On the Compressive Sensing Systems (Part II)

Prepared by:
Soheil Salari, Il-Min Kim - Department of Electrical and Computer Engineering, Queen's University
Francois Chan - Department of Electrical and Computer Engineering, RMC

Contractor name and address:
Francois Chan, RMC
13 General Crerar Crescent, Kingston, ON, K7K 7B4

Contract number: 2009-0302-SLA

Contract Scientific Authority's name, title, and phone number:
Dr. Sreeraman Rajan, Senior Defence Scientist 613-991-4138

Contract Report
DRDC-RDDC-2015-C026
February 2015

ON THE COMPRESSIVE SENSING SYSTEMS (PART II)

Prepared by [1]Soheil Salari, [1]Il-Min Kim, and [2]Francois Chan

1. Department of Electrical and Computer Engineering, Queen's University

2. Department of Electrical and Computer Engineering, RMC

Contractor name and address:

Francois Chan, RMC,

13 General Crerar Crescent, Kingston, ON, K7K 7B4

Contract number: 2009-0302-SLA

Contract Scientific Authority's name, title, and phone number:

Dr. Sreeraman Rajan, Senior Defence Scientist

613-991-4138

I. INTRODUCTION

Any signal in $\mathbb{R}^N$ can be represented in terms of a basis of $N \times 1$ vectors $\{\mathbf{\Psi}_i\}_{i=1}^N$. Using the $N \times N$ basis matrix $\mathbf{\Psi} = [\mathbf{\Psi}_1, \mathbf{\Psi}_2, ..., \mathbf{\Psi}_N]_{N\times N}$ with the vectors $\{\mathbf{\Psi}_i\}$ as columns, a signal $\mathbf{x}$ can be expressed as

$$\mathbf{x}_{N\times 1} = \sum_{i=1}^{N} s_i \mathbf{\Psi}_{iN\times 1} = \mathbf{\Psi}_{N\times N}\, \mathbf{s}_{N\times 1}, \tag{1}$$

where $\mathbf{s} = [s_1, s_2, ..., s_N]_{N\times 1}^T$. Clearly, $\mathbf{x}$ and $\mathbf{s}$ are equivalent representations of the signal, with $\mathbf{x}$ in the time or space domain and $\mathbf{s}$ in the $\mathbf{\Psi}$ domain.

The signal $\mathbf{x}$ is K-sparse if it is a linear combination of only K basis vectors; that is, only K of the s_i coefficients in (1) are nonzero and $(N - K)$ are zero. In the context of compress sensing (CS), the case of interest is when $K \ll N$. The signal $\mathbf{x}$ is compressible if the representation (1) has just a few large coefficients and many small coefficients.

Consider a general linear measurement process that computes $M \ll N$ inner products between $\mathbf{x}$ and a collection of $N \times 1$ vectors $\{\mathbf{\Phi}_j\}_{j=1}^M$, i.e. $y_j = \mathbf{\Phi}_j^T\ \mathbf{x}$. We first arrange the measurements y_j in an $M \times 1$ vector $\mathbf{y}$ and the measurement vectors $\mathbf{\Phi}_j$ as rows in an $M \times N$ matrix $\mathbf{\Phi}$. Then, by substituting $\mathbf{\Psi}$ from (1), $\mathbf{y}$ can be written as:

$$\mathbf{y}_{M\times 1} = \mathbf{\Phi}_{M\times N}\, \mathbf{x}_{N\times 1} = \mathbf{\Phi}_{M\times N}\ \mathbf{\Psi}_{N\times N}\, \mathbf{s}_{N\times 1} = \mathbf{\Theta}_{M\times N}\, \mathbf{s}_{N\times 1}. \tag{2}$$

The measurement process is not adaptive, meaning that $\mathbf{\Phi}$ is fixed and does not depend on the signal $\mathbf{x}$. In general, applying the technique of CS requires designing of [1]–[4]:

a) A stable measurement matrix $\mathbf{\Phi}$ such that the salient information in any K-sparse or compressible signal is not damaged by the dimensionality reduction from $\mathbf{x} \in \mathbb{R}^N$ to $\mathbf{y} \in \mathbb{R}^M$; and,

b) A reconstruction algorithm to recover $\mathbf{x}$ from only $M \approx K$ measurements $\mathbf{y}$.

The measurement matrix $\mathbf{\Phi}$ must allow the reconstruction of the length-N signal $\mathbf{x}$ from $M \ll N$ measurements $\mathbf{y}$, which is an ill-conditioned problem. If, however, $\mathbf{x}$ is K-sparse and the K locations of the nonzero coefficients in $\mathbf{s}$ are known, then the problem can be solved. For any vector $\mathbf{v}$ sharing the same K nonzero entries as $\mathbf{s}$ and for some $\varepsilon > 0$, a necessary and sufficient condition for perfect reconstruction was introduced in [1] as:

$$1 - \varepsilon \leq \frac{\|\mathbf{\Theta v}\|_2}{\|\mathbf{v}\|_2} \leq 1 + \varepsilon, \tag{3}$$

That is, the matrix $\mathbf{\Theta}$ must preserve the lengths of these particular K-sparse vectors. Of course, in general the locations of the K nonzero entries in $\mathbf{s}$ are not known. However, a ***sufficient condition*** for a stable

solution for both K-sparse and compressible signals is that $\mathbf{\Theta}$ satisfies (3) for an arbitrary 3K-sparse vector **v**. This condition is referred to as the **restricted isometry property** (**RIP**). A related condition, referred to as **incoherence**, requires that the rows $\{\mathbf{\Phi}_j\}$ of $\mathbf{\Phi}$ can not sparsely represent the columns $\{\mathbf{\Psi}_i\}$ of $\mathbf{\Psi}$ (and vice versa). In summary, within the framework of CS theory, due to the fact that **s** is K-sparse, signal recovery can actually be made possible when the matrix $\mathbf{\Theta}$ obeys the rule of RIP or the measuring matrix $\mathbf{\Phi}$ is incoherence with the basis matrix $\mathbf{\Psi}$.

Both the RIP and incoherence criteria can be achieved with high probability simply by selecting $\mathbf{\Phi}$ as a random matrix. For instance, let the matrix elements $\phi_{i,j}$ be independent and identically distributed (iid) Gaussian random variables with mean zero and variance 1/N. Then, matrix $\mathbf{\Phi}$ has following properties [1]–[4]:

1-The matrix $\mathbf{\Phi}$ is incoherent with the basis $\mathbf{\Psi} = \mathbf{I}$ of delta spikes with high probability. More specifically, an $M \times N$ iid Gaussian matrix $\mathbf{\Theta} = \mathbf{\Phi I} = \mathbf{\Phi}$ can be shown that satisfies the RIP property with high probability if $M \geq cKlog(N/K)$, where c is a small constant. Therefore, K-sparse and compressible signals of length N can be recovered from only $M \geq cKlog(N/K)$ random Gaussian measurements.

2- The matrix $\mathbf{\Phi}$ is universal in the sense that $\mathbf{\Theta} = \mathbf{\Phi\Psi}$ will be iid Gaussian and thus satisfies the RIP with high probability, regardless of the choice of orthonormal basis $\mathbf{\Phi}$.

The authors of [5] introduced another criteria entitled as uniform uncertainty principle (UUP). The UUP essentially state that the sensing matrix $\mathbf{\Theta}$ of size $M \times N$ obeys the restricted isometry hypothesis. Let $\mathbf{\Theta}_T$, $T \in \{1, ..., N\}$, be the $M \times |T|$ submatrix which is obtained by extracting the columns of $\mathbf{\Theta}$ corresponding to the indices in T. For all subsets T with $|T| \leq K$ and coefficient sequences $(c_j)_{j \in T}$, the K-restricted isometry constant δ_K is defined as the smallest quantity such that:

$$(1 - \delta_K) \|\mathbf{c}\|_2^2 \leq \|\mathbf{\Theta}_T \mathbf{c}\|_2^2 \leq (1 + \delta_K) \|\mathbf{c}\|_2^2 . \tag{4}$$

This property essentially requires that every set of columns with cardinality less than K approximately behaves like an orthonormal system. An important result is that if the columns of the sensing matrix $\mathbf{\Theta}$ are approximately orthogonal, then the exact recovery phenomenon occurs. As stated in [5], if every set of sensing matrix columns with cardinality less than the sparsity of the signal of interest is approximately orthogonal, then the sparse signal can be exactly recovered with high probability.

II. Compress Sampling Receiver

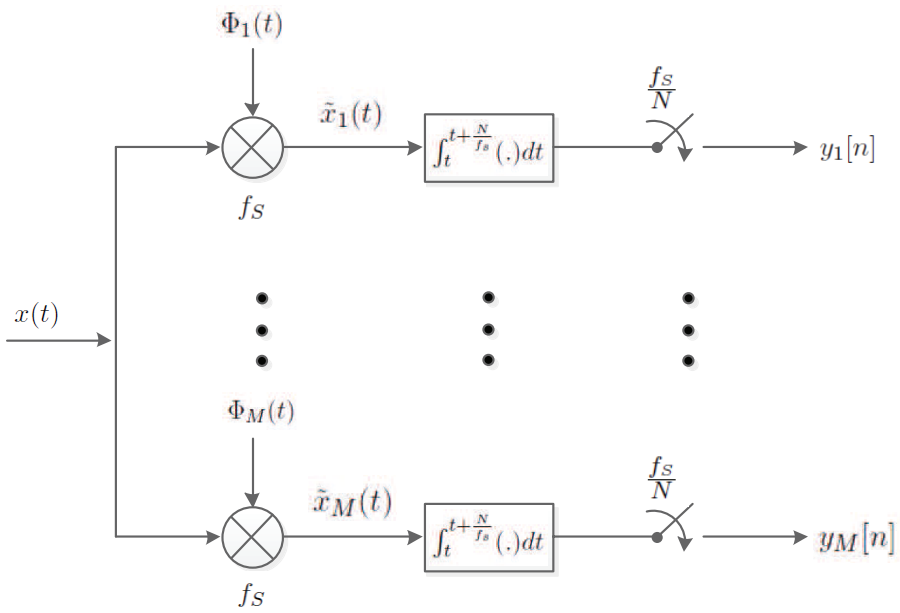

Fig. 1. Compress Sampling Receiver Schematic.

