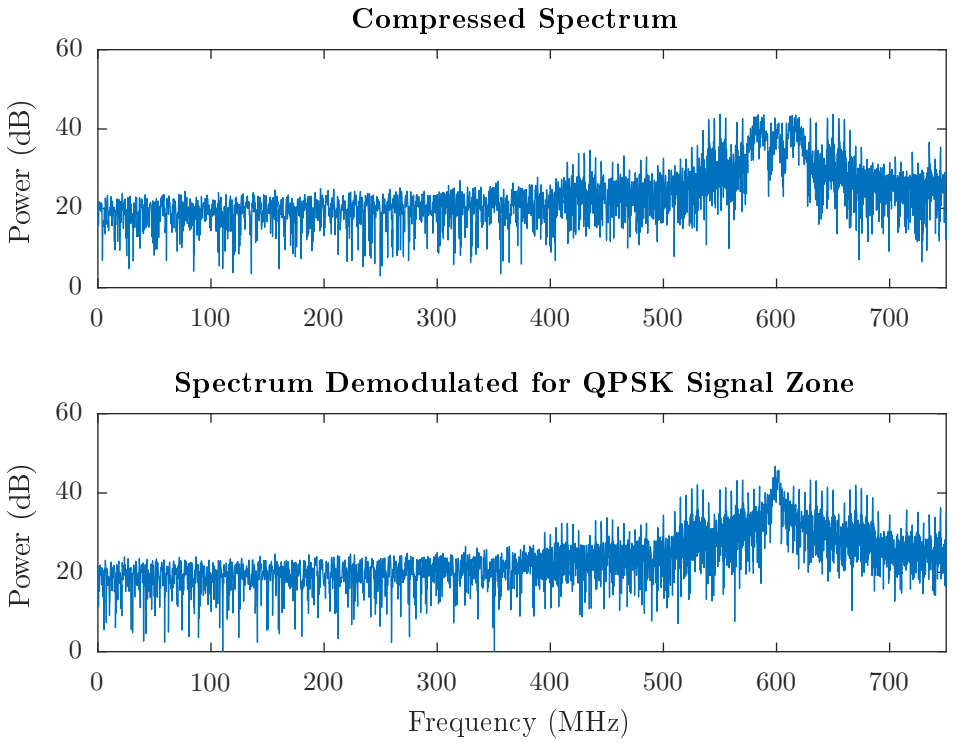

Figure 4.17: Power spectral density (PSD) of the compressed and uncompressed data with interferers. The upper panel shows the compressed spectrum. Note that the energy of the signals that were folded to 600 MHz is spread out over a wide bandwidth. The PSD of the demodulated data has a much narrower peak at 600 MHz from the decompressed QPSK signal.

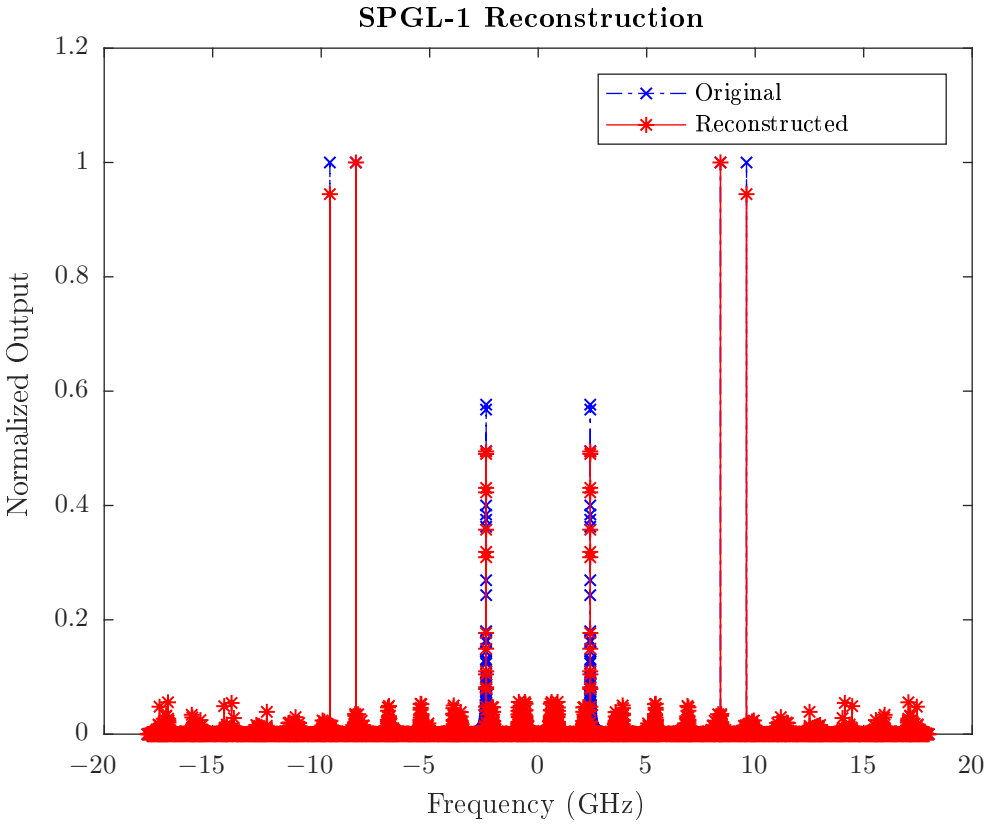

Figure 4.18: Reconstruction of the frequency support of the QPSK signal and two interferers. All three frequencies are reconstructed.

identifies that there is a signal at 2.4 GHz, in the second Nyquist zone. At this point the compressed data can be demodulated for this zone of interest. In a production system it would likely be necessary to demodulate each zone and side-band that the reconstruction reports as having a signal, in this example the three at $M = -2, -6, 6$, as which signals are particularly interesting might not be known.

The demodulation is then performed according to

$$z_{demod,M}(t) = z(t)e^{-jM\theta(t)} \tag{4.4}$$

where $\theta(t)$ is the sinusoidal modulation defined in Chapter 3, with an F_{mod} of 5 MHz and an F_{Δ} of 10 MHz. The demodulation operation has low computation cost as it it simply element-wise multiplication of two vectors, and each zone can be demodulated and processed in parallel, so decompressing multiple zones should not present any throughput issues in a real system. In this case we are interested in the QPSK signal, so the decompressed spectrogram for the $M = 2$ zone is plotted on the right of Figure 4.16. Upon comparison of the two spectrograms in Figure 4.16, it can be clearly seen that the frequency spread of the interfering signals is increased after decompression, as the right panel shows that the frequency deviation of the 9.6 GHz interferer is increased to $11F_{\Delta}$.

In a production system, a priori knowledge of expected communications protocols, or some algorithms to determine the scheme, would be needed. For this simulated case, with a priori knowledge of the bandwidth of the QPSK, and the center frequency recovered by the receiver, a digital FIR bandpass filter is dynamically created with the passband at $f_{recovered} \pm$ 5 MHz and used to isolate the decompressed data. The QPSK communication signal itself is then demodulated, with assumption of prior knowledge of the symbol timing and phase. Figure 4.19 plots the constellation diagram of the demodulated signal. The symbols are well defined, even with the added interference from the two overlapping RF signals of equal power. This simple recovery of the original communication signal displays the advantage of the structured sampling performed by the NYFR, largely preserving the original structure of measured signals.

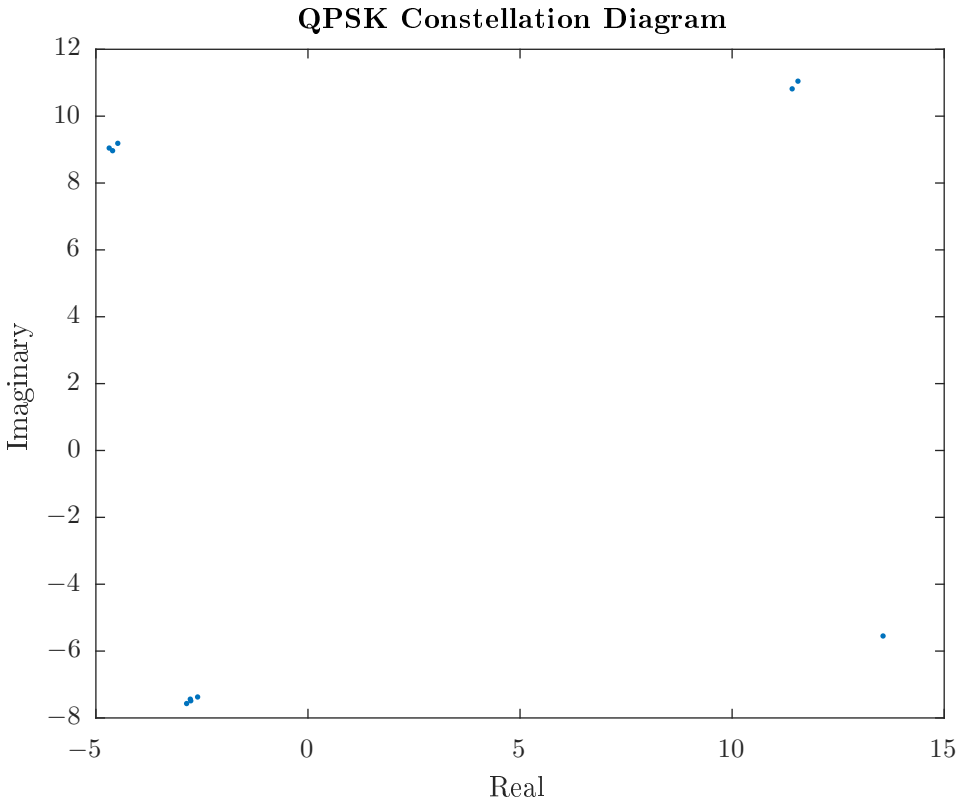

Figure 4.19: Constellation diagram for a simulated QPSK signal with the simulated receiver output decompressed for the QPSK signal's original zone. The two interferers that alias to the same center frequency have little effect, as the demodulated QPSK symbols are closely grouped into the four states that were transmitted. Thus the NYFR can potentially capture communications signals even if multiple signals with bandwidth fold to the same frequency.

