ESC-TR-2006-086

Project Report
A2I-1

Analog-to-Information Study Phase Final Report

K.W. Forsythe
J.I. Goodman
M.R. Green
B.A. Miller
G.M. Raz
J.H. Jackson

10 October 2007

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Defense Advanced Research Projects Agency under Air Force Contract FA8721-05 C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by the Defense Advanced Research Projects Agency, MTO, under Air Force Contract FA8721-05-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian
Administrative Contracting Officer
Plans and Programs Directorate
Contracted Support Management

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
10 October 2007	Project Report	

4. TITLE AND SUBTITLE
Analog-to-Information Study Phase Final Report

5a. CONTRACT NUMBER
FA8721-05-C-0002

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
K.W. Forsythe, J.I. Goodman, M.R. Green,
B.A. Miller, G.M. Raz, and J.H. Jackson

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420-9108

8. PERFORMING ORGANIZATION REPORT NUMBER
PR-A2I-1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Dennis Healy
Microsystems Technology Office, Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
ESC-TR-2006-086

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release: distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Many communications and Radar receivers must process data over a very wide band, which requires either high-rate analog-to-digital converters (ADCs) or multichannel receivers. The information content of that wideband data, however, is often sparse in some basis. Analog-to-Information (A2I) receivers exploit this sparseness in both the digital and analog domains by non-adaptively spreading the signal energy (analog) and using digital signal processing to recover the signal from an ADC sampling at a sub-Nyquist rate. A subsampled ADC implies the use of fewer receiver channels or less expensive, lower-rate devices. This report documents the signal processing techniques for such receivers developed by the MIT Lincoln Laboratory/GMR Research and Technology team in the study phase of the A2I program. We have developed two new A2I signal processing methods, both significantly outperforming compressed sensing (CS) techniques currently in the literature, which typically fail when signals occupy more than 15–20% of the downsampled band. One of our methods, Nonlinear Affine processing (NoLaff), uses a nonlinear front-end to spread signal energy before the sub-Nyquist ADC, and uses hypothesis testing to reconstruct the signal. In simulations, this technique has shown that it can reconstruct wideband signals occupying up to 72% of the downsampled basis. It is also much less sensitive to the difficulties CS has detecting signals with large magnitude variation in the compressible basis. Our other method, called Variable Projection and Unfolding (VPU), spreads the signal energy using random linear projections similar to those used in compressed sensing, but is able to reconstruct signals occupying nearly 100% of the downsampled basis. VPU achieves this using a technique similar to matching pursuit; the key difference being that VPU searches over blocks of consecutive columns rather than one column at a time. Performance bounds for NoLaff, VPU and traditional compressed sensing algorithms are also presented, supporting our experimental results. We also present a different, data-adaptive method for subsampled signal recovery: the use of dynamical systems. Dynamical systems provide a natural model for A2I receivers due to their randomization and memory properties. When driven by an unknown input signal that is sparse in some known basis, a dynamical system can use an ordinary differential equation to reconstruct the input signal. We present a description of dynamical system implementation, as well as some initial qualitative results. All three techniques show promise for use in future intelligence, surveillance and reconnaissance systems.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Same as report	113	19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified			

Massachusetts Institute of Technology
Lincoln Laboratory

Analog-to-Information Study Phase Final Report

J.I. Goodman
B.A. Miller
Group 102

K.W. Forsythe
Group 103

M.R. Green
Group 105

G.M. Raz
J.H. Jackson
GMR Research and Technology

Project Report A2I-1

10 October 2007

Lexington Massachusetts

ABSTRACT

Many communications and Radar receivers must process data over a very wide band, which requires either high-rate analog-to-digital converters (ADCs) or multichannel receivers. The information content of that wideband data, however, is often sparse in some basis. Analog-to-Information (A2I) receivers exploit this sparseness in both the digital and analog domains by non-adaptively spreading the signal energy (analog) and using digital signal processing to recover the signal from an ADC sampling at a sub-Nyquist rate. A subsampled ADC implies the use of fewer receiver channels or less expensive, lower-rate devices. This report documents the signal processing techniques for such receivers developed by the MIT Lincoln Laboratory/GMR Research and Technology team in the study phase of the A2I program. We have developed two new A2I signal processing methods, both significantly outperforming compressed sensing (CS) techniques currently in the literature, which typically fail when signals occupy more than 15–20% of the downsampled band. One of our methods, Nonlinear Affine processing (NoLAff), uses a nonlinear front-end to spread signal energy before the sub-Nyquist ADC, and uses hypothesis testing to reconstruct the signal. In simulations, this technique has shown that it can reconstruct wideband signals occupying up to 72% of the downsampled basis. It is also much less sensitive to the difficulties CS has detecting signals with large magnitude variation in the compressible basis. Our other method, called Variable Projection and Unfolding (VPU), spreads the signal energy using random linear projections similar to those used in compressed sensing, but is able to reconstruct signals occupying nearly 100% of the downsampled basis. VPU achieves this using a technique similar to matching pursuit; the key difference being that VPU searches over blocks of consecutive columns rather than one column at a time. Performance bounds for NoLAff, VPU and traditional compressed sensing algorithms are also presented, supporting our experimental results. We also present a different, data-adaptive method for subsampled signal recovery: the use of dynamical systems. Dynamical systems provide a natural model for A2I receivers due to their randomization and memory properties. When driven by an unknown input signal that is sparse in some known basis, a dynamical system can use an ordinary differential equation to reconstruct the input signal. We present a description of dynamical system implementation, as well as some initial qualitative results. All three techniques show promise for use in future intelligence, surveillance and reconnaissance systems.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Communications and Radar receivers are frequently tasked with processing very wideband signals. In many cases, the information content of the very wideband signal is compressible in some basis, equating to the signal having a sparse representation in the basis in which it is compressible. Analog-to-Information (A2I) receivers exploit a signal's sparse representation by non-adaptively spreading the RF/IF signal and then recovering the original signal from an analog-to-digital converter (ADC) sampling at a sub-Nyquist rate. The type of RF/IF spreading employed by an A2I receiver includes random linear projections used in many compressed sensing (CS) applications [1], as well as nonlinearities described in subsequent sections of this document. A2I receivers have application to electronic intelligence (ELINT) and signals intelligence (SIGINT) systems that would leverage the additional dynamic range of sub-Nyquist sampling ADCs to detect weak target signatures. For example, moving from a 1000 MSPS to a 100 MSPS ADC using present ADC technology would increase the spur free dynamic range (SFDR) by roughly 30 dB.

A2I receivers will also benefit intelligence, surveillance and reconnaissance (ISR) sensor networks. ISR sensor networks that pass unprocessed information back to a fusion center will be performance-limited both by power consumption, dynamic range and available communication bandwidth. The capacity of an ISR receiver to digitize a wide bandwidth with high dynamic range and pass the potentially unprocessed information onto a fusion center under tight power and communication bandwidth constraints is a significant challenge. Examples of such sensor networks include the Navy's Cooperating Engagement Capability, which is forced to exchange highly processed detection reports as opposed to unprocessed multimodality sensor data due to communication bandwidth immuration [2].

However, many ISR systems operate in environments populated with signals that are not so sparse in the basis in which they are compressible. Examples of not-so-sparse signals include those generated by many modern military communications systems (e.g., M-ary PSK/CPFSK/QAM) such as MP-CDL that are capable of supporting data rates of up to 250 Mb/s, and commercial 802.11 compliant communications devices with data rates in excess of 54 Mb/s. Current ISR systems operating in such environments are forced to implement multiple channel narrowband receivers that are expensive in terms of size, weight and power, or use high sampling rate analog-to-digital converters (ADCs) with a limited dynamic range.

To extend the application of A2I to SIGINT, ELINT and ISR platforms operating in dense signaling environments, the MIT Lincoln Laboratory/GMR Research and Technology team has developed novel signal processing algorithms and simulated analog IF circuits to enable next-generation ISR sensor networks with high information rate capacity, high dynamic range and low power operation. This report will detail the development, performance and preliminary mechanization plans for three unique program thrusts: Nonlinear Affine processing (NoLAff), Variable Projection and Unfolding (VPU) and dynamical systems. In the following sections, we will demonstrate both

VPU's and NoLAff's unique capability for identifying and reconstructing not-so-sparse signals that are digitized with a sampling rate *nearly an order of magnitude* lower than the Nyquist rate. In addition, we present NoLAff's unique approach to signal recovery using nonlinear analog circuitry and nonlinear digital signal processing, and a novel approach to an analog circuit implementation of dynamical systems for improving the signal identification capability of any A2I algorithm beyond what is currently possible using randomized linear projections or random sampling. The nearly one order of magnitude reduction in sample rate below the Nyquist rate using the A2I techniques outlined in this document holds the potential to enable the deployment of single-channel receivers with high dynamic range that are capable of operating in sensor networks and passing unprocessed information requiring nearly one order of magnitude less bandwidth than conventional systems.

NoLAff and VPU Undersampling NoLAff processing, specifically NoLAff hypothesis testing, is the first technique explored in this report for reduced rate sampling. NoLAff hypothesis testing utilizes an analog encoder consisting of a known nonlinearity into which the received signal is injected along with a strong, possibly *a priori*, known probe signal. One can view the addition of the probe signal as an affine transformation of the input signal. The now-ambiguous output from the sub-Nyquist (say M times undersampled) ADC is decoded to reconstruct the full spectrum of the received signal by employing hypothesis testing. The nonlinear artifacts are used to determine which of the M possible ambiguous copies of each input signal is the correct one. Hence, the nonlinear artifacts may be viewed as a form of signal diversity. This technique utilizes a form of nonlinear equalization to remove the nonlinear artifacts as part of the hypothesis testing and the final reconstruction. Section 2 details NonLinear Affine (NoLAff) signal processing, identification and reconstruction performance, computational complexity, and a first cut at an initial hardware implementation.

Like Compressed Sensing (CS), VPU leverages the property that a small number of random linear projections of a sparse signal contain most of the salient information. To reconstruct the original signal, VPU is used to find a sparse signal that matches the random linear projections of the original signal. Simply put, VPU sequentially shifts and coarsely increments a column pointer to locate the basis locations that the signal spans and then orthogonally projects the randomized and downsampled basis onto the received data to reconstruct the original transmitted signal. If the difference between the received signal and reconstructed signal is small, a detection is declared and those basis locations are used when search continues. After all the basis locations spanned by candidate signals have been identified, a fine search is conducted where basis locations are jointly incremented and decremented to more accurately pinpoint the locations in the basis that the signals span. We have found that in dense signaling environments, i.e., environments in which the signals occupy more than 15-20 percent of the downsampled basis in which the signal has a sparse representation, linear programming techniques are unable to effectively reconstruct the original signal. However, unlike linear programming techniques used in compressed sensing applications [1] for identification and reconstruction, VPU is capable of reconstructing signals that occupy up to

nearly 100 percent of the downsampled Nyquist band. We will present the VPU algorithm in Section 3 and compare its identification and reconstruction performance to the performance of linear programming used in compressed sensing. In Section 5 we will derive both the maximum likelihood and linear programming bound which we will use to support the empirical results in Section 3.

Dynamical Systems In applying the algorithmic techniques of compressed sensing to RF sensors, we utilize analog signal processing that incorporates randomization and memory. The processed analog data are sampled at a rate that is typically well below Nyquist, but full-band reconstruction of signals is possible given sparse signal environments.

Dynamical systems provide a natural model for analog signal processing with the requisite memory and randomness. A typical dynamical system is characterized by an ordinary differential equation (ODE) that models system behavior. Sampled outputs of the dynamical system can provide initial conditions for the ODE that allow the estimation of intersample values, supporting the concept of sampling below Nyquist rates.

When the dynamical system is driven by an unknown input signal, sampled values can be used to form an innovation process that is modeled by the ODE of the dynamical system. Given sparseness of the input in some known basis, we can use the sampled values to reconstruct the input signal. The precise manner in which we can achieve this reconstruction is discussed below.

The dynamical systems of practical interest are capable of handling wideband microwave signals. An important aspect of such a dynamical system can be signal delay in feedback loops. We consider a specific dynamical system with feedback delay and form an ODE model of what is inherently modeled by a delay differential equation (DDE). This ODE model is crucial since discrete sampling does not make sense in the context of DDE.

Most of the work presented here establishes an infrastructure for system modeling and signal recovery. Anecdotal results are shown for the successful recovery of undersampled wideband pulsed waveforms. Parameters of the dynamical systems have not been optimized in any way.

Performance Bounds: Compressed Sensing Much of the compressed sensing literature is focused on practical, iterative techniques that avoid the combinatorial searches inherent in straightforward approaches to sparse signal recovery. The typical formulations attribute a single response vector to each signal. Signal recovery amounts to finding these vectors and the corresponding signal amplitudes that explain the observed data. The sparsity required for perfect recovery in noiseless environments or stable recovery in noise is typically bounded asymptotically as the number of response vectors increases. In this limit, each signal is still a single response vector. This type of limiting performance does not take advantage of *a priori* information about signals that can potentially improve signal recovery.

Consider, for example, a fixed collection of finite bandwidth signals with unknown center frequencies. If the receiver bandwidth is also fixed, a limiting regime can involve increasing observation times, resulting in an increase in the effective number of samples for each signal. Abstractly, the number of response vectors associated with each signal grows in a structured manner. This structure and the additional sample support can be used to improve recovery. Note that recovery remains *prima facie* a combinatorial problem.

In Section 5.5 we introduce a more general formulation of basis pursuit that can handle the structured signals described above. For intuition, one should keep in mind the example involving finite bandwidth signals with unknown center frequencies and sparse total occupancy. We present lower bounds on the fraction of the band that can be perfectly reconstructed in the absence of noise. We will show that this bound is substantially larger than one commonly used for basis pursuit. We also propose a mixed norm functional to handle stable recovery in noise. An iterative algorithm to minimize this functional is also proposed.

Performance Bounds: Covariance Estimation We use a statistical signal model in order to assess the penalty associated with undersampling wideband data. Specifically, we assume that signals can be described by unknown covariances in a known signal basis. Assuming that the covariance is in some sense sparse, we can evaluate the loss in performance due to undersampling. The calculations are based on standard linear randomization and undersampling models.

Given sufficiently sparse covariances, a Cramer-Rao bound shows that the loss in performance associated with undersampling grows as the square of the undersampling factor. The Cramer-Rao bound does not address all loss mechanisms since it characterizes local (near the true value) behavior of unbiased estimators. Good estimators typically approach the bound at a sufficiently high signal-to-noise ratio (SNR), demarcating the boundary between global and local behavior of the estimator.

The results shown rely only on Cramer-Rao bounds for estimates of covariance parameters. Sparsity is used to evaluate the bounds but is not utilized by the receiver to improve performance. In some sense, sparsity is a form of *a priori* knowledge and thus could be incorporated into random parameter Cramer-Rao bounds to improve performance. For example, the mixed L_2, L_1 norm objective functions that are often used for signal recovery can be interpreted in terms of Laplacian priors [3] on the signal amplitudes. Results of this nature are not presented in this report.

2. NONLINEAR AND AFFINE SIGNAL PROCESSING

Nonlinearities in many receiver and signal processing applications are viewed as either a nuisance to be avoided or overcome. By engineering systems to have almost perfect linear characteristics, one can reduce nonlinear effects albeit at a price that usually involves compromise in performance otherwise (e.g., backing off from saturation in amplifiers and consequently paying a price in linear efficiency or SNR). Alternatively, one can live with some nonlinearities in the system and consequently reduce the nonlinear signal artifacts (e.g., intermodulation products and harmonics) after the fact (e.g., using digital signal processing equalization). In this case the price paid is one of computational complexity and the power consumption associated with it.

In this section we consider the possibility of exploiting nonlinearities and nonlinear signal artifacts to the benefit of the overall system. Rather than compromising performance or actively getting rid of distortions we use nonlinear signal artifacts as a source of additional information on the signals of interest. In particular we introduce the concept of NonLinear and Affine (NoLAff) [4] signal processing, which aims to exploit nonlinear interactions between various signals in the space of interest. Here we have a strong signal, possibly our own probe signal, which dominates other signals in the space of interest. The weaker signals may be viewed as small perturbations of signals existing on a differentiable manifold described by the nonlinear system. Here the strong probe controls the neighborhood on that manifold where the signals exist.

This form of nonlinear function - NoLAff - is nearly linear in the small signal with all the advantages that creates while having some of the benefits of a nonlinear system (e.g., signal diversity). The choice of probe signal gives the system design degrees of freedom, which may be exploited and easily controlled.

Figure 1 depicts a notional use of NoLAff where a weak signal of interest (shown here in blue) enters the system along with a strong signal (shown in red) which may exist in the environment or which may be a probe signal injected on purpose. The output is such that aside from the expected linear signal components we have higher dimensional nonlinear artifacts which contain copies of the original signals as well as cross products (shown in purple). These provide a form of signal diversity. While the strong signal is easy to detect and process, the weaker one now has multiple copies, which processed together may be used to increase its apparent dynamic range.

In the context of analog-to-information we use NoLAff processing as an approach to reduced rate sampling of signals. Here we consider the case where the signal of interest is known to be sparse in the sense that its rate of innovations is significantly lower than the Nyquist or Shannon sampling rate. In particular we assume that there is a linear decomposition of the signal of interest such that very few of the basis vectors used to represent it are nonzero. A somewhat deeper discussion of sparseness is given elsewhere in this report; hence, we concentrate here on a particular example of signals which have a sparse but *a priori* unknown representation in the frequency domain.

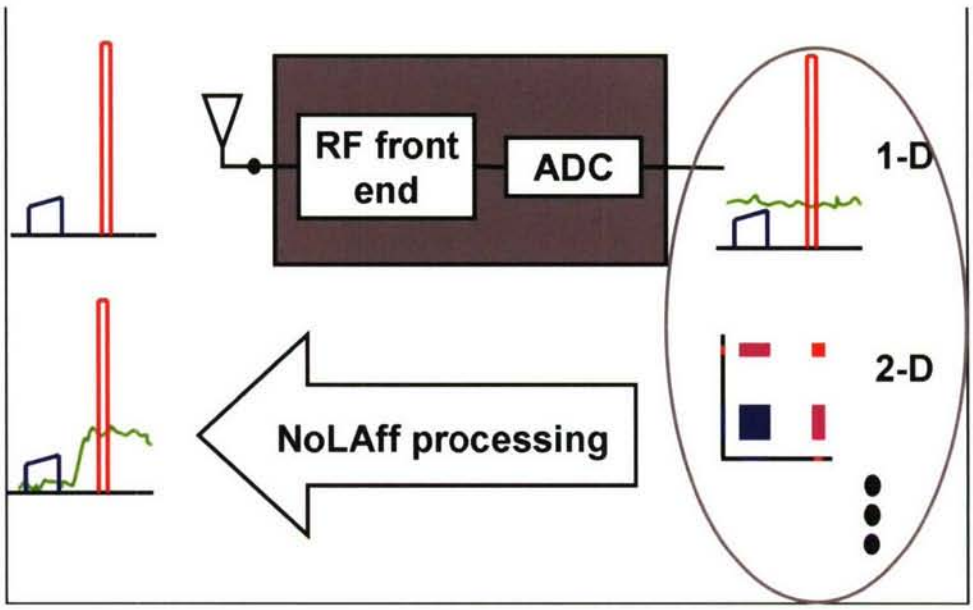


Figure 1: Signal (blue) and strong probe signal (red) pass through receiver with nonlinearities. Output contains linear components and nonlinear components (here only second order shown). This additional information on the signal from higher order components (purple) allows us to extract more information on the signal of interest than linear processing would.

Let's assume here that the sparse input signal $x(f)$ (where f represents the frequency variable) is such that $x(f) = 0$ for all f not in the band $[0, B]$ with the obvious extensions to any other passband. While the frequency support of x is sparse during any time interval of interest, we do not know the exact support. Hence classically we either have to sample at a rate of at least $2B$ or risk having unrecoverable aliasing distortions; or bandpass filter the input signal to a smaller support band and sample at the commensurate rate and potentially lose all information outside the band of the filter. We claim that NoLAff processing (both analog and digital) will allow us to sample at a rate significantly lower than $2B$ without losing any information. While sampling at $2B$ seems to be a good solution, we remember that there are practical limitations to sampling. For some bandwidths, there simply are no COTS ADCs available and even when there are ADCs, we recall that high-rate ADCs have significantly lower dynamic range than the slower counterparts and they consume more power and cost more typically.

A Nyquist rate ADC sampling receiver is depicted notionally in Figure 2. Here we notice the limited SNR of the signal after the ADC. While no information was lost in the sense of filtering or aliasing, some signals may be lower than the noise and hence unrecoverable. On the other hand, consider the NoLAff ADC sampling scheme depicted in Figure 3 where a low rate ADC is used, thereby creating aliasing distortion but having significantly higher dynamic range. Here NoLAff processing is used to remove the aliasing distortion while retaining the dynamic range of the low sample rate ADC. In this section we describe one actual method of using NoLAff processing to do just that.

With the remainder of Section 2, we present NoLAff signal processing in the context of A2I. In Section 2.1, we introduce sub-Nyquist sampling and bounds on unambiguous signal reconstruction. One method for undersampling and reconstructing a sparse signal, NoLAff hypothesis testing, is detailed in Section 2.2, and a discussion of its computational complexity follows in Section 2.3. Section 2.4 provides a brief explanation of how NoLAff undersampling relates to another A2I technique, Compressed Sensing. In Section 2.5, we present a simulated hardware implementation of NoLAff Hypothesis Testing, and conclude by detailing the simulations performed and discussing their results in Section 2.6.

2.1 NOLAFF SAMPLING FOR "SPARSE" SIGNALS

In the context of sub-Nyquist sampling of signals with a limited frequency support or more generally a limited representation in some basis D, we wish to explore the bounds on unambiguous signal reconstruction or extrapolation.

Let $\mathbf{x}$ be a signal of length m. We assume that all signals of interest (with the exception of noise) belong to a fixed subspace. Denote by $\mathbf{D}$ an $m \times n_D$ matrix whose columns form a basis spanning the fixed subspace. If only $n_\mathbf{T} < m$ components of the measured data vector $\mathbf{y} = \mathbf{Dx}$ are observed, we wish to identify $\mathbf{x}$ unambiguously from $\mathbf{y}$. Let $P_\mathbf{T}$ denote the projector onto the

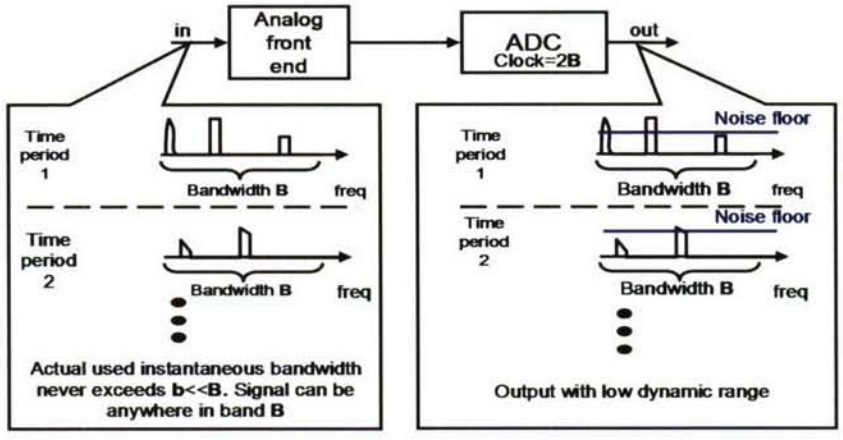


Figure 2: Standard Nyquist rate sampling. Notice the limited SNR of the signal after the ADC.

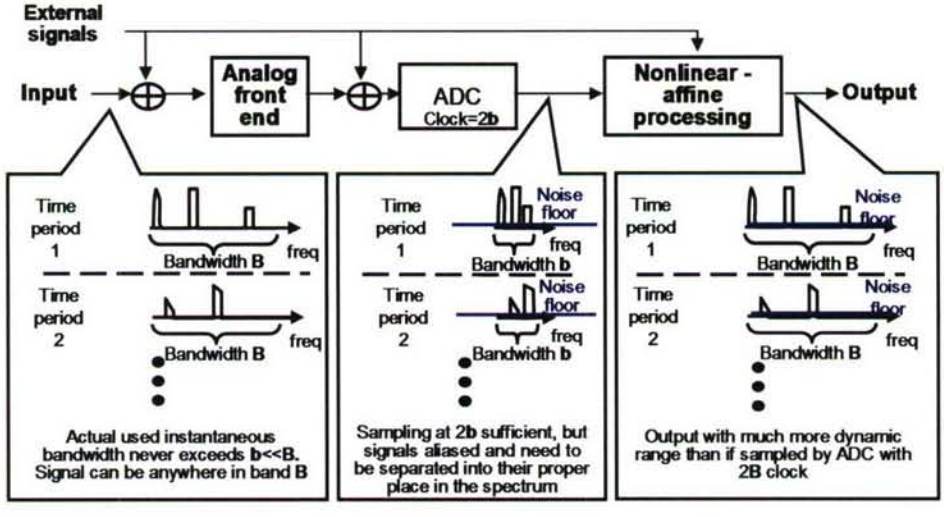


Figure 3: Sub-Nyquist NoLAff sampling. A low-rate ADC is used, thereby creating aliasing distortion but having significantly higher dynamic range. Here NoLAff processing is used to remove the aliasing distortion while retaining the dynamic range of the low sample rate ADC.

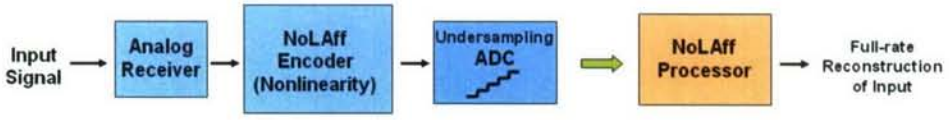


Figure 4: Diagram of the operation of a general NoLAff undersampling system.

subspace spanned by the $n_{\mathbf{T}}$ observed components of $\mathbf{y}$. This subspace is spanned by the columns of an $n \times n_T$ matrix $\mathbf{T}$. If another nonzero signal $\tilde{\mathbf{x}}$ satisfies $P_{\mathbf{T}}\mathbf{y} = P_{\mathbf{T}}\mathbf{D}\tilde{\mathbf{x}}$, then $\mathbf{D}(\mathbf{x} - \tilde{\mathbf{x}})$ is a nonzero element of the ortho-complement of $P_{\mathbf{T}}$.

Let $\mathcal{T}$ and $\mathcal{D}$ denote the subspaces spanned by the columns of $\mathbf{T}$ and $\mathbf{D}$ respectively. Then $\mathcal{T}^c \cap \mathcal{D} \neq \phi$ where $\cdot^c$represents the ortho-complement operator. Now

$$\dim \mathcal{T}^c \cap \mathcal{D} \geq \max(\dim \mathcal{T}^c + \dim \mathcal{D} - \dim n, 0)) = \max(n_{\mathbf{D}} - n_{\mathbf{T}}, 0) \tag{1}$$

with equality typically. That is, we would expect $\mathcal{T}^c \cap \mathcal{D} = \phi$ whenever $n_{\mathbf{T}} \geq n_{\mathbf{D}}$. Hence, when the sample support is greater or equal to the received signal's subspace dimension, we would expect unambiguous reconstruction typically. Specifically for bandlimited signals, this would achieve the Nyquist rate. The results in Section 2.6 indeed show that signals that occupy a subspace with dimensions commensurate with the sampling subspace dimension can be reconstructed uniquely.

2.2 UNDERSAMPLING USING NOLAFF

Traditionally, sampling requires a rate of at least twice the frequency support of the signal. However, in many circumstances, the desired signal is sparsely located within a large bandwidth with some *a priori* unknown support. Despite occupying only a small portion of the full bandwidth, traditional sampling requires such a signal to be sampled at the full Nyquist frequency. NoLAff undersampling provides a method for reducing the required sampling rate, closer to the information rate for a sparse signal.

The NoLAff undersampling system described here consists of two major parts, shown in Figure 4. First is an analog system front-end, which encodes signal information through a nonlinearity and samples the resulting signal below the traditional Nyquist rate. The second part processes the undersampled signal via hypothesis testing, decoding the signal information and placing the received signal into the proper location within the full bandwidth. Both parts are described below.

NoLAff Encoding Undersampling a signal results in aliasing. If a sparse signal is known to have been undersampled, it will contain an image of the original signal which was folded into

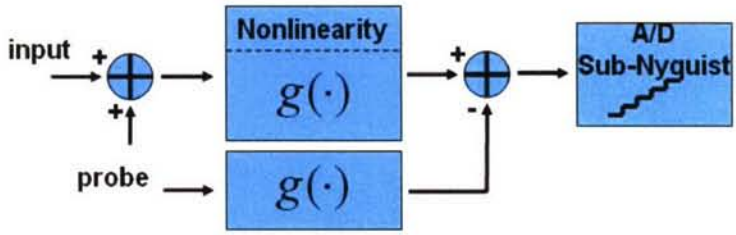


Figure 5: System diagram for a general NoLAff analog encoder front-end.

the smaller undersampled spectrum. In order to reconstruct the original signal at its full rate, the original frequency location of the aliased image must be determined. Without extra information from the signal and the system it passed through, it is impossible to do this. However, by spreading signal information into more frequency locations and filtering the results, NoLAff allows an undersampled sparse signal to be reconstructed.

Signal spreading creates extra copies of the original signal at deterministic frequency locations in the full spectrum. This alone is not enough, as these extra copies also have ambiguous frequency information after undersampling. These extra copies are not ambiguous after they have been filtered, as the filtering process imparts an alteration to the images which is unique based on frequency location in the full spectrum. These adjustments are deterministic, and are later used by hypothesis testing to choose the most likely origin of the signal.

The NoLAff encoder is an analog system which spreads a sparse input signal and filters the result to provide information for reassembling the signal after undersampling. Figure 5 shows a general NoLAff encoding system. This achieves both essential aspects for reconstruction of an undersampled signal using NoLAff encoding which are explained below.