Let $x(t)$ is the baseband equivalent of the received signal with bandwidth B. To avoid aliasing, $\Phi_j(t)$ $(j = 1, ..., M)$ needs to be applied at the Nyquist frequency (f_s) of the signal $x(t)$ or higher $(f_s \geq 2B)$. Further, let N is the number of integration samples per compression blocks. Then, we have:

$$\tilde{x}_j(t) = x(t)\Phi_j(t), \qquad j = 1, ..., M \tag{5}$$

$$y_j[n] = \int_t^{t+\frac{N}{f_s}} \tilde{x}_j(t)dt = \int_t^{t+\frac{N}{f_s}} x(t)\Phi_j(t)dt, \qquad j = 1, ..., M \tag{6}$$

It is worthy to mention that

$$\Phi_j(t)|_{t=\frac{m-1}{f_s}}^{t=\frac{m}{f_s}} = P_{j,m} \in \{+1, -1\}, \qquad j = 1, ..., M, \qquad m = 1, ..., N \tag{7}$$

Considering (7), we can rewrite (6) as

$$\begin{aligned} y_j[n] &= \int_t^{t+\frac{N}{f_s}} x(t)\Phi_j(t)dt \\ &= \int_t^{t+\frac{1}{f_s}} x(t)\Phi_j(t)dt + \int_{t+\frac{1}{f_s}}^{t+\frac{2}{f_s}} x(t)\Phi_j(t)dt + ... + \int_{t+\frac{N-1}{f_s}}^{t+\frac{N}{f_s}} x(t)\Phi_j(t)dt \\ &= P_{j,1}\underbrace{\int_t^{t+\frac{1}{f_s}} x(t)dt}_{x_1} + P_{j,2}\underbrace{\int_{t+\frac{1}{f_s}}^{t+\frac{2}{f_s}} x(t)dt}_{x_2} + ... + P_{j,N}\underbrace{\int_{t+\frac{N-1}{f_s}}^{t+\frac{N}{f_s}} x(t)dt}_{x_N} \\ &= \sum_{m=1}^{N} P_{j,m}x_m, \qquad j = 1, ..., M, \end{aligned} \tag{8}$$

where $x_m, m = 1, ..., N$, are simply the Nyquist samples of the signal $x(t)$. For our purposes, it is convenient to write (8) in matrix form as

$$\underbrace{\begin{bmatrix} y_1[n] \\ y_2[n] \\ \vdots \\ y_M[n] \end{bmatrix}}_{\mathbf{y}_{M\times 1}} = \underbrace{\begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,N} \\ P_{2,1} & P_{2,2} & \dots & P_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ P_{M,1} & P_{M,2} & \dots & P_{M,N} \end{bmatrix}}_{\mathbf{\Phi}_{M\times N}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{\mathbf{x}_{N\times 1}} \tag{9}$$

or, equivalently, as

$$\mathbf{y}_{M\times 1} = \mathbf{\Phi}_{M\times N}\mathbf{x}_{N\times 1} = \begin{bmatrix} \mathbf{\Phi}_1 \\ \vdots \\ \mathbf{\Phi}_M \end{bmatrix}_{M\times N} \mathbf{x}_{N\times 1}. \tag{10}$$

As before mentioned, any signal in $\mathbb{R}^N$ can be represented in terms of a basis of $N \times 1$ vectors $\{\mathbf{\Psi}_i\}_{i=1}^N$. Using the $N \times N$ basis matrix $\mathbf{\Psi} = [\mathbf{\Psi}_1, \mathbf{\Psi}_2, ..., \mathbf{\Psi}_N]_{N\times N}$ with the vectors $\{\mathbf{\Psi}_i\}$ as columns, $\mathbf{x}$ can be expressed as

$$\mathbf{x}_{N\times 1} = \sum_{i=1}^{N} s_i \mathbf{\Psi}_{iN\times 1} = \mathbf{\Psi}_{N\times N}\, \mathbf{s}_{N\times 1} = [\mathbf{\Psi}_1, \mathbf{\Psi}_2, ..., \mathbf{\Psi}_N]_{N\times N}\, \mathbf{s}_{N\times 1}, \tag{11}$$

where $\mathbf{s} = [s_1, s_2, ..., s_N]_{N\times 1}^T$. As shown in Fig. 2, it is possible that all of the vector $\mathbf{x}$ coefficients be non-zero values.

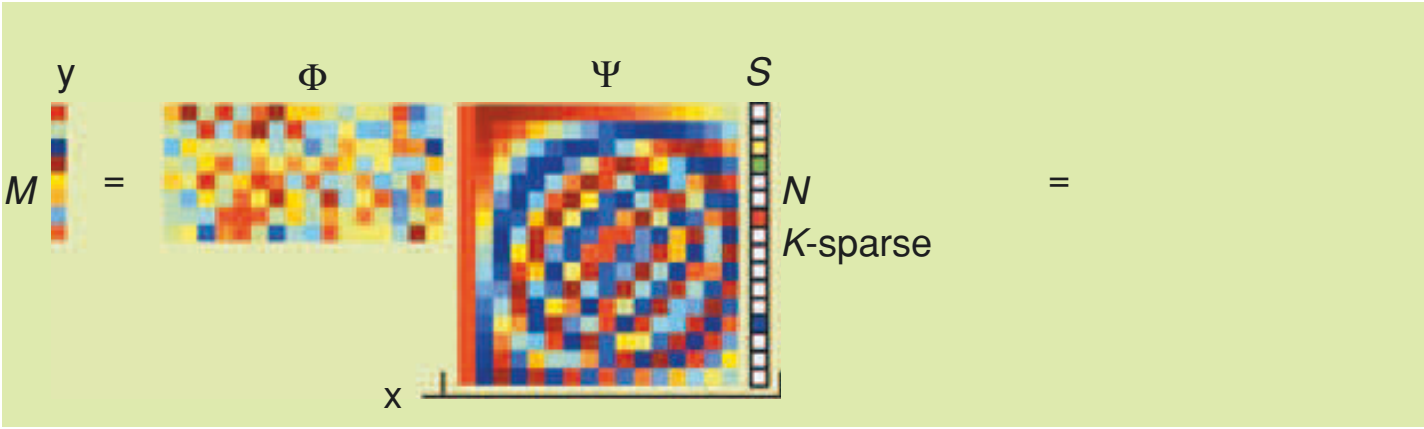

Fig. 2. Compress Sampling Process.

By substituting $\mathbf{x}$ from (11), $\mathbf{y}$ can be written as

$$
\begin{aligned}
\mathbf{y}_{M\times 1} &= \mathbf{\Phi}_{M\times N}\,\mathbf{x}_{N\times 1} = \mathbf{\Phi}_{M\times N}\,\mathbf{\Psi}_{N\times N}\,\mathbf{s}_{N\times 1} \\
&= \begin{bmatrix} \mathbf{\Phi}_1 \\ \vdots \\ \mathbf{\Phi}_M \end{bmatrix}_{M\times N} [\mathbf{\Psi}_1, \mathbf{\Psi}_2, ..., \mathbf{\Psi}_N]_{N\times N}\,\mathbf{s}_{N\times 1} \\
&= \underbrace{\begin{bmatrix} \mathbf{\Phi}_1\mathbf{\Psi}_1 & \mathbf{\Phi}_1\mathbf{\Psi}_2 & \dots & \mathbf{\Phi}_1\mathbf{\Psi}_N \\ \mathbf{\Phi}_2\mathbf{\Psi}_1 & \mathbf{\Phi}_2\mathbf{\Psi}_2 & \dots & \mathbf{\Phi}_2\mathbf{\Psi}_N \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{\Phi}_M\mathbf{\Psi}_1 & \mathbf{\Phi}_M\mathbf{\Psi}_2 & \dots & \mathbf{\Phi}_M\mathbf{\Psi}_N \end{bmatrix}}_{\mathbf{\Theta}_{M\times N}} \underbrace{\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix}}_{\mathbf{s}_{N\times 1}} = \mathbf{\Theta}_{M\times N}\,\mathbf{s}_{N\times 1}
\end{aligned}
\tag{12}
$$

III. Xampling: Signal Acquisition and Processing in Union of Subspaces

Signal processing methods have changed substantially over the last several decades. The number of operations that are shifted from analog to digital is constantly increasing, leaving amplifications and fine tunings to the traditional frontend. Sampling theory is the key enabling this revolution. Traditional sampling theorems assume that the input lies in a predefined subspace. Recently, nonlinear union of subspaces (UoS) models have been receiving growing interest in the context of analog sampling. The UoS setting captures uncertainty in the signal by allowing several possible subspace descriptions, where the exact signal subspace is unknown a-priori.

In contrast to classic subspace sampling, the theory of sampling over UoS is still developing. In particular, to date, there is no equivalent to the oblique projection operator which reconstructs the signal when its exact subspace is known for almost all sampling functions. The lack of a complete theory has not

withheld development of numerous stylized applications that aim to reduce the sampling rate below the Nyquist rate by exploiting the UoS model. The acquisition and reconstruction methods of UoS theory are substantially different from each other, raising the question of whether the apparent distinct approaches can be derived from a common framework.

The first and main contribution of [6] is defining a unified and pragmatic framework for acquisition and processing of UoS signal classes, referred to as Xampling. It consists of two main functions: low rate analog to digital conversion (X-ADC), in which the input is compressed in the analog domain prior to sampling with commercial devices, and low rate digital signal processing (X-DSP), in which the input subspace is detected prior to signal processing in the digital domain. As the second contribution, the authors of [6], studied the X-ADC block and addressed the choice of analog compression. They completed this task by examining the compression techniques used in the random demodulator (RD) and modulated wide band converter (MWC) systems, which employ CS technique to reduce the sampling rate of spectrally sparse signals below the Nyquist rate. As the last contribution, they studied the X-DSP stage and sub-Nyquist processing, which are challenging tasks since conventional DSPs assume their input data stream are given at the Nyquist rate. They developed a digital algorithm, named Back-DSP, that provides the MWC with a smooth interface to existing DSP softwares.

A. Xampling

1) Union of subspaces: Let $x(t)$ be an input signal in the Hilbert space $H = L_2(\mathbb{R})$. The signal $x(t)$ is assumed to lie in a UoS of H, namely U, within a parameterized family of subspaces

$$x(t) \in U \bigcup_{\lambda \in \Lambda} \triangleq A_\lambda \tag{13}$$

where Λ is a list of indices, and each individual subspace $A_\lambda \in H$. The key property of the UoS model is that the input $x(t)$ resides within A_λ^* for some $\lambda^* \in \Lambda$, but a-priori, the exact subspace index λ^* is unknown. Equation (13) typically represents a nonlinear set of possible inputs, which is a true subset of the linear sum of all subspaces A_λ, denoted hereafter by Σ.