Chapter 5

Experimental Results

Now that examination has been made of the trade-offs inherent in the design of $s_{LO}(t)$, and simulations have been performed showing several potential applications of the system, in this chapter we will confirm these findings on a physical NYFR prototype. For testing single channel frequency recovery, the simulations show that F_{Δ}=4 MHz and F_{mod} = 5 MHz should be a reasonable trade-off between the reconstruction dynamic-range, the probability of reconstructing high Nyquist zones, and a minimum reconstructable pulse duration of 0.2 µs, thus reconstructions of CW tones and pulses using these parameters will be demonstrated. First direct stimulation of the harmonic mixer input port will be shown, then later over-the-air (OTA) captures with a broadband antennas are given. The ability to use multiple channels for increased reconstruction performance and direction of arrival estimation will be tested. Finally an OTA communication signal will be decompressed and demodulated.

The complete experimental setup for testing the prototype receiver is pictured is pictured in Figure 5.1. The arbitrary waveform generator used for $s_{LO}(t)$ is at the bottom of the rack on the left, and host PC is to the right with the timing pulse generator sitting on top. The prototype receivers are on top of the rack that holds the AWG. Three copies of the receiver were packaged in two cases to

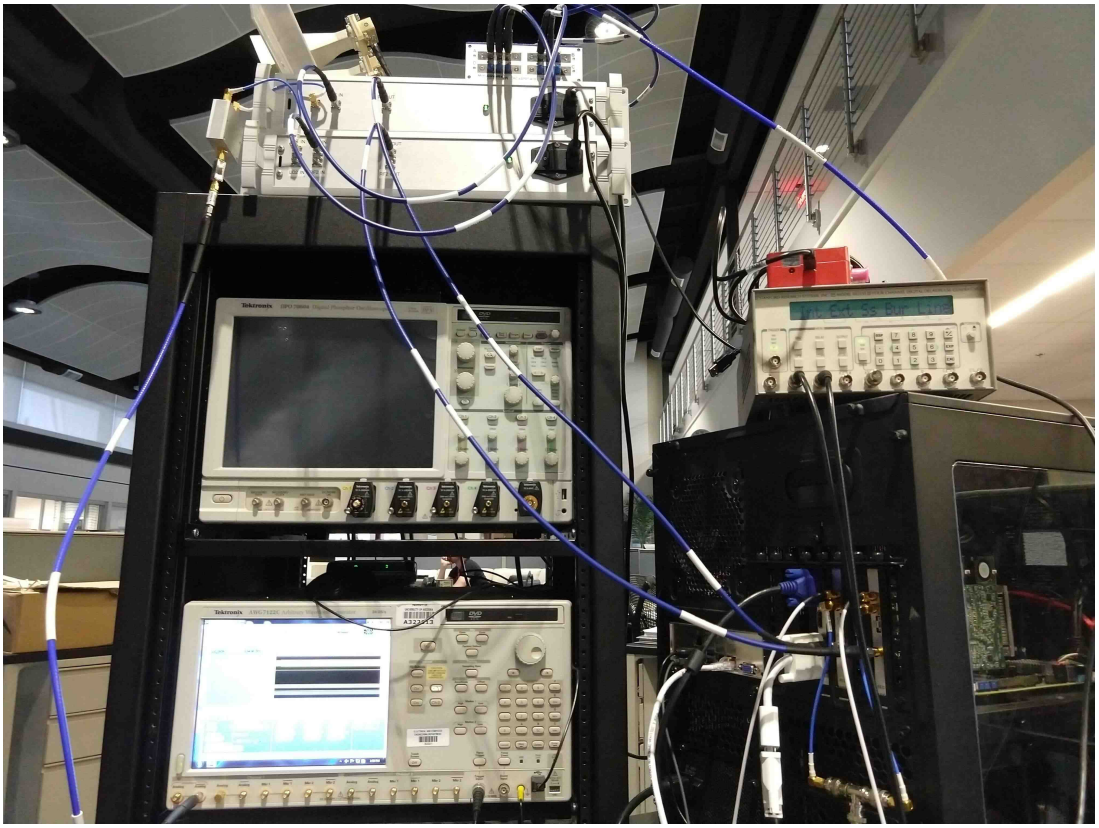

Figure 5.1: Full experimental setup with test equipment.

keep the contents protected but easily accessible.

5.1 Hardware Characterization

In order to measure the actual frequency response and useful bandwidth of the receiver chain, an automated testing suite was created in MATLAB. A CW tone with no modulation was generated at 1.5 GHz and fed into the LO port. Using a constant LO allowed the frequency response of the system to be measured separately from the magnitude roll-off caused by the usual spectrum spreading of higher Nyquist zones. A waveform generator controlled by MATLAB was attached to the RF input port and swept from 0 Hz to 18 GHz in increments of 10 MHz. At each step, a 10 ms ADC capture is taken and the PSD was

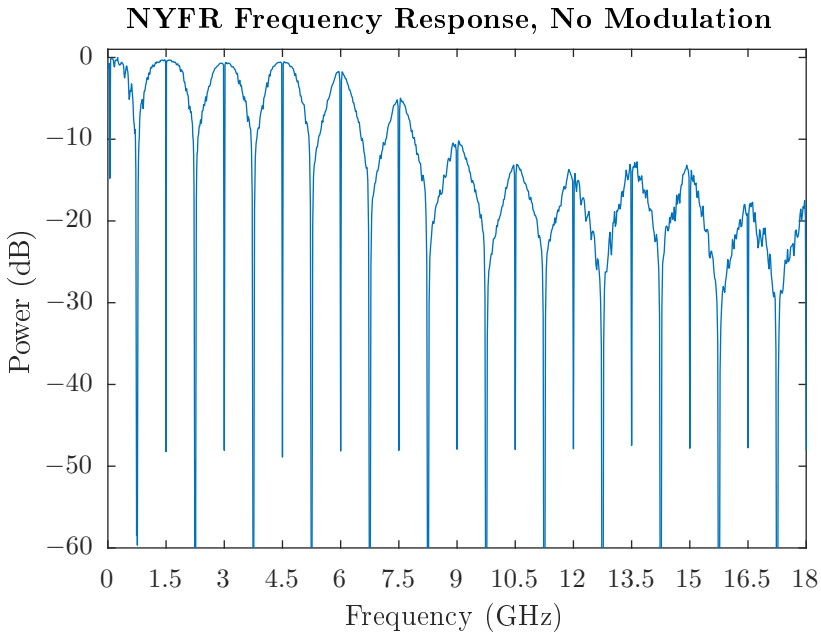

Figure 5.2: Frequency response of a single NYFR channel for $F_\Delta = 0$. The blind frequencies caused by the anti-aliasing filter appear at each multiple of 750 MHz. The roll-off of the harmonic mixer is visible in how the zones, each centered about multiples of 1.5 GHz, are attenuated at the higher frequencies.

estimated in MATLAB. The bin for the downconverted input RF frequency was then recorded. The resulting measure of the frequency response for one of the channels is pictured in Figure 5.2. The roll-off and notches caused by the low pass filter at 750 MHz clearly delineate the sub-bands of each Nyquist zone. The frequency response in each sub-band due to the two LPFs is far from flat; it would likely be useful in many situations to digitally correct for this. The ADCs themselves have about 1dB of attenuation at 750 MHz as well. The roll-off of harmonics generated by the harmonic mixer is shown by the weaker response to higher Nyquist zones. The goal for the receiver was to cover a range from 1.5 GHz to about 16 GHz. The 10 dB attenuation of signals in the higher frequency range did not prevent recovery of high frequency over-the-air signals that were

high power, but if weak signals are to be monitored, it should be improved.

5.2 Single-Channel Recovery

In the early stages of experimenting with the receiver, many tests were performed with the RF port stimulated by direct connection to waveform generators. Some work was required to account for the phase delay of the receiver chains and the delay between the ADC's receive trigger and the data samples. Another issue was that the AWG generated additional high-frequency harmonics in addition to $s_{LO}(t)$, which caused additional, weaker copies of the RF tones with higher frequency deviations in the receiver output. Adding an appropriate bandpass filter on the AWG output alleviated this. Once those two issues were solved, successful recovery of up to three CW tones was demonstrated on each of the receiver channels. Figure 5.3 shows one of the results of these experiments. The left panel shows a spectrogram of the compressed data captured by the signal. In the spectrogram the three input signals with varying peak levels of frequency modulation are visible. The right panel shows the results of reconstruction with the SPGL-1 algorithm, on top of the ideal spectrum of the input frequencies. Figure 5.4 gives another example, with the additional complication of two of the tones folding to the same alias frequency. The power levels of the inputs are also more closely matched in this example.