NoLAff Information Spreading A general NoLAff encoder is shown in Figure 5. This function is nonlinear and can be represented by

$$\mathbf{y} = \mathrm{f}(\mathbf{x}). \tag{2}$$

In general, this function adds a strong probe signal to the input, passes this result through a polynomial filter, then removes the probe and the harmonic distortions resulting from the probe, leaving the input signal, intermodulation products between the probe and input, as well as higher-order terms. All information regarding the probe characteristics is assumed to be known *a priori*. The NoLAff nonlinear function can be expressed as

$$\mathrm{f}(\mathbf{x}) = \mathbf{g}(\mathbf{x} + \mathbf{p}) - \mathbf{g}(\mathbf{p}), \tag{3}$$

where $\mathbf{p}$ represents the probe signal and the function $\mathbf{g}(\cdot)$ implements a general nonlinearity represented by a polynomial filter

$$\mathbf{g}(\cdot) = \sum_{i=1}^{\infty} a_i (\cdot)^{.i}. \tag{4}$$

Without loss of generality, we assume the nonlinearity to be memoryless except for the filter. The operations for element-wise multiplication and exponentiation are denoted by ".*" and "$()^{.x}$", respectively. In this derivation, we require

$$\|\mathbf{p}\| \gg \|\mathbf{x}\|,$$

such that

$$\begin{aligned} \mathrm{f}(\mathbf{x}) &= \mathbf{g}(\mathbf{x}+\mathbf{p}) - \mathbf{g}(\mathbf{p}) \\ &\approx \sum_{k=1}^{\infty} a_k \mathbf{x}. * \mathbf{p}^{.(k-1)} \\ &= \mathbf{x}. * \sum_{k=1}^{\infty} a_k \mathbf{p}^{.(k-1)}, \end{aligned} \tag{5}$$

Therefore, NoLAff is linear with respect to the input $\mathbf{x}$. This result may be interpreted as a modulation of the input signal by $k-1$ copies of the probe, where k is the order of the polynomial term. When, for example, the probe is a single tone, each modulation produces a copy of the input signal at a different frequency location, and the addition of several polynomial orders produces many copies of the original signal spread throughout the frequency spectrum. As a deterministic operation, these modulations can be repeated exactly during NoLAff decoding.

An illustrative example is the third-order system

$$y \approx \mathbf{x} + \alpha_3 \mathbf{x}. * \mathbf{p}^{.2}. \tag{6}$$

For simplicity, we let $\mathbf{x}$ and $\mathbf{p}$ be complex exponential vectors, with $\mathbf{x} = c_1 \begin{bmatrix} e^{-j\omega_1} \\ e^{-j2\omega_1} \\ \vdots \end{bmatrix} + \text{c.c.}$ and $\mathbf{p} = c_p \begin{bmatrix} e^{-j\omega_p} \\ e^{-j2\omega_p} \\ \vdots \end{bmatrix} + \text{c.c.}$ where $|c_p| \gg |c_1|$. From equation 6, we have

$$\begin{aligned} y &= \left(1 + 2\alpha_3 c_p^2\right) c_1 \begin{bmatrix} e^{-j\omega_1} \\ e^{-j2\omega_1} \\ \vdots \end{bmatrix} + \alpha_3 c_1 c_p^2 \begin{bmatrix} e^{-j(2\omega_p+\omega_1)} \\ e^{-j2(2\omega_p+\omega_1)} \\ \vdots \end{bmatrix} \\ &\quad + \alpha_3 c_1 c_p^2 \begin{bmatrix} e^{-j(2\omega_p-\omega_1)} \\ e^{-j2(2\omega_p-\omega_1)} \\ \vdots \end{bmatrix} + \text{c.c.} \end{aligned} \tag{7}$$

Clearly, the output frequency support is greater than the input frequency support.

NoLAff Signal Filtering Spreading of the signal, by itself, prior to undersampling is not enough to allow for the original signal to be reconstructed. To differentiate between the different aliased images, or the frequency zones from where they came, the nonlinear signal is filtered prior to undersampling. This filtering encodes the intermodulation products by altering amplitude and phase values of the distortions based on their frequency location in the full spectrum.

Two examples of filters used for encoding are given in Figure 6. Neither filter is ideal, but both work well and provide an intuitive feel for how the filtering encodes frequency information. Filter A is a digital high-pass filter defined by $h = \begin{bmatrix} 1 & -\frac{1}{2} \end{bmatrix}$; filter B is an all-pass delay filter given by $h = \begin{bmatrix} 0 & 1 \end{bmatrix}$. Filter A encodes a signal by causing a unique combination of amplitude and phase changes at each frequency location within the bandwidth. Encoding based on phase alone is demonstrated by Filter B. The all-pass nature of Filter B gives a constant magnitude response across the spectrum. This is advantageous in low SNR cases.

The all-pass filter also provides a convenient demonstration of the frequency encoding that takes place with NoLAff. The phase response for Filter B is re-examined in Figure 7, which shows the differences that occur in potentially ambiguous undersampled tones. The diagram assumes the signals were undersampled by a factor of 4, leading to four tones from the full spectrum that would alias to the same location in the low-rate spectrum. These are shown in green. Due to the phase delay imposed by the all-pass filter, the four tones from the full spectrum each receive a unique phase change. NoLAff decoding exploits these differences, which are applied to the nonlinear distortions,

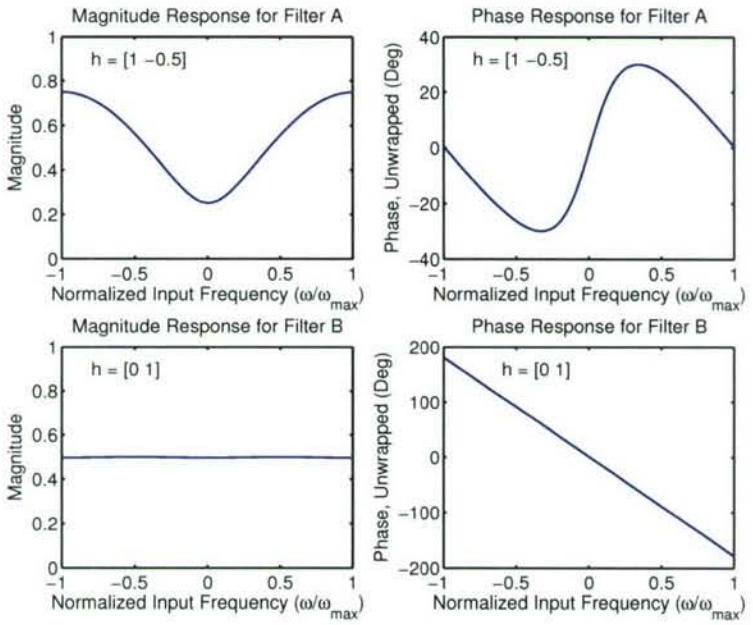


Figure 6: Magnitude and phase responses for two example filters. Filter A provides encoding of frequency location based on both amplitude (upper left) and phase (upper right) changes. Filter B, an all-pass filter, has a constant amplitude gain (lower left), but phase delay (lower right) provides the frequency encoding.

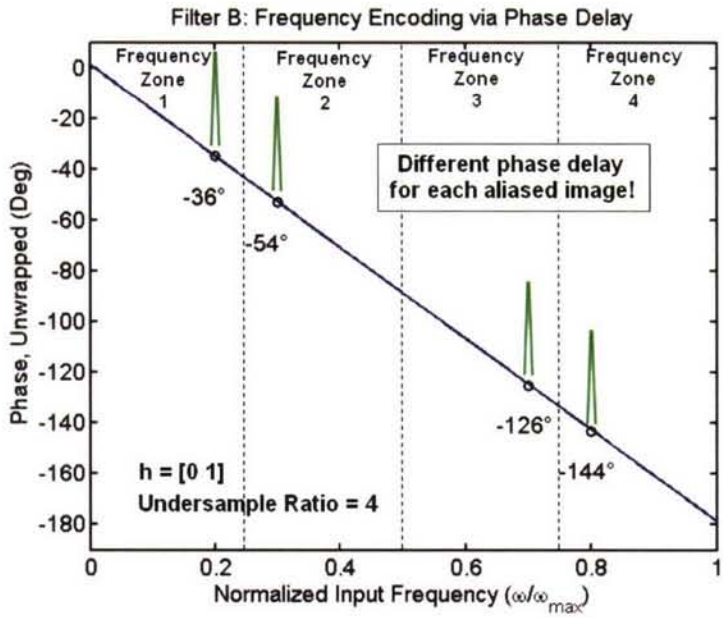


Figure 7: An all-pass filter can be used to encode the frequency location of aliased images. This example has an undersample ratio of 4, leading to four frequency zones from where the signal (green) could originate. The undersampled signal in this case would measure a center frequency of 0.2, normalized to the full-rate bandwidth. However, the measured phase delay would be different for each image, $-36°$,$-54°$,$-126°$, or $-144°$, corresponding to the full-rate signals centered at the normalized frequencies 0.2, 0.3, 0.7 or 0.8, respectively.

to decide which of the signal options from the full spectrum most closely resembles the measured, undersampled signal.

Hypothesis Testing After a sparse signal is encoded, it is sampled at a sub-Nyquist rate. The full encoding and undersampling process is depicted in Figure 8. The result is a measured signal with aliased images and encoded nonlinearities. The NoLAff processor reconstructs the original sparse input signal via hypothesis testing, depicted in Figure 9. Aliasing is viewed as splitting the full-rate bandwidth into equal-length frequency zones, which fold on top of each other during the undersampling process. Each hypothesis is passed through a copy of the nonlinear encoding system, and a Bayesian decision yields the ML unfolded signal.

Hypothesis testing requires several steps to arrive at the most likely input signal. Each hypothesis signal is generated from the measured signal parameters which are transferred into the appropriate full-rate frequency zone. Then, each hypothesis signal is individually passed though the encoding system: the probe signal is added, it is passed through the cubic function and filtered, and the probe is removed. Each hypothesis is then undersampled, resulting in the hypothesis output

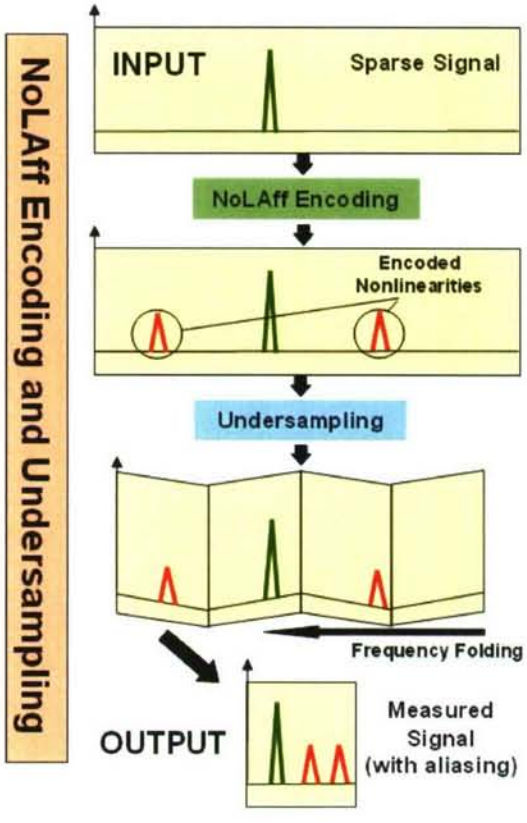


Figure 8: NoLAff undersampling spreads and encodes a sparse input signal. Although undersampling creates aliased images which fold on top of each other, the encoding process allows the original signal to be reconstructed.

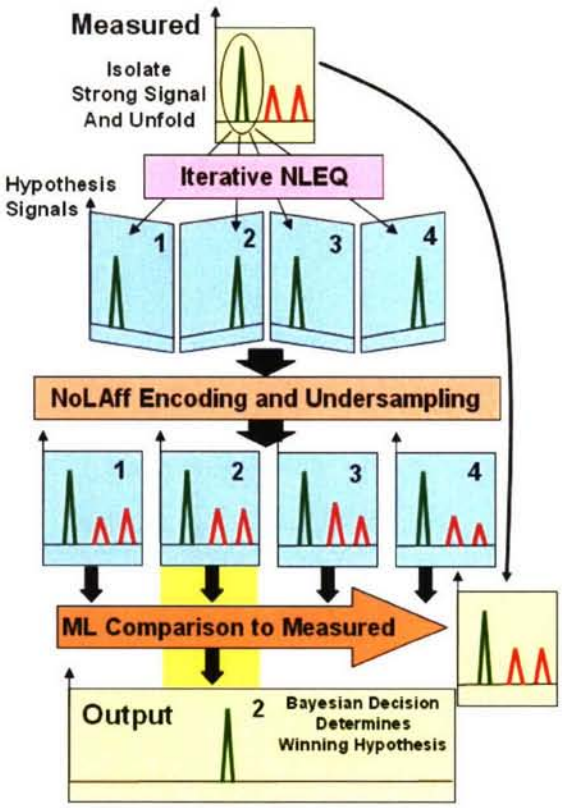


Figure 9: The decoding process for hypothesis testing requires several steps. First, the strongest portion of the measured signal is isolated and equalized by the iterative NLEQ process. This creates a set of hypothesis input signals (blue), each of which is processed through the NoLAff encoder, with the results compared against the measured signal. The closest matching hypothesis is taken to be the output and the most likely choice for the input prior to encoding.

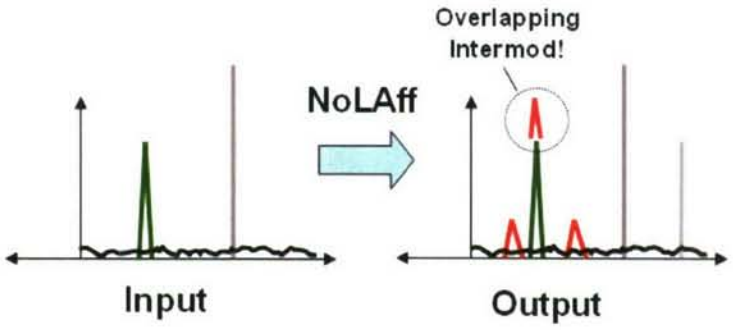


Figure 10: The NoLAff process generates a strong intermod which falls on top of the input signal. Iterative NLEQ is used to find the input signal amplitude.

signals. A winning hypothesis is chosen by the ML criterion. The ML is found by calculating the correlation metric [5]

$$C(\mathbf{y}, \mathbf{z}_m) = 2\mathbf{y} \cdot \mathbf{z}_m - \|\mathbf{z}_m\|^2 ,$$

where $\mathbf{y}$ is the measured signal and $\mathbf{z}_m$ is the mth hypothesis output signal, and then finding

$$k = \arg\max_m \left(C\left(\mathbf{y}, \mathbf{z}_m\right) \right),$$

giving the kth hypothesis as the most likely choice of the input signal.

Iterative NLEQ In the process of generating the NoLAff hypothesis signals, it is necessary to perform a nonlinear equalization (NLEQ) of the signal. The strongest frequency bins in the measured signal spectrum are translated into each frequency zone in the full spectrum as the first step in creating the hypotheses. However, the output of the NoLAff encoder has passed through the nonlinearity, while the hypothesis testing system requires an input signal prior to passing through the nonlinearity. This is important because any odd power nonlinearity will produce a strong intermodulation term located at the same frequencies as the input signal. Since this distortion term falls on top of the information needed to generate the hypothesis signals, an iterative NLEQ process is used to approximate the needed information.

An example of this can be seen with the cubic nonlinearity and linear passthrough from equation (7). It is clear that the cubic nonlinearity generates a tone at the same frequency ω_1 as the linear passthrough tone, and the result is a tone constructed from the combination of the two. This NoLAff operation is depicted in Figure 10, where the green signal represents the linear passthrough and the red signals are the nonlinear distortions. NoLAff needs an accurate measure of the green signal, and iterative NLEQ is used to separate it from the red signal.

```
y = NL(input + probe) - NL(probe)
Y = FFT(y)
loop
    Z_i(n+alias) = Y(n)
    z_i= iFFT(Z_i)
    ẑ_i= NL(z_i+ probe) - NL(probe)
    ŷ= y - (ẑ_i- z_i)
    Y = FFT(ŷ)
until z_i stabilizes
repeat for all i
```

Figure 11: Pseudocode for the iterative NLEQ process. Iterative NLEQ adjusts a hypothesis signal so that it will produce the closest estimate to the received signal after passing through the nonlinearity.

Pseudocode for the iterative NLEQ process is shown in Figure 11. The process begins by measuring the output signal $\mathbf{y}$ and generating the initial set of hypothesis signals $\mathbf{z_i}$. This was done by taking the DFT of $\mathbf{y}$, translating the frequency bins relating to the signal to their respective alias locations in the full spectrum, then performing an inverse FFT. Each hypothesis is then added to a copy of the probe, passed through the nonlinear system. The probe and its harmonics are subtracted off. The hypothesis signal prior to the nonlinear system is subtracted from the hypothesis signal after the nonlinear system, in order to isolate the intermodulation distortions. Those distortions are then subtracted from $\mathbf{y}$, which are used to generate the next iteration of $\mathbf{z_i}$. The process iterates until a stable hypothesis signal $\mathbf{z_i}$ is found. The process is repeated for each hypothesis signal.

We note two points regarding iterative NLEQ. First, this algorithm requires the nonlinear distortions to be weaker than the input signal in order to be stable. This counters the need for strong nonlinearities, which overcomes noise better and thus reduce the chance of error in the hypothesis decision. A good balance was found to be a 10 dB power reduction from the linear passthrough signal to the dominant 3rd-order intermod. Second, with low noise and simple inputs, only one hypothesis should equalize well. The wrong hypotheses will generate new nonlinear distortions while attempting to equalize the ones initially present. Despite this, all hypotheses reach a stable signal when the nonlinearity is of a proper strength.

2.3 COMPUTATIONAL COMPLEXITY

Analyzing the computational complexity of NoLAff decoding is essential for determining the feasibility of implementing hypothesis testing in a real system. In this section, an explanation and a rough estimate of the processing requirements for NoLAff hypothesis testing are described.

The complexity of the full NoLAff hypothesis testing algorithm builds from several subprocesses and is a function of several variables. The NoLAff variables affecting complexity are the undersample rate m, ranging from 1 (full rate) to 8 and beyond, and the bandwidth of the signal, indicated by the number of DFT frequency bins b the signal occupies. The number of frequency bins, b, relates to the signal bandwidth as a percentage of the full-rate bandwidth by $b = \text{ceil}\,(\%BW \times N/2)$. The primary operations in NoLAff are the iterative NLEQ (performed once), complex multiplication for creating hypothesis signals (repeated mb times), low-rate norm and maximum likelihood (ML) calculations (repeated mb times), and the nonlinearity, (repeated m times). Since the complex multiplications, needed for generating the hypothesis signals, require four multiplications per data point, this adds $4mbN$ operations to the complexity. The low-rate norm and ML calculations each require $\frac{N}{m}$ multiplications and are repeated three times for each hypothesis, giving $3N$ operations. Combining this information gives a complexity of

$$4mbN + 3N + \text{iterative NLEQ} + m \times \text{nonlinearity}$$

per sample. The computational complexity of the iterative NLEQ and nonlinearity are discussed below.

The complexity of the iterative NLEQ subprocess is based on one additional factor: the number of iterations I performed to equalize the nonlinearities. The number of iterations typically would range from 5 to 15. The calculation for the complexity of the iterative NLEQ operation requires a low-rate $\left(\frac{N}{m}\right)$ FFT (performed mI times), a complex multiplication for creating intermediary hypothesis signals (repeated mIb times), and the nonlinearity (repeated mI times). Combining this information shows that iterative NLEQ requires approximately $m\left(L + 4b + 2 + \frac{1}{m}\log\frac{N}{m}\right) IN$ calculations to complete.

The nonlinearity function block must be called multiple times during hypothesis testing. In a typical receiver NLEQ system, the nonlinearity calculation could require sixteen or more operations each of filters and multiplications for each sample [6, 7]. However, the nonlinearity in these simulations is assumed to be simple, consisting of only a simple cubic term, requiring two multiplies per data sample, and one filter operation per sample. The length of the filter L becomes an additional variable in the computational complexity. The nonlinearity results in $2N + LN$, or $(L+2)N$, calculations to complete.

The complexity of the full NoLAff hypothesis testing becomes

$$4mbN + 3N + m\left(L + 4b + 2 + \frac{1}{m}\log\frac{N}{m}\right) IN + m(L+2)N,$$

which simplifies to

$$N\left[I\log\frac{N}{m} + m\,(I+1)\,(L+4b+2)\right].$$

To provide a better understanding of how each of the factors affects the computational complexity, a set of typical values is applied, holding all but one factor to a reasonable value. Typical values used were $I = 10$ for NLEQ iterations, $L = 8$ for the filter length, $m = 8$ for the undersample ratio, and $b = 5$ for a sparse, 1% of full bandwidth signal when $N \approx 1000$. The computational complexity of hypothesis testing per input sample, as a function of iterations, becomes $f_1(I) \approx c_1 I$, where c_1 is a constant. Likewise, the complexity is approximately linear with respect to undersample rate, $f_2(m) \approx c_2 m$, and with respect to signal bandwidth, $f_3(b) \approx c_3 b + c_4$.

2.4 CONNECTING NOLAFF TO CS

It is instructive to compare the NoLAff approach to other A-to-I techniques. In particular, we compare the NoLAff encoder to the randomization-type encoders. While the mechanics of decoding are quite different, there are commonalities in the encoding. Much like randomization, nonlinear and affine transformations spread sparse signals over much of the dictionary representation in which they are sparse.

Let

$$\mathbf{x} \;=\; \mathbf{T}\theta, \tag{8}$$

be an input signal where $\mathbf{T}$ is a dictionary whose columns span the signal space of interest and the vector θ describes the linear combination of the columns in the dictionary.

For example, the $\mathbf{n} \times \mathbf{n}$ dictionary $\mathbf{T}$ is built of the columns of the $\mathbf{n} \times \mathbf{n}$ DFT matrix. The θ vector provides the complex amplitude values for each vector used in the signal $\mathbf{x}$.

In Section 2.2, it was shown that the output of the NoLAff function is approximately linearly related to the input when the condition

$$\|\mathbf{p}\| \gg \|\mathbf{x}\|$$

is met, where $\mathbf{p}$ is the probe and $\mathbf{x}$ is the input. This led to the expression,

$$\mathrm{f}(\mathbf{x}) \;\approx\; \mathbf{x}. * \sum_{k=1}^{\infty} a_k \mathbf{p}^{\cdot(k-1)}.$$

This equation can be developed into a left-hand factored form,

$$\mathrm{f}(\mathbf{x}) \;\approx\; \mathbf{N_L}\mathbf{T}\theta, \tag{9}$$

where $\mathbf{N_L} = \sum_{k=1}^{\infty} a_k \cdot \mathrm{diag}(\mathbf{p})^{(k-1)}$. Here, $\mathbf{N_L}$ serves a similar role to that of randomizing matrices used in CS–spreading the sparse signal $\mathbf{x}$. We note that here there is no low-pass filtering and hence we require hypothesis testing. CS techniques typically perform filtering.

The NoLAff function may also be described by a right-side factorization,

$$\mathbf{y} \approx \mathbf{T}\mathbf{N}_\mathbf{R}\theta, \tag{10}$$

where $\hat{\mathbf{T}} = \mathbf{T}\mathbf{N}_\mathbf{R}$ is the probe-dependent new dictionary resulting from applying the NoLAff transformation to the columns of $\mathbf{T}$. Because the output vector $\mathbf{y}$ is assumed to be sampled, then the aliased images all remain within the space spanned by the original dictionary $\mathbf{T}$. Thus, it is possible to describe every column of $\hat{\mathbf{T}}$ by a linear combination of the columns of $\mathbf{T}$. Here, $\mathbf{N}_\mathbf{R}$ represents (approximately) the spreading of information caused by NoLAff, and is thus called the nonlinear spreading matrix.

Either the left-hand or right-hand factorization can replace the random sampling matrix in the LASSO function for compressed sensing, giving $\tilde{J}(\theta) = \|\mathbf{y} - \mathbf{\Phi}\mathbf{N}_\mathbf{L}\mathbf{T}\theta\|_2^2 + \lambda\|\theta\|_1$ and $\tilde{J}(\theta) = \|\mathbf{y} - \mathbf{\Phi}\mathbf{T}\mathbf{N}_\mathbf{R}\theta\|_2^2 + \lambda\|\theta\|_1$ [8].

2.5 SIMULATION OF A NOLAFF IMPLEMENTATION IN AN RF SYSTEM

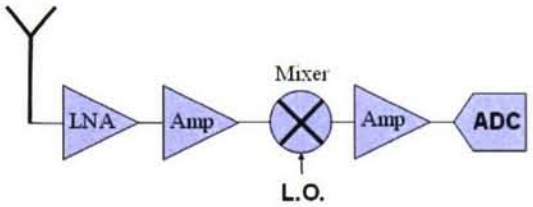
The goal of this section is to provide an initial approach and analysis for integrating NoLAff undersampling into an RF receiver system. A basic system diagram for a typical analog RF receiver, shown in Figure 12(a), can be converted to the NoLAff receiver system depicted in Figure 12(b) with the addition of an analog NoLAff subsystem. The analog NoLAff subsystem consists of a tone generator, a summing block for adding the strong probe tone to the received signal, a nonlinear filter subsystem, and several notch filters.

2.5.1 System components for NoLAff Undersampling

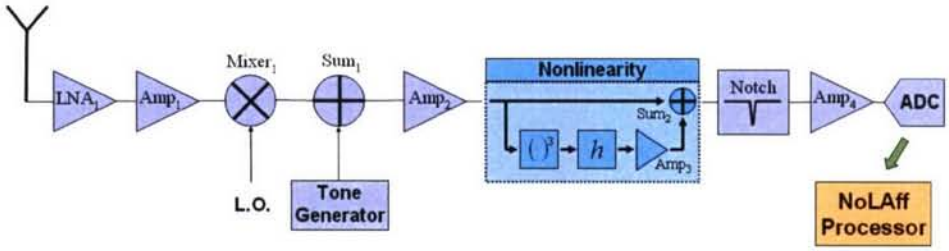
The additional components found in Figure 12(b) are required for NoLAff undersampling. It is the goal of this section to justify the inclusion of the extra components, and to analyze their impact on overall system performance.

Probe Generator and Summing Block The probe generator is required as an essential component in operation of a NoLAff system. To work well, the probe signal must be much stronger than the input signal to which it is added. A probe-to-target ratio of 60 dB was used for this simulation. Noise and distortions from the probe generator will have a comparatively large impact on system performance, and thus a low-noise and low-distortion component should be used. For this simulation, a high-quality source was assumed and noise and distortion from the probe generator were considered negligible.

The probe signal is added to the target signal by a summing block. The summing block can be implemented with an amplifier with gain and noise figure values similar to the other amplifier stages.



(a)



(b)

Figure 12: (a) Basic block diagram for a traditional receiver system. (b) System diagram for the hardware implementation simulation for NoLAff undersampling.

Nonlinear Filter Subsystem The most significant change to the receiver system is the inclusion of a nonlinear filter subsystem. This component is critical to the proper operation of NoLAff undersampling, and is composed of four individual parts: a cubic circuit, an all-pass filter, a gain stage and a summing block.

Cubic Circuit The first component in the nonlinear filter subsystem is the cubic circuit. The function of this circuit is to perform a cube operation on the signal entering it. The cubic circuit operates in the linear region of all of its components. The schematic for this circuit is provided in Figure 14. The circuit begins by converting the signal voltage into current, then performs the cubing operation in current. For the last step, the circuit converts the cubed current into a voltage by passing it through a resistor.

SPICE simulations were performed on the circuit of Figure 14 to verify the cubic nature of its operation. The output frequency response to a a single input tone at 25 MHz and at -16.48 dB input power is shown in Figure 13. This shows the cubic circuit produces the expected 3rd-order and fundamental tones, and shows that a weak 5th-order term is also present. All other harmonic tones were considered negligible for this simulation.

A further insight into the operation of the cubic circuit is the magnitude and phase responses from the SPICE simulation and included in Figure 15. The response curves show consistent behavior to about 200 MHz, beyond which performance degrades. However, with further design improvements

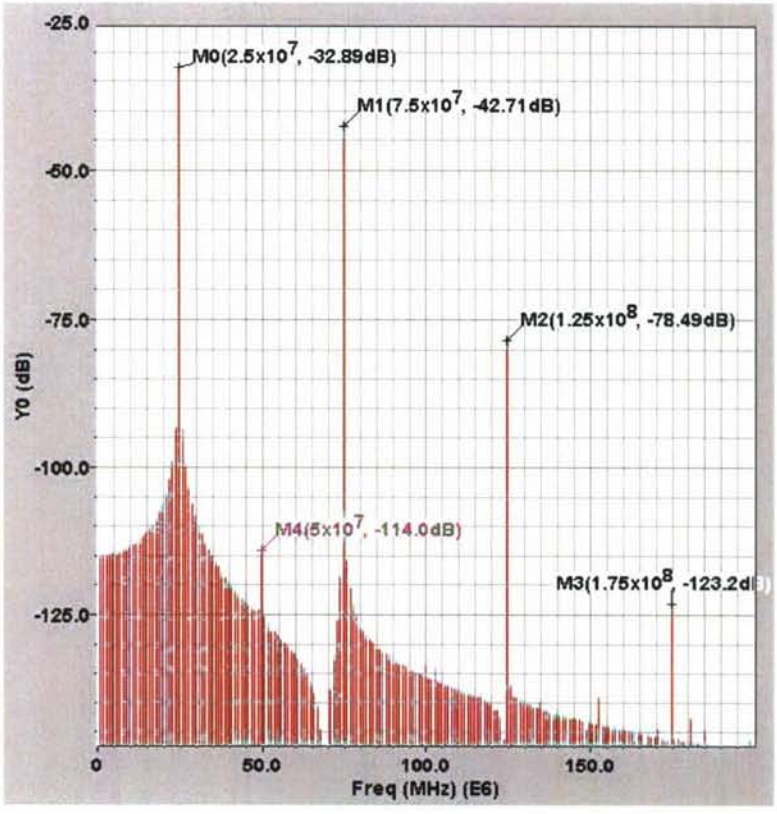


Figure 13: Output frequency response of the cubic circuit for an input tone of 25 MHz and relative input power or -16.48 dB.

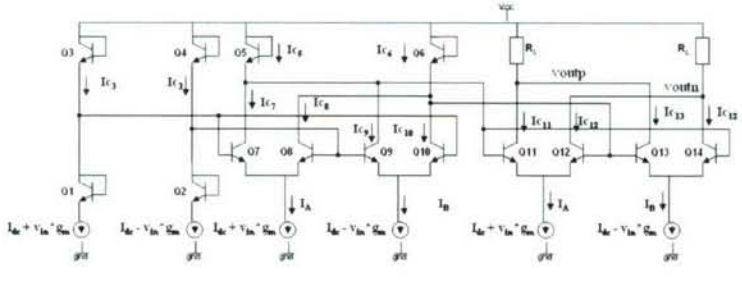


Figure 14: Diagram for the cubic circuit.