Authors of [6] have considered multiband sampling as an motivating example. Multiband sampling, encountered when a communication receiver intercepts multiple radio-frequency (RF) transmissions, but the knowledge of carrier frequencies f_i is not available. In this setting, the input $x(t)$ has multiband spectra with energy that is concentrated on N frequency intervals of individual widths B located anywhere below some maximal frequency f_{max}. Such a receiver faces a challenging sampling problem, since classic acquisition methods, such as RF demodulation or bandpass undersampling, require knowledge of f_i values. At first sight, it may seem that sampling at the Nyquist rate $f_{NYQ} = 2f_{max}$ is necessary. On

the other hand, since each specific fills only a portion of the Nyquist range (only NB Hz), one would intuitively expect to be able to reduce the sampling rate below f_{NYQ}.

A multiband model can be described in union terminology by indexing the possible band positions with $\lambda = f_i$ and letting A_λ capture the subspace of multiband signals on the chosen support. It is therefore expected that an input from a multiband union can be determined from sampling at a rate proportional to the actual bandwidth occupied by A_λ, namely NB, up to some rate increase needed to determine the unknown subspace index λ^* of the given $x(t)$. In principle, f_i lies in the continuum $f_i \in [0, f_{max}]$ in this modeling, so that the union contains infinitely many subspaces. A different viewpoint, is to divide the Nyquist range to M slices and enumerate the possible supports according to the slice indices that contain signal energy. This approach results in a finite union of bandpass subspaces, which enables efficient hardware and software implementation, as further discussed in detail.

2) Unified Goals: The authors of [6] defined the sampling problem for the union set (13) as the design of a system that provides:

e1) ADC: an acquisition operator which converts the analog input $x(t) \in U$ to a sequence $y[n]$ of measurements;

e2) DSP: a toolbox of processing algorithms, which uses $y[n]$ to perform classic tasks, e.g., estimation, detection, data retrieval etc., and

e3) DAC: a method for reconstructing $x(t)$ from the samples $y[n]$.

They adopted as a general design constraint that the above goals should be accomplished with minimum use of resources.

3) Architecture: The Xampling framework has the high-level architecture presented in Fig. 3. As highlighted, the Xampling architecture is driven by two main considerations: reducing analog bandwidth prior to sampling and gaining lowrate DSP, preferably backward-compatible with existing processing algorithms. We next describe the five functional blocks and explain these two considerations.

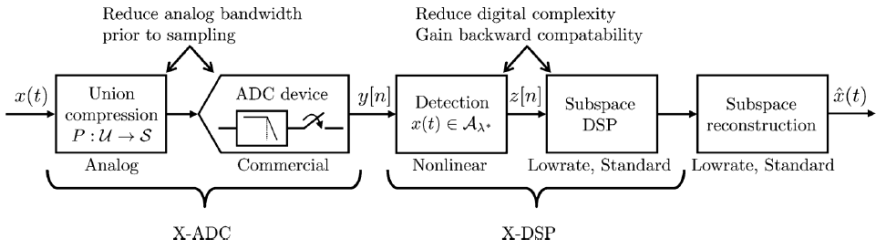

Fig. 3. Xampling - A functional framework for signal acquisition and processing in union of subspaces.

Analog bandwidth compression. The first two blocks, termed X-ADC, perform the conversion of $x(t)$ to digital domain. An operator P compresses the high-bandwidth input $x(t)$ into a signal with lower bandwidth, effectively capturing the entire union U by a subspace S with substantially lower sampling requirements. A commercial ADC device then takes pointwise samples of the compressed signal, resulting in the sequence of samples $y[n]$. The role of P in Xampling is to narrow down the analog bandwidth that enters the acquisition devices, so that lowrate ADC devices can be used.

Lowrate DSP. A second goal of Xampling is to translate the sampling rate reduction to a comparable decrease in processing speeds. Their proposal for achieving this goal consists of the three computational blocks in the digital part of Fig. 3. A nonlinear step detects the signal subspace A_{λ^*} from the lowrate samples. Once the index λ^* is determined, a low-rate sequence $z[n]$ of numbers that matches standard sampling of A_λ is computed. As a nice feature, this creates a seamless interface to existing DSP algorithms and interpolation techniques, hence provides backward compatibility. The combination of nonlinear detection and standard DSP is referred to as X-DSP.

Xampling is a generic template architecture. It does not specify the exact acquisition operator P or nonlinear detection method to be used. These are application-dependent functions. The authors goal in introducing Xampling was to propose a high-level system architecture and a basic set of guidelines:

1) An analog preprocessing unit to compress the input bandwidth;
2) Commercial lowrate ADC devices for actual acquisition at a low rate;
3) Subspace detection in software; and
4) Standard DSP and DAC methods.

B. X-ADC: SUB-NYQUIST SIGNAL ACQUISITION

In this section, we study the X-ADC stage of Fig. 3 and in particular the analog compression operator P. Please note that, analog compression involves mixing the input with certain waveforms prior to sampling, and reconstruction relies on CS techniques. The RD approach treats signals consisting of a discrete set of harmonic tones with the system that is depicted in Fig. 4.

Signal model. A multitone signal $f(t)$ consists of a sparse combination of integral frequencies:

$$f(t) = \sum_{w \in \Omega} a_w e^{j2\pi wt} \tag{14}$$

where Ω is a finite set of K out of an even Q number of possible harmonics. The model parameters are the tone spacing Δ, number of active tones K and grid length Q. The Nyquist rate is $Q\Delta$.

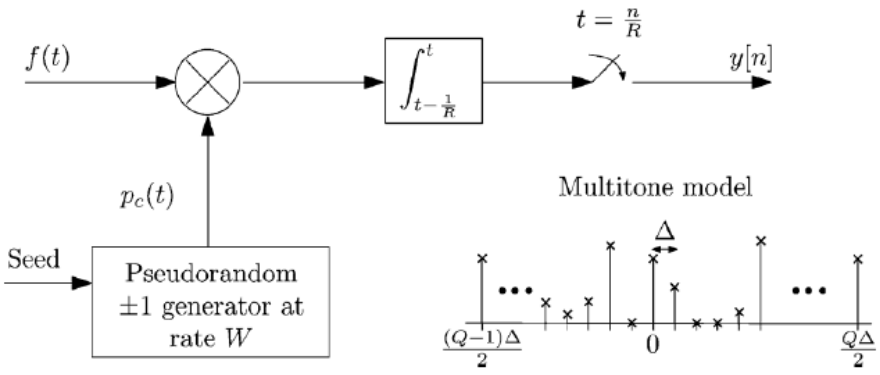

Fig. 4. Block diagram of the random demodulator.

Sampling. The input signal $f(t)$ is mixed by a pseudorandom chipping sequence $p_c(t)$ which alternates at a rate of W. The mixed output is then integrated and dumped at a constant rate R, resulting in the sequence $y[n], 1 \leq n \leq R$. The vector of samples can be written as

$$\mathbf{y} = \mathbf{\Phi x}, \qquad \mathbf{x} = \mathbf{Fs}, \qquad \|\mathbf{s}\|_0 \leq K \tag{15}$$

The matrix $\mathbf{\Phi}$ has dimensions $R \times W$, effectively capturing the mechanism of integration over $\frac{W}{R}$ Nyquist intervals, where the polarity of the input is flipped on each interval according to the chipping function $p_c(t)$. The W-squared discrete Fourier transform (DFT) matrix $\mathbf{F}$ accounts for the sparsity in the frequency domain. The vector $\mathbf{s}$ has Q entries s_w which are up to a constant scaling from the corresponding tone amplitudes a_w. Since the signal has only K active tones, $\|\mathbf{s}\|_0 \leq K$, where the ℓ_0-norm counts the number of nonzero entries.

Reconstruction. The unknown in (15) is $\mathbf{s}$. This type of problem has received extensive treatment in the CS literature, where $\mathbf{\Phi}$ is referred to as the sensing matrix and $\mathbf{F}$ is termed the sparsity basis of $\mathbf{x}$. Several polynomial-time CS techniques are known to coincide with the true $\mathbf{s}$ under certain conditions on $\mathbf{\Phi F}$. Example techniques include basis pursuit and greedy-type algorithms. Once the sparse $\mathbf{s}$ is found, the amplitudes a_w are determined from s_w by constant scaling, and the output $f(t)$ is synthesized according to (14).

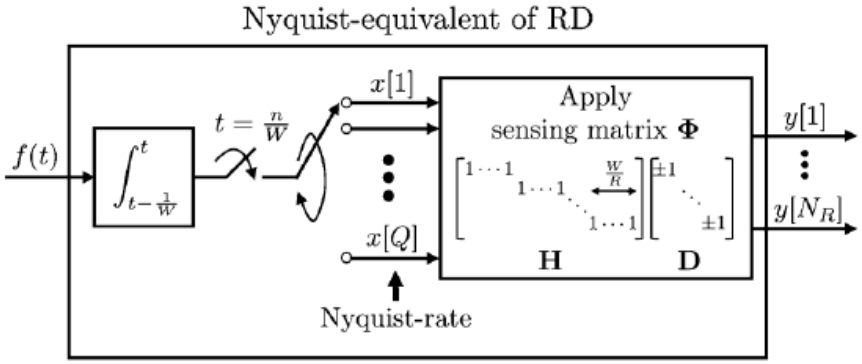

Fig. 5. The Nyquist-equivalents of RD sample the input at its Nyquist rate and apply the sensing matrix digitally.

Fig. 5 depicts the Nyquist-equivalent of RD method, which is the system that samples the input at its Nyquist rate and then computes the relevant sub-Nyquist samples by applying the sensing matrix digitally. The RD-equivalent integrates and dumps the input at rate W, and then applies $\mathbf{\Phi}$ on Q serial measurements, $\mathbf{x} = [x[1], ..., x[Q]]^T$. To coincide with the sub-Nyquist samples of Fig. 4, $\mathbf{\Phi} = \mathbf{HD}$ is used, where $\mathbf{D}$ is diagonal with ± 1 entries, according to the values $p_c(t)$ takes on $t = \frac{n}{W}$, and $\mathbf{H}$ sums over $\frac{W}{R}$ entries.