The timing sensitivity for clean reconstruction of one CW tone with the SPGL-1, SALSA, ROMP, and COSAMP algorithms was tested by increasing the delay between triggering the AWG and the ADCs until reconstruction failed. Both 4.3 GHz and 9.2 GHz tones were tested. For all algorithms, reconstruction fails when the misalignment with the model is greater than 8 ns nanoseconds for a 5 MHz modulation rate LO, slightly more tolerant than predicted in Section 4.1.3.

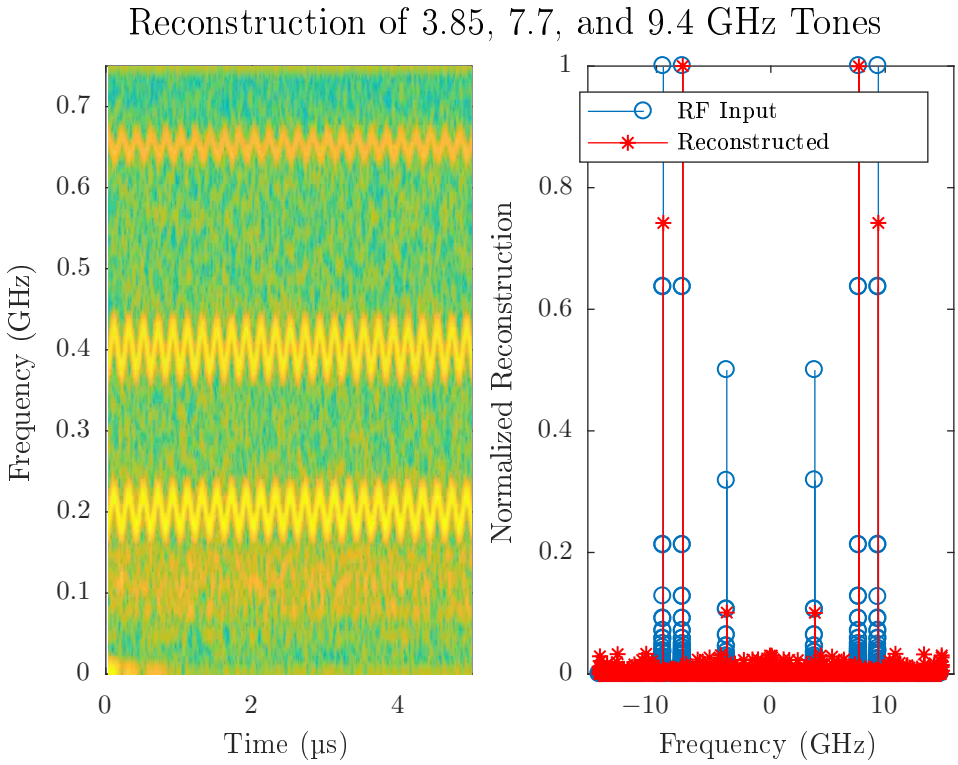

Figure 5.3: Reconstruction of three CW tones from signal generator connected to RF port. The left panel displays a spectrogram of the compressed data captured by the receiver. The compressed 3.85 GHz tone is at the top of the spectrogram, the 9.4 GHz in the middle, and the 7.7 GHz tone is at the bottom. The right panel shows that all three tones are recovered successfully. The 3.85 GHz tone is reconstructed as having very low amplitude, due to the input signal being generated at a lower power and the LPFs attenuating the 650 MHz frequency that the tone folds to.

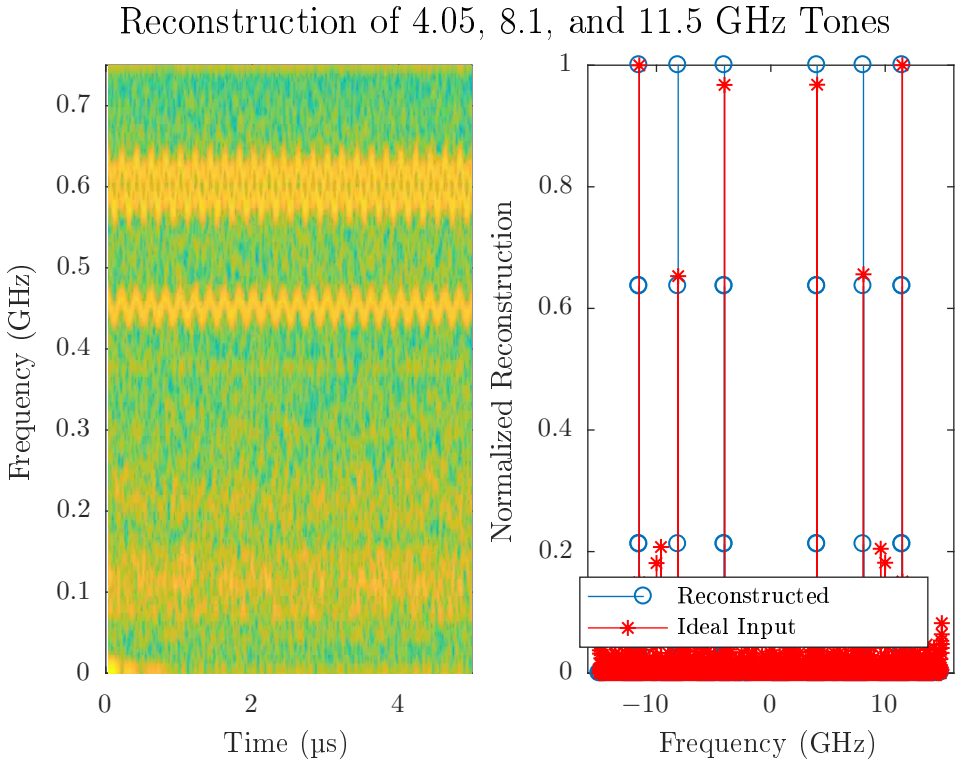

Figure 5.4: Reconstruction of three CW tones. Two fold to same intermediate frequency at 600 MHz. The right pane displays successful reconstruction of all three tones, thus it is shown that signals folding to the same frequency can indeed still be recovered.

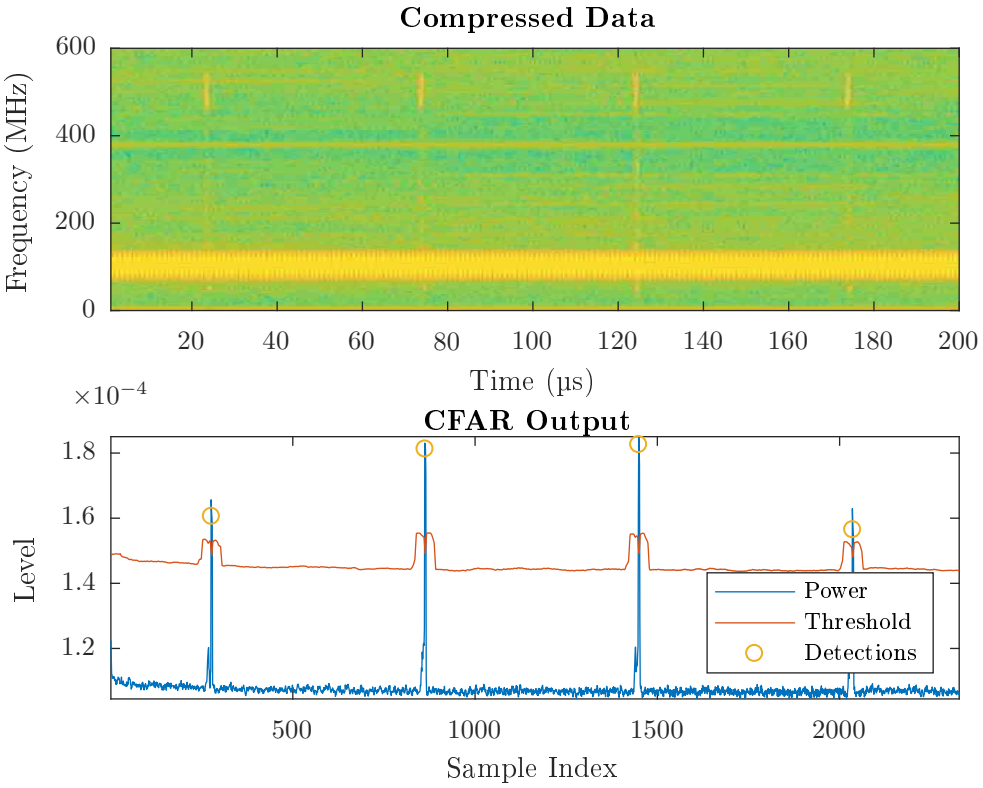

Figure 5.5: Example of CFAR pulse detection used on OTA data. The upper panel shows a spectrogram of the NYFR output. Both a CW tone and a series of 1 µs duration pulses are present. The lower panel shows the CFAR output. The threshold adjusts to detect each pulse once, and all pulses were detected as shown by the circles on the plot.

The timing requirement is not a huge constraint on the design of an NYFR, as the success of this system shows that timing on the order of a nanosecond is not difficult achieve. With the 5 MHz modulation rate commonly used for this implementation, pulses lasting only a fifth of a microsecond can be recovered.