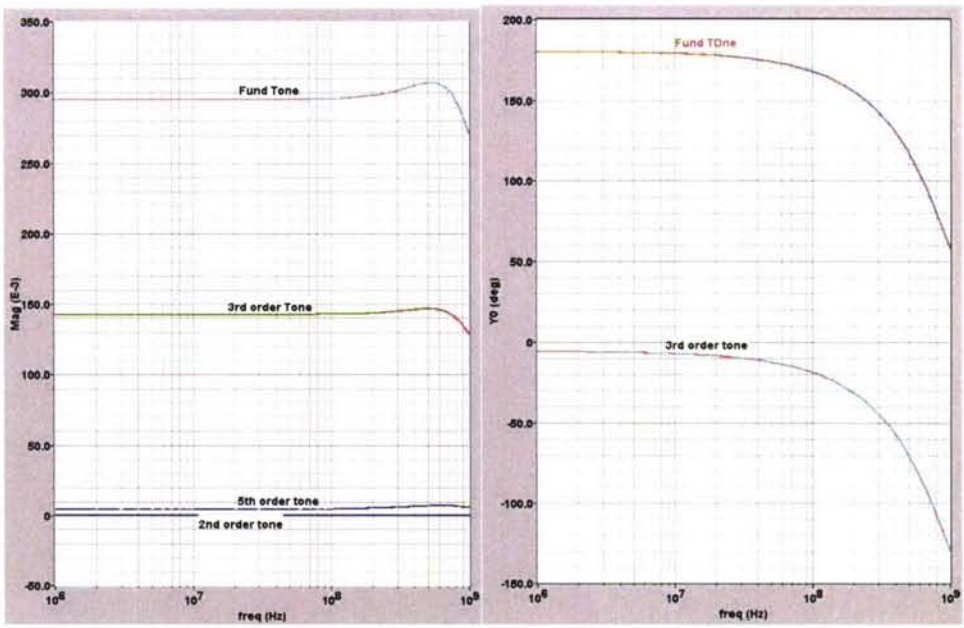


Figure 15: Magnitude and phase responses of the cubic circuit with respect to input tone frequency. The circuit performance begins to alter substantially above ~200 MHz.

to the voltage-to-current and current-to-voltage converters at the input and output to the circuit, the bandwidth of the cubic circuit is expected to extend well beyond 500 MHz. Simulation results demonstrated the cubic circuit would be expected to have a noise figure of about 6 dB, consistent with similar devices.

All-Pass Filter After the cubic circuit, the signal is filtered by an all-pass filter. This circuit provides a linear phase delay to the signal, which frequency-encodes the intermodulation products. A passive, 2nd-order all-pass circuit was used due to its simplicity of implementation and the relatively small amount of noise it adds to the signal. Figure 16 provides a circuit diagram for the passive second-order all-pass circuit. The transfer function for this filter is given by [9]

$$H(s) = \frac{V_o}{V_i} = 1 - \frac{2s(\omega_0/Q)}{s^2 + s(\omega_0/Q) + \omega_0^2}. \tag{11}$$

The initial filter design for NoLAff calls for a filter with a constant gain and a linear phase response. The 2nd-order all-pass provides a constant gain of 0.5 across all frequencies ideally, and up to 1 GHz by simulation. However, the phase response is not linear with frequency. By adjusting the value of Q in the transfer function design, the phase response was adjusted to provide a large bandwidth region with approximately linear operation. A SPICE simulation of this phase response, extending from 50 MHz to 250 MHz, is shown in Figure 17.

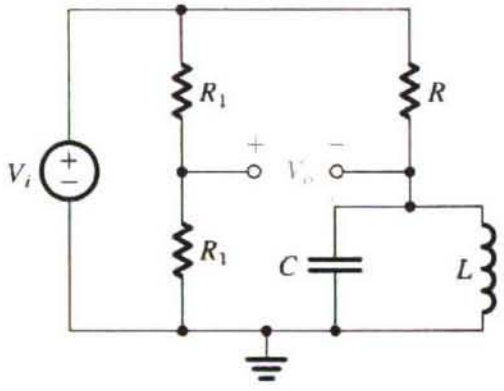


Figure 16: Circuit diagram for a passive 2nd-order all-pass filter [9].

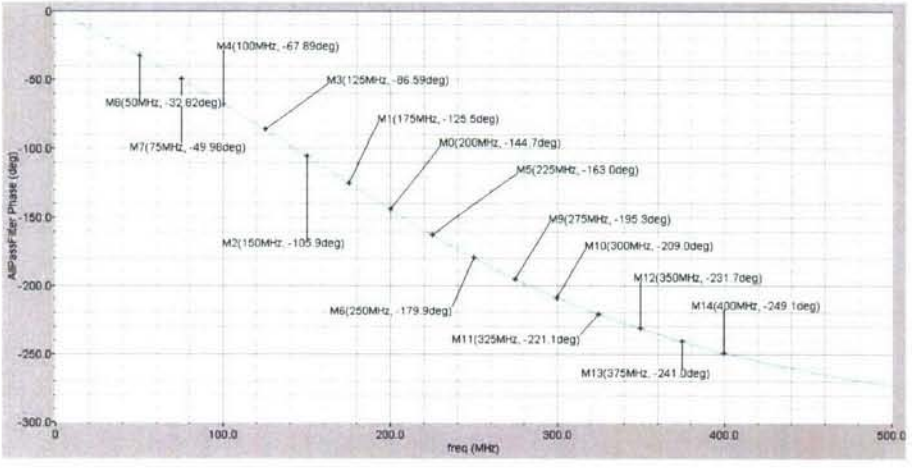


Figure 17: Spice simulation for the phase response of the 2nd-order all-pass circuit. The response is approximately linear with frequency in the region from 50 MHz to 250 MHz.

Because of the high output impedance of the passive filter, a gain stage is included at the output. The amplifier also allows the strength of the nonlinearity to be set to an appropriate level before it is added with the linear pass through. This amplifier is followed by another summing block, which adds the nonlinearity signal to the linear passthrough signal.

Notch Filters After adding the linear and nonlinear signals together, the last step for the analog system is to perform an A/D conversion. However, some filtering should be performed first. The probe tone that was added to the signal was much stronger than the original input signal, and by itself does not add any desired information to the signal. So, in order to improve the signal dynamic range prior to the ADC, the probe tone is removed with a notch filter. The harmonics of the probe are also quite strong, and so the 3rd and 5th probe harmonic are also notched. For this simulation, the notch filters were implemented as ideal, setting the power at the notch frequency to zero while not affecting any other frequency locations. This required careful decisions as to the input signal frequency locations, as the signal and the intermodulation products should not have crossed over the notch-filter frequencies.

Analog-to-Digital Converter The last part of the analog receiver NoLAff system performs an automatic gain function followed by the analog-to-digital conversion. Two ADCs were compared for this study: an ADC sampling at the full signal rate and an ADC sampling at the undersample rate. For the full rate, a MAX108 operating at 1 Gsps was simulated. The MAX108 is an 8-bit ADC, and according to Maxim, it operates with 7.5 ENoBs at 1 Gsps. For the undersample rate ADC, the Analog Devices AD9461, which is a 16-bit converter with 12 ENoBs when running at the sample rate of 125 Msps, was simulated.

2.5.2 Derivation of System Noise and the Impact on SNR

One impact of the additional components in the NoLAff receiver is the effect on the noise. It will be shown that the system nonlinearity will have negligible effect on the overall system noise and that the low-noise amplifier (LNA) will remain the dominant noise source. Using Friis' formula [10], the noise figure up to the second summing block, with the nonlinear branch removed, is

$$F_L = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + \frac{F_5 - 1}{G_1 G_2 G_3 G_4}, \tag{12}$$

while the noise figure through the nonlinearity with the linear passthrough removed, is

$$F_N = F_L + \frac{1}{G_1 G_2 G_3 G_4}\left(\frac{F_6 - 1}{G_5} + \frac{F_7 - 1}{G_5 G_6} + \frac{F_8 - 1}{G_5 G_6 G_7}\right). \tag{13}$$

This leads to a noise figure entering the second summing block of

$$F_{Sum2} = \frac{1}{2}\left(F_L + F_N\right). \tag{14}$$

Component	Gain	Typical Gain (dB)	Noise Figure	Typical Noise Figure (dB)
LNA_1	G_1	12	F_1	3
Amp_1	G_2	12	F_2	6
$Mixer_1$	G_3	2	F_3	8
Sum_1	G_4	2	F_4	6
Amp_2	G_5	30	F_5	8
Cubic	G_6	-16	F_6	-10
h filter	G_7	-6	F_7	-6
Amp_3	G_8	12	F_8	6
Sum_2	G_9	2	F_9	6
Amp_4	G_{10}	20	F_{10}	8

TABLE 1

Gain and Noise Figure values for each component in the simulated receiver system of Figure 12.

Substituting gain and noise figure values from Table 1 into equations (12–14) gives $F_L = 3.49$ dB and $F_N = 3,49$dB, so $F_{Sum2} = 3.49$dB as well. Because the noise figure through the linear passthrough gives the same result as the noise figure through the nonlinearity, the circuitry for the nonlinearity contributes no substantial noise to the overall system. The reason for this is the high gain provided by Amp_2. Removing Amp_2 from the system, for example, would cause F_N to increase to 4.84 dB. However, with it in place, the LNA, with its 3 dB noise figure, is the dominant component in the receiver noise.

2.6 SIMULATIONS AND RESULTS

Three categories of simulations were conducted for NoLAff hypothesis testing. The first tested the algorithm performance, in regards to ambiguity error rate and reconstruction error, over a set of input signals that varied by SNR and signal bandwidth. Tests were also performed varying system parameters, such as the search bandwidth when the actual bandwidth is unknown. The second simulations focused on the performance due to limitations arising from the use of real system components. An analysis of the effect of different sources of noise on hypothesis testing is also considered. The final tests examined performance of recovering a small signal in the presence of a large signal. This large dynamic range case presents specific challenges which other A-to-I techniques, specifically compressed sensing, do not handle well. We present promising results for the use of NoLAff in such situations.

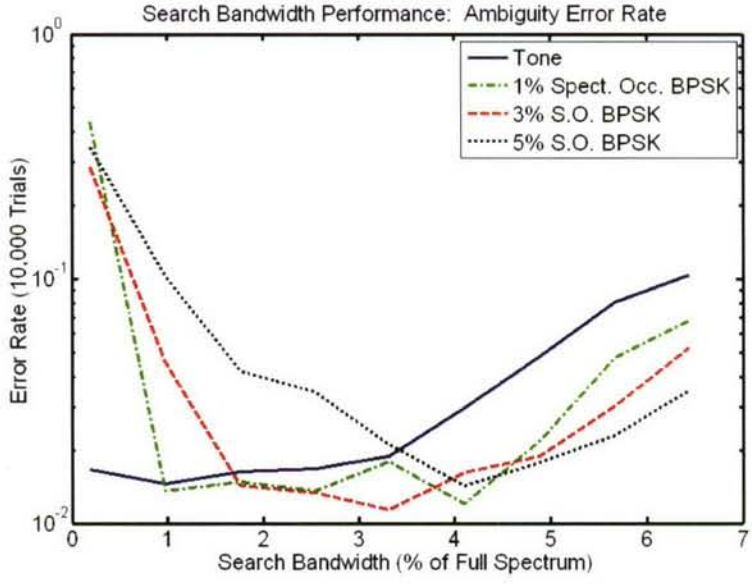


Figure 18: Ambiguity error rate versus hypothesis search bandwidth. Results for a tone signal and 1%, 3% and 5% spectral occupancy binary phase-shift keyed (BPSK) signal are shown. The lowest error rates appear when the search bandwidth is near the signal bandwidth.

For all of the simulations that follow, several system parameters remained consistent throughout. First, datasets began with 1024 samples at full rate; an undersample rate of 8 was used for all simulations, giving 128 samples in the low-rate data. The NoLAff nonlinearity was taken to be cubic—or near-cubic for the hardware simulation—added to a linear passthrough. Response curves of the filters used are found in Figure 6 for the algorithm tests and in Figure 17 for the hardware simulation. Other parameters and conditions are discussed separately for each simulation below.

Search Bandwidth Performance The first set of results presented show the performance of NoLAff hypothesis testing for a variety of input signals. Input signals were varied with input SNR, ranging from 5 to 25 dB, and with bandwidth, ranging from a simple tone to a BPSK signal occupying 9% of the full bandwidth. The decision error rates over 100,000 trials for these signals are shown in Figure 20 as a function of SNR. The reconstruction MSE for this trial set is provided in Figure 21. This shows the average reconstruction error for only the correctly chosen hypotheses from the trial set. NoLAff undersampling is able to recover signals that occupy a large percentage of the undersample bandwidth, as shown in Figures 20 and 21. At the largest bandwidth in the experiment, 9%, or 72 % of the folded spectrum, NoLAff achieved no less than 99.9% accuracy with an input SNR of at least 20 dB. Performance improved as the signal bandwidth was decreased, both in errant decision and in reconstruction error. The reconstruction error decreased as the

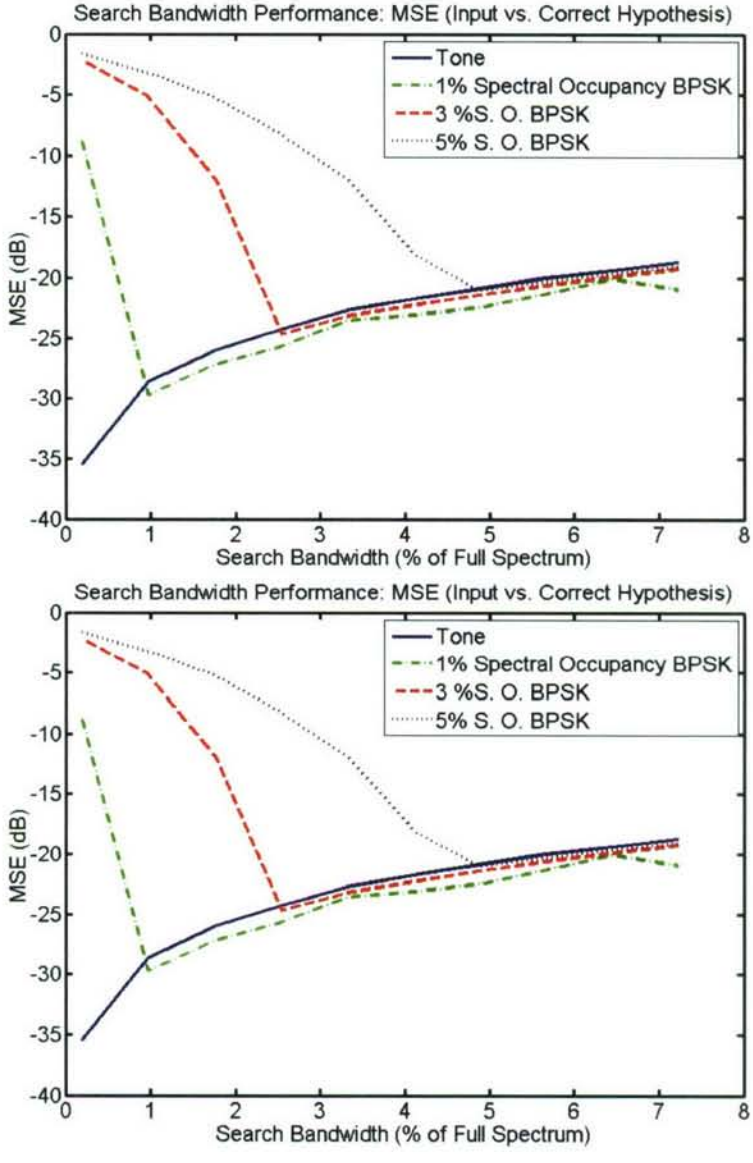


Figure 19: Reconstruction error (MSE) is shown as the hypothesis search bandwidth increases. The error is lowest for each dataset when the search bandwidth is near the signal bandwidth.

signal bandwidth decreased, which is possibly a result of less noise falling within the reconstruction bandwidth used by NoLAff.

Signal Bandwidth Performance Simulation results pertaining to the portion of the undersampled spectrum used for hypothesis testing and reconstruction were also collected. If the exact bandwidth of the undersampled, sparse signal is unknown, then NoLAff hypothesis testing can be applied using a larger or smaller portion of the measured bandwidth. This is called the search bandwidth, and the effects of choosing a bandwidth that was either too large or too small, on overall performance of NoLAff undersampling, were collected. Figure 18 shows the ambiguity decision error rate, and Figure 19 gives the reconstruction error for these simulations. Both figures clearly suggest that NoLAff works best when the search bandwidth covers exactly the band which contains the signal, as the error rate and reconstruction error reach a minimum when the search bandwidth equals the signal bandwidth. The reconstruction error shows a consistent pattern that repeats for all bandwidth trials. The pattern shows large errors when the search bandwidth is smaller than the signal and is likely due to a portion of the signal information being excluded from the search. As the search bandwidth grows beyond the size of the signal, the reconstruction error increases along a very similar path for all four trials. This may be explained by the inclusion of noise outside the signal bandwidth in the reconstructed signal.

Hardware Simulation Performance The hardware simulation of Figure 12 was simulated and compared to the initial NoLAff design and to the full-rate sampling of the signal. Results for the decision error rates and reconstruction errors are provided in Figures 22 and 23, respectively. Figure 22 shows the rate of errors encountered over a test of 100,000 trials for several SNR values. Figure 23 provides average mean-square-error reconstruction error over the 100,000 trials, but choosing only trials where the chosen hypothesis was correct. The 'initial' system has an ideal cubic polynomial filter with $h = \begin{bmatrix} 0 & 1 \end{bmatrix}$. The hardware system is based on the hardware circuit simulations for the cubic nonlinearity and the all-pass filter. Two different input signals were also used: simple tone inputs and a BPSK signal, which occupied no more than 2% of the full spectrum. For all trials, the probe signal was a tone located at a consistent frequency location, and the center frequency of the input signal was chosen at random. Also, care was taken to avoid situations of signal folding or overlapping, and the input signal was chosen in order to avoid the notch frequencies. Data for the reconstruction of a full-rate sampled signal was collected and is provided in Figure 23.

These results show a trend similar to the results found for the algorithm tests, with error rate decreasing rapidly with increasing SNR, and average reconstruction error improving with increasing SNR. However, a surprising result was found when comparing the hardware error rate to the initial design error rate in Figure 22: the hardware simulation outperformed the initial system design, showing a lower error rate at all SNR values. This indicates several points. First, the differences between the hardware and initial design performace are most likely related to the encoding scheme,

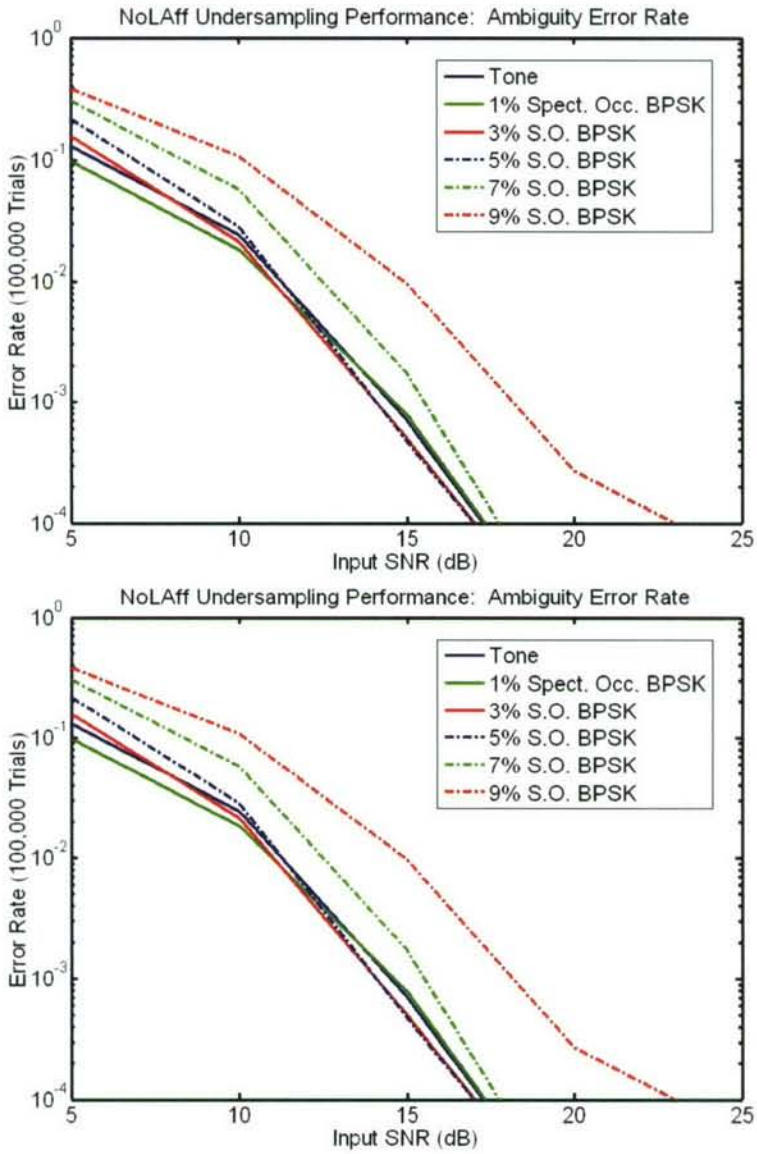


Figure 20: NoLAff undersampling error rate over 100,000 trials for tone signals and 1%, 3%, 5%, 7%, and 9% spectral occupancy BPSK signals.

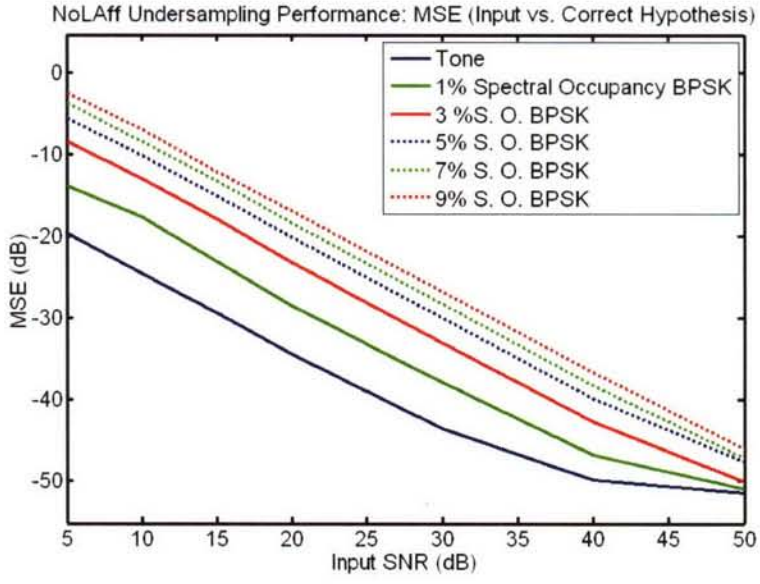


Figure 21: Reconstruction error (MSE) for NoLAff undersampling of tone signals and 1%, 3%, 5%, 7% and 9% spectral occupancy BPSK signals.

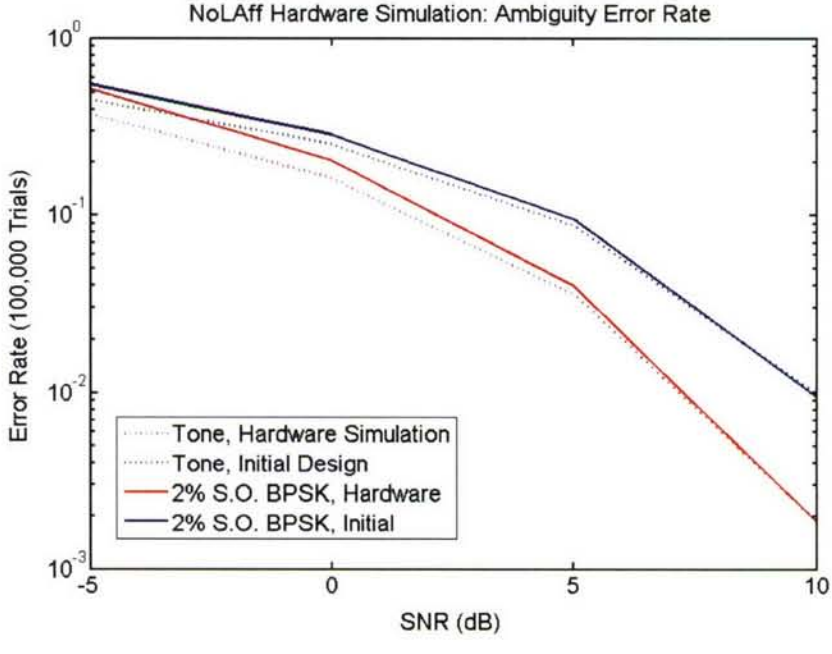


Figure 22: Ambiguity error rates for hardware and initial design simulations. Error rates are shown for both tone signals and 2% spectral occupancy BPSK signals.

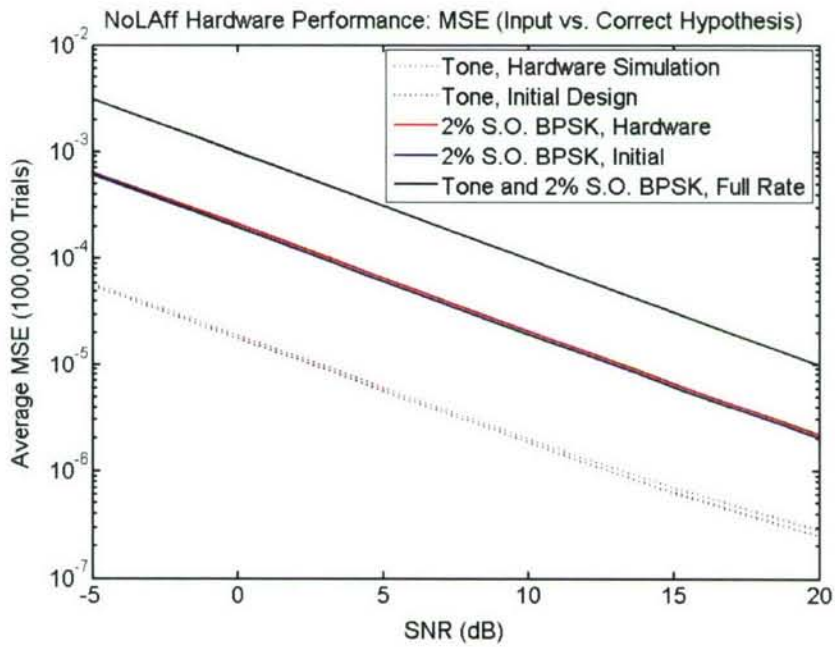


Figure 23: Mean square error for signal reconstruction in hardware, initial design and full-rate simulations. MSE for both tones and a 2% spectral occupancy BPSK signal is shown.

which is controlled by the all-pass filter within the nonlinearity. Second, this suggests that neither the initial design, nor the hardware simulation, implements an "ideal" encoding scheme. The explanation of NoLAff encoding provided in Figure 7, while instructive, is overly simplistic and does not accurately reflect the more complicated interactions between the filtered nonlinear distortions, the linear passthrough signals, and the effects of iterative NLEQ. Exploring the impact of different group delays, nonlinear phase delays and phase offsets is important future work, which should lead to performance gains by reducing overall error rates for both software and hardware system designs.

Noise: LNA vs. ADC Two components in the receiver system are expected to dominate the noise–the Low Noise Amplifier (LNA) and the Analog-to-Digital Converter (ADC). The ADC is expected to dominate when the signals present in the measured data have a large dynamic range requirement; the LNA is expected to dominate otherwise. Any steps that can be taken to lower the dynamic range of the signals help reduce the effect of the ADC, and this is why the probe and its harmonics–the strongest signals present in the system–were removed with analog notch filters prior to sampling. When the ADC is the dominant source of noise, NoLAff can produce stunning improvement in overall SNR by allowing a lower-rate converter to be used, gaining as much as 30 dB in dynamic range. Figure 24 demonstrates that the lower-rate ADC reduces the noise caused by the converter, leading to improvements in SNR. The benefits of NoLAff when the LNA is the dominant component are not as dramatic, but do show some improvement as shown in Figure 23. However, with the gains in SNR offered by the NoLAff approach, some of the regained SNR is required in order to make correct decisions for the signal reconstruction.

Large Dynamic Range Environments Detection and reconstruction of a weak signal from a spectrum crowded with stronger signals is a difficult problem arising in many real-world applications. NoLAff undersampling is explored as a method to facilitate the removal of a strong interfering signal, allowing greater access to a weaker signal. The improved dynamic range from undersampling the signal, depicted in Figure 24, provides improvement in SNR. In CS approaches based on random projections, the residual error of reconstructed large signals is effectively noise for small signal detection and reconstruction. NoLAff applied properly does not have this problem.

Let

$$\mathbf{x} = \mathbf{x_s} + \mathbf{x_w},$$

be the received signal, where $\mathbf{x_s}$ and $\mathbf{x_w}$ are sparse signals, non-overlapping in frequency, and $\|\mathbf{x_s}\| \gg \|\mathbf{x_w}\|$. The input signal is encoded by the NoLAff system as described in Section 2.2. The decoding strategy is to first remove the strong signal $\mathbf{x_s}$, then redefine the probe as $\mathbf{p_{p+s}} = \mathbf{p} + \mathbf{x_s}$ to resolve the weak signal $\mathbf{x_w}$. Results are provided in Figures 25 and 26. Figure 25 shows the ambiguity error rate for choosing the correct hypothesis for the weak signal for a given SNR for the weak signal. The strength of the strong signal varies from 10 to 50 dB above the weak signal, and

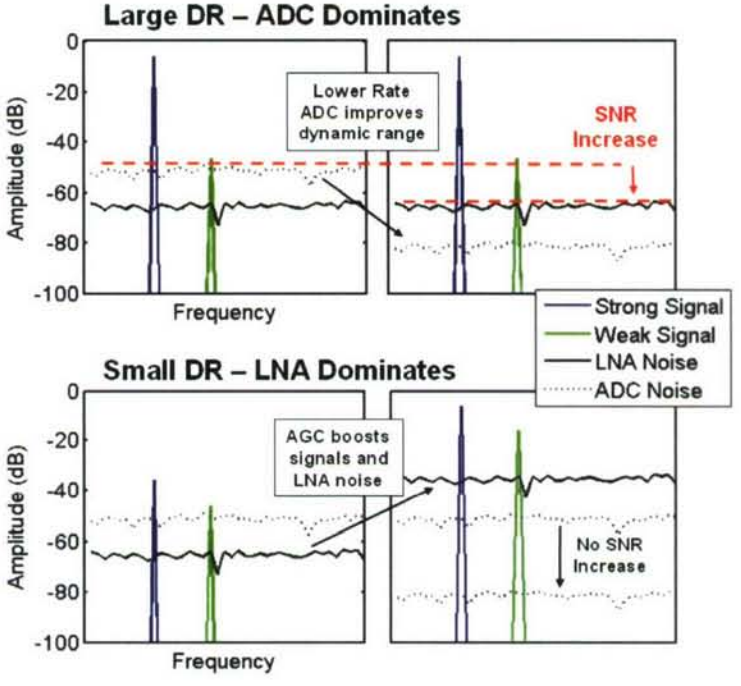


Figure 24: The relationship of ADC noise and LNA noise to overall system SNR. The top diagram shows that for signals with large dynamic range, the lower-rate ADC afforded by NoLAff undersampling results in lower ADC noise and an overall improvement in SNR. For signals with a smaller dynamic range, as in the lower diagram, an Automatic Gain Controller is assumed to boost the signals and LNA noise prior to the A/D conversion, and improvements in ADC noise do not have a substantial effect on the overall SNR.