C. X-DSP: SUB-NYQUIST SIGNAL PROCESSING

Whilst the streaming measurements enter the digital domain at a low rate, they often cannot be used directly for DSP purposes. More broadly, the difficulty in directly processing the X-ADC output stems from the fact that popular DSP algorithms assume an input stream at the Nyquist rate. A fundamental reason for processing at the Nyquist rate is the clear relation between the spectrum of $x(t)$ and that of its pointwise values $x(nT)$, so that digital operations can be easily substituted for their continuous counterparts. Digital filtering is an example where this relation is successfully exploited. Since the power spectral densities of continuous and discrete random processes are associated in a similar manner, estimation and detection of parameters of analog signals can be performed by DSP. When sampling below Nyquist, this key relation no longer holds in general.

Therefore, the authors of [6] addressed the challenge of lowrate DSP and developed the Back-DSP algorithm, which completes the X-DSP functionality for another system (MWC system), with lowrate processing options via a smooth interface to standard DSP packages.

IV. Signal Processing With Compressive Measurements

In recent decades, the digital signal processing (DSP) community has enjoyed enormous success in developing algorithms for capturing and extracting information from signals. Capitalizing on the early work of Whitaker, Nyquist, and Shannon on sampling and representation of continuous signals, signal processing has moved from the analog to the digital domain. Digitization has enabled the creation of sensing and processing systems that are more robust, flexible, cheaper and, therefore, more ubiquitous than their analog counterparts.

As a result of this success, the amount of data generated by sensing systems has grown from a trickle to a torrent. We are thus confronted with the following challenges:

1) acquiring signals at ever higher sampling rates;

2) storing the resulting large amounts of data;

3) processing/analyzing large amounts of data.

Until recently, the first challenge was largely ignored by the signal processing community, with most advances being made by hardware designers. Meanwhile, the signal processing community has made tremendous progress on the remaining two challenges, largely via research in the fields of modeling, compression, and dimensionality reduction. However, the solutions to these problems typically rely on having a complete set of digital samples. Only after the signal is acquired, presumably using a sensor tailored to the signals of interest, could one distill the necessary information ultimately required by the user or the application. This requirement has placed a growing burden on analog-to-digital converters. As the required sampling rate is pushed higher, these devices move inevitably toward a physical barrier, beyond which their design becomes increasingly difficult and costly.

Thus, in recent years, the signal processing community has also begun to address the problem of signal acquisition more directly by leveraging its successes in addressing the last two above-mentioned challenges. In particular, CS method has emerged as a framework that can significantly reduce the acquisition cost at a sensor. Up to now, most of the CS literature has focused on improving the speed and accuracy of the signal recovery, i.e., the problem of recovering the signal from the measurements.

However, signal recovery is not actually necessary in many signal processing applications. Very often we are only interested in solving an inference problem (extracting certain information from measurements) or in filtering out information that is not of interest before further processing. While one could always attempt to recover the full signal from the compressive measurements and then solve the inference or filtering problem using traditional DSP techniques, this approach is typically suboptimal in terms of both accuracy and efficiency.

Recently, [7] takes some initial steps towards a general framework for what the authors call **compressive signal processing (CSP)**, an alternative approach in which signal processing problems are solved directly in the compressive measurement domain without first resorting to a full-scale signal reconstruction. In espousing the potential of CSP, the authors of [7] focus on four fundamental signal processing problems: detection, classification, estimation, and filtering. The first three enable the extraction of information from the samples, while the last enables the removal of irrelevant information and separation of signals into distinct components. While these choices do not exhaust the set of canonical signal processing operations, the authors of [7] believe that they provide a strong initial foundation.

As has been demonstrated in the CS literature, for example, random measurements can be used to acquire any sparse signal without requiring advance knowledge of the locations of the nonzero coefficients. Thus, compressive measurements are agnostic in the sense that they capture the relevant information for the entire class of possible K-sparse signals. The authors of [7] have extended this concept to the CSP framework and demonstrated that it is possible to design agnostic measurement schemes that preserve the necessary structure of large signal classes in a variety of signal processing settings.

A. Compressive Sensing and Restricted Isometries

In the standard CS framework, we acquire a signal $\mathbf{x} \in \mathbb{R}^N$ via the linear measurements

$$\mathbf{y} = \mathbf{\Phi x} \tag{16}$$

where $\mathbf{\Phi}$ is an $M \times N$ matrix representing the sampling system and $\mathbf{y}$ is the vector of measurements. The CS theory, allows for $M \ll N$ as long as the signal $\mathbf{x}$ is sparse or compressible in some basis.

To understand how many measurements are required to enable the recovery of a signal $\mathbf{x}$, we must first examine the properties of $\mathbf{\Phi}$ that guarantee satisfactory performance of the sensing system. In [8], Candes and Tao introduced the **restricted isometry property (RIP)** of a matrix $\mathbf{\Phi}$ and established its important role in CS. We say that a matrix $\mathbf{\Phi}$ satisfies the RIP of order K if there exists a constant $\delta \in (0, 1)$, such that

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \tag{17}$$

holds for all $\mathbf{x} \in \Sigma_K$, where Σ_K is the set of all K-sparse vectors.

The following Theorem bounds the recovery error of $\mathbf{x}$ with respect to the measurement noise and with respect to the ℓ_1-distance from $\mathbf{x}$ to its best K-term approximation denoted $\mathbf{x}_K$ which is defined as:

$$\mathbf{x}_K = arg \min_{\acute{\mathbf{x}}} \|\mathbf{x} - \acute{\mathbf{x}}\|_1.$$

Theorem 1 [Candes]: Suppose that $\mathbf{\Phi}$ satisfies the RIP of order $2K$ with isometry constant $\delta < \sqrt{2}-1$. Given measurements of the form $\mathbf{y} = \mathbf{\Phi x} + \mathbf{e}$, where $\|\mathbf{e}\|_2 \leq \epsilon$, the solution to

$$\hat{\mathbf{x}} = arg \min_{\acute{\mathbf{x}} \in \mathbb{R}^N} \|\acute{\mathbf{x}}\|_1 \qquad \text{subject to } \|\mathbf{\Phi}\acute{\mathbf{x}} - \mathbf{y}\|_2 \leq \epsilon \tag{18}$$

obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \epsilon + C_1 \frac{\|\mathbf{x} - \mathbf{x}_K\|_1}{\sqrt{K}} \tag{19}$$

where

$$C_0 = 4\frac{\sqrt{1+\delta}}{1-(1+\sqrt{2})\delta}, \qquad C_1 = 2\frac{1-(1-\sqrt{2}\delta)}{1-(1+\sqrt{2})\delta}.$$

Note that in practice we may wish to acquire signals that are sparse or compressible with respect to a certain sparsity basis $\mathbf{\Psi}$, i.e., $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\alpha}$, where $\mathbf{\Psi}$ is represented as a unitary $N \times N$ matrix and $\boldsymbol{\alpha} \in \Sigma_K$. In this case, we would require instead that $\mathbf{\Phi\Psi}$ satisfy the RIP, and the performance guarantee would be on $\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_2$.

Before we discuss how one can actually obtain a matrix $\mathbf{\Phi}$ that satisfies the RIP, we observe that we can restate the RIP in a more general form. Let $\delta \in (0,1)$ and $U, V \subset \mathbb{R}^N$ be given. We say that a mapping $\mathbf{\Phi}$ is a δ-stable embedding of (U, V) if

$$(1-\delta)\|\mathbf{u}-\mathbf{v}\|_2^2 \leq \|\mathbf{\Phi u} - \mathbf{\Phi v}\|_2^2 \leq (1+\delta)\|\mathbf{u}-\mathbf{v}\|_2^2 \tag{20}$$

for all $\mathbf{u} \in U$ and $\mathbf{v} \in V$.

We now turn to the more general question of how to construct linear mappings $\mathbf{\Phi}$ that satisfy (20) for particular sets U and V. While it is possible to obtain deterministic constructions of such operators, at present the most efficient designs (i.e., those requiring the fewest number of rows) rely on random matrix constructions.

Several hardware architectures have been proposed that enable the acquisition of compressive measurements in practical settings. Examples include the random demodulator [9], random filtering and random convolution [10], and several compressive imaging architectures [11].

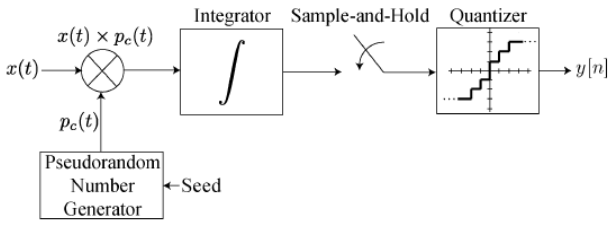

Fig. 6. Random demodulator for obtaining compressive measurements of analog signals.

We briefly describe the random demodulator as an example of such a system. Fig. 6 depicts the block diagram of the random demodulator. The four key components are a pseudo-random ± 1 chipping sequence operating at the Nyquist rate or higher, a low-pass filter, represented by an ideal integrator with reset, a low-rate sample-and-hold, and a quantizer. An input analog signal $x(t)$ is modulated by the chipping sequence and integrated. The output of the integrator is sampled and quantized, and the integrator is reset after each sample.

Mathematically, such a system implements a linear operator that maps the analog input signal to a discrete output vector followed by a quantizer. It is possible to relate this operator to a discrete measurement matrix $\mathbf{\Phi}$ which maps, for example, the Nyquist-rate samples of the input signal to the discrete output vector. The resulting $\mathbf{\Phi}$ matrices, while randomized, typically contain some degree of structure. For example, a random convolution architecture gives rise to a $\mathbf{\Phi}$ matrix with a subsampled Toeplitz structure. While theoretical analysis of these matrices remains a topic of active study in the CS community, there do exist guarantees of stable embeddings for such practical architectures.

B. Detection with Compressive Measurements

We begin by examining the simplest of detection problems. We aim to distinguish between two hypotheses:

$$H_0 : \mathbf{y} = \mathbf{\Phi n}$$

$$H_1 : \mathbf{y} = \mathbf{\Phi}(\mathbf{s} + \mathbf{n})$$

where $\mathbf{s} \in \mathbb{R}^N$ is a known signal, $\mathbf{n} \sim N(0, \sigma^2 \mathbf{I}_N)$ is i.i.d. Gaussian noise, and $\mathbf{\Phi}$ is a known (fixed) measurement matrix. We will consider instead the case where $\mathbf{\Phi}$ is designed without knowledge of $\mathbf{s}$.