Successful reconstruction was demonstrated with mixed CW tones and microsecond-scale pulses. CFAR thresholding was used as in Chapter 4 to detect processing windows that contain a pulse, then reconstruction was performed for each detected pulse. The CFAR output is shown in Figure 5.5. Figure 5.6 shows an example of one of the detected reconstruction windows. The right pane of the

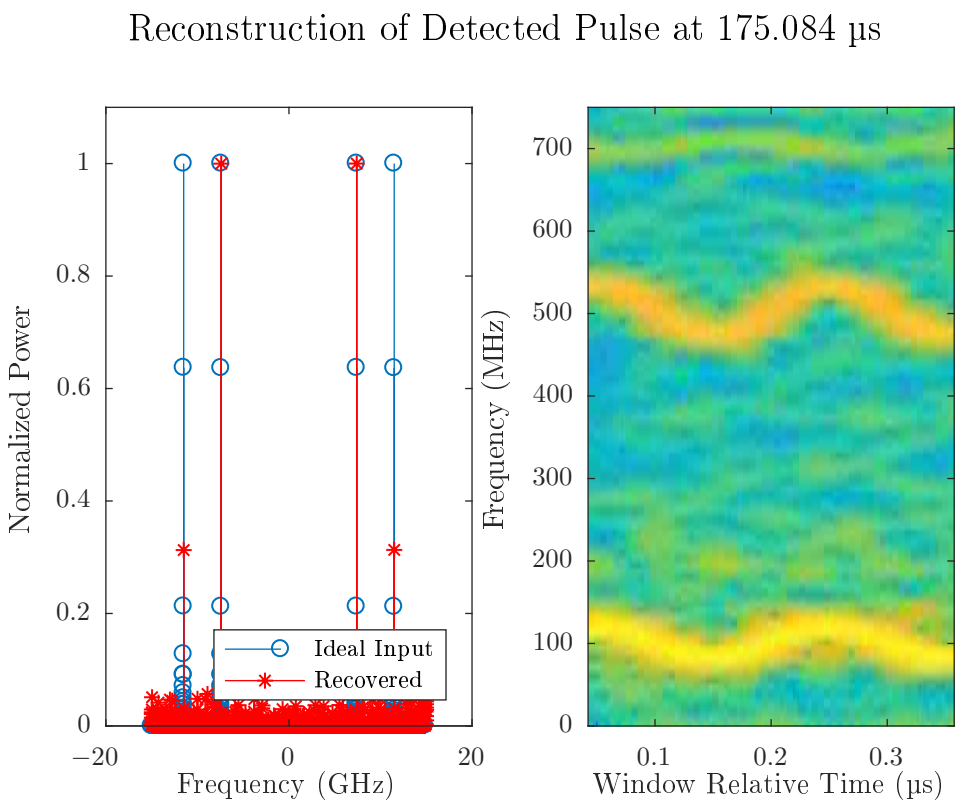

Figure 5.6: Example of OTA pulse reconstructed by CFAR detection and processing window selection. Both the CW tone and pulse frequencies were recovered without performing reconstruction for the whole 200 µs capture, greatly lowering the number of computations performed.

figure displays a spectrogram of the window of compressed data that the reconstruction is performed on. As the spectrogram clearly shows two signals, both the CW tone and at the pulse frequency, it is clear that CFAR detected a window containing a pulse. The reconstruction in the left panel shows that both frequencies were reconstructed.

Next, reconstruction is performed on over-the-air captures. A broadband 2-20 GHz double ridge horn was connected to the RF input of one channel. Captures were performed in a busy lab environment containing multiple WiFi, Bluetooth, and cellular devices. As most of these operate at frequencies that are in lower

Nyquist zones, the wide frequency coverage is demonstrated with two X-band transmitters purposely pointed toward the receiving antenna. A spectrogram and recovered spectrum for one of the more interesting captures, featuring input from a cell phone and a WiFi device, is shown in Figure 5.7. The RF environment in the building was rather sparse, as it contained a limited number of communication devices that share narrow spectrum allocations and transmit infrequently, and the building itself is well shielded from outside transmitters. Figure 5.8 shows another example, with reconstruction performed on a 1 µs window. The dynamic range of the signals present is wide enough that the low-power Bluetooth signal was below the harmonics of the other sources, so an example of predicting and removing harmonics from the recovered spectrum is given.

Another set of data was captured with a high-gain, 1.0-2.5 GHz antenna with the intent of detecting weaker signals in the lower industrial, scientific, and medical (ISM) bands. Figure 5.9 shows one example from this dataset with successful recovery of the center frequencies of two different WiFi channels. Identification of individual channels demonstrates potential usage for the system in cognitive radio; for example a single receiver channel could simultaneously detect which channels are in use in the 900 MHz, 2.4 GHz, or 5 GHz bands in order to select an empty one to use for transmission.

Successful recovery of the frequency support has been demonstrated for a sparse environment and for up to five simultaneous sources. It is significant that the communication signals recovered have bandwidth, proving that the system can perform recovery of signals other than ideal, frequency-bin-centered, CW tones.

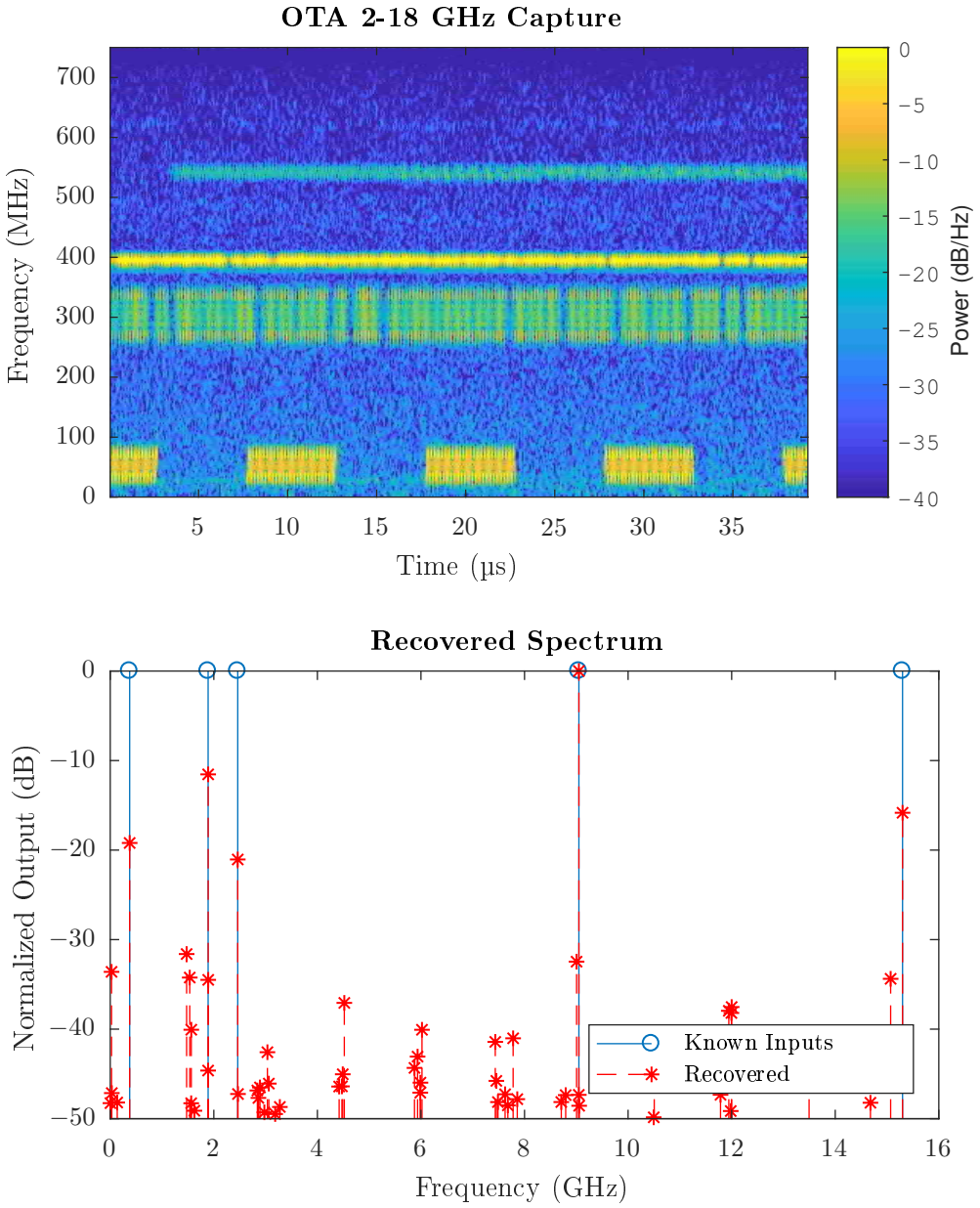

Figure 5.7: Example of reconstruction of over-the-air capture with a 2 to 18 GHz antenna. A high power QPSK signal is present at 15.3 GHz, 5 µs pulses at 9.05 GHz, LTE cell phone signal at 1.9 GHz, and a WiFi QPSK signal at 2.462 GHz. A spur from the ADC sampling clock at $f_s/4$, 375 MHz, was also reconstructed. This gives a useful example of recovery in an actual sparse-RF environment that contains several communication signals with some amount of bandwidth.