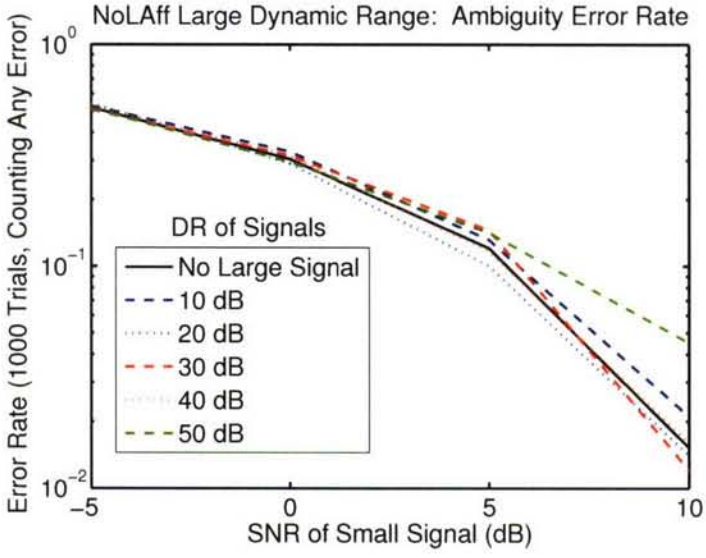


Figure 25: Decision error rate for a weak signal with a strong signal present is plotted for the SNR of the weak signal, and for the dynamic range between the strong and weak signals. Also provided for comparison is the error rate when there is no large signal present. Error rates closely match the case with no strong signal. Large dynamic range cases begin to exhibit inaccuracies from insufficient iterations of the iterative NLEQ.

the probe was set to 90 dB stronger than the weak signal. Performance for the weak signal recovery in the presence of the strong signal closely matches that of the weak signal alone. At the larger dynamic range, the effect of the iterative NLEQ inaccuracy can be seen by an increase in the error rate. Increasing the number of iterations would likely improve performance at the higher dynamic range.

Figure 26 presents the large dynamic range reconstruction error results. These results include only the cases where the correct hypothesis signal is successfully chosen. Reconstruction of a weak signal is not significantly affected by the presence of a strong signal once the correct hypothesis is chosen.

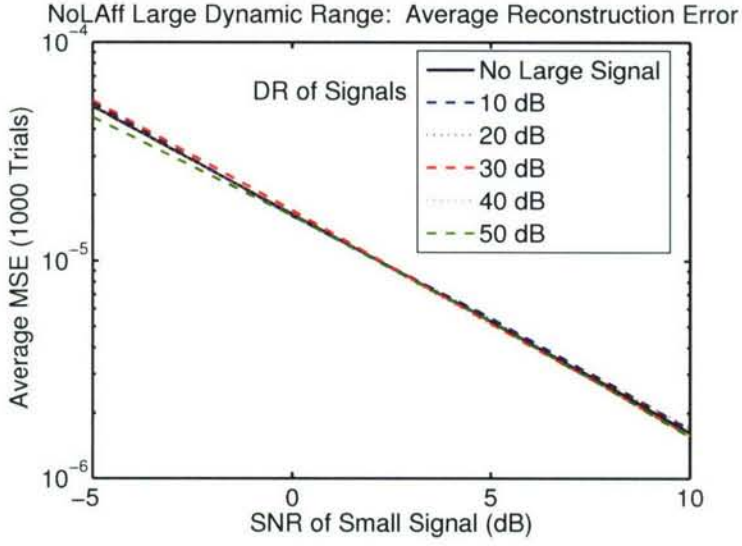


Figure 26: The reconstruction error, as average mean square error, is provided for weak signals in the presence of a stronger signal. This suggests the strong signal has little to no effect on the reconstruction of a weak signal once it has been correctly identified from the hypothesis signals.

3. VARIABLE PROJECTION AND UNFOLDING

Basis Pursuit (BP) is remarkably efficient in reconstructing signals that randomly occupy a small fraction of the positions in the discrete basis where the signals have a sparse representation. However, we have found that in dense signaling environments, i.e., environments in which the signals occupy more than 15-20 percent of the downsampled basis, BP is unable to effectively reconstruct the original signal. To reconstruct not-so-sparse signals that occupy up to nearly 100 percent of the downsampled basis, we have developed Variable Projection and Unfolding (VPU), which is a maximum likelihood (ML) sequential detection and estimation technique that searches for and identifies the contiguous columns in the discrete basis matrix that the signals span. VPU is similar to orthogonal matching pursuit (OMP) [11]; however, unlike OMP that searches over rank-1 subspaces to identify the basis support, VPU is capable of searching over subspaces of any rank. In this section, we present the VPU algorithm and its computational complexity and highlight the results of sensor network simulations comparing VPU and BP peformance. In Section 5, we derive and compare the ML/minimum-mean-square-error (MMSE) bounds to BP bounds to support the results presented here.

3.1 VARIABLE PROJECTION AND UNFOLDING

Variable Projection and Unfolding is similar to OMP as both first try to directly identify the basis support of the received signal. OMP proceeds by first correlating each of the columns of the basis with the received signal and selects the column that has largest absolute correlation. Using the selected basis column, the original signal (prior to randomization and downsampling) is reconstructed and subtracted from the received signal, and the basis column is added to the reconstruction basis column set. This process continues iteratively as illustrated in Figure 27. Although computationally efficient, the principal drawback to OMP is its restrictive search over rank-1 subspaces which preclude it from exploiting the correlation between coefficients in the basis in which the signal has a sparse representation.

VPU is nearly identical to OMP with the exception that the search process is not restricted to identify basis support only over rank-1 subspaces. VPU sequentially searches over rank-n subspaces to identify the basis support. As opposed to correlation, VPU projects the signal onto the subspace spanned by the candidate n columns. If the magnitude of the projection (mean square error) is below a prescribed threshold, the candidate columns are added to the reconstruction basis set, and this set will be used to form new projections when search commences. VPU proceeds by sequentially moving to a new rank-n subspace (e.g., adjacent or overlapping basis columns) and periodically incrementing the subspace size. VPU processing is graphically illustrated in Figure 28. Unlike OMP, VPU exploits the correlation between coefficients to enable it to significantly outperform either OMP or BP in reconstructing signals that occupy more than 15 percent of the downsampled basis, which is demonstrated in this section. The details of the algorithm are presented below.

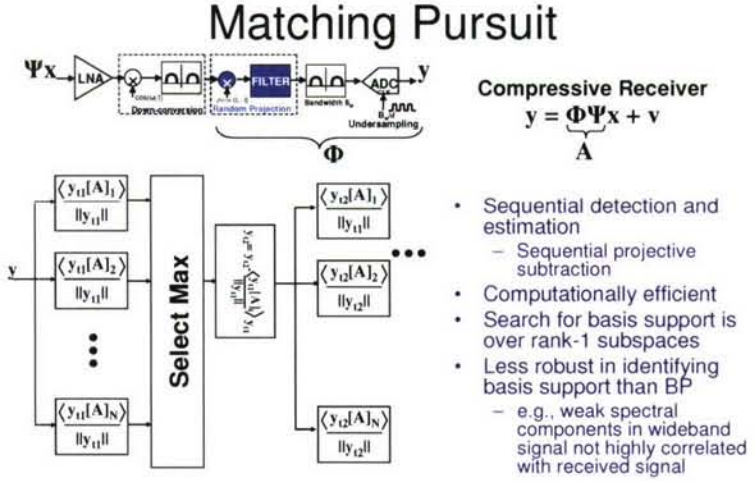


Figure 27: Block diagram of orthogonal matching pursuit processing.

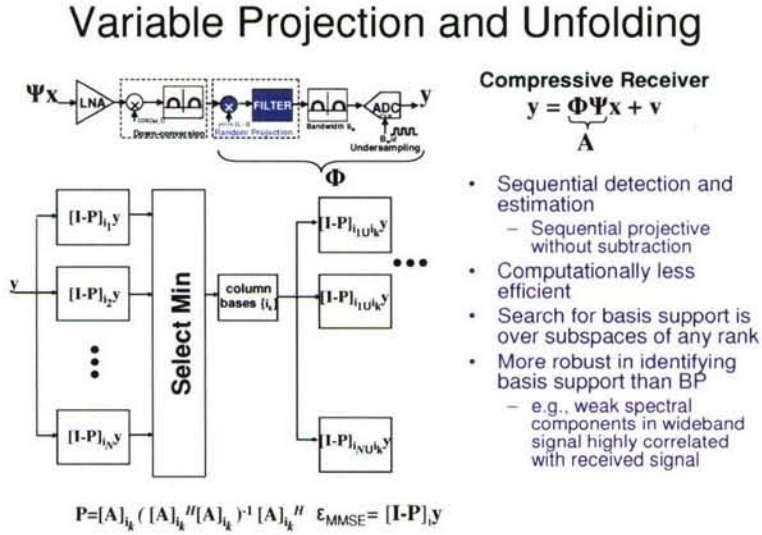


Figure 28: Block diagram of VPU processing.

Variable Projection and Unfolding (Coarse Search)

```
s_cols = {ø}, search = 0;
while (search < signals_present)
{
 for (chunk = min_swath; chunk = chunk + β; chunk ≤ max_swath)
 {
  s = {1,2,···,chunk}
  for (slide = 0; slide ≤ N; slide = slide + chunk)
  {
    error[slide,chunk] ← ε_MMSE for union(s+slide, s_cols);
    [min_slide, min_chunk ] ← argmin error
                                slide, chunk
  }
 s_cols = union(s_cols, {1,2,..., min_chunk}+min_slide};
 search = search + 1;
}
```

Figure 29: VPU pseudocode for identifying and reconstructing signals using a coarse search.

Consider a signal $x \in \mathbb{C}^{N\times 1}$ occupying unknown frequency locations in some bandwidth B that is digitized at a rate $2B/D$, with $D \equiv \Theta\Psi$ where $\Psi \in \mathbb{C}^{N\times N}$ is the basis matrix in which x has a sparse representation, d is the downsample factor and $\Theta \in \mathbb{R}^{\lfloor \frac{N}{d} \rfloor \times N}$ is a randomizing downsample matrix [12]. The digitized signal can then be described by $y = Dx + Dv$ where the noise $v \in \mathbb{R}^{\lfloor \frac{N}{d} \rfloor \times 1}$ and signal are normally distributed as $v \sim \mathcal{N}(0, C_v)$ and $x \sim \mathcal{N}(\bar{x}, C_x)$, respectively, and the symbol $\lfloor z \rfloor$ represents the greatest integer that is less than or equal to z. VPU is capable of reconstructing *not-so-sparse* signals in the presence of noise, where the term *not-so-sparse* refers to signals that can occupy up to $\lfloor \frac{N}{d} \rfloor$ elements in the basis in which the signal is sparse. VPU is a sequential estimation and detection process which searches over the columns $s = \{s_1, s_2, \cdots, s_q\}$ of $[D]_s$ for the reconstruction error that minimizes

$$\begin{aligned} &\arg\min_s \|\varepsilon_{MMSE}\|_2^2 = \arg\min_s \|(I - D(s))\, y\|_2^2 \\ &\text{where } D(s) = [D]_s \left([D]_s^H C_v^{-1} [D]_s + C_{x_s}^{-1}\right)^{-1} [D]_s^H C_v^{-1}. \end{aligned} \tag{15}$$

The pseudocode for VPU is listed below. VPU sequentially searches for the columns that minimize the error in equation (15); the columns of $D(s)$ that minimize the error in (15) are then fixed and the search continues for other signals present.

Note that the coarseness of the search is controlled by the three parameters: min_swath, β, and max_swath in which max_swath is typically set equal to $\lfloor \frac{N}{d} \rfloor$, and the parameter β controls how many excess locations in the signal basis are used in the detection and estimation process. A large β translates into reduced sensitivity due to the inclusion of excess noise, while a small β improves detection and estimation performance at the expense of computational complexity. Once

VPU (Recursive Fine Search)

```
signal = 1; s = s – perturb; error = 0;
double fine_search(signal, signals_present, s, error)
{
 if (signal ≤ signals_present)
  {
   for (offset = 0; offset=offset+1; offset ≤ 2×perturb)
    {
     error = fine_search(signal+1, signals_present, s, error)
     s(signal) = s(signal) + 1;
     if (signal == signals_present) {
      error(config(s))← find s that minimizes ε_MMSE;
     }
    }
  }
}
s ← s associated with min(error(config(s)));
```

Figure 30: VPU pseudocode for refining the identification and reconstruction of signals located in coarse search.

all the signals are found, a fine search is conducted by jointly varying the column widths of $D(s)$ that the detected signals span over a very small search radius to find the minimum error.

For cases where the channel coefficients and signal are unknown and for relatively high signal-to-noise ratio (SNR), equation (15) can be simplified using $D(s) \approx [D]_s \left([D]_s^H C_v^{-1} [D]_s\right)^{-1} [D]_s^H C_v^{-1}$ so that

$$\arg\min_s \|\varepsilon_{MMSE}\|_2^2 \approx \arg\min_s \; y^T P_{[D]_s}^{\perp}(s) y \tag{16}$$

where $P_{[D]_s}^{\perp}(s) = (I - [D]_s \left([D]_s^H C_v^{-1} [D]_s\right)^{-1} [D]_s^H C_v^{-1}))$ is the projector onto the null space of the columns of $[D]_s$. Because the projector in (16) is varied over the columns of the matrix $D(s)$, the signal is detected and reconstructed via *variable projection* and *unfolding* by locating the column positions in the basis matrix that the signal vector spans prior to downsampling and aliasing.

Once the column locations that the signal(s) span have been identified, equation (16) can be reformulated for joint synchronous center frequency estimation and baseband signal reconstruction. As an example, consider the simple case of a single signal that is sparse in the frequency basis, then

$$y = \underbrace{\begin{bmatrix} \theta & \theta \end{bmatrix} \begin{bmatrix} \cos\omega_k n & \\ & -\sin\omega_k n \end{bmatrix} \begin{bmatrix} [\Psi_{IDFT}]_s & \\ & [\Psi_{IDFT}]_s \end{bmatrix}}_{[D(\omega_k)]_s} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + Dv \tag{17}$$

$$\text{with } \phi = \tan^{-1}\left(\frac{x_1}{x_2}\right) \text{ and } \mathrm{x} = 0.5 \cdot \left(\frac{\mathrm{x}_1}{\cos\phi} + \frac{\mathrm{x}_2}{\sin\phi}\right),$$

where the carrier phase offset ϕ , baseband signal x and center frequency ω_k are simultaneously recovered using fine search VPU over ω_k, where the initial estimate of ω_k is the frequency associated with center column of $[\Psi_{IDFT}]_s$. Extending (17) to handle multiple signals is a straightforward extension of the formulation above.

The computational complexity of VPU is dominated by iterative matrix inversion requiring $\mathrm{O}(k \cdot q^3)$ operations per iteration, with $k = \sum_{i=1}^{\lfloor N/\beta \rfloor} \frac{\mathrm{max\_swath} - (\beta \times i) \cdot \mathrm{min\_swath}}{\beta} \times \mathrm{signals\_present}$ iterations. The computational complexity of VPU is higher than basis pursuit (BP) $\mathrm{O}(k_{BP} \cdot N^2)$ and matching pursuit (MP) $\mathrm{O}(\lfloor \frac{N}{d} \rfloor \cdot N \cdot q)$ [11], where k_{BP} is the number of iterations in the basis pursuit optimization. However, we will show that the VPU is capable of reconstructing signals that occupy up to $\lfloor \frac{N}{d} \rfloor$ positions of the discrete signal basis while both MP and BP fail far before reaching this limit.

As demonstrated in the next section, BP is very well suited for identifying and reconstructing signals that occupy only a small fraction of signal basis, while VPU is capable of identifying and reconstructing signals that occupy a significantly higher fraction of the basis. It is interesting to consider combining VPU and BP; the more computationally efficient BP would run initially and identify sparse signals in the basis; VPU would build on BP's coarse identification and extend the capacity of A2I to reconstruct signals well beyond the BP limits as expressed in equation (51). Combining BP and VPU is currently an active area of research.

3.2 VPU PERFORMANCE

Two simulation scenarios were used to test the efficacy of VPU and to compare its basis support identification and reconstruction performance against BP. In the first test scenario, VPU and BP detected and reconstructed a random signal without noise that occupied from 8 to 62 contiguous positions in a 520-dimensional frequency basis, i.e., with $\theta \in \mathbb{R}^{64\times520}$, $\Psi_{IDFT} \in \mathbb{C}^{520\times520}$. The second simulation scenario modeled a signal intelligence (SIGINT) Ku-band receiver operating in an environment with one to three binary phase-shift keyed (BPSK) signals present and power levels at the receiver that ranged from -67 dBm to -77 dBm in a 500 MHz band, yielding roughly a 10 dB-20 dB SNR. The receiver employed an ADC with a sampling rate of 125 MHz; in all cases the aggregate BPSK signal spectral occupancy was 52 MHz with each of the BPSK transmitters employing a pulse shaping root-raised cosine filter with a roll-off factor of 0.3. For all scenarios we ran ten thousand Monte Carlo simulations and measured the performance of both VPU and BP;

for BP we used an $\ell_2 - \ell_1$ mixed-norm optimization

$$\arg\min_x \|y - \theta\Psi_{IDFT}x\|_2^2 + \lambda \|x\|_1 , \tag{18}$$

where the regularization parameter λ was chosen to optimize BP performance [13]. In all cases the randomizing downsampling matrix θ had eight unique elements per row that were chosen from an i.i.d. Gaussian distribution.

3.2.1 Test Scenario 1: Identification and Reconstruction of Noiseless Wideband Signal

The performance results for the first test scenario are tabulated below.

	8 locations	16 locations	24 locations	62 locations
BP	100%	62%	17%	0%
VPU	100%	100%	100%	100%

TABLE 2

BP and VPU identification performance with a single noiseless wideband signal occupying exactly 8, 16, 24 and 62 frequency locations in a 520-point frequency basis with a downsampling factor of 8.

Identification performance in Table 2 corresponds to the percentage of times in our simulations that BP and VPU correctly identified all of the locations in the frequency basis that the signal randomly occupied. Note that in all cases VPU was able to simultaneously identify and reconstruct the signal with virtually perfect performance. As will be predicted in (51), BP never was able to perfectly identify the frequency support and reconstruct signals that occupied any more than $\frac{1}{4}$ of the downsampled Nyquist band. Although VPU was able to perfectly identify and reconstruct wideband signals that occupied up to 62 positions in the frequency basis, it was unable to uniquely identify the frequency support when the signal occupied 64 frequency positions. This is because the projector in (16) becomes an all-zero matrix, making it impossible to discriminate which basis locations used in reconstruction minimize the MSE. Using ML or MMSE techniques for simultaneous identification and reconstruction means that the signals need to occupy no more than one (for single-sided spectrum) or two (double-sided spectrum) fewer positions from the entire downsampled Nyquist band to successfully identify the basis support.

3.2.2 Test Scenario 2: SIGINT

The identification and reconstruction/demodulation performance of the SIGINT receiver is illustrated in Figures 31, 32, and 33 below. In Figure 31, we illustrate the basis support misidentification rate of VPU for the cases where one, two or three BPSK signals were present and whose

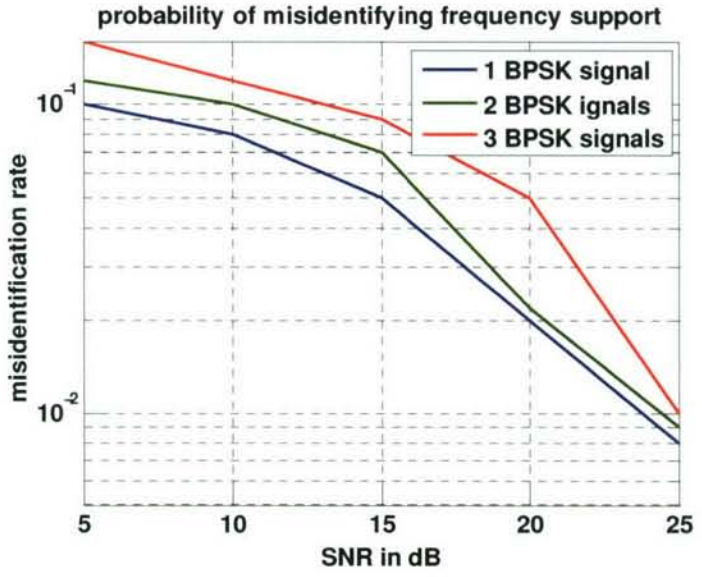


Figure 31: The rate of the misidentification of the frequency support of BPSK signals occupying an aggregate of 52 MHz in the frequency basis. The 500 MHz bandwidth received was downsampled by a factor of 8.

aggregate spectral occupancy prior to the 3 dB pulse-shaping filter roll-off was approximately 52 MHz. BP performance was difficult to gauge as it was incapable of isolating only those frequencies spanned by the BPSK signals present. Given correct frequency support identification, the measured demodulation bit error rate (BER) after reconstruction was compared to the theoretical optimum performance of an ML soft-decision demodulator. Figure 32 illustrates the difference between theoretical optimum and VPU performance with the gap graphically illustrating the effects of the folding loss. In Figure 32, VPU BER performance is averaged over reconstructing from one to three BPSK signals, while for BP the average is only over one reconstructed BPSK signal; this is because after one signal BP performance was not noticeably better than random selection. Finally, Figure 33 illustrates BPSK reconstruction at a digital carrier frequency of $\pi/8$ and an SNR of 15 dB prior to carrier frequency estimation and synchronous baseband demodulation using equation (17).

3.3 WIDEBAND CHIRP SIGNAL WITH INTERFERENCE

Like BP, VPU is capable of identifying and reconstructing signals in bases other than frequency. As an example, consider the test scenario for the identification of a chirp signal illustrated in Figure 34. A 500 MHz linear FM chirp signal was received in the presence of two strong narrowband interferers. The signal-to-interference+noise (SINR) ratio was -1 dB. The received signals were subject to random linear projection at the RF front-end and digitized at a rate of 250 MHz. We

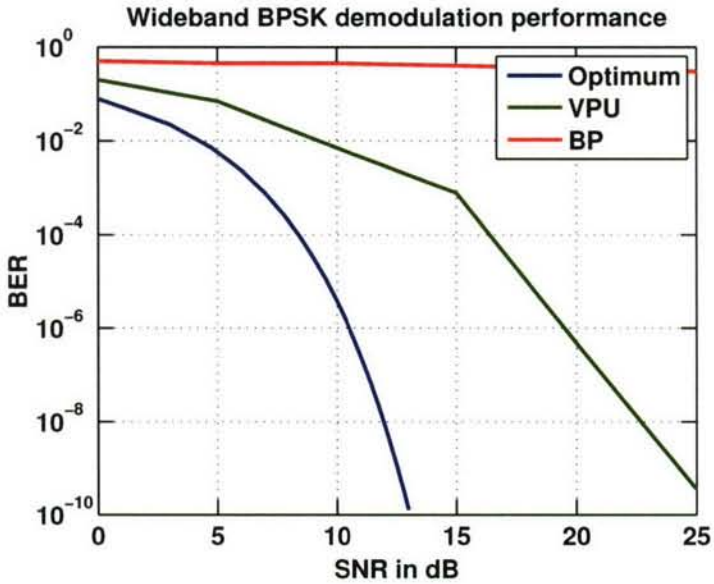


Figure 32: BER after BPSK demodulation using VPU and BP compared to theoretical optimum. Note that the performance gap between the optimum and VPU is due to the folding loss.

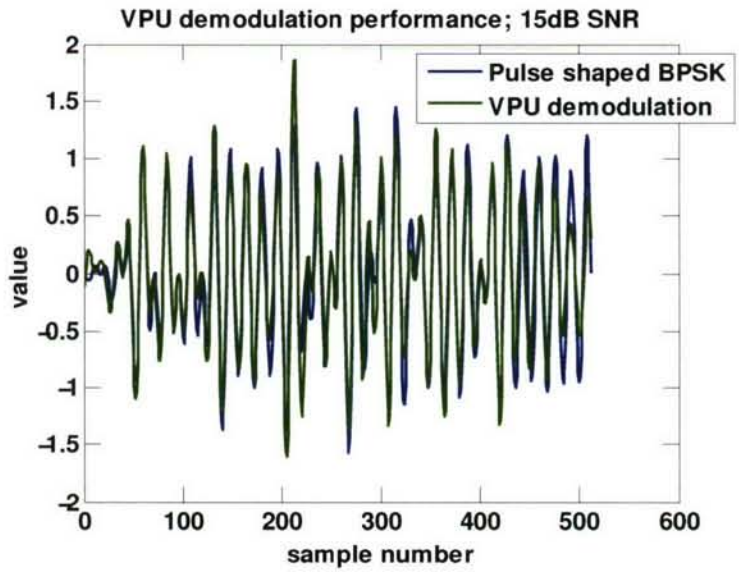


Figure 33: Example of VPU reconstruction of a BPSK signal centered at the carrier frequency $\pi/8$.

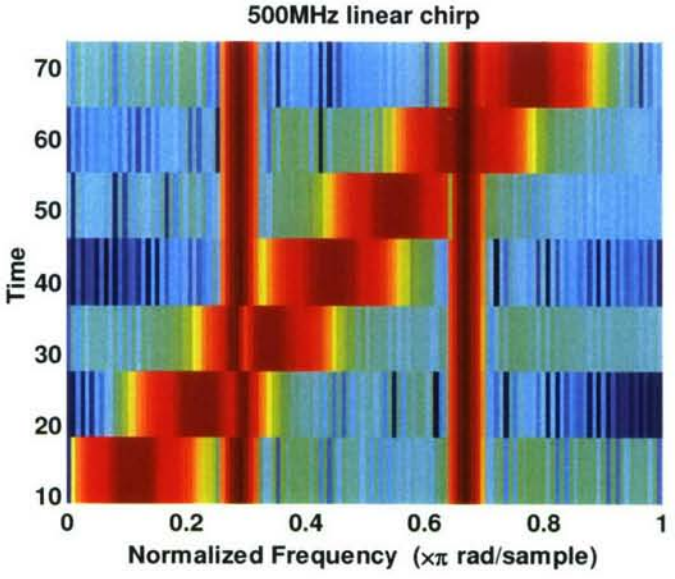


Figure 34: Chirp signal at input to the receiver prior to randomization and downsampling in the presence of two strong narrowband interferers with SINR of -1 dB.

used VPU to identify the locations in the chirp basis that yielded the lowest MSE. In 10,000 Monte Carlo runs, VPU always correctly detected and identified (see Figure 35) the chirp signal buried in interference plus noise at the output of the analog-to-digital converter.

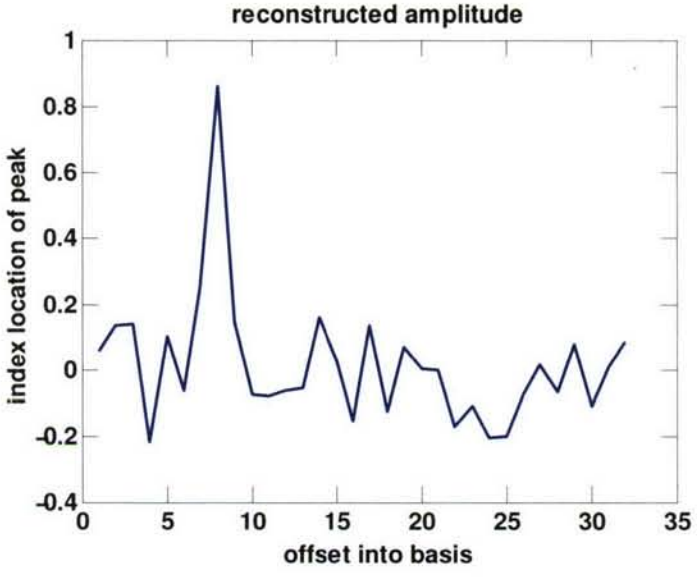


Figure 35: Detection of a 500 MHz chirp signal in the presence of two strong narrowband interferers.

4. IMPLEMENTATIONS OF COMPRESSED SENSING USING DYNAMICAL SYSTEMS

This section treats the use of dynamical systems for compressed sensing in sparse signal environments. In applying the algorithmic techniques of compressed sensing to RF sensors, we utilize analog signal processing that incorporates randomization and memory. The processed analog data are sampled at a rate that is typically well below Nyquist, but full-band reconstruction of signals is possible given sparse signal environments.

Dynamical systems provide a natural model for analog signal processing with the requisite memory and randomness. A typical dynamical system is characterized by an ordinary differential equation (ODE) that models system behavior. Sampled outputs of the dynamical system can provide initial conditions for the ODE that allow the estimation of intersample values, supporting the concept of sampling below Nyquist rates.

When the dynamical system is driven by an unknown input signal, sampled values can be used to form an innovation process that is modeled by the ODE of the dynamical system. Given sparseness of the input in some known basis, we can use the sampled values to reconstruct the input signal. The precise manner in which we can achieve this reconstruction is discussed below.

There is an extensive literature on dynamical systems outside of the present context. A few book treatments are mentioned here. A standard, fairly complete, modern reference is [14]. The nonlinear nature of dynamical systems suggests that the study of specific cases can provide insight into more general behavior. Both [15] and [16] provide numerous useful examples. A theory of the global behavior of dynamical systems has been a work in progress for a number of decades. A collection of significant results, emphasizing entropic properties, is presented in [17].

Section 4.1 discusses discrete-time dynamical systems, illustrating the randomization and memory that are important for compressed sensing. The role of dynamical systems in compressed sensing is described and lays the foundation for the practical physical realizations described later in Section 4.2. Section 4.3 describes the compressed sensing algorithms used in the examples. Section 4.4 presents a practical realization of a low-power, wideband dynamical system. The techniques introduced in Section 4.2 are used as tools in modeling this system, which is described by a delay differential equation.