To set notation, let

$$P_F = Pr(H_1 \text{ chosen when } \mathrm{H}_0 \text{ true}) \text{ and}$$

$$P_D = Pr(H_1 \text{ chosen when } \mathrm{H}_1 \text{ true})$$

denote the false alarm rate and the detection rate, respectively. The Neyman-Pearson (NP) detector is the decision rule that maximizes P_D subject to the constraint that $P_F < \alpha$. In order to derive the NP detector, we first observe that for the hypotheses H_0 and H_1, we have:

$$f_0(y) = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^T(\sigma^2\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{y}\right)}{(2\pi)^{M/2}|\sigma^2\mathbf{\Phi}\mathbf{\Phi}^T|^{1/2}} \text{ and}$$

$$f_1(y) = \frac{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{\Phi}\mathbf{s})^T(\sigma^2\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}(\mathbf{y}-\mathbf{\Phi}\mathbf{s})\right)}{(2\pi)^{M/2}|\sigma^2\mathbf{\Phi}\mathbf{\Phi}^T|^{1/2}}$$

It is easy to show that the NP-optimal decision rule is to compare the ratio $f_1(y)/f_0(y)$ to a threshold η, i.e., the likelihood ratio test as:

$$\Lambda(y) = \frac{f_1(y)}{f_0(y)} \gtrless \eta,$$

where η is chosen such that

$$P_F = \int_{\Lambda(y)>\eta} f_0(y)d(y) = \alpha.$$

By taking the logarithm, we obtain an equivalent test that simplifies to:

$$\mathbf{y}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{s} \gtrless \underbrace{\sigma^2 log(\eta) + \frac{1}{2}\mathbf{s}^T\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{s}}_{\gamma}.$$

We now define the compressive detector as:

$$t \triangleq \mathbf{y}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{s}. \tag{21}$$

It can be shown that t is a sufficient statistic for our detection problem, and thus contains all of the information relevant for distinguishing between H_0 and H_1.

Using $P_{\mathbf{\Phi}^T} = \mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}$ and $\mathbf{s}^T\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi}\mathbf{s} = \|P_{\mathbf{\Phi}^T}\mathbf{s}\|_2^2$, it is easy to prove

$$\gamma = \sigma\|P_{\mathbf{\Phi}^T}\mathbf{s}\|_2 Q^{-1}(\alpha) \tag{22}$$

and

$$P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \frac{\|P_{\mathbf{\Phi}^T}\mathbf{s}\|_2}{\sigma}\right) \tag{23}$$

Let us now define $\mathrm{SNR} \triangleq \frac{\|\mathbf{s}\|_2^2}{\sigma^2}$. We can bound the performance of the compressive detector as follows.

Theorem 2: Suppose that $\sqrt{N/M}P_{\mathbf{\Phi}^T}$ provides a δ-stable embedding of $(\boldsymbol{S}, \{0\})$. Then, for any $\mathbf{s} \in \mathbf{S}$, we can detect $\mathbf{s}$ with error rate

$$P_D(\alpha) \leq Q\left(Q^{-1}(\alpha) - \sqrt{1+\delta}\sqrt{\frac{M}{N}}\sqrt{\mathrm{SNR}}\right)$$

and

$$P_D(\alpha) \geq Q\left(Q^{-1}(\alpha) - \sqrt{1-\delta}\sqrt{\frac{M}{N}}\sqrt{\mathrm{SNR}}\right).$$

Theorem 2 tells us in a precise way how much information we lose by using random projections rather than the signal samples themselves, not in terms of our ability to recover the signal as is typically addressed in CS, but in terms of our ability to solve a detection problem. Specifically, for typical values of δ

$$P_D(\alpha) \simeq Q\left(Q^{-1}(\alpha) - \sqrt{\frac{M}{N}}\sqrt{\mathrm{SNR}}\right) \tag{24}$$

which increases the miss probability by an amount determined by the SNR and the ratio M/N.

Corollary 1: Suppose that $\sqrt{N/M}P_{\mathbf{\Phi}^T}$ provides a δ-stable embedding of $(\boldsymbol{S}, \{0\})$. Then, for any $\mathbf{s} \in \mathbf{S}$, we can detect $\mathbf{s}$ with success rate

$$P_D(\alpha) \geq 1 - C_2 e^{-C_1 M/N}$$

where C_1 and C_2 are absolute constants depending only on α, δ, and the SNR. Thus, for a fixed SNR and signal length, the detection probability approaches 1 exponentially fast as we increase the number of measurements.

C. Experiments and Discussion

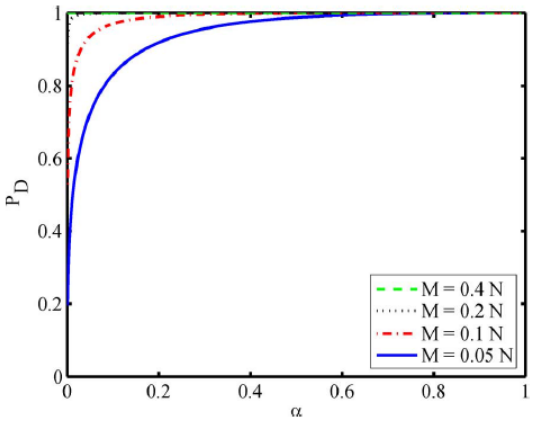

Fig. 7. Effect of M on P_D.

We first explore how M affects the performance of the compressive detector. As described above, decreasing M does cause a degradation in performance. However, in certain cases (relatively high SNR; 20 dB in this example), the compressive detector can perform almost as well as the traditional detector with a very small fraction of the number of measurements required by traditional detection. Specifically, in Fig. 7 we illustrate the receiver operating characteristic (ROC) curve, i.e., the relationship between P_D and P_{FA} predicted by (24). Observe that as M increases, the ROC curve approaches the upper-left corner, meaning that we can achieve very high detection rates while simultaneously keeping the false alarm rate very low. As M grows we see that we rapidly reach a regime where any additional increase in M yields only marginal improvements in the tradeoff between P_D and P_{FA}.

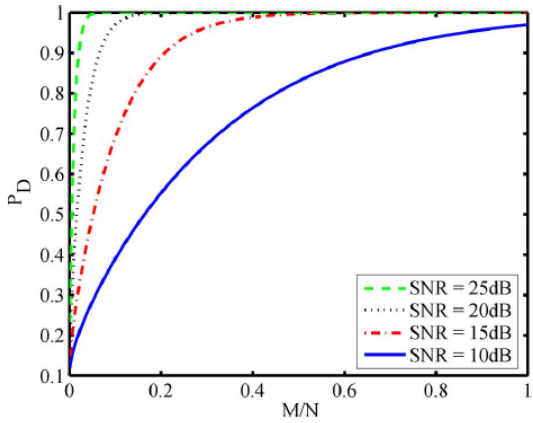

Fig. 8. Effect of M on P_D at different SNR values.

Furthermore, the exponential increase in the detection probability as we take more measurements is illustrated in Fig. 8, which plots the performance predicted by (24) for a range of SNRs with $\alpha = 0.1$. However, we again note that in practice this rate can be significantly affected by the SNR, which determines the constants in the bound of Corrolary 1.

V. MIMO Radar Using Compressive Sampling

The authors of [12]–[14] proposed a MIMO radar system that can be implemented by a small-sized wireless network. Network nodes serve as transmitters or receivers. Transmit nodes transmit uncorrelated waveforms. Each receive node applies compressive sampling to the received signal to obtain a small number of samples, which the node subsequently forwards them to a fusion center. Assuming that the targets are sparsely located in the angle-Doppler space, the fusion center formulates an l_1-optimization problem, the solution of which yields target angle and Doppler information. The proposed approach can extract target information based on a small number of measurements from one or more receive nodes. In particular, for a mild jammer, the proposed method has been shown to be at least as good as the Capon, APES, GLRT, and MUSIC techniques while using a significantly smaller number of samples. In the case of strong thermal noise and strong jammer, the proposed method performs slightly worse than the GLRT method. In that case, its performance is still acceptable, especially if one takes into account the fact that it uses significantly fewer samples than GLRT.

An important feature of the proposed approach is energy saving. The obtained energy saving would be significant in prolonging the life of the wireless network. The proposed approach assumes that nodes are synchronized and the fusion center has perfect node location information. The effects of localization and synchronization errors and ways to mitigate them need to be further studies.

A. System Model

i) M_t transmit nodes, location of i-th transmit node (r_i^t, α_i^t)

ii) N_r receive nodes, location of i-th receive node (r_i^r, α_i^r)

iii) K point targets, k-th target is at azimuth angle θ_k and moves with constant radial speed v_k, its range is $d_k(t) = d_k(0) - v_k t$, where $d_k(0)$ denotes the distance between k-th target and the origin at time equal to zero.

The i-th transmit node transmits the continues-time waveform $x_i(t)e^{j2\pi ft}$ (complex baseband equivalent signal), where $x_i(t)$ is a periodic pulse with duration T_p and pulse repetition interval (PRI) T and $T_p \ll T$. Throughout [12], the following assumptions are considered:

1- Far field assumption, $d_k(t) \gg r_i^{t/r}$, which results in:

$$d_{ik}^{t/r}(t) \approx d_k(t) - \eta_i^{t/r}(\theta_k) = d_k(0) - v_k t - \eta_i^{t/r}(\theta_k), \tag{25}$$

where $\eta_i^{t/r}(\theta_k) = r_i^{t/r} cos(\theta_k - \alpha_i^{t/r})$.