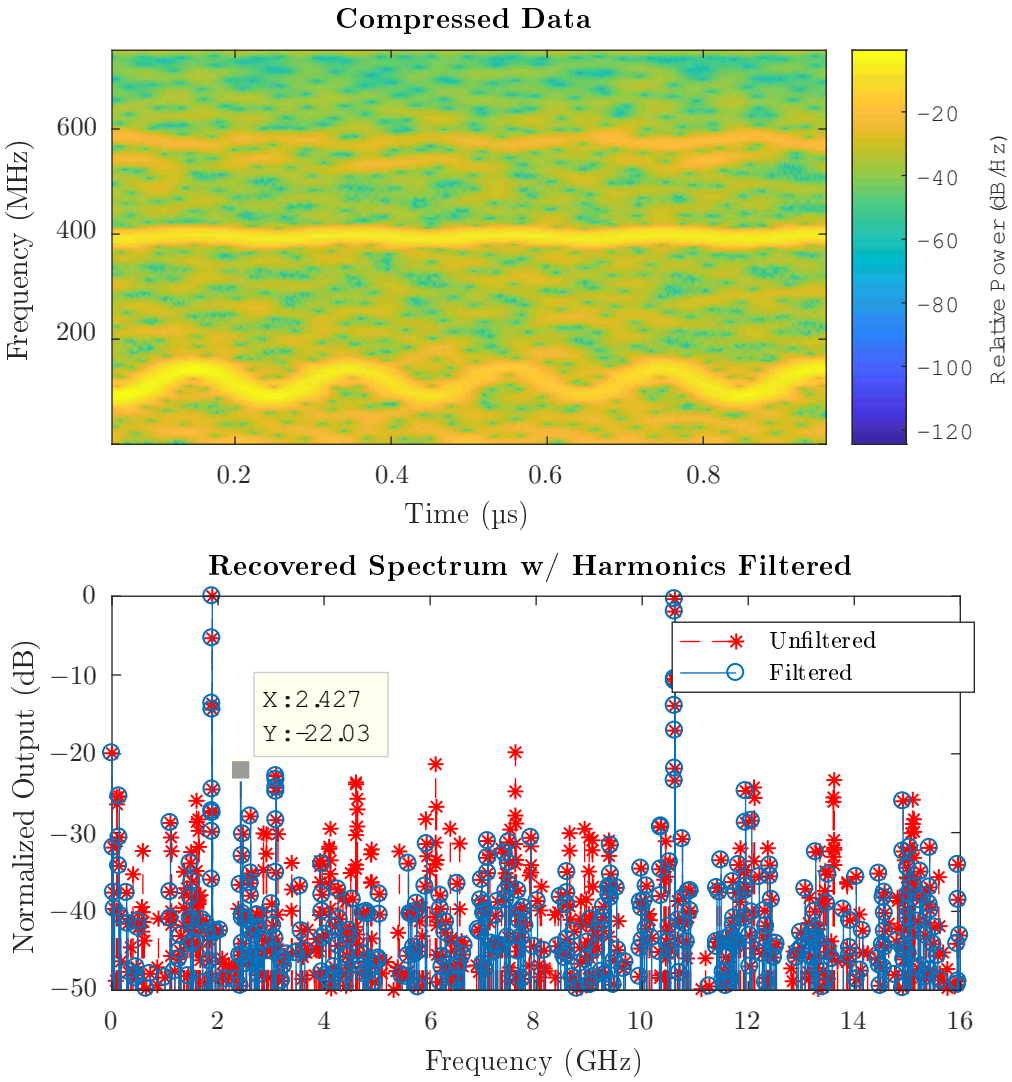

Figure 5.8: Example of reconstruction of over-the-air capture with a 2 to 18 GHz antenna. A high power QPSK signal is present at 10.6 GHz, LTE cell phone signal at 1.9 GHz, and a low-power Bluetooth signal at 2.42 GHz. The relative power of the Bluetooth signal is so low compared to the other two signals that it is outside of the dynamic range for the $s_{LO}(t)$ used, as can be seen by the red stars. An example of filtering the harmonics by iterating through the strongest reconstructions, removing their predicted harmonics until a threshold of -30 dB is hit, is given by the blue circles.

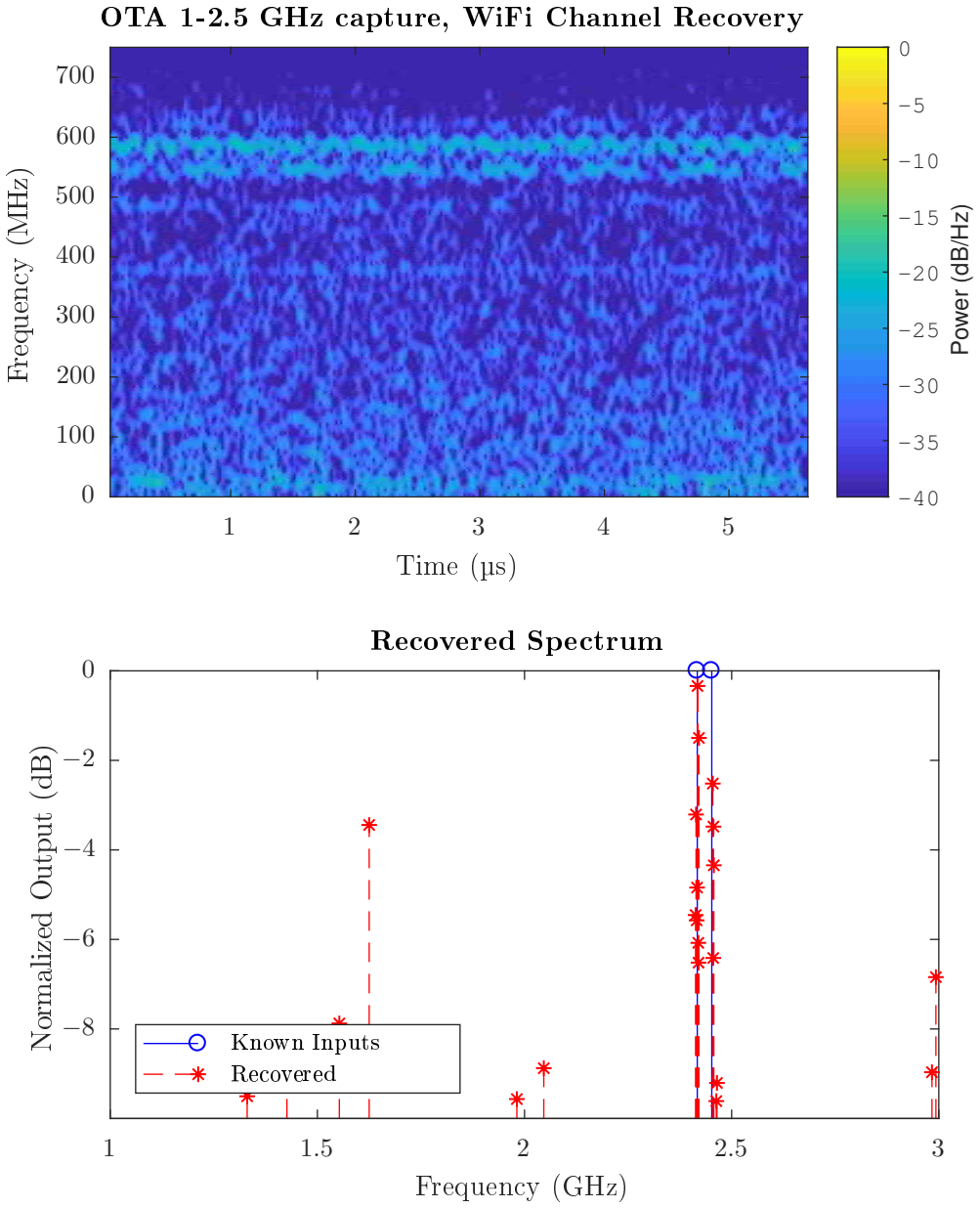

Figure 5.9: Example spectrogram and reconstruction from the dataset captured with a high gain 1-2.5 GHz antenna. The spectrogram clearly shows two communication signals between 2.4 and 2.5 GHz, as well as the faint ADC sample clock spur at 350 MHz. A scan of available WiFi networks from a smartphone shows that devices are transmitting on WiFi Channel 1 and 11 at about 2.412 and 2.462 GHz, matching within ±1 MHz of the peak frequencies in the reconstructed spectrum shown at the bottom of the figure.

5.3 Multichannel Recovery

After single-channel recovery was demonstrated, experiments were performed with multiple receiver channels. It is predicted in Section 4.3 that simultaneously using the data from multiple channels with different modulation schemes has a positive impact on reconstruction performance, as more incoherent measurements are being used to reconstruct the signal. For the ideal case of using the same modulation, there should be no benefit, as the additional samples from one channel would be identical to those from the other and thus add no additional information. In reality the ADCs will not sample at the exact same time, and thus multiple channels with the same modulation scheme would provide more measurements, but in order to actually realize benefit the phase difference between the channels would have to be known and accounted for in the sensing model.