4.1 RELATIONSHIP OF DYNAMICAL SYSTEMS TO COMPRESSED SENSING: DISCRETE-TIME SYSTEMS

Discrete-time dynamical systems are based on the deterministic sequence $x_{n+1} = f(x_n)$ generated by the nonlinearity $f(x)$. For many choices of f, including those motivated by physical realizations, the sequence $\{x_n\}$ has an apparent statistical behavior that can be related to that of

a limiting form of the random process $\{x_{n+1}(y) = f(x_n(y)) + y_n\}$, where $\{y_n\}$ is stochastic. When $\{y_n\}$ approaches 0, the limiting statistical behavior can be attributed to the deterministic $\{x_n(0)\}$. The random process $\{x_n(y)\}$ is described physically as a driven form of the deterministic dynamical system with drive $\{y_n\}$. In the continuous case, such random processes can be associated with random, Markovian flows on a manifold and have an Ito-like calculus.

Detection and estimation of $\{y_n\}$ can be performed in a differential sense based on the observations $\{x_n(y)\}$. If the dynamical system determined by $f(x)$ is highly structured (e.g., has zero entropy), one can work instead with a partially known drive of the form $y_n = \tilde{y}_n + \psi_n$, where $\tilde{y}_n$ is known.

Some benefits of differential detection and estimation using a dynamical system involve the ability to work with subsampled data, given certain assumptions on the input $\{y_n\}$. For convenience, we avoid a statistical model for $\{y_n\}$ here and treat the observations deterministically and perturbatively.

To see the benefits of the dynamical system and the connection with compressed sampling, consider subsampling the data by observing only $\{x_{n+ml}(y)\}$ for fixed n and l. Iterating the mapping $f(x)$ l times, we obtain the sequence of innovations

$$\rho_{n+ml} \equiv x_{n+ml}(y) - f^l(x_{n+(m-1)l}(y)) \equiv x_{n+ml} - \overbrace{(f \circ \ldots \circ f)}^{l \text{ times}}(x_{n+(m-1)l}(y)), \tag{19}$$

which represents the difference between the observed datum x_{n+ml} and the predicted datum based on the last observation $x_{n+(m-1)l}$. Linearizing about the deterministic solution $\{x_{n+ml}(0)\}$, one has the approximation

$$\rho_{n+ml} \approx \sum_{s=0}^{l-1} A_{ms} y_{n+(m-1)l+s}, \tag{20}$$

where

$$A_{ms} \stackrel{\text{def}}{=} \begin{cases} \prod_{k=s+1}^{l-1} f'(f^k(x_{n+(m-1)l})) & 0 \le s \le l-2 \\ 1 & s = l-1 \end{cases}. \tag{21}$$

The latter can be summarized in the compact form

$$\vec{\rho} \equiv \vec{x} - f^l(\vec{x}) \approx A \cdot \vec{y}, \tag{22}$$

given a finite vector $\vec{x} \leftrightarrow \{x_{n+ml}\}_{m=1}^{M}$ of subsampled observations and a corresponding finite input vector $\vec{y} \leftrightarrow \{y_{n+s}\}_{s=0}^{Ml-1}$. Here, the matrix A represents the linearized relation between the input $\{y_n\}$ and the innovations $\vec{x} \leftrightarrow \{\rho_{n+ml}\}_{m=1}^{M}$. A is a rectangular matrix with more columns than rows ($M \times Ml$), reflecting the subsampling. The notation $f(\vec{x})$ indicates component-by-component application of f.

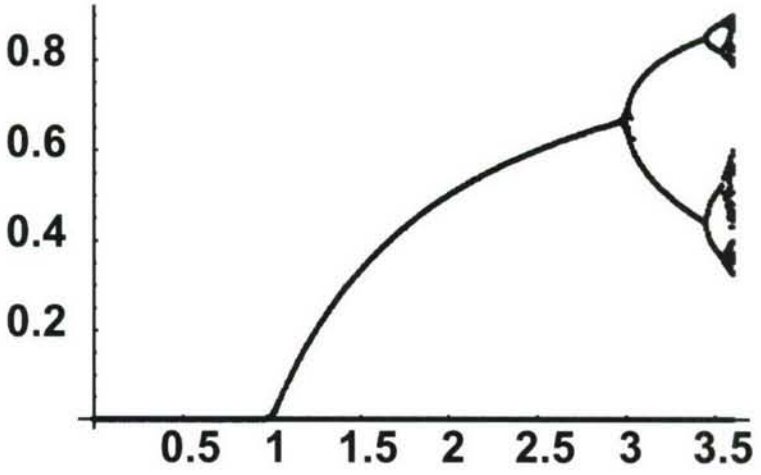


Figure 36: The bifurcation diagram of the logistic system is shown. Each point on the plot belongs to a periodic orbit. As the bifurcation parameter increases, the orbit shifts location until, at certain thresholds, it doubles in length.

Calculating A requires only the subsampled observations expressed by $\vec{x}$. However, these observations depend on $\vec{y}$ implicitly. Thus the apparently linear equation (22) is in fact highly nonlinear. However, equation (22) can be used along with sparsity models for $\vec{y}$ to solve the underdetermined system.

The dynamical system associated with the logistic map defined by

$$f(x) \stackrel{\text{def}}{=} rx(1-x) \tag{23}$$

has been used to represent population dynamics of biological systems [16] from epoch to epoch. The parameter r, $0 \leq r \leq 4$, characterizes a family of systems that exhibit substantially different, but stable, behavior. The logistic map has a single attracting fixed point at zero for $0 \leq r \leq 1$. As r increases, an additional stable fixed point appears and, with further increases in r, bifurcates into a stable, attracting orbit of two points. Further increases in r lead to more and more orbit doubling bifurcations until, at around $r = 3.6$, the dynamical system exhibits quasi-random, but stable behavior. The bifurcation diagram (the traces of points involved in periodic orbits) is shown schematically in Figure 36.

Some simple examples of the performance of a discrete-time dynamical system are shown below.

Figure 37 shows the time-domain response of the dynamical system based on the logistic map, downsampled by a factor of 4. Tonal inputs are shown along with the system response. When

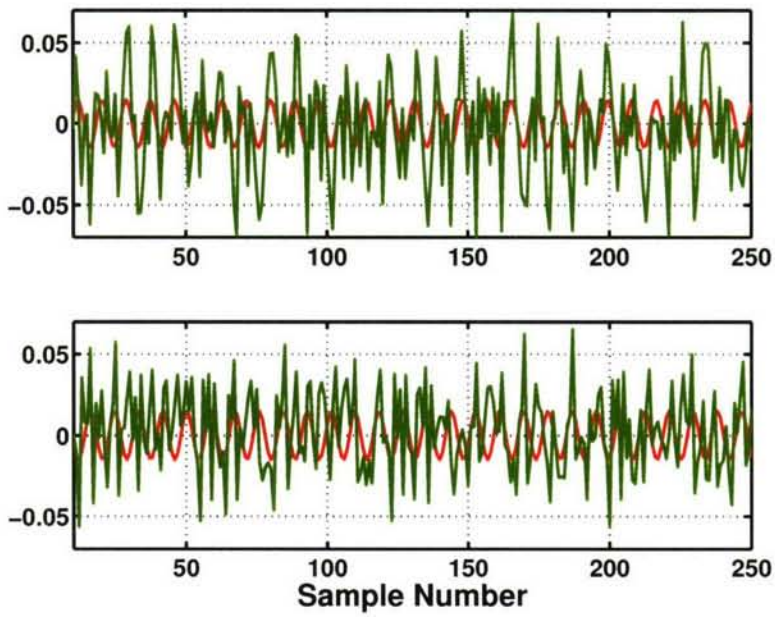


Figure 37: The logistic discrete-time dynamical system is downsampled by a factor of 4. In the top figure, a tone, shown in red, drives the system, which exhibits the response shown in green. Note that the response appears random. The second figure shows the system response when an ambiguous tone (one whose downsampled frequency response is identical to the first) drives the same system. The system's response is again somewhat random and, more significantly, quite different from that shown in the top figure.

tonal inputs are ambiguous given the downsampling factor, the sampled systems remain quite distinguishable. Figure 38 shows the innovations of the discrete-time logistic map both as modeled and under the linearized model required for compressed sensing. There is good agreement between the two models. Figure 39 shows the successful identification of superimposed tones for the logistic map dynamical system.

4.2 RELATIONSHIP OF DYNAMICAL SYSTEMS TO COMPRESSED SENSING: CONTINUOUS-TIME SYSTEMS

Ordinary Differential Equations Continuous-time dynamical systems, as studied extensively in the literature, are characterized by an ordinary differential equation of the form $\dot{x} = f(x)$. In analogy to the discrete case, we consider a driven system of the form

$$\dot{x}(t) = f(x(t)) + y(t), \tag{24}$$

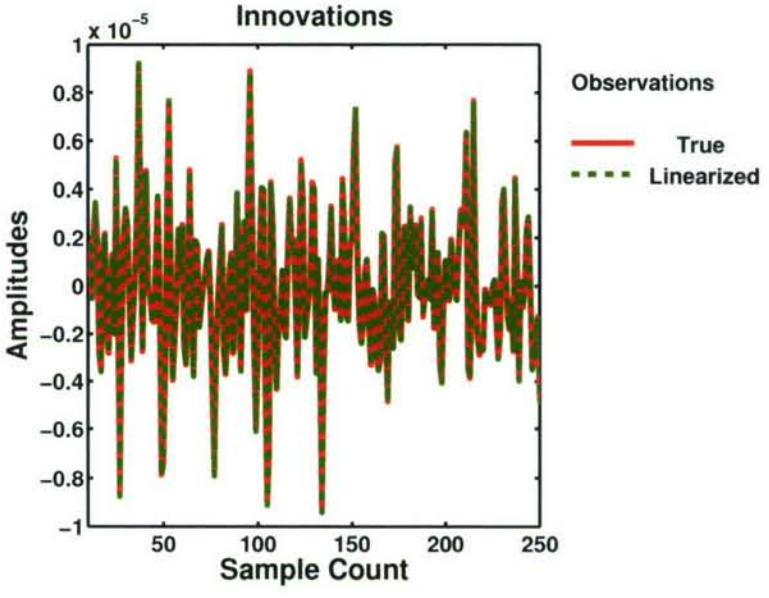


Figure 38: The logistic discrete-time dynamical system is downsampled by a factor of 4. Eight tones randomly located across the band drive the system. The red trace shows the time series of the driven dynamical system. The green trace shows the modeled behavior after linearization. There is good qualitative agreement between the true and linearized models.

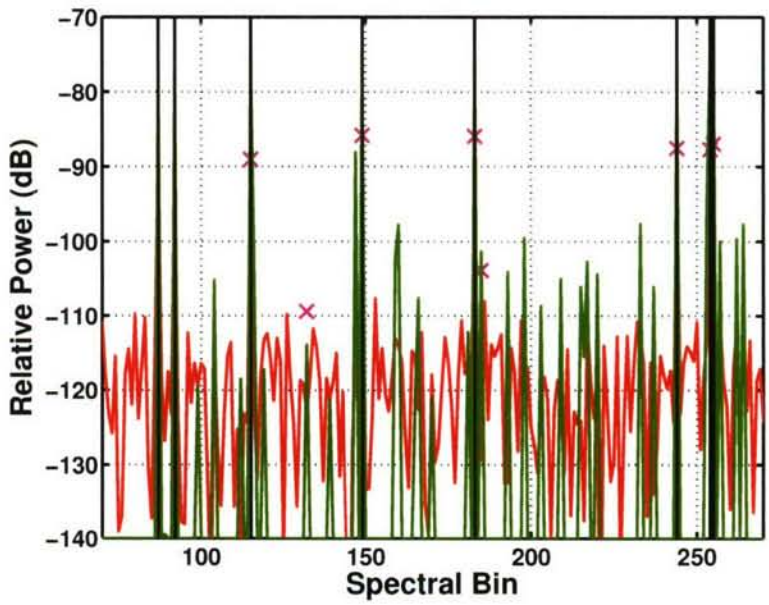


Figure 39: The logistic discrete-time dynamical system is downsampled by a factor of 4. Eight tones randomly located across the band are successfully identified (green trace). The signals occupy 2 % of the band. For comparison, the crosses mark the locations of signals identified with matching pursuits. The red trace shows the spectrum of the downsampled data, which has a noise-like appearance. The black vertical lines show the signal's true locations.

where $y(t)$ represents the signal input. This input can consist of known $\tilde{s}(t)$ and unknown $\sigma(t)$ components under the relationship $s(t) = \tilde{s}(t) + \sigma(t)$. To simplify the discussion, we assume that the known component vanishes.

Given sampled observations of the state vector $x(t)$, we wish to determine the input vector $y(t)$. This can be accomplished if the input is suitably constrained. For example, the input may be bandlimited, as assumed below. The sample rate required to estimate the input depends on the input signal's bandwidth. This sample rate can be reduced if the input is sparse in some sense.

We sketch one path toward reconstruction of the input signal. Consider the ancillary autonomous system $\frac{d}{dt}\tilde{x}(t) = f(\tilde{x}(t))$ for state vector $\tilde{x}$ over the interval $[nT, (n+1)T)$ given the initial condition $\tilde{x}(nT) = x(nT)$. The variable T indicates the time between sampled values $x(nT)$ of the dynamical system. This autonomous system is a predictor of $x(t)$. The discrepancy $\epsilon(t) \stackrel{\text{def}}{=} x(t) - \tilde{x}(t)$ between $\tilde{x}(t)$ and $x(t)$ measures the effect of the input $y(t)$ on the dynamical system during the interval $[nT, (n+1)T)$. Let

$$\rho(nT) \stackrel{\Delta}{=} \lim_{t \to nT} x(t) - \tilde{x}(t) \tag{25}$$

so that $\rho(nT)$ can be interpreted as a discrete-time innovations process.

To first order, $\epsilon(t)$ satistfies a linear ODE for the form

$$\dot{\epsilon}(t) = Df(\tilde{x}(t))\epsilon(t) + y(t) \tag{26}$$

in the interval $[nT, (n+1)T)$ with the initial condition $\epsilon(nT) = 0$. The expression $Df(\tilde{x}(t))$ represents the Jacobian matrix of the mapping $f(x)$ evaluated along the trajectory $\tilde{x}$ of the predictor. The values of $\tilde{x}(t)$ are determined by the initial condition $\tilde{x}(nT) = x(nT)$, which is the last sampled observation. The linearity of the ODE for $\epsilon(t)$ allows us to express the innovations as a solution to the linear algebra system

$$\begin{pmatrix} \rho(T) \\ \vdots \\ \rho(nT) \end{pmatrix} = A \begin{pmatrix} \vdots \\ a_1 \\ \vdots \\ a_n \\ \vdots \end{pmatrix} \tag{27}$$

where

$$A_{nm} \stackrel{\Delta}{=} \int_{nT}^{(n+1)T} \text{T.O.}\{e^{\int_\tau^{(n+1)T} Df_{\tilde{x}(t)}\, dt}\} p(\tau - m\delta)\, d\tau, \tag{28}$$

given the cardinal series representation of the bandlimited signal $s(t)$

$$s(t) = \sum_m a_m p(t - m\delta) \tag{29}$$

for some δ. The entries A_{mn} of the coefficient matrix A are shown in the form of the general solution to a system of linear ODE's given inputs of the form $p(t - m\delta)$. The notation "T.O." indicates a time ordering of the matrix products (integrating factor), which is required when the ODE system is not a single scalar equation (see [18] for an interesting discussion of these solutions).

4.3 APPLICABLE COMPRESSED SENSING ALGORITHMS

In evaluating the performance of compressed sensing for dynamical systems, we use the mixed L_2, L_1 norm as an objective function. Specifically, let Φ represent a known $m \times n$ real matrix, g a length m observed vector, and f a length n unobserved vector. The objective function

$$\|\Phi f - g\|_2^2 + \mu \|f\|_1 \tag{30}$$

is minimized over f to obtain a sparse solution to an underconstrained least-squares fit given $m \leq n$. The procedure we follow is based on [19] and [3]. Taking the gradient of the objective function yields

$$\nabla \|\Phi f - g\|_2^2 = 2\Phi^T(\Phi f - g). \tag{31}$$

If the shrinkage function S_μ is defined on n-vectors as

$$S_\mu(x) \stackrel{\text{def}}{=} \begin{cases} x - \mu/2 & x > \mu/2 \\ x + \mu/2 & x < -\mu/2 \\ 0 & \text{otherwise} \end{cases}, \tag{32}$$

then the update step of an iterative minimizer for equation (30) is given by

$$f \leftarrow S_\mu(f + \Phi^T(g - \Phi f)). \tag{33}$$

In use, the iterative step is initialized with $f = 0$ and iterated a fixed number of times. If the update f is still zero (as is often the case when μ is large), the value of μ is reduced and the algorithm is repeated until a nonzero f is reached. To ease implementation, it is useful to run the iterations on a low dimensional subspace which is spanned by a smaller number $n' < n$ of components of f. These components can be selected using basis pursuit [20].

Another type of approach [21] involves the use of a sequence of surrogate functions that bound the objective function and are easier to minimize. For example, the L_1 norm has quadratic upper bounds expressed by the one-dimensional inequality (see Figure 40)

$$|x| \leq |x_0| + \frac{x_0}{|x_0|}(x - x_0) + \frac{1}{2}\frac{(x - x_0)^2}{|x_0|}. \tag{34}$$

Added to the quadratic least-squares term, the resulting upper bound is quadratic and can be minimized explicitly. Then new upper bounds can be determined, *et cetera*. The resulting iterations are expressed by

$$f \leftarrow (\Phi^T\Phi + \mu \text{diag}(f)^{-1})^{-1}\Phi^T g. \tag{35}$$

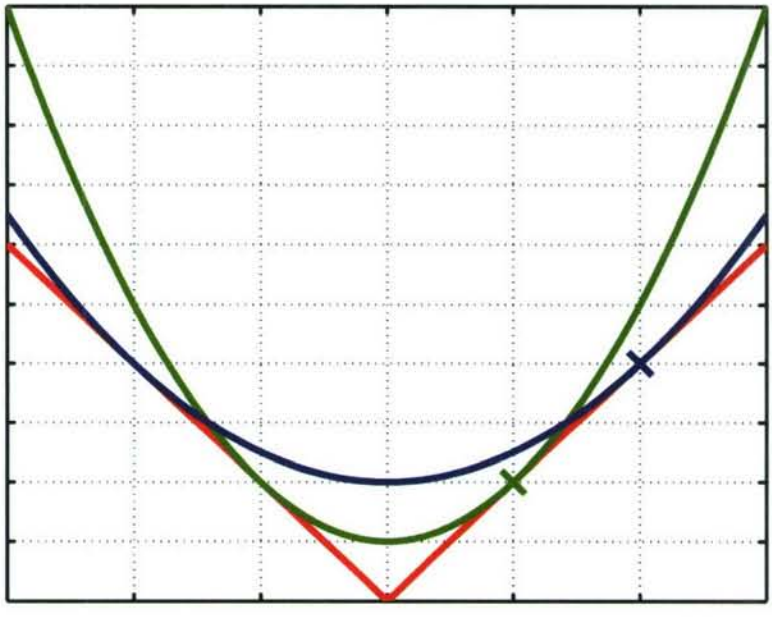


Figure 40: The L_1 norm in one dimension can be upper-bounded by quadratics that are tangent to L_1 contours at the coordinates of points from successive stages in the iteration. Added to the quadratic least-squares term, the resulting upper bound on the objective function is quadratic and easily minimized, providing the next point in the iteration and the next set of bounding quadratics.

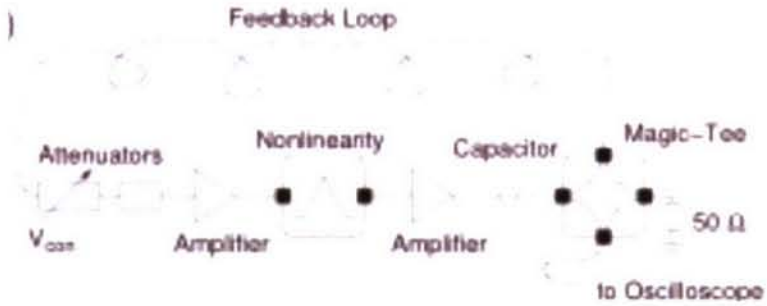


Figure 41: Block diagram of a practical, wideband dynamical system with a tent-map nonlinearity in the loop.

It's worth noting that the iterative approach in equation (33) is also based on a surrogate function. In this case the surrogate does not have the quadratic term contributed by the least-squares portion of the objective function. The resulting surrogate can be optimized component by component [19].

4.4 EXAMPLES OF A BROADBAND IMPLEMENTATION

4.4.1 Broadband Circuit and Model

A broadband microwave dynamical system based on a simple nonlinearity has been studied in [22]. This dynamical system is described by the block diagram of Figure 41, which shows a tent-map nonlinearity within a loop with delayed feedback. The nonlinearity is implemented as shown in Figure 42. The implemented system exhibits a bandwidth on the order of 500 MHz given feedback delays of about 10 nsec. This low-power, wideband implementation is extremely simple. It appears that operating bandwidths can be extended to multiple GHz without difficulty by miniaturizing the circuitry.

4.4.2 Phase Space Dynamical Model

The dynamical system is modeled in [22] by an integro-differential equation for the voltage $v(t)$. This equation incorporates a model for the bandpass response of the feedback loop. Here a source $s(t)$ is added, resulting in

$$v(t) + a\dot{v}(t) + b\int^{t} v(l)\,dl + s(t) = dF((cv(t-\delta)), \tag{36}$$

where $F(v)$ is the tent-map nonlinearity given by

$$F(v) \stackrel{\text{def}}{=} v_0 - \sqrt{F_1^2(v) + a_F^2} \tag{37}$$

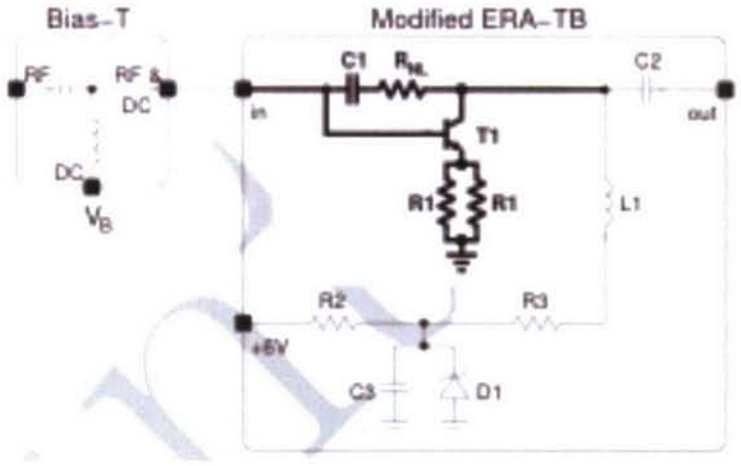


Figure 42: Circuit diagram of the tent-map nonlinearity.

and

$$F_1(v) \stackrel{\text{def}}{=} \begin{cases} A_l(v - v^\star) & v \leq v^\star \\ A_r(v - v^\star) & v > v^\star \end{cases} . \tag{38}$$

For the example treated here, the parameters in the integro-differential equation become $a \stackrel{\text{def}}{=} \Delta^{-1}$, $b \stackrel{\text{def}}{=} \omega_0^2/\Delta$, $c \stackrel{\text{def}}{=} -\gamma/\alpha$, and $d \stackrel{\text{def}}{=} \alpha = 9$, which in turn depend on the parameters

$$\begin{aligned} A_l &= .47 \\ A_r &= -.62 \\ a_F &= .05 \\ v^\star &= .12 \\ v_0 &= 6 \\ \gamma &= 3.7 \\ \omega_0 &= 1.5\,10^9 \\ \Delta &= 2.5\,10^9 \end{aligned} \tag{39}$$

A detailed physical decription of the circuit model and interpretation of the parameters are presented in [22]. It suffices to note here that the loop delay is expressed by δ.

There are several ways to convert the integro-differential equation (36) into a delay-differential equation (DDE). One is pursued here. Defining the state variables

$$\begin{aligned} x(t) &\stackrel{\text{def}}{=} v(t) \\ y(t) &\stackrel{\text{def}}{=} \dot{v}(t) \end{aligned}$$

and differentiating, we have the DDE

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y(t) \\ -y(t)/a - bx(t)/a - \dot{s}(t)/a + dc\dot{F}(cx(t-\delta))y(t-\delta)/a \end{pmatrix}. \tag{40}$$

4.4.3 ODE Approximation

By defining auxilliary variables, a DDE can be a approximated by an ODE. We will illustrate the modeling process for the DDE of equation (40). Let $x_k(t) \stackrel{\text{def}}{=} x(t + k\delta_s)$ and similarly for $y_k(t)$. Assume that the loop delay δ is a multiple of the sample interval δ_s. In other words, assume $m\delta_s = \delta$. Then

$$\begin{pmatrix} \dot{x}_k(t) \\ \dot{y}_k(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\Delta \end{pmatrix} \begin{pmatrix} x_k(t) \\ y_k(t) \end{pmatrix} \tag{41}$$
$$+ \begin{pmatrix} 0 & 0 \\ 0 & -\gamma\Delta\dot{F}(\frac{\gamma}{\alpha}x_{k-m}(t)) \end{pmatrix} \begin{pmatrix} x_{k-m}(t) \\ y_{k-m}(t) \end{pmatrix} - \Delta\dot{s}(t + k\delta_s) \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

This system of equations in the augmented state variables $\{(x_k, y_k)\}$ is certainly not finite. The loop delay creates an unbounded memory in the DDE system that is not modeled exactly with a finite system of ODEs. Note that the additional state variables $\{(x_k, y_k)\}$ form a system with a banded coefficient matrix that contains only two bands: the diagonal and a single off-diagonal band reflecting the delay coupling.

Since the loop gain γ limits the coupling expressed by the off-diagonal terms, it is plausible to expect that a truncated version of the ODE system equation (41) can approximate a solution of the DDE equation (40) sufficiently far from the truncation boundary.

As an example of an approximating ODE model, truncate equation (41) so that the ODE has 2×100 state variables corresponding to 100 sample times (indexed by k above). Assign the value zero to coefficients in the ODE system involving variables that are out of bounds. Values at the sample times provide initial conditions for the system of ODE's, which are solved to determine the between-sample values that are required for compressed sensing. The level of effort in the computation is limited by the very sparse nature (2 bands) of the system matrix. In principle, solving the ODE system over one sample interval results in a time series valid over 100 intervals; in practice, the ODE only approximates the solution to the DDE. More specifically, the initial sample intervals, which are near the truncation boundary, do not accurately model the DDE solution.

The difference between the DDE and ODE solutions is shown, for example, in Figure 43. Time, spanning 100 sample intervals, is labeled on the x-axis. Relative error is shown on the y-axis. Note that the ODE-modeling error is very large for initial samples, but settles rapidly to between -40 and -50 dB. The latter residual is determined by the accuracy of the DDE and ODE integrators, which can be tuned to some extent.

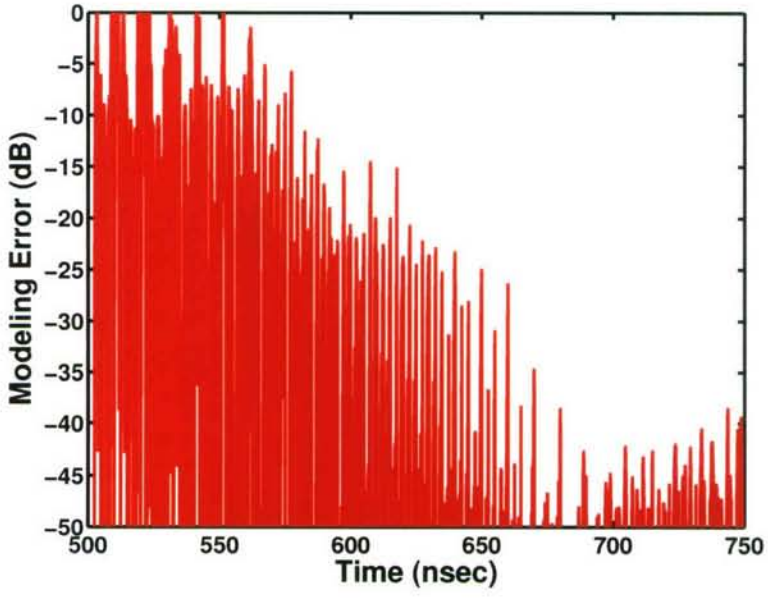


Figure 43: The DDE system can be approximated by an ODE system. The approximation suffers errors due to truncation of the inherently infinite dimensional ODE model required by the infinite memory of the DDE system. Loss due to truncation is shown. Large losses occur near the truncation boundary. These losses decrease rapidly as time progresses.

4.4.4 Linearization of ODE Model

Coefficient matrices relating basis expansions of signals (e.g., cardinal series) and observations (e.g., innovations) can be derived for ODE models of dynamical systems using the linearization methodology of Section 4.2. We examine linearization for the ODE model of the DDE system of Section 4.4.2.

Introduce the notation $x_k(t) \stackrel{\text{def}}{=} \tilde{x}_k(t) + \chi_k(t)$, and $y_k(t) = \tilde{y}_k(t) + \phi_k(t)$ where $(\tilde{x}_k(t), \tilde{y}_k(t))$ satisfies equation (41) with $s(t) \equiv \tilde{s}(t)$. Then equation (41) can be linearized about the solution with input $\tilde{s}(t)$. The resulting linear ODE with input $s(t) = \tilde{s}(t) + \sigma(t)$ and state variable $(\chi_k(t), \phi_k(t))$ becomes

$$\begin{aligned}
\begin{pmatrix} \dot{\chi}_k(t) \\ \dot{\phi}_k(t) \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ -\omega_0^2 & -\Delta \end{pmatrix} \begin{pmatrix} \chi_k(t) \\ \phi_k(t) \end{pmatrix} \\
&+ \begin{pmatrix} 0 & 0 \\ \frac{\Delta\gamma^2}{\alpha} \ddot{F}(\frac{-\gamma}{\alpha}\tilde{x}_{k-m}(t))\tilde{y}_{k-m}(t) & -\gamma\Delta\dot{F}(\frac{-\gamma}{\alpha}\tilde{x}_{k-m}(t)) \end{pmatrix} \begin{pmatrix} \chi_{k-m}(t) \\ \phi_{k-m}(t) \end{pmatrix} \\
&- \Delta\dot{\sigma}(t + k\delta_s) \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \qquad (42)
\end{aligned}$$

The coefficient matrix is a function of the solution $(\tilde{x}_k(t), \tilde{y}_k(t))$, resulting in a linear ODE with nonconstant coefficients. Again, the linear system is truncated upon application to the DDE system of Section 4.4.2.