2- Clutter-free environment

3- Perfect synchronization

4- The location of all nodes is available

5- $x_i(t)$ is a narrow-band signal

6- Slowly moving targets, i.e., $\frac{v_k f T_p}{c} \ll 1$

Based on above assumptions, the received signal at k-th target is modeled as:

$$y_k(t) = \beta_k \sum_{i=1}^{M_t} x_i \left(t - \frac{d_{ik}^t(t)}{c} \right) e^{j2\pi f \left(t - \frac{d_{ik}^t(t)}{c} \right)} \tag{26}$$

where $\beta_k, k = 1, ..., K$ are complex amplitudes proportional to the RCS and are assumed to be the same for all the receivers. In addition, the received signal at l-th receive antenna is given by:

$$\begin{aligned} z_l(t) &= \sum_{k=1}^{K} y_k \left(t - \frac{d_{lk}^r(t)}{c} \right) + \varepsilon_l(t) \\ &= \sum_{k=1}^{K} \beta_k \sum_{i=1}^{M_t} x_i \left(t - \frac{d_{ik}^t(t) + d_{lk}^r(t)}{c} \right) e^{j2\pi f \left(t - \frac{d_{ik}^t(t) + d_{lk}^r(t)}{c} \right)} + \varepsilon_l(t) \end{aligned} \tag{27}$$

Considering the fact that $d_{ik}^t(t) + d_{lk}^r(t) \approx 2d_k(0) - 2v_k t - \eta_i^t(\theta_k) - \eta_l^r(\theta_k)$, and recalling the narrow-band assumption of transmit waveforms along with slowly moving target and far-field assumptions, leads us to the following equations:

$$x_i \left(t - \frac{d_{ik}^t(t) + d_{lk}^r(t)}{c} \right) = x_i \left(t - \frac{2d_k(0) - 2v_k t - \eta_i^t(\theta_k) - \eta_l^r(\theta_k)}{c} \right) \approx x_i \left(t - \frac{2d_k(0)}{c} \right) \tag{28}$$

and

$$\begin{aligned} j2\pi f \left(t - \frac{2d_k(0) - 2v_k t - \eta_i^t(\theta_k) - \eta_l^r(\theta_k)}{c} \right) &= j2\pi f t + j2\pi \frac{2v_k f}{c} t + j\frac{2\pi}{\lambda} \left(-2d_k(0) + \eta_i^t(\theta_k) + \eta_l^r(\theta_k) \right) \\ &= j2\pi f t + j2\pi \frac{2v_k f}{c} t - j\frac{2\pi}{\lambda} 2d_k(0) + j\frac{2\pi}{\lambda} \eta_i^t(\theta_k) + j\frac{2\pi}{\lambda} \eta_l^r(\theta_k) \\ &= j2\pi f t + j2\pi f_k t - j\frac{2\pi}{\lambda} 2d_k(0) + j\frac{2\pi}{\lambda} \eta_i^t(\theta_k) + j\frac{2\pi}{\lambda} \eta_l^r(\theta_k), \end{aligned} \tag{29}$$

where $f_k = \frac{2v_k f}{c}$ is the doppler shift caused by the k-th target. Removing $e^{j2\pi f t}$ from $x_i(t) e^{j2\pi f t}$, the

demodulated baseband signal can be approximated by:

$$\begin{aligned} z_l(t) &\approx \sum_{k=1}^{K}\beta_k\sum_{i=1}^{M_t}x_i(t)e^{j2\pi f_k t}e^{-j\frac{2\pi}{\lambda}2d_k(0)}e^{j\frac{2\pi}{\lambda}\eta_i^t(\theta_k)}e^{j\frac{2\pi}{\lambda}\eta_l^r(\theta_k)}+\varepsilon_l(t) \\ &= \sum_{k=1}^{K}\beta_k e^{-j\frac{2\pi}{\lambda}2d_k(0)}e^{j\frac{2\pi}{\lambda}\eta_l^r(\theta_k)}e^{j2\pi f_k t}\mathbf{x}^T(t)\mathbf{v}(\theta_k)+\epsilon_l(t) \end{aligned} \tag{30}$$

where

$$\mathbf{x} = [x_1(t), x_2(t), ..., x_{M_t}(t)]^T_{M_t\times 1} \tag{31}$$

$$\mathbf{v}(\theta_k) = \left[e^{j\frac{2\pi}{\lambda}\eta_1^t(\theta_k)}, e^{j\frac{2\pi}{\lambda}\eta_2^t(\theta_k)}, ..., e^{j\frac{2\pi}{\lambda}\eta_{M_t}^t(\theta_k)}\right]^T_{M_t\times 1} \tag{32}$$

B. Nyquist Sampling

The l-th node samples the return signal to obtain L samples per pulse, where $L=\frac{T_p}{T_s}$ denotes the number of T_s-spaced samples of the transmitted waveforms within one pulse. The Received samples collected during the m-th pulse are given by:

$$\mathbf{z}_{lm}=\begin{bmatrix} z_l[(m-1)T] \\ z_l[(m-1)T+T_s] \\ z_l[(m-1)T+2T_s] \\ \vdots \\ z_l[(m-1)T+(L-1)T_s] \end{bmatrix}_{L\times 1} = \sum_{k=1}^{K}\gamma_k e^{j\frac{2\pi}{\lambda}\eta_l^r(\theta_k)}e^{j2\pi f_k(m-1)T}\mathbf{D}(f_k)\mathbf{X}\mathbf{v}(\theta_k)+\boldsymbol{\epsilon}_{lm}, \tag{33}$$

where $\gamma_k=\beta_k e^{-j\frac{2\pi}{\lambda}2d_k(0)}$,

$$\mathbf{D}(f_k) = diag\left(\left[1, e^{j2\pi f_k T_s}, ..., e^{j2\pi f_k(L-1)T_s}\right]\right)_{L\times L} \tag{34}$$

$$\mathbf{X} = [\mathbf{x}[0], \mathbf{x}[T_s], ..., \mathbf{x}[(L-1)T_s]]^T_{L\times M_t} \tag{35}$$

$$\boldsymbol{\epsilon}_{lm} = [\epsilon_l[(m-1)T], \epsilon_l[(m-1)T+T_s], ..., \epsilon_l[(m-1)T+(L-1)T_s]]^T_{L\times 1} \tag{36}$$

C. Compress Sampling

Let us discretize the angle-Doppler plane on a fine grid

$$\mathbf{a} = [(a_1, b_1), (a_2, b_2), ..., (a_N, b_N)]\,. \tag{37}$$

We can rewrite (33) as

$$\mathbf{z}_{lm} = \sum_{n=1}^{N} s_n e^{j\frac{2\pi}{\lambda}\eta_l^r(a_n)}e^{j2\pi b_n(m-1)T}\mathbf{D}(b_n)\mathbf{X}\mathbf{v}(a_n)+\boldsymbol{\epsilon}_{lm}, \tag{38}$$

where

$$s_n = \left\{ \begin{array}{c} \gamma_k, \text{ if the k} - \text{th target is at } (\text{a}_\text{n}, \text{b}_\text{n}) \\ 0, \qquad \text{otherwise} \end{array} \right\} \tag{39}$$

In matrix form we have

$$\mathbf{z}_{lm} = \mathbf{\Psi}_{lm}\mathbf{s} + \boldsymbol{\epsilon}_{lm}, \tag{40}$$

where $\mathbf{s} = [s_1, s_2, ..., s_N]^T_{N\times 1}$ and

$$\mathbf{\Psi}_{lm} = \left[e^{j\frac{2\pi}{\lambda}\eta^r_l(a_1)} e^{j2\pi b_1(m-1)T} \mathbf{D}(b_1)\mathbf{X}\mathbf{v}(a_1), ..., e^{j\frac{2\pi}{\lambda}\eta^r_l(a_N)} e^{j2\pi b_N(m-1)T} \mathbf{D}(b_N)\mathbf{X}\mathbf{v}(a_N) \right]_{L\times N} \tag{41}$$

It is worthy to mention that $\eta^r_l(a_k) = r^r_l cos(a_k - \alpha^r_l)$, where (r^r_l, α^r_l) is the location of the l-th Receiver in the polar coordinates and

$$\mathbf{D}(b_k) = diag\left(\left[1, e^{j2\pi b_k T_s}, ..., e^{j2\pi b_k (L-1)T_s} \right] \right)_{L\times L} \tag{42}$$

$$\mathbf{X} = \begin{bmatrix} x_1(0) & x_1(T_s) & \dots & x_1[(L-1)T_s] \\ x_2(0) & x_2(T_s) & \dots & x_2[(L-1)T_s] \\ \vdots & \vdots & \vdots & \vdots \\ x_{M_t}(0) & x_{M_t}(T_s) & \dots & x_{M_t}[(L-1)T_s] \end{bmatrix}^T_{L\times M_t} \tag{43}$$

$$\mathbf{v}(a_k) = \left[e^{j\frac{2\pi}{\lambda}\eta^t_1(a_k)}, e^{j\frac{2\pi}{\lambda}\eta^t_2(a_k)}, ..., e^{j\frac{2\pi}{\lambda}\eta^t_{M_t}(a_k)} \right]^T_{M_t\times 1} \tag{44}$$

where $\eta^t_i(a_k) = r^t_i cos(a_k - \alpha^t_i)$. Here, (r^t_i, α^t_i) is the location of the i-th Transmitter in the polar coordinates.

Assuming that there are only a small number of targets, the positions of targets are sparse in the angle-Doppler plane, i.e., $\mathbf{s}$ is a sparse vector. Let us measure linear projections of $\mathbf{z}_{lm}$ as

$$\mathbf{r}_{lm} = \mathbf{\Phi}_{lm}\mathbf{z}_{lm} = \mathbf{\Phi}_{lm}\mathbf{\Psi}_{lm}\mathbf{s} + \mathbf{\Phi}_{lm}\boldsymbol{\epsilon}_{lm} = \mathbf{\Phi}_{lm}\mathbf{\Psi}_{lm}\mathbf{s} + \tilde{\boldsymbol{\epsilon}}_{lm}, \tag{45}$$

where $\mathbf{r}_{lm}$ denotes an $M \times 1$ vector, $\mathbf{\Phi}_{lm}$ is an $M \times L$ $(M < L)$ zero-mean Gaussian random matrix that has small correlation with $\mathbf{\Psi}_{lm}$, and M must be larger than the number of targets $(M > K)$.

If the l-th node in the network knows which nodes are the transmit nodes and also knows the transmitters coordinates relative to a fixed point in the network, then the node is able to construct the matrix $\mathbf{\Psi}_{lm}$. Hence, it can recover the sparse vector via l_1-optimization approach, based on the nodes own received data. If no such location information is available to the node, or the interference is strong, then the receive node will pass the linear projections $\mathbf{r}_{lm}$ to a fusion center, which has global and local information.