First for comparison identical modulation schemes are used on each channel. The external triggers used to time the LO waveform generator and the PCIe ADCs start the capture on each ADC card within ± 1 sample of each other, or approximately within 1 ns, which is within the requirements for matching the sensing matrix model. As the ADCs used were not specifically designed to be well matched for use in multichannel applications, the frequency response of the two differ somewhat. The signal centered at 340 MHz in the baseband pictured in Figure 5.10 is about 3 dBm lower in channel 2 than channel 1. This difference did not prevent recovery in the experiments performed, but it is possible that a system designed to have more closely matched channels would have better reconstruction performance.

Next experiments were performed with different modulation schemes applied

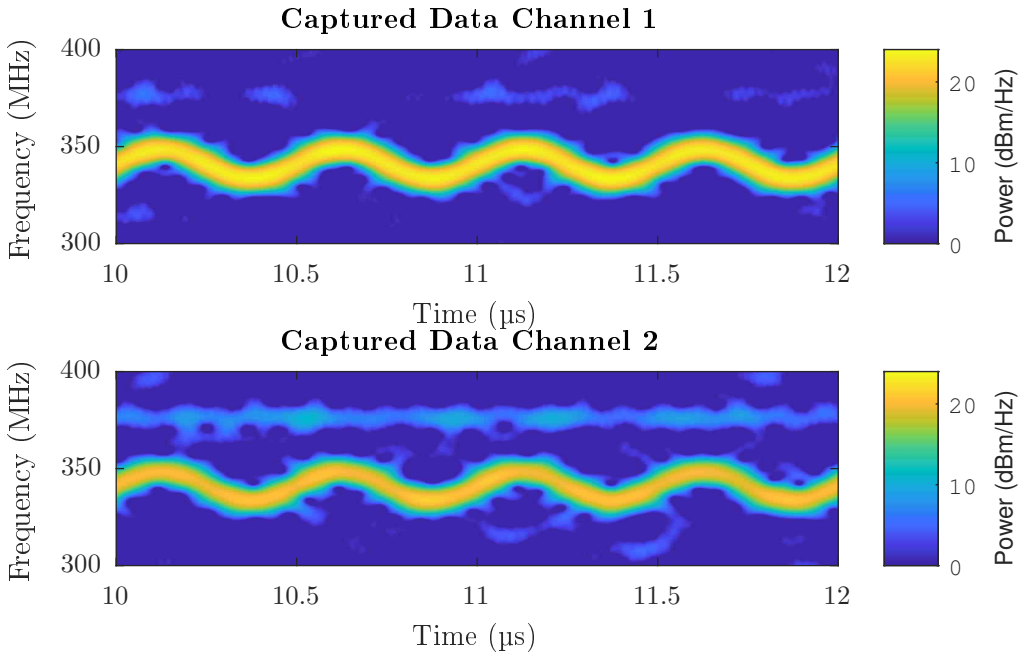

Figure 5.10: Comparison of two receiver channels with same modulation. The signal in Channel 2 is about 3 dBm lower and the ADC $f_s/4$ spur is slightly worse. The channels are reasonably well matched though, as reconstruction with identical CS models succeeded on each and successful angle of arrival estimation will be shown later.

to each channel, specifically combining linear and sinusoidal frequency modulation. Figure 5.11 gives an example of using sinusoidal modulation on channel one with a peak deviation of 4 MHz and modulation rate of 5 MHz, and triangular modulation on channel two with a peak deviation of 6 MHz and 0.5 µs period. The reconstructed spectrum using the combined measurements from both channels shows improvement over the single-channel reconstructions with a decreased peak spurious frequency magnitude. This matches the theoretical improvements shown in Section 4.3.

Now to test direction finding, an experiment was created with a two element array of X-band horn antennas. The horns are physically large, limiting the minimum element separation to 2.75 cm, giving a maximum grating lobe free look angle of only $\pm 5.2°$ for a 10 GHz source. This would limit utility in a practical

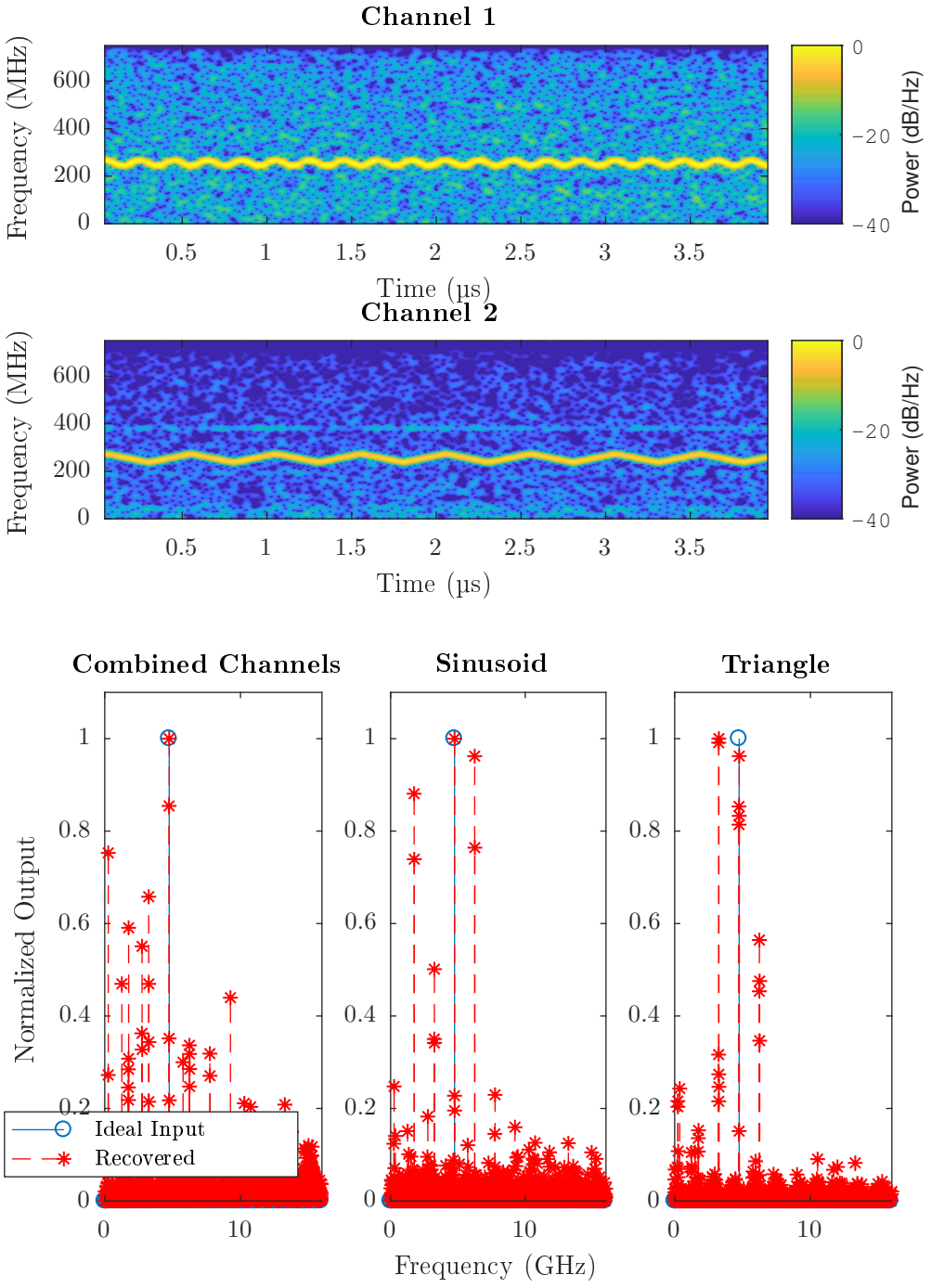

Figure 5.11: Combined multichannel recovery with different modulation schemes. Channel 1 has sinusoidal modulation and Channel 2 has a triangle waveform applied. The recovery resulting from the combined channels has a larger difference between the correct signal at 4.75 GHz and the strongest false result. Thus it is shown that combining measurements from multiple modulation schemes can give improved reconstruction quality.

system unless mechanical scanning is used, but for the purpose of showing that the direction of arrival estimation is viable, the limited maximum look angle is sufficient. The coordinate system is set up such that when facing from the array, negative angles are to the left of the array center and positive to the right. A pulsed source at 7.845 GHz is to the right of a pulsed source at 8.02 GHz, both sources are less than 5 degrees to the right of the array center. When a single-channel is used the two source frequencies are identified in Figure 5.12. The sources are actually close enough to the center of the array that the time delay between the two elements is lower than 4.25 ns, specifically

$$\tau \leq 0.0275 \frac{\sin(5^\circ)}{c} = 8ps \quad , \tag{5.1}$$

low enough that combining measurements from both channels for reconstruction does not fail, but due to the phase delay the reconstruction quality degraded. As the difference between the source angle and the array center increases, reconstruction over multiple channels should fail. The MVDR angle-frequency spectrum is calculated around the two source frequencies and the peak values selected. Figure 5.13 shows that the 7.845 GHz source is correctly identified as being right of the 8.02 GHz source, and the estimated angles for each are to the right of the array center as expected. Thus application of the NYFR to joint direction of arrival and frequency estimation is shown to be viable.