4.4.5 Performance

Some brief examples of performance are presented for the DDE system. The intent is to illustrate qualitatively the performance of the dynamical system. System parameters are not optimized in any significant sense. In this stage the results are primarily anecdotal, but the system itself is practical, wideband, low power and capable of extension to larger bandwidths on the order of multiple GHz.

Although the results are mainly qualitative, a simple quantitative analysis based on a linearized system model is presented for the case of impulsive inputs.

The input to the dynamical system consists of the combination of a known signal (a tone in the examples) and an unknown superposition of wideband pulses. The pulse bandwiths are 1 GHz. The tone is a large signal that is used to ensure sufficient signal-dependent randomization even if the input signals are weak. The system is downsampled by a factor of 3 to 5. The 3 dB bandwidth of the dynamical system suggests the factor of three downsampling, but the slow spectral roll off shown in Figure 45 makes moot the choice of a particular downsampling factor.

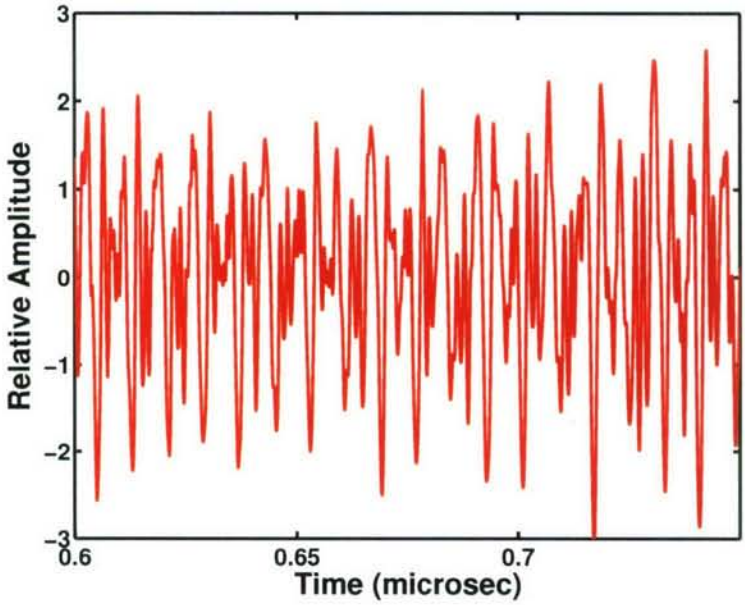


Figure 44: The response of the dynamical system chosen as an example is shown. The input to the dynamical system consists of three wideband pulses, none of which are apparent in the system's response.

Figure 44 shows the time-domain response of the dynamical system. The interval shown in the figure includes all three wideband impulses, but none are evident in the plot. Figure 45 shows the spectrum of the same dynamical system. Note the prominent discretes in the wideband spectrum. The character of the spectrum is highly dependent on the values of circuit parameters and in particular on the amount of feedback.

The coefficient matrix describing the linearized ODE approximation to the DDE system is based on equation (42). Graphically, the linearized system results in a coefficient matrix with the support shown in Figure 46. The matrix entries are based on a cardinal series model (train of fixed pulses with unknown coefficients) of the signals. Note that the coefficient support is sparse and banded, reflecting the delay structure in the DDE.

When 3 pulses of an impulsive waveform are used to drive the DDE, compressed sensing applied to the linear ODE model can be used to recover the unknown input armed only with knowledge of the signal basis (impulses, in this example), as shown in Figure 47. The true pulse locations are shown with vertical lines and the estimated powers are shown by the green curves. The compressed sensing techniques used for recovery are described briefly in the beginning of Section 4.3.

The coefficient matrices used for signal recovery are based on the linearized ODE model of the DDE system. This model provides accurate predictions of system response that can be used to

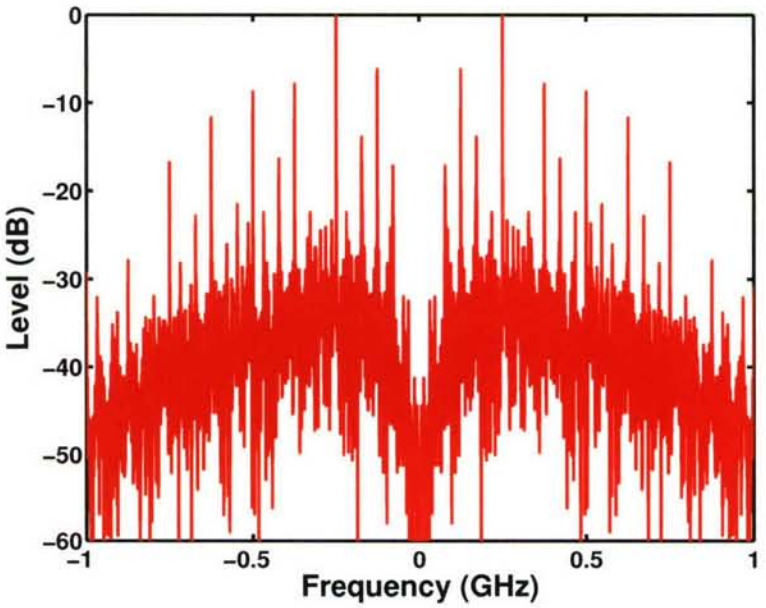


Figure 45: The dynamical system chosen as an example exhibits the spectrum shown here. Note the notch at low frequencies due to the bandpass character of the circuit. The spectrum is wideband with significant discretes.

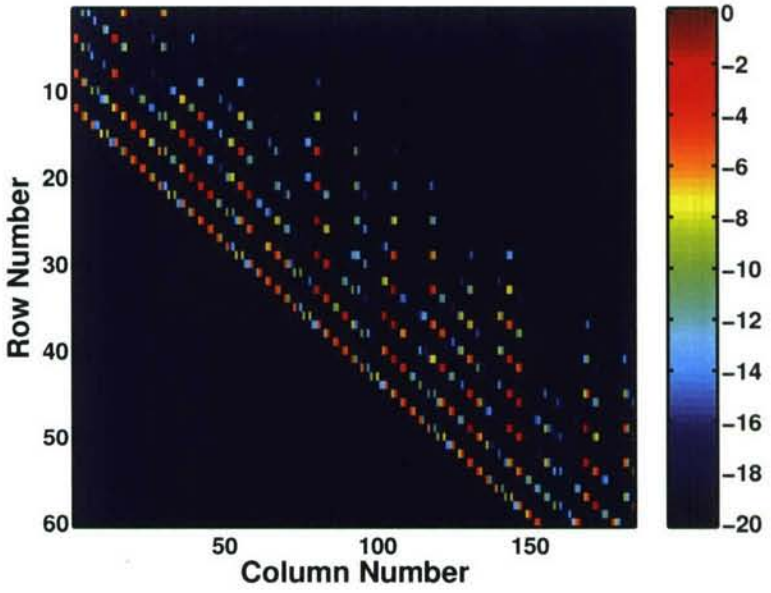


Figure 46: The coefficient matrix used to determine the input signal is shown as an image, revealing the typical banded structure that reflects the bulk delay in the DDE (gaps between bands) as well as fading memory related to the size of the feedback.

form statistics of performance under various densities of input signals. In Figure 48, the cumulative distribution of the loss of SNR is shown under various system loadings. For example, with 1% of the time line occupied by wideband (1 GHz) pulses, each additional wideband pulse suffers very little loss associated with interference from other pulses. As the time line fills, the loss becomes significantly larger. These losses do not include the typical folding loss associated with downsampling, which is about the same as the downsampling factor. This loss is suffered even in the absence of other interference, as long as the RF device noise dominates the LSB of the quantizer.

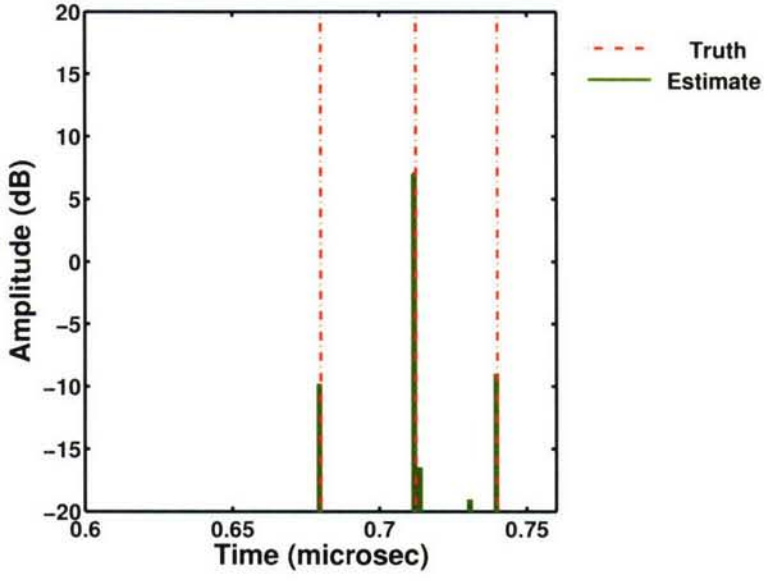


Figure 47: Multiple pulses are detected successfully after downsampling by a factor of 3 to 5.

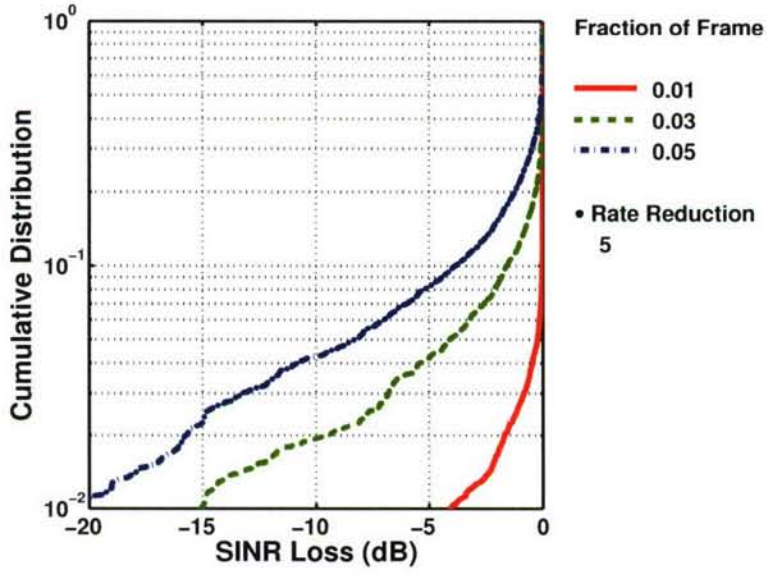


Figure 48: For various pulse densities, the loss in SNR incurred due to downsampling by a factor of 3 to 5 is shown. Background pulses, which are normally well-separated from each other in time, create self-interference within the dynamical system. The amount of interference depends on the downsampling factor as well as the fraction of the time line that is filled. This phenomenon occurs with any signal basis. Not included in the loss is the folding loss associated with downsampling.

5. PERFORMANCE BOUNDS

In this section, we discuss performance bounds for the techniques presented in Sections 2 and 3.

In Sections 5.1–5.3, we present and derive A2I reconstruction performance bounds for BP and VPU, respectively. The analog receiver configuration for the scenarios presented are illustrated in Figure 49 and the details of the hardware implementation are presented in [12], [23] and Section 5.1. In all cases and with no loss in generality, the signals are subject to random linear projections by circuitry in the analog RF front-end after which the signals are digitized using an undersampling analog-to-digital converter (ADC). First, the rationale and requirements for random linear projection are presented, after which performance bounds are derived for undersampled signals both with and without noise present.

In order to understand the penalty in performance associated with compressed sensing, in Section 5.4 we examine accuracy of the estimates of some signal parameters. Specifically, we consider quasi-stationary signals with unknown covariance. The accuracy of covariance estimates is evaluated under the assumption of a sparse signal environment. Accuracy is characterized as a function of the amount of undersampling.

Finally, with Section 5.6 we describe the performance bounds for NoLAff hypothesis testing.

5.1 RANDOMIZATION REQUIREMENTS

Downsampling induces correlation between the columns of the basis matrix of the received signal. This can be expressed as $D = H_{DS}\Psi$, where the downsample matrix

$$H_{DS} = \begin{bmatrix} 1 & 0 & \cdots & 0 & & & \cdots & 0 & \cdots \\ 0 & & \cdots & 1 & 0 & \cdots & 0 & & \\ & \vdots & & & \vdots & & 1 & 0 & \\ & & & & & & & & \ddots \end{bmatrix}, \quad H_{DS} \in \mathbb{R}^{\lfloor \frac{N}{d} \rfloor \times N}, \ \Psi \in \mathbb{C}^{N \times N} \tag{43}$$

selects every dth element from each column of the basis matrix Ψ, where d is the downsample factor and the symbol $\lfloor x \rfloor$ represents the greatest integer that is less than or equal to x. As an example, consider a signal that is sparse in an orthonormal frequency basis. Each of the columns in the basis matrix (i.e., IDFT matrix) after downsampling is identical to d other columns corresponding to the effects of aliasing. An ideal anti-aliasing filter prior to downsampling would place zeros in the aliased signal vector locations. That is, for the received signal vector

$$y = \underbrace{H_{DS}\Psi_{IDFT}}_{D_{IDFT}} \underbrace{Fx}_{\hat{x}} = D_{IDFT}\hat{x} \ \text{ where } F = \text{diag}(\ 1\text{s} \quad 0\text{s} \quad 1\text{s}\ ), \tag{44}$$

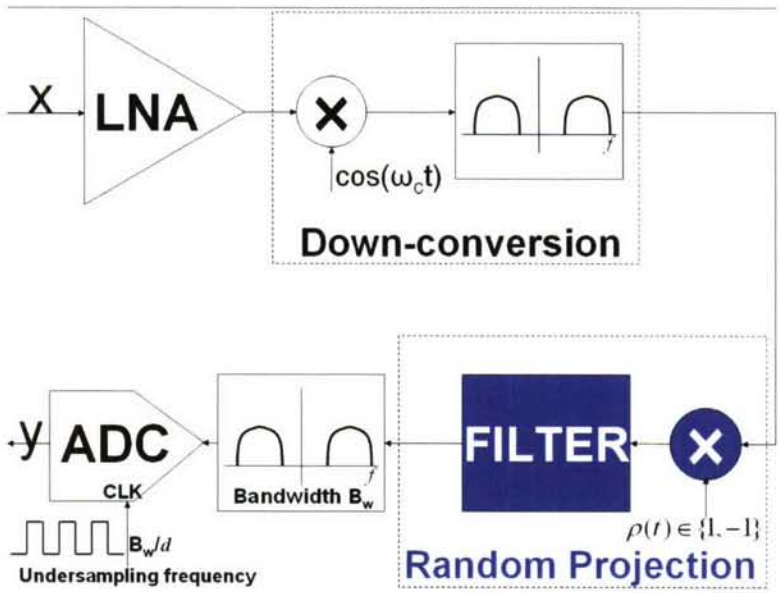


Figure 49: Block diagram of the analog front-end of a A2I receiver. The bandwidth of the received spectrum is undersampled by a factor of d.

columns $\lfloor \frac{N}{2d} \rfloor$ to $N - \lfloor \frac{N}{2d} \rfloor$ of the downsampled IDFT matrix D_{IDFT} do not contribute to the received signal vector y because the data x have been brick-wall filtered by the diagonal matrix F. This is the traditional view of alias-free downsampling, where the brick-wall filtered baseband signal is ideally confined to only have nonzero support that spans the first and last $\lfloor \frac{N}{2d} \rfloor$ columns of the IDFT matrix Ψ_{IDFT}. If, however, the signal locations are unknown at the receiver and not confined to baseband, brick-wall filtering will not facilitate the undistorted recovery of the signal x from the downsampled signal y.

A2I mitigates the effects of aliasing with random linear projections prior to downsampling [12]. The key idea is to replace the downsample matrix H_{DS} with a randomizing downsample matrix θ given by

$$\theta = \begin{bmatrix} \omega_{1,1} & \omega_{1,2} & \cdots & \omega_{1,d+1} & & & \cdots & 0 & \cdots \\ 0 & & \cdots & \omega_{2,d+1} & \omega_{2,d+2} & \cdots & & & \\ 0 & \vdots & 0 & 0 & \vdots & & \omega_{3,2d+1} & \omega_{3,2d+2} & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ddots \end{bmatrix} \tag{45}$$

where each of the $\omega_{i,j}$ entries in the matrix above is an i.i.d. random variable drawn from a Gaussian or $\pm$ Bernoulli distribution [24]; e.g., see Figure 49. The key idea is that if the dot product between the columns of the matrix $D \triangleq \theta\Psi$ is very low, then in principle it should be possible to recover the signal x in the basis in which it is sparse because it is unlikely that an element from x will be mistakenly identified with more than one column of D. In general, the number of random nonzero

entries in each row of θ needs to be greater than or equal to the downsample factor d to ensure that all of the full-rate sampling phases of the signal are represented in the measurement matrix θ; that is,

$$\left\| \left[\theta^T\right]_i \right\|_0 \geq d \text{ for } i = 1, 2, \cdots, \left\lfloor \frac{N}{d} \right\rfloor, \tag{46}$$

where the notation $[X]_i$ represents the ith column of matrix X, the superscript ^{T} is the matrix transpose operator and $\|x\|_0$ is the so-called ℓ_0-norm of the vector x.

5.2 A2I VIA BASIS PURSUIT

Constrained ℓ_1-programming techniques (basis pursuit) employed in compressed sensing are remarkably efficient in identifying and reconstructing signals that occupy sparse and random locations in the orthogonal basis in which the signals are compressible. However, for signals that occupy any more than a very small fraction of contiguous basis positions, the basis pursuit performance bound offers no guarantee on identification and reconstruction performance. In the following paragraphs we will outline the basis pursuit performance bound as first presented in [24] and [25], an then offer an interpretation of the bound in the context of identifying not-so-sparse signals at the output of a downsampling A2I receiver.

Consider a sparse signal occupying unknown frequency locations in some bandwidth B that is digitized at a rate $2B/d$. The digitized signal can then be described by

$$y = Dx + v \tag{47}$$

where $v \in \mathbb{R}^{\lfloor \frac{N}{d} \rfloor \times 1}$ is the noise with power inequality $E\{v^T v\} \leq \eta$, and $E\{\cdot\}$ is the expectation operator. In [1] it was demonstrated that it is possible to identify and reconstruct the sparse signal x by solving the basis pursuit (BP) problem formulated as

$$\min ||x||_1 \ s.t. \ \|y - Dx\|_2^2 \leq \eta. \tag{48}$$

The successful recovery of x by BP is guaranteed [25] if the following conditions of the matrix D obeys the uniform uncertainty principle. Defining $\Lambda \subset \{1, \cdots N\}$ and $[D]_\Lambda$ to consist of the columns of D indexed by Λ, then the local isometry constant $\delta_\Lambda(D)$ is the smallest number satisfying

$$(1 - \delta_\Lambda(D))||x_\Lambda||_2^2 \leq \| [D]_\Lambda x_\Lambda\|_2^2 \leq (1 + \delta_\Lambda(D))||x_\Lambda||_2^2, \tag{49}$$

where $x_\Lambda \in \mathbb{C}^{|\Lambda|}$ and the global restricted isometry constant is defined as

$$\delta_q(D) \triangleq \sup_{|\Lambda|=q} \delta_\Lambda(D) \tag{50}$$

with $|\Lambda|$ denoting the cardinality of the set Λ and q denoting the number of nonzero terms in the vector x. One can show that it is possible to recover an estimate of signal x to within a Euclidean distance no greater than $C\eta$ [24] where C is a constant if

$$\delta_{3q}(D) + 3\delta_{4q}(D) < 2. \tag{51}$$

An interpretation of the equations (49) through (51) above is that if the downsampling factor d is increased multiplicatively by some factor α, then the recovery of the signal x to within some Euclidean distance $C\eta$ is now only guaranteed if x spans 4α fewer positions in the basis Ψ . As an example, consider the toy problem of a signal received without noise, i.e., $v = 0$, where it should theoretically be possible to perfectly recover the signal x if the conditions as expressed in (51) are met. With λ denoting an eigenvalue of the matrix $\left([D]_\Lambda^H [D]_\Lambda\right)$, then from (49) it is easy to see that

$$\lambda_{\min} \leq \frac{\| [D]_\Lambda x_\Lambda \|_2^2}{\|x_\Lambda\|_2^2} \leq \lambda_{\max} \tag{52}$$

so that $\delta_{4q}(D)$ is bounded from below by the value 1 when $\lambda_{min} = 0$ because $4q > \lfloor \frac{N}{d} \rfloor$ which violates the inequality of equation (51). In other words, even with no noise present there is no guarantee that BP can identify and reconstruct the signal x if the number of nonzero values in the basis Ψ is greater than $\frac{1}{4}\lfloor \frac{N}{d} \rfloor$ under the most ideal of conditions. As we will shortly show, this will be a limiting factor in BP identifying and reconstructing signals that occupy only a fraction of the locations in the downsampled Nyquist band.

5.3 A2I VIA MAXIMUM LIKELIHOOD TECHNIQUES

In this section we will derive a maximum likelihood bound for identifying and reconstructing signals that occupy a much larger number of contiguous basis locations than BP has a guaranteed bound for. In doing so, we will set the stage for the derivation and application of Variable Projection and Unfolding presented in Section 3.1 to not-so-sparse signals at the output of a sub-Nyquist sampling A2I receiver.

Let $s = \{s_1, s_2, \cdots, s_q\}$, $s^c = \{s_{q+1}, s_{q+2}, \cdots, s_N\}$ where $s_i \in \{1, 2, \cdots, N\}$, $s_i \neq s_j$, and $D = [[D]_s \mid [D]_{s^c}]$ where $[D]_s \in \mathbb{C}^{\lfloor \frac{N}{d} \rfloor \times q}$ is the matrix whose columns are spanned by the signal(s) in the basis passband[1] with $\lfloor \frac{N}{d} \rfloor \geq q$. Then the received signal vector can be expressed as

$$y = [D]_s x_s + [D]_{s^c} x_{s^c} + Dv, \tag{53}$$

where $x_s \in \mathbb{C}^{q \times 1}$ is the signal energy in the passband, $x_{s^c} \in \mathbb{C}^{(N-q)\times 1}$ is the signal energy in the stop band, and $v \in \mathbb{C}^{N \times 1}$ is the receiver noise. It is easy to see from equation (53) that if the matrix D is

[1] We are liberally using the term 'passband' to denote the locations in the basis where signal energy is principally confined.

orthonormal, i.e., perfect mutual incoherence holds, then the signal x_s may be recovered from y in the presence of noise by matched filtering, i.e., when $([D]_s)^H y = x_s + v_s$. However, downsampling will generate a matrix D with more columns than rows; therefore, all of the columns will never be perfectly orthogonal. In [26] and [27], reconstruction performance is directly related to the degree of coherence between the columns in D. We will take a statistical signal processing viewpoint of reconstruction performance under the following conditions: the noise is independent and normally distributed as $v \sim \mathcal{N}(0, \sigma_v^2 I)$, and the stopband signal energy is negligible, e.g., $x_{s^c} \approx 0$. Then the maximum likelihood estimator (MLE) is given by

$$\begin{aligned} &\hat{x}_s^{MLE} = \left([D]_s^H C_v^{-1} [D]_s\right)^{-1} [D]_s^H C_v^{-1} y \; = \\ &x_s + v_s + \underbrace{\left([D]_s^H C_v^{-1} [D]_s\right)^{-1} [D]_s^H C_v^{-1} [D]_{s^c} v_{s^c}}_{Folding\ Loss}, \end{aligned} \tag{54}$$

where the covariance $C_v = D(\sigma_v^2 I) D^H$. Equation (54) consists of three terms: the desired signal, the noise at the locations in the basis that the signal occupies, and a folding loss which corresponds to noise outside these locations *folding* back in. The folding loss is the penalty that is paid in A2I for using reduced rate sampling and ML reconstruction. Notice that the degree of the folding loss is directly related to the coherence between the columns in the matrix D; as the cosine of the angle between the columns of $[D]_s$ and $[D]_{s^c}$ approaches zero, so does the folding loss. The MLE formulation in equation (54) may significantly amplify some of the noise components. Another cost function, the minimum mean square error MMSE [28], is often used to balance the quality of the estimate against the potential for noise amplification. Using equation (53) with $x_{s^c} \approx 0$ and $x_s \sim \mathcal{N}(0, C_{x_s})$, e.g., the signal is subject to frequency selective Rayleigh fading, then the MMSE estimator is given by

$$\hat{x}_s^{MMSE} = \left([D]_s^H \, C_v^{-1} \, [D]_s + C_{x_s}^{-1}\right)^{-1} [D]_s^H \, C_v^{-1} y. \tag{55}$$

Using the matrix inversion lemma, it is easy to show that (55) is equivalent to

$$\begin{aligned} \hat{x}_s^{MMSE} \;\; = \;\; & x_s + v_s + \underbrace{\left([D]_s^H C_v^{-1} [D]_s\right)^{-1} [D]_s^H C_v^{-1} [D]_{s^c} v_{s^c}}_{Folding\ Loss} \\ & - \underbrace{\tilde{D}^{-1} (\tilde{D}^{-1} + C_{x_s})^{-1} \tilde{D}^{-1} [D]_s^H C_v^{-1} ([D]_s x_s + Dv)}_{MMSE\ Folding\ Bias}, \end{aligned} \tag{56}$$

where $\tilde{D} = [D]_s^H \, C_v^{-1} \, [D]_s$.

The folding loss associated with ML/MMSE identification and reconstruction is illustrated in Figure 50.

The scenario illustrated in Figure 50 employed a measurement matrix θ with entries randomly chosen from a $\pm$ Bernoulli distribution with signals whose basis coefficients covered random but

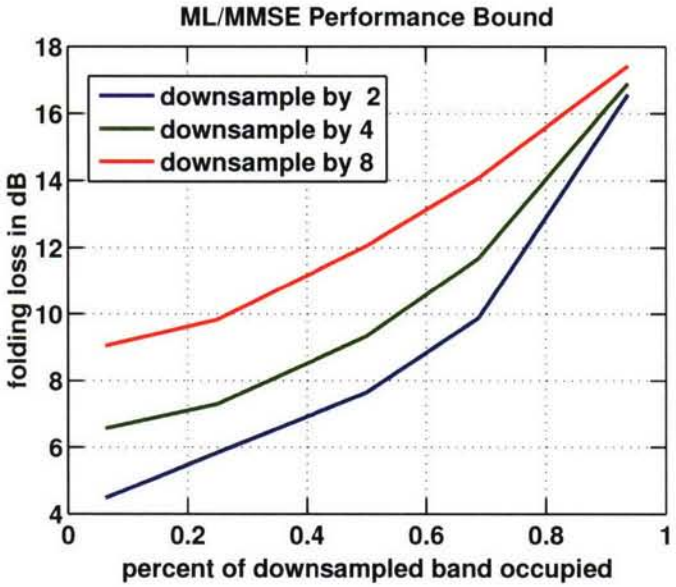


Figure 50: Folding loss associated with an ML/MMSE receiver. Note that noise folding loss corresponds to an overall loss in SNR as a function of the signal's bandwidth (sparse in frequency) and the downsampling factor.

contiguous positions in frequency. The folding loss in ML/MMSE identification and reconstruction corresponds to the loss in SNR due to random linear projections followed by an undersampling (sub-Nyquist) ADC. The folding loss may be explained in two ways: first, as explained previously, there is a loss due to noise from ambiguous basis locations folding back onto the basis locations that the signal occupies. This is reflected by the fact that folding loss is greater when the downsampling is more aggressive, e.g., from a factor of 2 to a factor of 8. Second, as the bandwidth of the signal increases, the variability in the frequency response of the random linear projections makes it more likely that there will be noise amplification in the identification and reconstruction process. This is illustrated in Figure 50, which shows that as the bandwidth increases, noise amplification, as reflected by the change in folding loss, also increases. The choice of an ML or MMSE estimator is dependent on the signal bandwidth; at one extreme, narrowband spikes (tones) are best reconstructed using an ML estimator, while a wider band signal should be identified and reconstructed using an MMSE estimator to obtain the best trade between noise amplification and reconstruction performance.

5.4 COVARIANCE ESTIMATION

The ability to estimate signal parameters is often well characterized by Cramer-Rao bounds. For compressed sensing applications, these bounds have little direct relevance since estimation occurs on the basis of a single vector observation. Mixed signal components are extracted using models

of the observations that rely on known signal bases. When more data are available, incoherent integration of the additional samples can support better estimation of signal parameters that include powers and even signal bases themselves. It is also well known that model-based parameter estimates are sensitive to modeling errors. These errors can have less deleterious effects on signal statistics such as covariances. Covariances support estimates of signal powers and signal bases (more precisely, signal subspaces).

Below, some initial results on the estimation of signal covariances are presented. The quality of parameter estimates is characterized by Cramer-Rao bounds on estimation error. The effects of different amounts of signal knowledge are discussed as is the impact of undersampling on estimator performance.

Let us assume that the Cramer-Rao bound is formulated using a known signal basis with unknown signal powers. The model of the data is expressed by the length m observed vector $y_k = A_k z_k$, given the known $m \times n$ matrix A_k and the unobserved data of length n expressed by the vector z_k. Here, k is the sample number. Each sample vector z_k is assumed to have the covariance R. The matrices A_k are assumed to vary randomly from one vector sample to the next. We assume that the $n \times n$ covariance R is diagonal. For the moment, assume in addition that the $s \ll m$ diagonal entries associated with signals are represented by known (e.g., the first) s diagonal components of R. Let e_{jk} denote the matrix whose sole nonzero entry is unity in the $(j,k)^{th}$ component. Then the approximate Cramer-Rao bound matrix for the coordinates $\{e_{jj}\}$ becomes, in the limit of large m and n

$$\frac{1}{2l}\left(\frac{n}{m}\right)^2 I_s, \tag{57}$$

where l is the number of vector samples. When the location of the signal components is unknown, the result changes only slightly, resulting in the approximate Cramer-Rao bound matrix

$$\frac{1}{2l}\left(\frac{n}{m}\right)^2 I_n. \tag{58}$$

The terminology "approximate" reflects the sparseness of the true covariance ($s \ll m$) and the asymptotic nature of the expressions (large n and m).