Combining the output of N_p pulses at N_r receive antennas, the fusion center can formulate the equation

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_{11} \\ \vdots \\ \mathbf{r}_{1N_p} \\ \vdots \\ \mathbf{r}_{N_r1} \\ \vdots \\ \mathbf{r}_{N_rN_p} \end{bmatrix}_{MN_pN_r\times 1} = \begin{bmatrix} (\mathbf{\Phi}_{11}\mathbf{\Psi}_{11}) \\ \vdots \\ (\mathbf{\Phi}_{1N_p}\mathbf{\Psi}_{1N_p}) \\ \vdots \\ (\mathbf{\Phi}_{N_r1}\mathbf{\Psi}_{N_r1}) \\ \vdots \\ (\mathbf{\Phi}_{N_rN_p}\mathbf{\Psi}_{N_rN_p}) \end{bmatrix}_{MN_pN_r\times N} \mathbf{s} + \begin{bmatrix} \tilde{\boldsymbol{\epsilon}}_{11} \\ \vdots \\ \tilde{\boldsymbol{\epsilon}}_{1N_p} \\ \vdots \\ \tilde{\boldsymbol{\epsilon}}_{N_r1} \\ \vdots \\ \tilde{\boldsymbol{\epsilon}}_{N_rN_p} \end{bmatrix}_{MN_pN_r\times 1} = \mathbf{\Theta}\mathbf{s} + \mathbf{E} \quad (46)$$

Thus, the fusion center can recover $\mathbf{s}$ by applying the Dantzig selector to the convex problem of (46) as follows:

$$\widehat{\mathbf{s}} = \min \|\mathbf{s}\|_1 \qquad \text{Subject to } \|\mathbf{\Theta}^{\mathrm{H}}(\mathbf{r} - \mathbf{\Theta}\mathbf{s})\|_\infty < \mu, \quad (47)$$

where $\mu < \|\mathbf{\Theta}^H\mathbf{r}\|_\infty$.

The schematic diagram of the receiver is shown in Fig. 9.

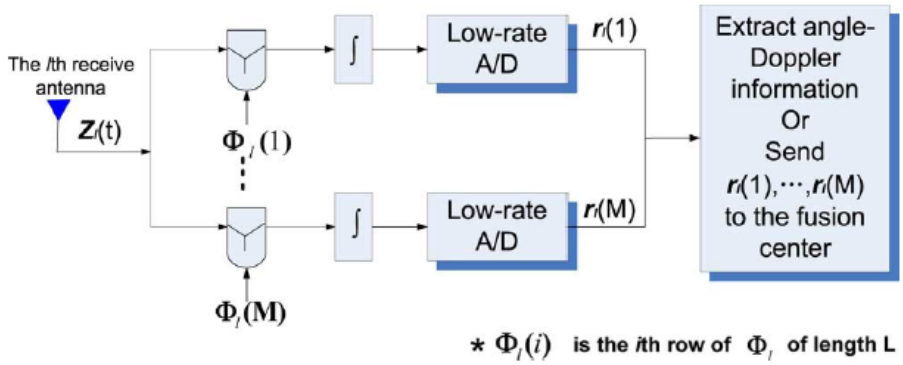

Fig. 9. Schematic diagram of the receiver. $\mathbf{\Phi}_l$ denotes the measurement matrix for the l-th receive node.

1) Resolution: As before mentioned, the UUP indicates that if every set of columns with cardinality less than the sparsity of the signal of interest of the sensing matrix are approximately orthogonal, then the sparse signal can be exactly recovered with high probability. For a fixed M, the correlation of columns of the sensing matrix can be reduced if the number of pulses N_p and/or the number of receive nodes N_r is increased. Intuitively, the increase in N_p and N_r increases the dimension of the sensing matrix columns, thereby rendering the columns less similar to each other. Moreover, increasing the number of transmit nodes, i.e., M_t, also reduces the correlation of columns.

In general, to achieve high resolution, a fine grid is required. However, for fixed M_t, N_p and N_r, decreasing the distance between the grid points would result in more correlated columns in the sensing matrix. Also, based on the theory of CS, the effects of a higher column correlation can be mitigated by using a larger number of measurements, i.e., by increasing M.

One might tend to think that in order to achieve good resolution one has to involve a lot of measurements, or trasnmit/receive antennas, or pulses, which in turn would involve high complexity. However, extensive simulations suggest that this is not the case. In fact, the proposed approach can match the resolution that can be achieved with conventional methods, while using far fewer received samples, than those used by the conventional methods.

2) Maximum Grid Size for the Angle-Doppler Space: The grid in the angle-Doppler space must be selected so that the targets that do not fall on the chosen grid points can still be captured by the closest grid points. This requires sufficiently high correlation of the signal reflected by each target with the columns of $\mathbf{\Theta}$ corresponding to grid points close to the targets in the angle-Doppler plane. However, this requirement goes against the UUP, which requires that every set of columns with cardinality less than the sparsity of the signal of interest be approximately orthogonal. Thus, there is a tradeoff of the correlation of between the columns of the sensing matrix and the grid size.

D. Performance Analysis in the Presence of a Jammer Signal

It is clear that, in the presence of a jammer that transmits a waveform which is uncorrelated with the radar transmit waveforms, the performance of the proposed CS method will deteriorate. Suppose that each transmitter transmits N_p pulses. Recalling (28) and (29), in the presence of a jammer at location (d, θ), the received signal at the l-th receive antenna during the m-th pulse can be expressed as:

$$\mathbf{r}_{lm} = \mathbf{\Phi}_{lm}\mathbf{z}_{lm} = \sum_{k=1}^{K} \gamma_k e^{j\frac{2\pi}{\lambda}\eta_l^r(\theta_k)} \mathbf{\Phi}_{lm} e^{j2\pi f_k(m-1)T} \mathbf{D}(f_k)\mathbf{X}\mathbf{v}(\theta_k) + \beta e^{-j\frac{2\pi}{\lambda}(d-\eta_l^r(\theta))} \mathbf{\Phi}_{lm}\widetilde{\mathbf{x}}_m + \mathbf{\Phi}_{lm}\boldsymbol{\epsilon}_{lm}, \quad (48)$$

where $\widetilde{\mathbf{x}}_m = [\widetilde{x}_m(0), \widetilde{x}_m(T_s), ..., \widetilde{x}_m[(L-1)T_s]_{L\times 1}^T$ contains the samples of the signal transmitted by the jammer during the m-th pulse, and β denotes the square root of the power of the jammer over the

duration of one signal pulse. We can rewrite (48) as

$$\mathbf{r}_{lm} = \begin{bmatrix} \mathbf{r}_{l_1} \\ \vdots \\ \mathbf{r}_{lN_p} \end{bmatrix}_{MN_p \times 1} = \sum_{k=1}^{K} \gamma_k e^{j\frac{2\pi}{\lambda}\eta_l^r(\theta_k)} \begin{bmatrix} \mathbf{\Phi}_{l1} \\ \vdots \\ \mathbf{\Phi}_{lN_p} e^{j2\pi f_k(N_p-1)T} \end{bmatrix} \mathbf{D}(f_k)\mathbf{X}\mathbf{v}(\theta_k) + \beta e^{-j\frac{2\pi}{\lambda}(d-\eta_l^r(\theta))} \begin{bmatrix} \mathbf{\Phi}_{l1}\widetilde{\mathbf{x}}_1 \\ \vdots \\ \mathbf{\Phi}_{lN_p}\widetilde{\mathbf{x}}_{N_p} \end{bmatrix} + \begin{bmatrix} \mathbf{\Phi}_{l1}\boldsymbol{\epsilon}_{l1} \\ \vdots \\ \mathbf{\Phi}_{lN_p}\boldsymbol{\epsilon}_{lN_p} \end{bmatrix} \quad (49)$$

The following assumptions are considered for jammer:

1 - Jammer is located at (d, θ).

2 - For all m, $E\{\widetilde{x}_m^*(i)\widetilde{x}_m^*(j)\} = 1/L$, and 0 otherwise. Thus, $E\{\widetilde{\mathbf{x}}_m^H(i)\widetilde{\mathbf{x}}_m^*(j)\} = 1$.

3 - $\widetilde{x}_m, m = 1, ..., N_p$ are uncorrelated with the main period of the transmitted waveforms. Thus, the effect of the jammer signal is similar to that of additive noise. In the related analysis it is assumed that the jammer contribution is much stronger than that of additive noise, and therefore the third term on the right-hand side of (49) is ignored.

4 - All receive nodes use the same random measurement matrix over N_p pulses, i.e., $\mathbf{\Phi}_l = \mathbf{\Phi}_{l1} = ... = \mathbf{\Phi}_{lN_p}$.

Based on the above-mentioned assumptions, the average power of the desirable signal taken over the positions of TX/RX nodes can be approximated by:

$$P_s(l) \approx N_p \sum_{k=1}^{K} |\beta_k|^2 \sum_i q_{i,i}^{k,k} + \sum_{k \neq k'} \beta_k^* \beta_{k'} e^{j4\pi/\lambda(d_k(0)-d_{k'}(0))} \varsigma_{kk'}{}^2 \mu_{kk'} \sum_i q_{i,i}^{k,k'}$$

Similarly, the average power of the jammer signal over TX/RX locations is given by:

$$P_j(l) = E\{\mathbf{r}_{lj}^H \mathbf{r}_{lj}\} = |\beta|^2 \sum_{m=1}^{N_p} \widetilde{\mathbf{x}}_m^H \mathbf{\Phi}_l^H \mathbf{\Phi}_l \widetilde{\mathbf{x}}_m. \quad (50)$$

The SJR is defined as the ratio of the power of the signal to the power of the jammer. Since the denominator does not depend on node locations, the average SJR equals

$$SJR = \frac{P_s(l)}{P_j(l)} \quad (51)$$

SJR Based on a Modified Measurement Matrix:

Since the jammer signal is uncorrelated with the transmitted signal, the SJR can be improved by correlating the jammer signal with the transmitted signal. Therefore, the following measurement matrix is proposed:

$$\widetilde{\mathbf{\Phi}}_l = \acute{\mathbf{\Phi}}_l \mathbf{X}^H \tag{52}$$

where $\acute{\mathbf{\Phi}}_l$ is an $M \times M_t$ Gaussian random matrix. Note that $\widetilde{\mathbf{\Phi}}_l$ is also Gaussian. A random measurement matrix with i.i.d. entries, e.g., Gaussian or ± 1 random variables, is nearly incoherent with any fixed basis matrix. Therefore, the proposed measurement matrix exhibits low coherence with $\mathbf{\Psi}_l$, thus guaranteeing a stable solution to (47). Applying some minor changes, the average power of the desirable signal given before is usable. The average power of the jammer signal is given by (50) where $\mathbf{\Phi}_l$ is replaced by $\widetilde{\mathbf{\Phi}}_l$.

In summary, the use of $\acute{\mathbf{\Phi}}_l$ instead of $\acute{\mathbf{\Phi}}_l$ improves SJR by a factor of L/M_t when $L > M_t$. The SJR can be improved by an increase in L. However, increasing L will require a smaller T_s when the pulse duration is fixed. This will increase the bandwidth of the signal. Finally, it is important to note that the obtained SJR do not depend on the number of measurements, M.