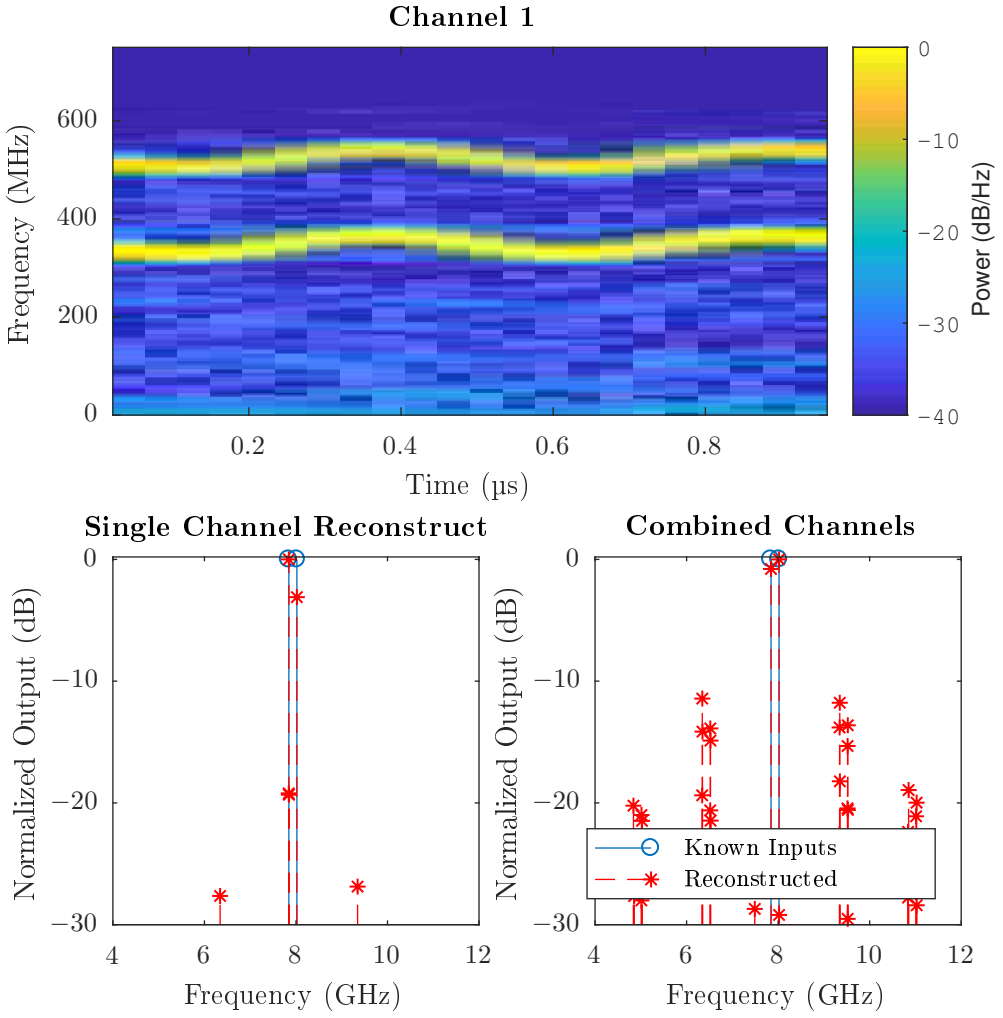

Figure 5.12: Recovery of source frequencies impinging on the two element array. Single-channel reconstruction is successful. The combined channel reconstruction is worse due to the time delay between the array elements, but does not fail outright as in this case the time delay is much less than a nanosecond.

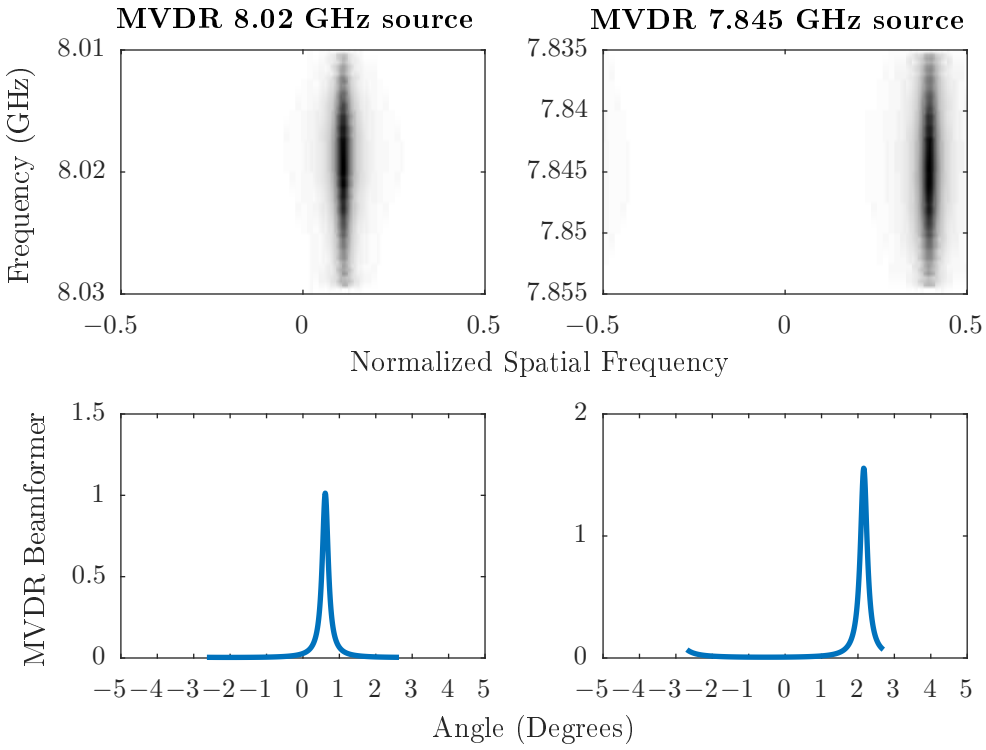

Figure 5.13: Plots of the normalized angle-frequency spectrum slice around the estimated frequency. The bottom plots of are the MVDR beamformer output for the peak frequency slice found from the spectrum above. The 8.02 GHz is correctly identified as being on the left of the 7.45 GHz source.

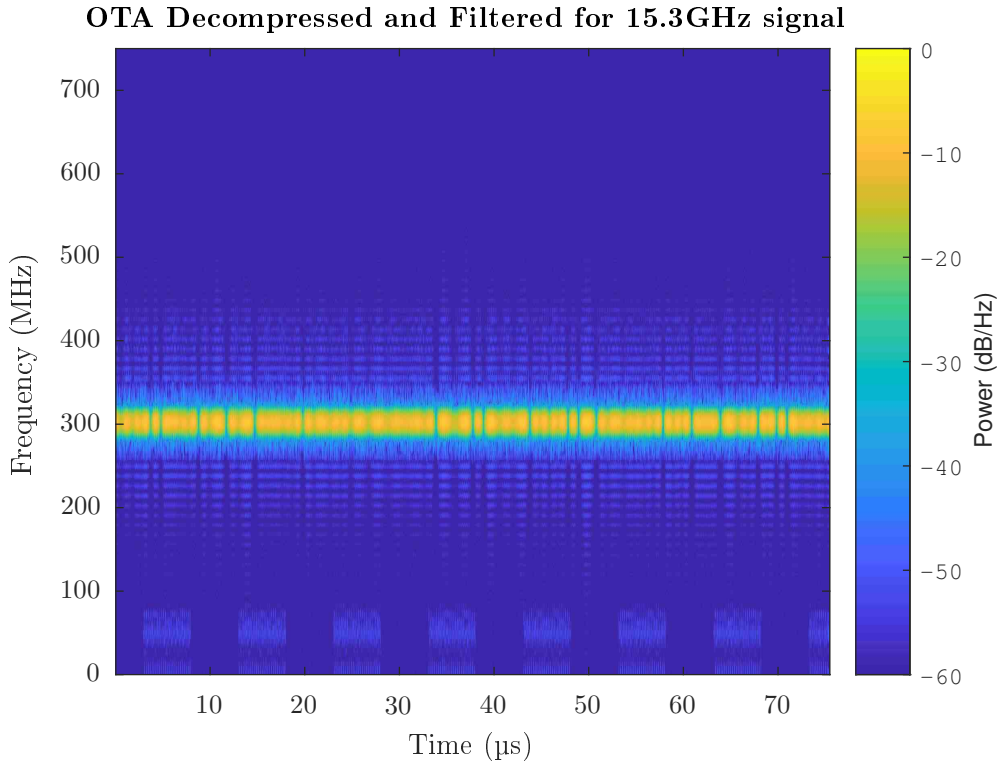

Figure 5.14: The OTA capture in Figure 5.7 was decompressed for M=10, then bandpass filtered with a center frequency at 300 MHz in order to extract the signal detected at 15.3 GHz. The spectrogram of the decompressed and filtered output clearly shows one communication signal at 300 MHz. The pulsed signal that was present in the capture is faintly visible at about -50 dB below the comm signal.