A few comments on these results are in order. First, it is clear that the effect of downsampling the data by a factor of n/m results in an increase in the required sample support by a factor of $(n/m)^2$ in order to maintain the same level of performance. Second, performance is independent of the signal basis, as long as this basis is known to the receiver. Third, the additional knowledge of the active signal components does not improve estimates of signal powers. It should be recalled that these results assume that the signal model is sparse ($s \ll m$).

Other aspects of covariance estimation can also be treated. For example, signal bases can be estimated. These estimates can be shown to decouple from those of the signal powers. A brief treatment of basis estimation will be presented in future drafts.

It is interesting to examine the effect of *a priori* information on the performance of covariance estimation. The random parameter Fisher matrix [29] can be used to assess the benefits of additional information. In particular, this approach can be applied to the exponential priors associated with L_1 penalty functions used in conjunction with maximum likelihood parameter estimators. Results indicate some benefits from prior information in the sense that the required sample support can be reduced somewhat. These results will also be reported in a future draft.

Cramer-Rao Bounds When the environment is quasistationary it is possible to characterize the performance of certain estimators in terms of Cramer-Rao lower bounds on mean-squared errors of parameter estimates. Specifically, consider vector-valued Gaussian data with probability density function

$$p(z|R) = \pi^{-n}|R|^{-1}e^{-z^\dagger R^{-1}z}. \tag{59}$$

We assume that the data obeys a complex, circular Gaussian distribution, which is discussed in many references (see [30] for a self-contained discussion of these statistics as well as complex versions of some standard distributions required below). When more than a single observation of z is available, the covariance R can be estimated and used for detection, identification, and reconstruction. The benefits of stationarity are significant. For example, the model-based techniques used with a single observation can be sensitive to modeling errors that are less of a problem with covariance-based estimators. Sparsity still plays a role in characterizing performance and in the structure of the estimator, but the theory is less developed in this case.

The performance of covariance-based estimators is characterized below in terms of Cramer-Rao bounds. Sparsity is assumed in order to arrive at a simple approximation for the bounds. Of particular interest is the loss in performance due to downsampling the observations. Some variants of the bounds are discussed as they relate to different signal hypotheses. Although Cramer-Rao lower bounds are only achievable in an asymptotic sense, they are typically approached with small sample support (i.e., small numbers of observations) given good estimators, as long as the appropriate SNR's are above certain thresholds. Determining these thresholds is a much more difficult problem, but for many practical examples, the thresholds are in operationally interesting regions.

For our applications, the data are observed indirectly through an $m \times n$ matrix A in the sense that only $y = Az$ is observed, then the Fisher matrix for covariance parameters is given by (defining $R_A \stackrel{\text{def}}{=} ARA^\dagger$)

$$F = 2l\mathrm{E}[\mathrm{tr}(R_A^{-1}\dot{R}_A R_A^{-1}\dot{R}_A R_A^{-1} yy^\dagger)]. \tag{60}$$

The variable l denotes the number of vector-valued observations. The notation $\dot{R}_A$ is short for $AXA^\dagger$ with X a tangent vector to R (conceptually, $X = \dot{R}$). Note that the covariances R lie in the real n^2 dimensional vector space of Hermitian matrices so that all tangent vectors can be viewed as elements of a real vector space. Expectation (denoted $\mathrm{E}[\cdot]$) is taken over the random variates A and

y. The distribution of y, conditioned on A, is given by $p(y|R_A)$ in the notation of equation (59). Defining $y^{(w)} \stackrel{\text{def}}{=} R_A^{-1/2} y$, and noting that the conditional expectation of $y^{(w)}$ is mean-zero complex circular Gaussian with covariance I_m, we can reexpress the Fisher as

$$2l\mathrm{E}[\ \mathrm{E}[\mathrm{tr}(R_A^{-1/2}\dot{R}_A R_A^{-1}\dot{R}_A R_A^{-1/2} y^{(w)} y^{(w)\dagger})\ |A]] = 2l\mathrm{E}[\mathrm{tr}(R_A^{-1}\dot{R}_A R_A^{-1}\dot{R}_A)]. \tag{61}$$

The measure on the $m \times n$ matrix A is assumed to be invariant under unitaries acting on the left or right of A. For example, we can assume that all entries of A are i.i.d. complex circular Gaussians with unity complex covariance. Define $T \stackrel{\text{def}}{=} (AA^\dagger)^{-1/2} A$. We assume $m \leq n$ so that the rows of T form an orthonormal basis for the row space of A. In particular, $TT^\dagger = I_m$. We can rewrite the Fisher in terms of the induced measure μ on T as

$$\langle X, Y \rangle_R = F(X, Y) = 2l \int \mathrm{tr}[(TRT^\dagger)^{-1}(TXT^\dagger)(TRT^\dagger)^{-1}(TYT^\dagger)]\, d\mu(T). \tag{62}$$

This integral emphasizes the fact that only the statistics of the row space of A matter in the Fisher. Note that the invariance properties of μ imply

$$\langle UXU^\dagger, UYU^\dagger \rangle_{URU^\dagger} = \langle X, Y \rangle_R. \tag{63}$$

At this stage we assume that the true covariance matrix R is sparse. Specifically, sparseness is assumed to be equivalent to a representation of R as

$$U^\dagger R U = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} + I_n, \tag{64}$$

given the $s \times s$ matrix C, with $s \ll m$. Here U is a unitary matrix. If we write $TU = (T_1\ T_2)$ so that $TRT^\dagger = I_m - T_1 C T_1^\dagger$, then

$$(TRT^\dagger)^{-1} = I_m - T_1(C^{-1} + T_1^\dagger T_1)^{-1} T_1^\dagger \tag{65}$$

and hence

$$I_m \geq (TRT^\dagger)^{-1} \geq I_m - P_{T_1}, \tag{66}$$

where $P_{T_1} = T_1(T_1^\dagger T_1)^{-1} T_!^\dagger$ is the projector onto the column space of T_1, provided $s \leq m$. This inequality can be used to bound the integrand of equation (62), obtaining

$$\begin{aligned}
\mathrm{tr}(TXT^\dagger TXT^\dagger) &\geq \mathrm{tr}((TRT^\dagger)^{-1}(TXT^\dagger)(TRT^\dagger)^{-1}(TXT^\dagger)) \\
&= \mathrm{tr}((TXT^\dagger)^2) - 2\mathrm{tr}(P_{T_1}(TXT^\dagger)^2) \\
&\quad + \mathrm{tr}(P_{T_1}(TXT^\dagger) P_{T_1}(TXT^\dagger)) \\
&\geq \mathrm{tr}((TXT^\dagger)^2) - 2\mathrm{tr}(P_{T_1}(TXT^\dagger)^2).
\end{aligned}$$

Next, we use the bound

$$\mathrm{E}[\mathrm{tr}(P_{T_1}(TXT^\dagger)^2)] \leq \mathrm{E}[\lambda_{\max}^2(X)\mathrm{tr}(P_{T_1})] = \lambda_{\max}^2(X)\frac{s}{m} = O(\frac{s}{m}) \tag{67}$$

to conclude that

$$\langle X, Y\rangle_R = \langle X, Y\rangle_{I_n} + O(\frac{s}{m}). \tag{68}$$

A sharper result can be conjectured based on the weak dependence of P_{T_1} and $(TXT^\dagger)$, namely

$$\langle X, Y\rangle_R = \langle X, Y\rangle_{I_n} + O(\frac{sm}{n^2}). \tag{69}$$

A proof seems rather technical and the sharper result does not help us here.

The inner product $\langle X, Y\rangle_{I_n} \equiv \langle X, Y\rangle$ is invariant under unitaries:

$$\langle UXU^\dagger, UYU^\dagger\rangle = \langle X, Y\rangle. \tag{70}$$

All real, symmetric, invariant inner products on Hermitian matrices have the form

$$\langle X, Y\rangle = a\mathrm{tr}[(X - \frac{\mathrm{tr}X}{n}I_n)(Y - \frac{\mathrm{tr}Y}{n}I_n)] + b\frac{\mathrm{tr}X}{n}\cdot\frac{\mathrm{tr}Y}{n} \tag{71}$$

for nonnegative coefficients a, b. This fact is well known. For completeness, we present a self-contained argument that can be skipped. The slick argument notes that the Hermitian matrices of trace zero are orthogonal to the matrices that are multiples of the identity matrix and both subspaces, that of trace zero matrices and that of multiples of the identity, are separately invariant. The subspace of trace zero Hermitian matrices, when multiplied by the square root of -1, can be identified with the skew-adjoint matrices that form the Lie algebra of the Lie group $SU(n)$ of unitary matrices with unity determinant. The action $U(iX)U^\dagger$ of $SU(n)$ on its Lie algebra is the so-called adjoint representation, which is known to be irreducible (no nontrivial invariant subspaces). For convenience and more generality, let the notation $g \cdot X = gX$ denote the action of a group element g on a vector space element X. In our context, if g represents the unitary U, then gX corresponds to $UXU^\dagger$. The relationship between invariant inner products and irreducible representations is cemented by Schur's lemma as follows. Assume we are given two invariant inner products $\langle\cdot,\cdot\rangle_k$ for $k = 1, 2$. Also assume that the $k = 2$ inner product is nondegenerate. Then there exists a linear operator ϕ such that $\langle X, \cdot\rangle_1 = \langle\phi(X), \cdot\rangle_2$. But $\langle gX, \cdot\rangle_1 = \langle\phi(gX), \cdot\rangle_2$ and

$$\langle gX, \cdot\rangle_1 = \langle X, g^{-1}\cdot\rangle_1 = \langle\phi(X), g^{-1}\cdot\rangle_2 = \langle g\phi(X), \cdot\rangle_2, \tag{72}$$

where the first and last equalities follow from invariance of the inner products. Since the $k = 2$ inner product is nondegenerate, we have $\phi(gX) = g\phi(X)$. By Schur's lemma and the fact that the action by g is irreducible, we have that $\phi = c\,Id$. In other words, ϕ is a multiple of the identity. Thus $\langle\cdot,\cdot\rangle_1 = c\langle\cdot,\cdot\rangle_2$. The nondegeneracy of $\mathrm{tr}(XY)$, restricted to traceless Hermitian matrices is

easily shown and this fact completes the proof that $\mathrm{tr}(XY)$ is a unique invariant inner product on these matrices, up to a scalar factor.

The unknown coefficients a, b can be identified by plugging in specific matrices into the inner product. By letting $X = I_n$, we find that $b = m$. The remaining coefficient is more difficult to evaluate. To proceed, let

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{73}$$

with scalar X_1. In addition, let $E_m = (I_m \, 0_{m,n-m})$ and express $T = E_m U$ for some unitary U. Then

$$TXT^\dagger TYT^\dagger = E_m U X U^\dagger E_m^\dagger E_m U Y U^\dagger E_m^\dagger. \tag{74}$$

Let $U = (U_1 \, U_2)$ with U_1 representing the first column of U, so that $UX = (U_1 X_1 \, 0_{n,n-1})$. Then

$$U^\dagger E_m^\dagger E_m U = \begin{pmatrix} U_1^\dagger E_m^\dagger E_m U_1 & U_1^\dagger E_m^\dagger E_m U_2 \\ U_2^\dagger E_m^\dagger E_m U_1 & U_2^\dagger E_m^\dagger E_m U_2 \end{pmatrix}. \tag{75}$$

Putting the notation together, equation (74) can be written

$$E_m U_1 X_1 U_1^\dagger E_m^\dagger E_m U_1 X_1 U_1^\dagger E_m^\dagger \tag{76}$$

with trace $\|E_m^\dagger U_1\|^4 X_1^2$. Thus the integral equation (62) becomes

$$2l \int \mathrm{tr}[(E_m U X)(U^\dagger E_m^\dagger E_m U)(X U^\dagger E_m^\dagger)] \, d_{\mathrm{vol}} U = 2l X_1^2 \int \|E_m^\dagger U_1\|^4 \, d_{\mathrm{vol}} U. \tag{77}$$

The measure on the unitaries is normalized Haar measure. Conceptually, if we form an $n \times n$ matrix filled with i.i.d., mean zero, complex circular Gaussian random variates, then the Q factor in a QR decomposition will have the homogeneous density of the normalized Haar measure. In particular, the first column U_1 of U has a distribution modeled by that of a complex, circular Gaussian random vector of length n with i.i.d. components, normalized by the L_2 norm of the same vector. To evaluate the integral, we need to identify it with the complex version of a canonical random variate. Let the random variables $\{w_k\}$ have i.i.d. complex circular Gaussian distributions with zero mean. Then the random variable associated with

$$\frac{\sum_{k=1}^{\nu} |w_k|^2}{\sum_{k=1}^{\nu+\mu} |w_k|^2} \tag{78}$$

is a complex beta random variable with density [30]

$$p(\beta; \nu, \mu) = \frac{(\nu + \mu - 1)!}{(\nu - 1)!(\mu - 1)!} \beta^{\nu-1} (1 - \beta)^{\mu-1}. \tag{79}$$

Furthermore, $\|E_m^\dagger U_1\|^2$ has the same distribution as this beta variate with parameters $\mu = m$, $\nu = n - m$, as the unitary U varies uniformly over all unitaries (normalized Haar measure). Thus

$$\rho \stackrel{\text{def}}{=} \mathrm{E}[\|E_m^\dagger U_1\|^4] = \int \beta^2 p(\beta; m, n-m)\, d\beta = \frac{(m+1)m}{(n+1)n}. \tag{80}$$

Finally, we use the value of ρ to solve for a:

$$\langle X, X\rangle = X_1^2 \rho = a X_1^2 [1 - \frac{1}{n}] + m\frac{X_1^2}{n^2} \tag{81}$$

so that

$$a = \frac{n}{n-1}\left(\rho - \frac{m}{n^2}\right) = \frac{m(mn-1)}{n^2(n+1)}. \tag{82}$$

Summarizing,

$$\begin{aligned}\frac{\langle X, Y\rangle}{2l} &= \int \mathrm{tr}[TXT^H\, TYT^H]\, d\mu(T) \\ &= \tfrac{m(mn-1)}{n^2(n+1)} \mathrm{tr}\left[\left(X - \tfrac{\mathrm{tr}X}{n} I_n\right)\left(Y - \tfrac{\mathrm{tr}Y}{n} I_n\right)\right] + m\tfrac{\mathrm{tr}X}{n}\tfrac{\mathrm{tr}Y}{n}.\end{aligned}$$

Let us assume that the Fisher matrix is based on a known signal basis with unknown signal powers. Without loss of generality (due to the invariance of the inner product), we can assume that the covariance R is diagonal. For the moment, assume in addition that the s diagonal entries associated with signals are represented by the first s diagonal components of R. If e_{jk} denotes the matrix whose sole nonzero entry is the $(j,k)^{th}$, then the approximate Fisher matrix for the coordinates $\{e_{jj}\}$ becomes

$$(\langle e_{jj}, e_{kk}\rangle) = 2l\frac{m(mn-1)}{n^2(n+1)}\left[I_s + \frac{n(n-m+1)+1}{n(nm-1)} 1_s 1_s^\dagger\right], \tag{83}$$

where 1_s is the vector of all ones. Let $m = \lambda\mu$ and $n = \lambda\nu$. Then the inverse of this approximate Fisher matrix, in the limit of large λ becomes

$$\frac{1}{2l}\left(\frac{\nu}{\mu}\right)^2 I_s = \frac{1}{2l}\left(\frac{n}{m}\right)^2 I_s. \tag{84}$$

When the location of the signal components is unknown, the argument changes only slightly, resulting in the inverse of the approximate Fisher (for large λ)

$$\frac{1}{2l}\left(\frac{n}{m}\right)^2 I_n. \tag{85}$$

5.5 A GENERALIZATION OF BASIS PURSUIT

5.5.1 Summary of Results

Basis pursuit, in the noiseless case, is formulated as the L_1 minimization problem

$$\min_{\Phi f = g} \|f\|_1, \tag{86}$$

which can be solved by linear programming. If $\Phi f = g$ has a sufficiently sparse solution, then basis pursuit will find that solution, avoiding what would normally be an intractable combinatorial search. A degree of sparseness that guarantees success for basis pursuit has been calculated, in an asymptotic limit, in [31].

The asymptotic upper bound on sparseness is pessimistic, given the simulated performance of related algorithms in noise. Furthermore, the limiting regime where the bound applies does not always accurately model important applications. For example, consider a fixed receiver band sparsely occupied by finite bandwidth signals. As the sample support increases, we would like to hold the number of signals fixed as the number of samples per signal increases due to increased frequency resolution. Instead, the asymptotic bound would increase the number of signals (to maintain fixed spectral occupancy) while decreasing the bandwidth of individual signals.

A modified version of basis pursuit can handle a fixed number of finite bandwidth signals in the limit of increasing sample support. Abstractly, we partition the columns of the $p\lambda \times m\lambda$ matrix Φ in blocks of length λ. The blocks represent response vectors associated with the same signal. We can, as is convenient, interpret f as a $m \times \lambda$ matrix with j^{th} block the j^{th} row $f_{j\cdot}$. Then the modified form of basis pursuit becomes

$$\min_{\Phi f = g} \sum_k \|f_{k\cdot}\|_2. \tag{87}$$

In words, the sum of the L_2 norms of the blocks of f is minimized under the given equality constraint. The objective function is L_2 within blocks and L_1 between them. The solution can be found by solving a quadratic programming problem.

A potential advantage of the quadratic programming approach, in applicable signal models, is a much larger bound on the fraction of the band that can be handled. Shown in Figure 51 is the fraction of the band that can be perfectly reconstructed from noiseless observations under two different types of signal models. The random model comes from [31]. It is valid in the limit of large receiver bandwidth or in the limit of vanishing signal bandwidth, given fixed receiver bandwidth. In effect, each signal is sampled once, corresponding to the basis vector associated with the signal. A more common situation occurs when the receiver and signal bandwidths are fixed while the observation time increases. In this case, the number of samples per signal increases with the observation time. As a result, performance improves substantially as indicated in the figure. The band occupied by the signals is unknown, so reconstruction remains a potentially combinatorial

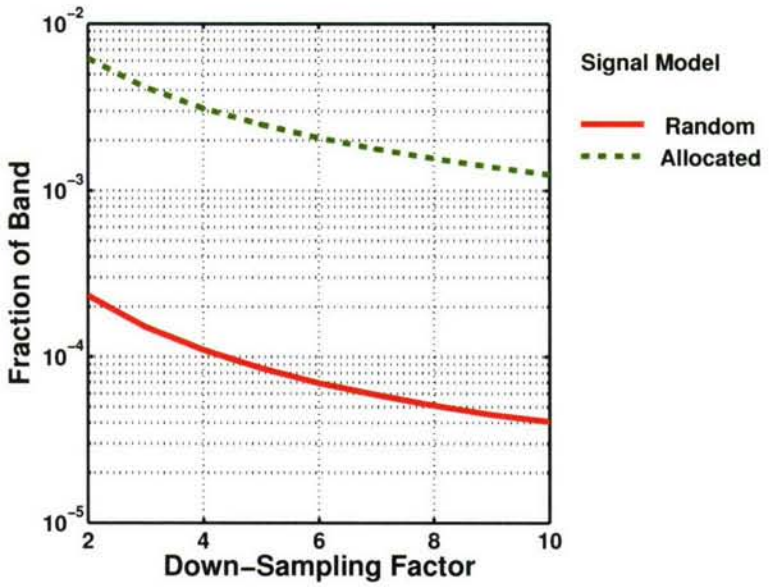


Figure 51: Two types of asymptotic bounds on the fractional band occupancy that permits perfect reconstruction are shown. The random bound is consistent with signal models that have fixed bandwidth in the limit of increasing receiver bandwidth, or with models that have decreasing signal bandwidth in the limit of increasing observation time with fixed receiver bandwidth. The allocated bound pertains to signals with fixed bandwidth in the limit of increasing observation time and fixed receiver bandwidth. These signals can be viewed as allocated to certain subbands that are not known a priori. *With the allocated model, as observation time increases, so does the number of samples per signal. In contrast, the random model maintains a single sample per signal in any limit.*

problem; but the quadratic program of equation (87) can be used instead of a combinatorial search to achieve exactly the same result in the noiseless case.

At the end of Section 5.5.2 some comparisons, with and without noise, are made between different versions of basis pursuit.

5.5.2 Gory Details

The proof that equation (87) can recover a sufficiently sparse signal follows in detail the argument of [31]. The ancillary results needed are stated below; their proofs are essentially the same as the special cases stated in [31] and are not presented here. We will show that equation (87) follows as a consequence. Since the argument closely mirrors that in [31], we will use, for the most part, the notation in that paper.

All vectors (sometimes called blocked vectors) c are assumed have length $m\lambda$. Their components are grouped in consecutive blocks of length λ. It is convenient to index the components as c_{jk}, where j is the block number and k the component in the block. The notation $c_{j\cdot}$ refers to the vector of length λ consisting of the components in the j^{th} block. A *block subset* T of the indices $\{1, 2, \ldots, m\}$ is a subset of all block indices. The length of T is denoted $|T|$. If Φ is a $p\lambda \times m\lambda$ matrix, the notation Φ_T denotes the matrix of size $p\lambda \times |T|\lambda$ consisting of the $|T|\lambda$ columns with block indices in the subset T. Similarly, c_T denotes the vector of length $|T|\lambda$ supported on the blocks indexed by T. Denote by $b(j) = \{(j-1)\lambda + 1, \ldots, j\lambda\}$ the set of indices in the j^{th} block. Let $\Phi_{b(j)}$ consist of the columns of Φ in the j^{th} block.

The Euclidean inner product is denoted $\langle \cdot, \cdot \rangle$ with associated Euclidean norm $\| \cdot \|$. For convenience, we let $\langle w, \Phi_{b(j)} \rangle$ denote the vector of length λ whose components consist of the inner product between w and columns of $\Phi_{b(j)}$. A vector c is said to have block support T if its non-zero components exactly span the blocks indexed by T.

Some of the definitions of [31] are slightly altered to accommodate the block structure of the signals.

Definition 1 *Define δ_S as the infimium of all δ satisfying*

$$(1-\delta)\|c\| \leq \max_{|T| \leq S} \|\Phi_T c\| \leq (1+\delta)\|c\| \tag{88}$$

where T is any block subset of length at most S and c is a blocked vector of the appropriate size.

Only cases where $\delta_S < 1$ are of interest.

Definition 2 *Define $\theta_{S,S'}$ as the infimium of all θ satisfying*

$$\max_{|T| \leq S, |T'| \leq S'} \langle \Phi_T c, \Phi_{T'} c' \rangle \leq \theta \|c\| \|c'\| \tag{89}$$

where T (T') is any block subset of length at most S (S').

We briefly state some of the lemmas of [31] in the slightly more general form needed below. The second lemma follows from the first by an essentially equivalent argument. For convenience, we let $\theta_S \equiv \theta_{S,S}$.

Lemma 1 *Assume that $\delta_S < 1$. There exists a vector w of length $m\lambda$ and block length λ such that $\langle w, \Phi_{b(j)} \rangle = c_{j\cdot}$, $j \in T$, where T is a block subset consisting of at most S blocks and c is a blocked*

vector supported on T. If $S + S' \leq m$, there exists an exceptional block set E, with block length $|E| \leq S'$, such that

$$\begin{aligned} \|\langle w, \Phi_{b(j)} \rangle\| &\leq \frac{\theta_{S,S'}}{(1-\delta_S)\sqrt{S'}} \|c\|, \quad j \notin T \cup E \\ \left(\sum_{j \in E} \|\langle w, \Phi_{b(j)} \rangle\|^2 \right)^{1/2} &\leq \frac{\theta_{S,S'}}{1-\delta_S} \|c\|, \quad j \in E \\ \|w\| &\leq K\|c\|. \end{aligned}$$

The following is the result of whittling away at the exceptional set above.

Lemma 2 *Assume that $\delta_S + \theta_{S,2S} < 1$. With c as above, there exists a w satisfying $\langle w, \Phi_{b(j)} \rangle = c_{j\cdot}$, $j \in T$, and*

$$\|\langle w, \Phi_{b(j)} \rangle\| \leq \frac{\theta_S}{(1-\delta_S-\theta_{S,2S})\sqrt{S}} \|c\|, \quad j \notin T.$$

To see an immediate consequence of the last lemma, let c be supported on T with $c_{j\cdot}$ a unit vector for each $j \in T$. Then

$$\|\langle w, \Phi_{b(j)} \rangle\| \leq \frac{\theta_S}{1-\delta_S-\theta_{S,2S}}. \tag{90}$$

when $j \notin T$.

Assume that $\delta_S + \theta_{S,S} + \theta_{S,2S} < 1$. Let f be a blocked vector with support T, $|T| \leq S$, that solves $\Phi f = g$. Let f' solve equation (87). Define the blocked vector u with support T so that $u_{j\cdot} = f_{j\cdot}/\|f_{j\cdot}\|$. We can find a weight w such that $\langle w, \Phi_{b(j)} \rangle = u_{j\cdot}$, $j \in T$, and with $\|\langle w, \Phi_{b(j)} \rangle\| < 1$ for $j \notin T$. Then

$$\begin{aligned} \sum_j \|f'_{j\cdot}\| &= \sum_{j \in T} \|f_{j\cdot} + (f'_{j\cdot} - f_{j\cdot})\| + \sum_{j \notin T} \|f'_{j\cdot}\| \\ &\geq \sum_{j \in T} u_{j\cdot} \cdot (f_{j\cdot} + f'_{j\cdot} - f_{j\cdot}) + \sum_{j \notin T} \langle w, \Phi_{b(j)} \rangle \cdot f'_{j\cdot} \\ &= \sum_{j \in T} \|f_{j\cdot}\| + \sum_{j \in T} \langle w, \Phi_{b(j)} \rangle \cdot (f'_{j\cdot} - f_{j\cdot}) + \sum_{j \notin T} \langle w, \Phi_{b(j)} \rangle \cdot f'_{j\cdot} \\ &= \sum_{j \in T} \|f_{j\cdot}\| + \langle w, \sum_{1 \leq j \leq m} \Phi_{b(j)} f'_{j\cdot} - \Phi_T f_T \rangle \\ &= \sum_{j \in T} \|f_{j\cdot}\| + \langle w, \Phi f' - \Phi f \rangle \\ &= \sum_j \|f_{j\cdot}\|. \end{aligned} \tag{91}$$

Since f' minimizes equation (87) under the linear constraints, it follows that each inequality must be an equality and hence $f'_{j\cdot} = 0$ for $j \notin T$. Thus $\Phi_T(f_T - f'_T) = 0$ and hence $f = f'$ since Φ_T has a trivial null space.

To complete the argument, first note that $\theta_{S,S'} \leq \delta_{S+S'}$ is established in [31] and the same proof can be applied to the block versions of these variables. Next, note that [31] informs us that, in our context,

$$\text{Prob}\left(\sup_{T:|T|=S} \lambda_{\max} > (1 + \sqrt{S/p} + \eta_{p\lambda} + t)^2\right) \leq \binom{m}{S} e^{-p\lambda t^2/2}, \tag{92}$$

where $\lambda_{\max}$ is the principal eigenvalue of $\Phi_T^\dagger \Phi_T$. Since $\lambda_{\max} \leq 1 + \delta_S$, we have, with high probability as $\lambda \to \infty$,

$$\delta_S < -1 + \left(1 + \sqrt{S/p}\right)^2. \tag{93}$$

Let d denote the downsampling factor m/p and $r = S/m$ the fraction of the band occupied by signals. Define

$$\rho(r, d) \stackrel{\text{def}}{=} \sum_{k=1}^{3} -1 + \left(\sqrt{kr} \cdot \sqrt{d} + 1\right)^2 > \delta_S + \delta_{2S} + \delta_{3S}. \tag{94}$$

In the limit of large λ, a lower bound on the fraction of the band r that can be recovered by equation (87) is determined by $\rho(r, d) = 1$.

It's worth noting that the fixed bandwidth per signal could have been variable, but known in advance. In general terms, the procedure equation (87) is used to exploit *a priori* structural knowledge of signals.

5.5.3 The Noisy Case

Although no results are presented here, a potential practical algorithm for solving equation (87) can be based on equation (33). Define

$$\sigma_\mu(x) \stackrel{\text{def}}{=} \begin{cases} x - \frac{\mu}{2}\frac{x}{\|x\|} & \|x\| \geq \mu/2 \\ 0 & \|x\| < \mu/2 \end{cases} \tag{95}$$

for each block component x of a blocked vector. Extend this definition to all blocks using

$$S_\mu(x_{1\cdot}, \ldots, x_{m\cdot}) \stackrel{\text{def}}{=} (\sigma_\mu(x_{1\cdot}), \ldots, \sigma_\mu(x_{m\cdot})). \tag{96}$$

Then

$$f \leftarrow S_\mu(f + \Phi^T(g - \Phi f)). \tag{97}$$

provides a modified gradient algorithm for the mixed norm problem

$$\|\Phi f - g\|^2 + \mu \sum_k \|f_{k\cdot}\|. \tag{98}$$

That this is an appropriate modification of equation (33) for the new mixed norm problem (98) can be established using surrogate functions in much the same way that equation (33) is derived in [19]. We make no claims about convergence, but it seems reasonable to believe this is not a problem.

As an example of the performance of equation (98), consider a comparison with the usual form of basis pursuit, as shown in Figure 52. The example uses 100 samples for standard basis pursuit, which utilizes equation (98) for blocks of size one, and 500 samples for generalized basis pursuit utilizing blocks of size 5. No noise is present in this example. The figure demonstrates that signals represented by multiple basis vectors with a known pattern can potentially occupy larger fractions of the band than randomly located signals, each represented by a single basis vector. The probability of identification, in both cases, refers to the number of signals correctly detected and located. If both curves are based on 500 samples, the standard basis pursuit degrades slightly while the generalized version is, of course, unchanged. The latter comparison corresponds to the asymptotic bounds treated above.

When noise is present, the general conclusions drawn from Figure 52 still hold, as Figure 53 illustrates.

5.6 NOLAFF HYPOTHESIS TESTING PERFORMANCE BOUNDS

NoLAff hypothesis testing undersampling performance bounds are directly related to the performance of the two main algorithmic components of the technique; namely, Bayesian decision which underlies the hypothesis testing to resolve ambiguity and iterative nonlinear equalization (NLEQ) which undoes the nonlinear distortion of the NoLAff encoder. We note that while performance bounds for nonlinear processing are notoriously hard to come by, we can make a few simplifying assumptions which allow us to ignore this factor for most practical implementations.

In the low SNR input signal case, the nonlinear artifacts after equalization will be below the noise and hence ignorable. At the other extreme of very high SNR, we can safely assume that the equalized nonlinear artifacts are significantly lower than the signal of interest, albeit possibly stronger than the noise, and hence not a limitation to the ambiguity resolution. Assuming perfect ambiguity resolution, residual NLEQ distortion levels are typically well below the noise level of the ADC itself (See e.g., [32,33]).

We note as an aside that only the nonlinear artifacts that fall onto the support of the input signal itself concern us here in terms of linearization, since the other artifacts (prior to linearization) are in fact used for ambiguity resolution. Hence, the stronger (but still linearizable) the nonlinear

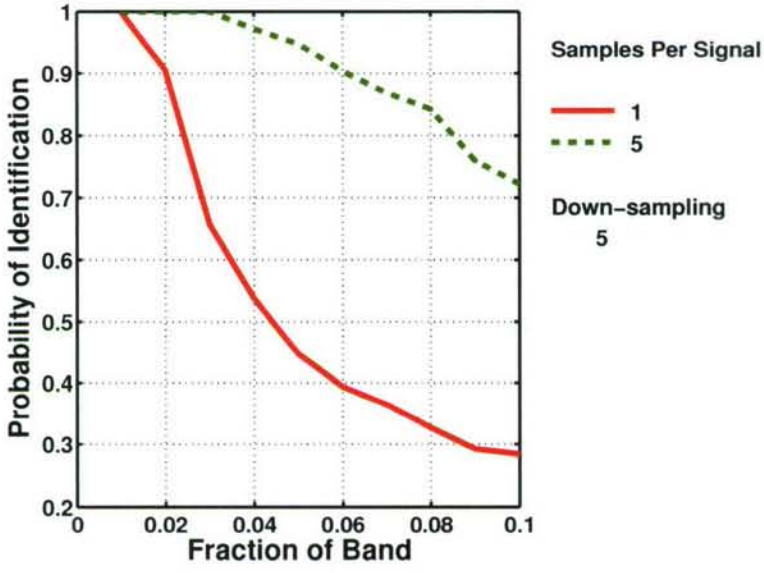


Figure 52: The standard form of basis pursuit associates one basis vector to each signal. When the signals are more naturally associated with multiple basis vectors in a known pattern, more general versions of basis pursuit can offer better performance given the same measure of sparsity. Shown above is a comparison of two cases: one with a single basis vector for each signal and another with 5 for each signal. By using more basis vectors, sparsity can be reduced by a factor of 3 to 4. The example has no noise.

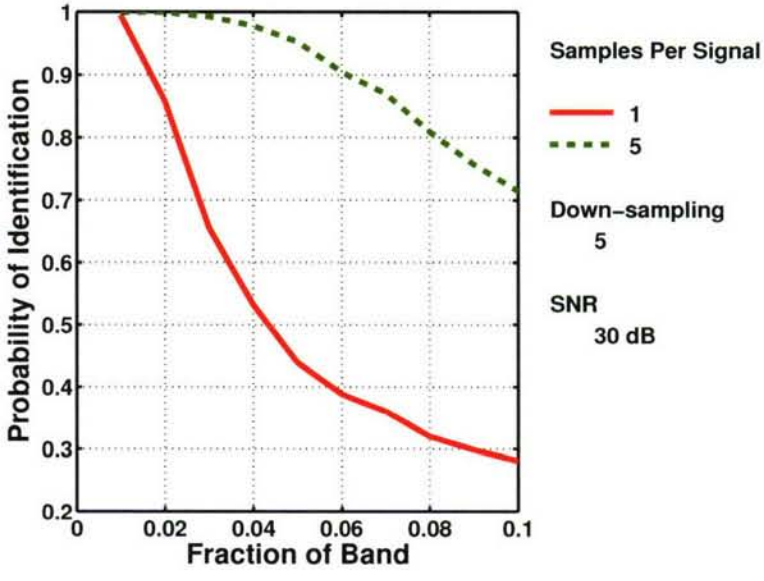


Figure 53: The standard form of basis pursuit associates one basis vector to each signal. When the signals are more naturally associated with multiple basis vectors in a known pattern, more general versions of basis pursuit can offer better performance given the same measure of sparsity. Shown above is a comparison of two cases: one with a single basis vector for each signal and another with 5 for each signal. By using more basis vectors, sparsity can be reduced by a factor of 3 to 4. All signals have an SNR of 30 dB.

distortions are, the better our ambiguity resolution will be. In practice a nonlinear distortion level of 10-15 dB below the signal is invertible.

Under the assumption that the signal can be recovered up to its undersampling ambiguity with SNR that is at least as good as that imposed by the ADC used to sample the NoLAff-encoded received signal, we are left with an analysis of the ambiguity resolution itself.

Let $\mathbf{y}$ denote the measured (undersampled) nonlinear distortions of the signal we wish to resolve. That is, the probe and its nonlinear artifacts with itself and the received signal itself are removed. Let $\mathbf{z}_T$ denote the true nonlinear distortions and let $\mathbf{z}_{F,i}$ denote the i'th false nonlinear distortions signal, that is, the one of the distortions signals due to an aliased copy of the received signal. A bound on the probability of error is found from

$$\text{prob}\left(\cup A_1\right) \leq \sum_i \text{prob}\left(A_i\right), \tag{99}$$

where A_i is the event

$$\frac{|\mathbf{y}^H \mathbf{z}_{F,i}|}{\|\mathbf{z}_{F,i}\|^2} > \frac{|\mathbf{y}^H \mathbf{z}_T|}{\|\mathbf{z}_T\|^2}. \tag{100}$$

6. FUTURE WORK

The topics described in this report offer new and innovative approaches to problems of minimizing sample rate while maximizing the information output of receiver systems. In this section we discuss some continuing challenges and areas for future research and development leading to realistic implementations of these techniques.

6.1 FUTURE WORK FOR NOLAFF

6.1.1 Algorithmic Development

Although NoLAff hypothesis testing has been tested in various simulations, some algorithmic challenges remain. The first of these involves the overlapping of signal information onto itself, the probe, or other signals or distortions present in the undersampled spectrum. Overlapping degrades the ability of NoLAff to create and disambiguate hypotheses. NoLAff has been tested as a means of spreading sparse information into unused basis vectors. While some overlap can be currently tolerated, a strategy must be developed for all signal overlapping.

Overlapping of information may arise from several situations. The simplest occurs when the linear portions of the signals and probe are nonoverlapping in the undersampled spectrum, but the nonlinearity leads to overlapping distortions after undersampling. Figure 54 depicts this situation, with yellow blocks indicating frequency locations where distortion will overlap. More difficult situations can occur when a narrowband signal happens to cross a fold location in the full frequency spectrum, leading a portion of the signal to fold on top of itself in the undersampled spectrum. Similarly, multiple signals with clear spectral separation in the full spectrum may overlap when undersampled because of aliasing. Algorithmic solutions to these challenges need to be explored.

Further performance benefits may be achieved by exploring the use of an adaptive probe. Figure 55 shows an adaptive probe could be used to avoid a problematic scenario of overlapping distortions. Aside from avoiding overlap, an adaptive probe may be able provide other performance improvements, such as steering distortions into more advantageous spectral locations to provide a clearer separation between hypotheses.

All NoLAff simulations to this point have utilized probe signals that are injected by the system designer and entirely known *a priori*. However, there is no reason that the probe must be a signal injected by the system designer; a very strong signal within the detected environment, such as a television broadcast or a jammer, may also serve the same mathematical role as an injected probe. Successfully implementing NoLAff with an environmental probe would turn traditionally interfering signals into beneficial sources of signal diversity.

Another area for further research lies in optimizing the nonlinearity for NoLAff encoding. More modalities that are created by the nonlinearity provide more diversity in the spreading of the

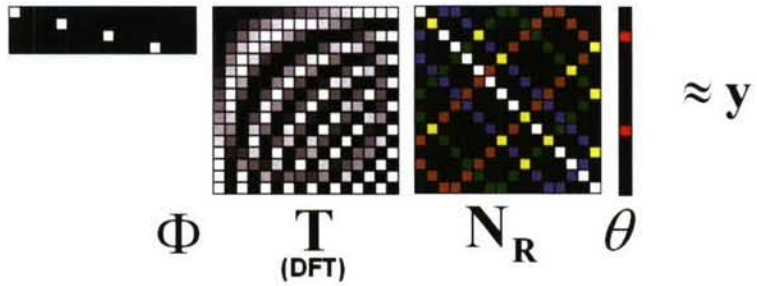


Figure 54: Overlapping distortions occur with NoLAff information spreading. This figure graphically depicts the undersampling matrix Φ, the DFT dictionary matrix $\mathbf{T}$, the nonlinear spreading matrix $\mathbf{N_R}$, and the sparse information vector θ. The spreading matrix is shown with several signals present. The yellow squares represent locations with overlap. As more distortions are created, filling out the signal spaces, more of these overlap locations begin to exist.

signal. However, as more spreading occurs, the likelihood of overlapping information increases. The encoding filter can also be developed to emphasize differences between hypothesis signals across the full signal spectrum.

6.1.2 NoLAff Mechanization

This report demonstrated the feasibility of implementing a NoLAff encoder with relatively simple circuit components. Future work for NoLAff mechanization would involve the construction of a testbed to analyze performance from actual hardware components. Initial encoding tests would verify performance with smaller bandwidths using the circuits described in this report. However, it would be desirable to demonstrate the ability of sampling a sparse, very wide band environment, such as 8 GHz or wider, using a high-speed COTS analog-to-digital converter. In order to perform such tests, the analog encoding hardware must be designed to handle much wider bandwidths beyond the capabilities of the circuits presented in this report.

The implementation of a NoLAff decoder in digital hardware should also be explored. The hardware design trade-offs between speed, performance, bandwidth and hardware capabilities should be analyzed.

6.1.3 NoLAff Applications

NoLAff undersampling has shown very good performance for signals sparse in frequency. Applying NoLAff techniques to signals sparse in other bases would be an interesting and logical direction for further research. Developing NoLAff approaches for exploiting sparsity in time and sparsity in time-frequency are two such possibilities. NoLAff demonstrated promising results for large dynamic range environments, and we would seek out applications requiring large dynamic

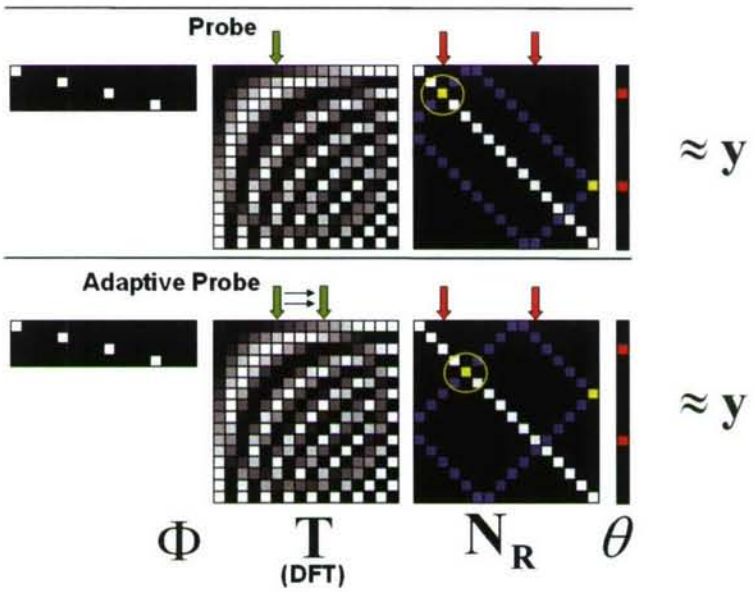


Figure 55: The upper diagram represents the creation of an undersampled encoded signal with a single probe. The probe basis vector is indicated by the green arrow, and the sparse signal vectors are indicated by the red arrows. In the top diagram, the first signal vector contains overlapping distortions, which reduces NoLAff decoding performance. The lower diagram shows how the probe can be shifted to a different basis vector, resulting in the overlapping distortion location moving away from the basis vector used by the sparse signal.

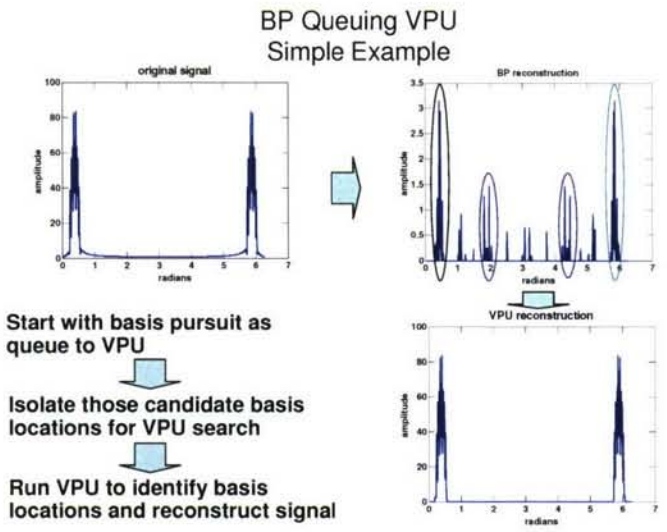


Figure 56: Basis Pursuit cuing VPU to identify and reconstruct a signal that occupies roughly 80 percent of downsampled frequency band.

range. One such application is the Angle-of-Arrival problem, involving space-time sparseness and requiring good performance over a large dynamic range. AoA is described below, and it is noted that CS techniques fail at certain dynamic range levels. NoLAff may provide better performance for this problem.

6.2 COMBINING BP AND VPU

Although BP is not well suited for reconstructing signals that occupy any more than 15-20 percent of the downsampled basis, it can be used as cue to VPU to reconstruct signals that can occupy up to nearly 100 percent of the downsampled basis. The process of BP cuing VPU is graphically illustrated in Figure 56; first BP roughly identifies the basis support and reconstitutes undesirable spurious signals. When VPU is cued by BP as illustrated in Figure 56, VPU not only identifies and reconstructs signals that are present with very high fidelity but also rejects spurious signals. Combining the computational efficiency of BP with the performance of VPU yields a powerful algorithm far more robust than either VPU or BP operating independently.

Both BP and VPU suffer from the so-called near-far problem in which signals received with high SNR mask low SNR signals from being detected or reconstructed with reasonable fidelity. For example, in angle-of-arrival applications, ML techniques are frequently outperformed by subspace techniques such as root-music which are much less sensitive to power variation between two received signals. However, it may be possible to recast both VPU and BP in a subspace framework; for

example, consider the following formulation $Y = AX + V$ where $Y = \{y(t_1), y(t_2) \ldots, y(t_T)\}$, so that the BP problem can be formulated as

$$\begin{aligned} &\arg\min \|Y - AX\|_f^2 + \lambda \|x_{\ell_2}\|_1 \\ &x_{\ell_2} = \{x_1, x_2, \ldots, x_N\} \text{and } x_i = \|x_1, x_2, \ldots, x_N\|_2, \end{aligned} \tag{101}$$

where Y is the received signal matrix, A the randomized downsampled sparsity basis and X is the signal matrix. The intuition behind the formulation above is that if one can formulate the ℓ_2 problem in matrix form, the problem can be solved using subspace techniques. For example, consider the case where the SVD of Y is given by UDV^T, and letting $Y_{SV} = Y[V_k]_k$, and $X_{SV} = X[V_k]_k$, then the BP problem can be formulated as

$$\arg\min \|Y_{SV} - AX_{SV}\|_f^2 + \lambda \|\tilde{x}_{\ell_2}\|_1, \tag{102}$$

where $[V]_k$ are the singular vectors principally associated with the signal subspace. This approach reduces the variation of the noise eigenvalues and may mitigate the near-far problem faced by both BP and VPU.

6.3 ANGLE-OF-ARRIVAL IN ARRAY PROCESSING

It is possible to architect a unique *spatial* compressed sensing receiver consisting of an array of antenna elements that *sample* the signals arriving at the face of the array from a countably finite number of directions. The spatial A2I receiver is illustrated in Figure 57, where an 128-element array is broken up into 32 subarrays consisting of 4 elements per subarray. The output of each element is subject to random linear projections with coefficients drawn from a $\pm$ Bernoulli distribution. This type of compressed sensing receiver applies random linear projections and then downsamples in the spatial (as opposed to temporal) domain.

We tested the performance of BP in an environment with two wideband CDMA signals present, received with an SNR of 5 dB per channel. In all cases we compared BP performance to root-music operating with an analog front-end receiver almost identical to the receiver illustrated in Figure 57, with the one exception that the root-music array was steered, i.e., the random linear projections in Figure 57 were replaced by complex exponentials. BP was modified for super-resolution capability; this involved refining the columns of Ψ to super-resolve the source after an initial identification of the coarse spatial basis support. The performance of super-resolution BP as compared to root-music is illustrated in Figure 58. We found that in general super-resolution BP was extremely competitive with root-music, and in many cases, BP slightly outperformed root-music in determining the AoA (in azimuth) of the two sources present.

We also compared the performance of BP and root-music when the two signal sources had unequal signal power, and the results of this test are illustrated in Figure 59. With one source 40 dB stronger than the other—the so-called near-far problem—BP was unable to uniquely identify

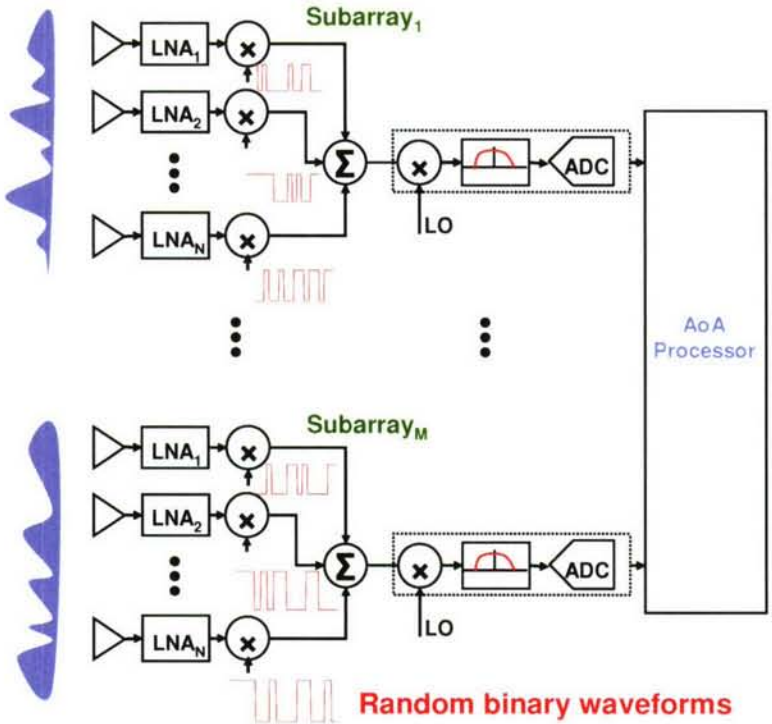


Figure 57: Subarray A2I receiver for angle-of-arrival processing.

the AoA of the second weaker source. With an array calibration error model, we are currently investigating the possibility of employing estimation and subtraction to remove the effect of the strong signal, as well as linear projections that place the strong interferer into the null space of D.

6.4 DYNAMICAL SYSTEMS

A key aspect of future endeavors is the continued study of dynamical systems (DS) for A2I. We have established the utility of DS for compressed sensing at GHz bandwidths with practical, low-power circuits. What remains is establishing a robust modeling approach for such circuits and assessing the impact of modeling errors on performance.

6.5 DENSER ENVIRONMENTS

Sparse environments often have additional structure that can be exploited to enhance the performance of A2I. For example, signals can have allocated bandwidths, as considered in Section 5.5. It is also possible for signals to be differentially sparse in the sense that their environment is partially

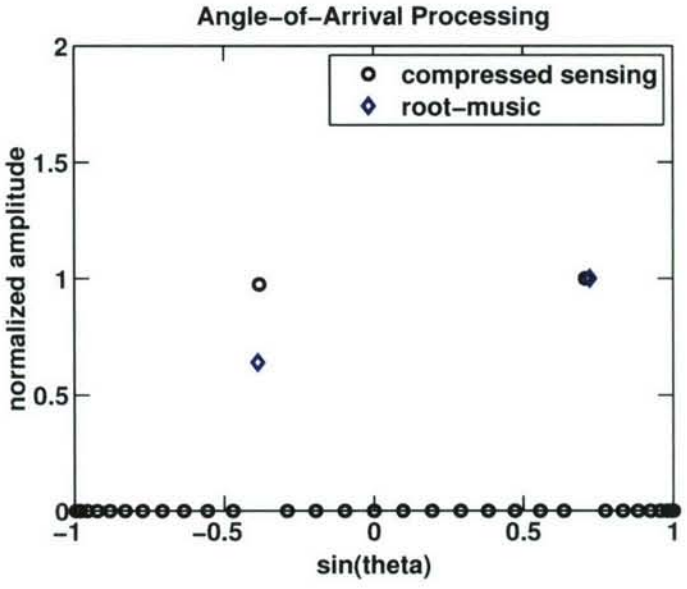


Figure 58: Angle-of-Arrival processing from a 128-element array broken up into 32-element subarrays. A2I AoA super-resolution via modified BP performance is compared to root-music where in the case of root-music each of the subarrays is steered to point at a specific angular location. Two sources located at -0.3827 and 0.7071 are present.

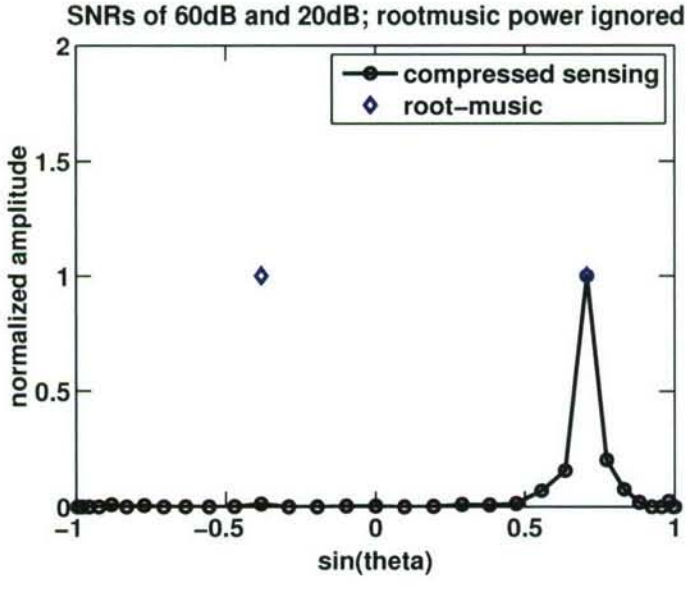


Figure 59: Near-far problem associated with BP processing. With one source having a four orders of magnitude higher power with respect to the second, the second source becomes undetectable. Note that root-music is insensitive to the near-far problem.

known. There are obvious extensions of compressed sensing to this case as well as other approaches that may even be useful in dense environments if detection alone is satisfactory.

6.6 IMPROVED DYNAMIC RANGE

Finding low-level signals in the presence of strong interferers is one of the weaknesses of the standard random linear model of compressed sensing. Some implementations, such as NoLAff, can avoid this problem. It may also be possible to use more robust estimation techniques to mitigate the dynamic range problems associated with conventional compressed sensing algorithms.

7. SUMMARY

In this report, we presented significant developments on three Analog-to-Information (A2I) research activities at the heart of the collaborative effort of the MIT Lincoln Laboratory and GMR Research & Technology team: Nonlinear Affine processing (NoLAff), Variable Projection and Unfolding (VPU), and dynamical systems. Substantial effort has been made to demonstrate the practical capabilities of these new techniques in realistic signal environments and to provide a path toward mechanization. We presented results indicating strong potential for NoLAff hypothesis testing, a novel technique invented by the team which exploits nonlinear distortions as a new form of signal diversity, enabling undersampling and digital reconstruction of sparse signals. NoLAff showed very accurate performance on both sparse and non-sparse signals, even detecting and reconstructing signals with bandwidth approaching the entire undersampled spectrum with less than a 0.1% error rate. Environments with multiple signals requiring a large dynamic range (>40 dB) were shown to be effectively undersampled and reconstructed with NoLAff as well. A candidate analog circuit for NoLAff encoding, robust to implementation variations, was presented along with simulation results that suggest a hardware NoLAff system may realistically perform very well. The complexity of the NoLAff hypothesis digital decoder was also investigated and shown to be reasonably low.

A new approach to solving the A2I problem was presented in Variable Projection and Unfolding (VPU). VPU is similar to Orthogonal Matched Pursuit (OMP), where both capture the most important information from a sparse signal by using a subspace containing a small number of random linear projections from the original signal space. VPU, however, is not restricted to rank-1 subspaces and can exploit correlations between coefficients in the signal basis. A dual-stage search process accurately locates the mostly likely position of the downsampled spectrum within the full spectrum. This process showed excellent performance on signals considered too dense for standard CS or Basis Pursuit (BP) reconstruction, signals with bandwidths at nearly 100% of the undersampled spectrum. VPU also showed good performance in reconstructing signals sparse in bases other than frequency. Applied to a chirp signal buried in interference and noise, VPU consistently identified the undersampled signal correctly.

Our work on dynamical systems offers a unique mechanization approach for compressed sensing. Dynamical systems inherently provide the randomness and memory needed for undersampling with CS, and allow reconstruction of undersampled signals by applying sampled values as initial conditions to an ordinary differential equation characterizing the system. We demonstrated the capability of this dynamical system approach to CS in recovering an undersampled pulse signal using a practical, low-power, wideband circuit. Nonlinear models were formulated to support the application of dynamical systems to CS.

The work resulting from this collaborative effort has advanced the understanding of new and unique technologies for the A2I problem. We presented groundbreaking research into NoLAff processing, and NoLAff hypothesis testing in particular, and this program has led to a better under-

standing of the capabilities of this technology. Our results suggest NoLAff may outperform other A2I techniques, particularly in sparseness and dynamic range. VPU is also an original technology, developed through A2I research, with promising results for both non-sparse signals and signals with non-frequency-based sparseness. This report provides a strong indication that these technologies may hold practical benefit for the future of ELINT and SIGINT systems and ISR sensor networks, for both detection and reconstruction, by offering dramatic improvements in the ability to process information more quickly and with greater accuracy.

REFERENCES

[1] A. S. Sedra and K. C. Smith, *Microelectronic Circuits.* USA: Oxford University Press, 1991.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] J. Goodman, A. Reuther, and D. Martinez, "Next-generation technologies to enable sensor networks," in *Handbook of Sensor Networks*, pp. (2)1–(2)21, CRC Press, 2004.

[4] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Proc.*, vol. 12, no. 8, pp. 906–916, 2003.

[5] G. M. Raz, "Method and systems for nonlinear and affine signal processing." U.S. patent 7,173,555 B2, Feb. 6 2007.

[6] J. G. Proakis, *Digital Communications.* McGraw-Hill, 1983.

[7] G. M. Raz and B. D. Van Veen, "Baseband Volterra filters for implementing carrier based nonlinearities," *IEEE Transactions on Signal Processing*, vol. SP-46, no. 1, pp. 103–115, 1998.

[8] G. M. Raz, "Nonlinear signal processing for wideband high dynamic range systems," project report, GMR Research & Technology, July 2006.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal Royal Statistical Society B*, vol. 58, pp. 267–288, 1996.

[10] R. E. Ziemer and W. H. Tranter, *Principles of Communications.* New York: John Wiley & Sons, Inc, 1995.

[11] J. Tropp and A. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit." Preprint, 2005.

[12] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 872–875, 2006.

[13] D. Malioutov, M. Cetin, and A. Wilsky, "Homotopy continuation for sparse signal representation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, 2005.

[14] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, vol. 54 of *Encyclopedia of Mathematics and its Applications.* Cambridge: Cambridge University Press, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[15] S. Wiggins, *Global bifurcations and chaos*, vol. 73 of *Applied Mathematical Sciences*. New York: Springer-Verlag, 1988. Analytical methods.

[16] A. J. Lichtenberg and M. A. Lieberman, *Regular and chaotic dynamics*, vol. 38 of *Applied Mathematical Sciences*. New York: Springer-Verlag, second ed., 1992.

[17] L. A. Bunimovich, I. P. Cornfeld, R. L. Dobrushin, M. V. Jakobson, N. B. Maslova, Y. B. Pesin, Y. G. Sinaĭ, Y. M. Sukhov, and A. M. Vershik, *Dynamical systems. II*, vol. 2 of *Encyclopaedia of Mathematical Sciences*. Berlin: Springer-Verlag, 1989. Ergodic theory with applications to dynamical systems and statistical mechanics, edited and with a preface by Sinaĭ, translated from the Russian.

[18] J. A. Oteo and J. Ros, "From time-ordered products to Magnus expansion," *Journal of Mathematical Physics*, vol. 41, no. 5, pp. 3268–3277, May, 2000.

[19] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with sparsity constraint," *Communications in Pure and Applied Mathematics*, vol. LVII, pp. 1413–1457, 2004.

[20] V. Guigue, A. Rakotomamonji, and S. Canu, "Kernel basis pursuit." Preprint, 2005.

[21] T. J. Kragh and A. A. Kharbouch, "Monotonic iterative algorithms for SAR image restoration," in *Image Processing, IEEE 2006 International Conference on*, pp. 645–648, Oct. 2006.

[22] L. Illing and D. J. Gauthier, "Ultra-high-frequency chaos in a time-delay electronic device with band-limited feedback," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Aug., 2006.

[23] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragbeh, Y. Massoud, and R. Baraniuk, "Analog-to-information conversion via random demodulation," in *Proc. IEEE Dallas Circuits and Systems Workshop (DCAS)*, vol. 1, pp. 71–74, 2006.

[24] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1233, 2006.

[25] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[26] M. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity," in *Proc. Allerton Conference on Communication, Control, and Computing*, vol. 1, 2006.

[27] M. Elad, "Optimized projections for compressed sensing." Preprint, 2006.

[28] T. Krauss and M. Zoltowski, "Chip-level MMSE equalization at the edge of the cell," in *Wireless Comm. and Networking Conference*, vol. 1, pp. 386–392, 2000.

[29] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Part I.* Wiley, New York, 1968.

[30] E. J. Kelly and K. W. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Tech. Rep. 848, M.I.T. Lincoln Laboratory, April, 1989.

[31] E. Candes and T. Tao, "Decoding by linear programming," *arXiv:math.MG/0502327*, vol. v1, Feb. 15, 2005.

[32] G. M. Raz, "Preliminary report of nonlinear equalization of radar receivers," Project Report NLEQ-1, Lincoln Laboratory, Massachusetts Institute of Technology, Oct. 2002.

[33] C. Chan and G. M. Raz, "Computational complexity versus performance using a new architecture for nonlinear equalization," Internal memorandum NLEQ 5-1, M.I.T. Lincoln Laboratory, May 2003.