5.4 Communication Signal Recovery

Final experiments were performed to show that recovery of communication signals by the NYFR is possible. The OTA capture from Figure 5.7 is decompressed for the modulation index M=10 and bandpass filtered to extract the QPSK signal at 15.3 GHz. The QPSK signal is a randomly generated bitstream with a known symbol rate of 1 Msym/s. Figure 5.14 shows the decompressed and filtered QPSK signal. Root-raised-cosine pulse shaping is used on the transmitter, so matched filtering is applied to the decompressed data. Timing recovery was done by sam-

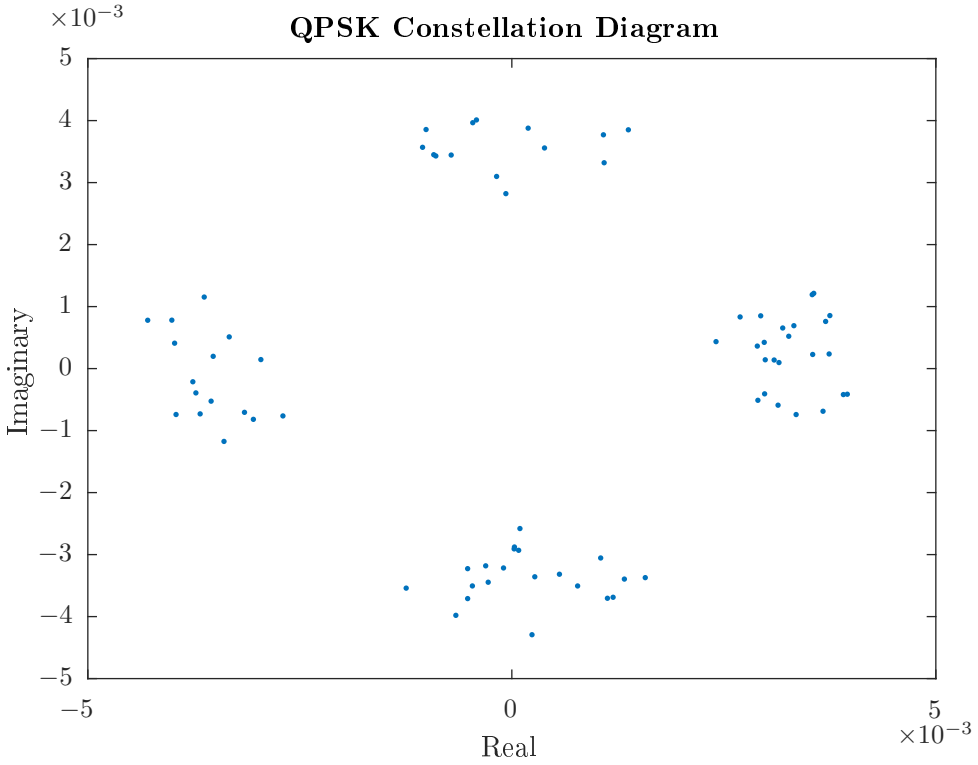

Figure 5.15: Constellation diagram of the QPSK signal recovered from the over-the-air capture. The constellation is well defined, with the 80 symbols captured all clumped relatively close one of the four transmitting states. Now if the pilot signal or sequence were known by the system, the digital information that was transmitted has been recovered by the NYFR.

pling the matched filter peaks, as the processing was done offline. The recovered constellation diagram is plotted in Figure 5.15, showing that the symbols are well identified. Thus the viability of recovering communication signals with the NYFR is shown.

Chapter 6

Conclusions and Future Work

The trade-offs of design parameters for $s_{LO}(t)$ which controls the modulation pattern induced on RF signals by the Nyquist Folding Receiver were identified and discussed. In particular, a trade-off must be made between correctly identifying the original Nyquist zone and the ability to recover low SNR signals from higher Nyquist zones. The LO signal modulation rate must be selected for the shortest pulse time that is to be recoverable while not being such a high rate of modulation that the phase difference of the real hardware and the compressive sensing model is not too great for reconstruction.

Multichannel NYFR systems were also considered. A system model was created that combines measurements from multiple channels, and it is found that if each channel has a unique modulation scheme, more incoherent measurements are collected in the same time window, improving reconstruction quality. Conversely, combining multiple channels with identical modulation does not affect reconstruction. Multiple channels are also used for joint angle of arrival and frequency estimation.

A full process for identifying a communication burst, recovering the frequency support, and then decompressing the data with a zone-dependent decompression (demodulation) operation for detected signals of interest was given. It was

demonstrated that when decompressing a particular zone, the spectrum of signals originating from other Nyquist zones remains spread, so that even if multiple signals fold to the same intermediate frequency, the particular communication signal of interest can be demodulated.

Experiments were performed with a hardware prototype. Successful recovery of up to five simultaneous signals was shown, including identification of the center frequencies for LTE, WiFi, and Bluetooth communication signals. The advantage of combining measurements from multiple channels was demonstrated. Joint angle and frequency estimation is shown to be viable with an antenna array connected to the prototype. These results show that the Nyquist Folding Receiver is an architecture with very promising potential for wide instantaneous bandwidth applications.

Interesting future topics not explored by this thesis include modifying or formulating compressive reconstruction algorithms to be optimized for the structured sampling performed by the receiver. Dictionary learning of signals that are to be identified in a real environment would be useful for monitoring purposes. An improved sensing model with a phase-independent representation basis could potentially be derived and used to combine measurements from multiple channels in a direction finding array, or a group sparsity algorithm might be used to promote sparsity in each channel. It would also be interesting to test the receiver in an environment with a larger number of controlled signal sources, in determine what the maximum sparsity level is that can be accommodated by the receiver subject to a required reconstruction probability.

Bibliography

[1] G. L. Fudge, R. E. Bland, M. A. Chivers, S. Ravindran, J. Haupt, and P. E. Pace, "A nyquist folding analog-to-information receiver," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pp. 541–545, October 2008.

[2] H. Nyquist, "Certain factors affecting telegraph speed," *Transactions of the American Institute of Electrical Engineers*, vol. XLIII, pp. 412–422, January 1924.

[3] E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3, no. 12, pp. 1433–1452, 2006.

[4] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the $\ell 1$ norm," *GEOPHYSICS*, vol. 44, no. 1, pp. 39–52, 1979.

[5] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, December 1993.

[6] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[7] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003. [Online]. Available: http://www.pnas.org/content/100/5/2197

[8] R. Maleh, G. L. Fudge, F. A. Boyle, and P. E. Pace, "Analog-to-information and the nyquist folding receiver," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 564–578, September 2012.

[9] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee, "Compressive sensing: From theory to applications, a survey," *Journal of Communications and Networks*, vol. 15, no. 5, pp. 443–456, October 2013.

[10] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University, 2005.

[11] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, Oct 2007.

[12] E. Arias-Castro and Y. C. Eldar, "Noise folding in compressed sensing," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 478–481, Aug 2011.

[13] M. S. Asif and J. Romberg, "Streaming measurements in compressive sensing: $\ell 1$ filtering," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, pp. 1051–1058, October 2008.

[14] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589 – 592, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1631073X08000964

[15] A. S. Bandeira, E. Dobriban, D. G. Mixon, and W. F. Sawin, "Certifying the restricted isometry property is hard," *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3448–3450, June 2013.

[16] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, 2007.

[17] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[18] ——, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008. [Online]. Available: http://link.aip.org/link/?SCE/31/890

[19] J. Yoo, S. Becker, M. Monge, M. Loh, E. Candès, and A. Emami-Neyestanak, "Design and implementation of a fully integrated compressed-sensing signal acquisition system," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5325–5328, March 2012.

[20] M. Wakin, S. Becker, E. Nakamura, M. Grant, E. Sovero, D. Ching, J. Yoo, J. Romberg, A. Emami-Neyestanak, and E. Candès, "A nonuniform sampler for wideband spectrally-sparse environments," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 516–529, Sept 2012.

[21] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *CoRR*, vol. abs/0902.4291, 2009. [Online]. Available: http://arxiv.org/abs/0902.4291

[22] ——, "Xampling: Analog data compression," in *2010 Data Compression Conference*, pp. 366–375, March 2010.

[23] F. Uysal, J. C. Martin, and N. A. Goodman, "Single channel rf signal recovery for nyquist folding receiver," forthcoming in *IEEE International Radar Conference 2017*, October 2017.

[24] R. Maleh and G. L. Fudge, "RIP analysis of modulated sampling schemes for recovering spectrally sparse signals," *CoRR*, vol. abs/1207.7347, 2012. [Online]. Available: http://arxiv.org/abs/1207.7347

[25] P. P. Gandhi and S. A. Kassam, "Analysis of cfar processors in nonhomogeneous background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 427–445, July 1988.

[26] C. Qian, L. Huang, Y. Shi, and H. C. So, "Joint angle and frequency estimation using structured least squares," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2972–2976, May 2014.

[27] A. N. Lemma, A. J. van der Veen, and E. F. Deprettere, "Analysis of joint angle-frequency estimation using esprit," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1264–1283, May 2003.

[28] S. S. Reddi, "Multiple source location - a digital approach," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-15, no. 1, pp. 95–105, Jan 1979.

[29] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, March 1986.

[30] B. Friedlander and A. J. Weiss, "Direction finding for wide-band signals using an interpolated array," *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1618–1634, April 1993.

[31] J. L. Moll, S. Krakauer, and R. Shen, "P-N junction charge-storage diodes," *Proceedings of the IRE*, vol. 50, no. 1, pp. 43–53, January 1962.

[32] R. J. Weber and Y. Huang, "A joint frequency and doa estimation algorithm with diagonal load," in *2012 IEEE Aerospace Conference*, pp. 1–6, March 2012.