VI. Target Estimation Using Sparse Modeling for Distributed MIMO Radar

MIMO radar is typically used in two antenna configurations, namely **distributed** and **co-located**. In distributed MIMO radar the antennas are widely separated. This enables viewing the target from different angles. Hence, if the target returns between a particular transmitter and receiver are weak, then it is highly likely that they will be compensated by the returns between other antenna pairs. While distributed MIMO radar exploits the spatial diversity, co-located MIMO radar exploits the waveform diversity. In co-located configuration, all the antennas are closely spaced and hence the target radar cross section (RCS) values are the same for all transmitter-receiver pairs. RCS denotes the transformation undergone by the transmitted signal while reflecting from the surface of the target. Since all the antennas are closely spaced in the co-located configuration, each antenna pair will have the same RCS value as all the signals bounce off the target surface at the same angle.

A. System Model

The authors of [15] assume that there are M_T transmitters, M_R receivers, and K targets. Further, they assume that all the targets are moving in a two dimensional plane. The RCS center of gravity of the target is located at $\overrightarrow{\mathbf{p}^k} = [p_x^k, p_y^k]$ on a Cartesian coordinate system and it moves with a velocity $\overrightarrow{\mathbf{v}^k} = [v_x^k, v_y^k]$. The i-th transmitter and j-th receiver are located at $\overrightarrow{\mathbf{t}^i} = [t_{i_x}, t_{i_y}]$ and $\overrightarrow{\mathbf{r}^j} = [r_{j_x}, r_{j_y}]$, respectively. Let

$w_i(t)$ be the complex baseband waveform transmitted from the transmitter. Then, the bandpass signal emanating from the i-th transmit antenna is given as

$$\tilde{w}_i(t) = Real\left(w_i(t)e^{j2\pi f_c t}\right) \tag{53}$$

Under a narrow band assumption on the waveforms, the bandpass signal arriving at the j-th receiver can be expressed as

$$y_j(t) = Real\left(\sum_{k=1}^{K}\sum_{i=1}^{M_T} a_{ij}^k(t)w_i(t-\tau_{ij}^k)e^{j2\pi\left(f_{D_{ij}}^k(t-\tau_{ij}^k)+f_c(t-\tau_{ij}^k)\right)}\right). \tag{54}$$

The received signals at each receiver are first down converted from the radio frequency and then passed through a bank of matched filters, each of which corresponds to a particular transmitter. The sampled outputs of the i-th matched filter at the j-th receiver are given as:

$$y_{ij}(n) = \sum_k a_{ij}^k(n)e^{j2\pi\left(f_{D_{ij}}^k(nT_s-\tau_{ij}^k)-f_c\tau_{ij}^k\right)} + e_{ij}(n). \tag{55}$$

We define the target state vector $\boldsymbol{\zeta} = [p_x, p_y, v_x, v_y]^T$. Hence, the important properties of the target (position, velocity) are specified by $\boldsymbol{\zeta}$. The goal is to estimate $\boldsymbol{\zeta}$ for all the K targets. Now, the target state space is discretized into a grid of L possible values $\boldsymbol{\zeta}^l, l = 1, ..., L$. Hence, each of the targets is associated with a state vector belonging to this grid. If the presence of a target at $\boldsymbol{\zeta}^l$ would contribute to the matched filter output at n, then we have

$$\psi_{ij}^l(n) = e^{j2\pi\left(f_{D_{ij}}^l(nT_s-\tau_{ij}^l)-f_c\tau_{ij}^l\right)} \tag{56}$$

otherwise, $\psi_{ij}^l(n) = 0$. Also, if $\boldsymbol{\zeta}^l$ is the state vector of the k-th target, we define

$$s_{ij}^l(n) = a_{ij}^k(n) \tag{57}$$

Otherwise, $s_{ij}^l(n) = 0$. For each j, vectors stack as:

$$\mathbf{s}_j^l(n) = [s_{1j}^l(n), s_{2j}^l(n), ..., s_{M_T j}^l(n)]^T \tag{58}$$

$$\mathbf{y}_j(n) = [y_{1j}(n), y_{2j}(n), ..., y_{M_T j}(n)]^T \tag{59}$$

$$\mathbf{e}_j(n) = [e_{1j}(n), e_{2j}(n), ..., e_{M_T j}(n)]^T \tag{60}$$

$$\boldsymbol{\Psi}_j^l(n) = diag[\psi_{1j}^l(n), \psi_{2j}^l(n), ..., \psi_{M_T j}^l(n)] \tag{61}$$

Further, it is possible to arrange them as:

$$\mathbf{s}^l(n) = [(\mathbf{s}_1^l(n))^T, (\mathbf{s}_2^l(n))^T, ..., (\mathbf{s}_{M_R}^l(n))^T]^T \tag{62}$$

$$\mathbf{y}(n) = [(\mathbf{y}_1(n))^T, (\mathbf{y}_2(n))^T, ..., (\mathbf{y}_{M_R}(n))^T]^T \tag{63}$$

$$\mathbf{e}(n) = [(\mathbf{e}_1(n))^T, (\mathbf{e}_2(n))^T, ..., (\mathbf{e}_{M_R}(n))^T]^T \tag{64}$$

$$\mathbf{\Psi}^l(n) = diag[\mathbf{\Psi}_1^l(n), \mathbf{\Psi}_2^l(n), ..., \mathbf{\Psi}_{M_R}^l(n)] \tag{65}$$

Finally, the received vector can be expressed as:

$$\mathbf{y}(n) = \mathbf{\Psi}(n)\mathbf{s}(n) + \mathbf{e}(n) \tag{66}$$

where

$$\mathbf{s}(n) = [(\mathbf{s}^1(n))^T, (\mathbf{s}^2(n))^T, ..., (\mathbf{s}^L(n))^T]^T \tag{67}$$

$$\mathbf{\Psi}(n) = diag[\mathbf{\Psi}^1(n), \mathbf{\Psi}^2(n), ..., \mathbf{\Psi}^L(n)] \tag{68}$$

In the above formulation, $\mathbf{\Psi}(n)$ is known and only $\mathbf{s}(n)$ depends on the actual targets present in the illuminated area. The nonzero entries of $\mathbf{s}(n)$ represent the target attenuation values and the corresponding indices determine the positions and velocities. Further, note that in order to obtain the measurement vector in the above equation, each of the receivers sends their measurements to a common processor that stacks them appropriately to obtain $\mathbf{y}(n)$.

B. Sparse Support Recovery

In order to find the properties of the targets (position, velocity), we need to recover the sparse vector $\mathbf{s}(n)$ from the measurements $\mathbf{y}(n)$. There are many approaches to perform the recovery. Two approaches are **basis pursuit** (BP) and **matching pursuit** (MP). These algorithms are well known in the field of sparse signal processing.

1) Basis Pursuit: BP is an unconstrained minimization problem which can be expressed as:

$$\min \frac{1}{2} \left(\|\mathbf{y}(n) - \mathbf{\Psi}(n)\|_2\right)^2 + \lambda \|\mathbf{s}(n)\|_1 \tag{69}$$

BP problems can be solved by CVX.

2) Block-Matching Pursuit: If the columns of $\mathbf{\Psi}$ are divided into blocks of size $M_T M_R$, L blocks are obtained. In each subsequent iteration, we project the residual vector onto all of these L blocks and pick the block that gives the highest energy after projection. Now, the estimated reconstructed vector is

updated by adding the projections onto each of the columns of this block. The residual is also updated accordingly. Note that the main difference when compared with BP is that here the updates are done one block (columns corresponding to different transmitter and receivers) at a time whereas in BP, the updates are done one column at a time. The improvement in using the block sparsity based recovery algorithms is maximum when all the columns within a block are orthogonal. It can be easily checked that the columns corresponding to different transmitter-receiver pairs in the basis matrix are orthogonal by the definition of $\mathbf{\Psi}$ in the above mentioned signal model.

C. Optimal Adaptive Energy Allocation

Adaptive energy allocation has been shown to provide improved detection performance in distributed MIMO radar systems. In this section, we present a novel adaptive energy allocation scheme to improve the estimation performance.

Let E_i be the energy of the waveform transmitted from the i-th transmitter. The authors of [15] initialize the system by transmitting multiple pulses of unit energy waveforms $E_i = 1$, from all the transmitters. Hence, the total energy transmitted per pulse is $\sum_{i=1}^{M_T} E_i = M_T$. After collecting a vector of outputs at the multiple receiver matched filters, the processor performs the sparse recovery to estimate the attenuations a_{ij}^k using the algorithms mentioned in the previous section. Since the different antenna pairs view the targets from different angles, these attenuations and their corresponding estimates will be different from each other. Hence, equal energy allocation to all the transmitters does not necessarily give the best performance. After the estimation, the energy allocation scheme is applied to decide upon the transmit energies for the next set of transmit pulses while keeping the total transmitted energy constant. The goal of this scheme is to maximize the minimum target returns. This is naturally motivated from the performance metric defined earlier in this section. The numerator in the performance metric denotes the minimum target returns. The authors of [15] solved the following optimization problem and found the optimal E_i such that

$$\max_{E_i} \min_{k} \sum_{i=1}^{M_T} \sum_{j=1}^{M_R} E_i \|\hat{a_{ij}^k}\|^2 \tag{70}$$

Since this problem depends only on the dimensionality of the MIMO radar configuration and the number of targets and not on the huge dimensionality of the basis dictionary, it can be solved quickly. This makes it amenable to use in practical systems in an online manner. After solving this problem, the processor feeds back this information to the transmitters which send the next set of pulses with these optimally selected values of energies. Hence, the system operates in a closed loop. The energy allocation mechanism

discussed above is not only applicable to the sparsity based estimation method mentioned in this paper but it is relevant in any multiple target scenario. Note that it only requires us to have estimates of the attenuations. These estimates are computed using sparse support recovery. In principle, these estimates can be computed using any other approach and still this energy allocation scheme will be relevant.

REFERENCES

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

[2] E. Candes, N. Braun, and M. Wakin, "Sparse signal and image recovery from compressive samples," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, 2007, pp. 976–979.

[3] R. Baraniuk, "Compressive sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, 2008, pp. iv–v.

[4] R. Baraniuk, E. Candes, R. Nowak, and M. Vetterli, "Compressive sampling [from the guest editors]," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 12–13, 2008.

[5] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling (a sensing/sampling paradigm that goes against the common knowledge in data acquisition)," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[6] M. Mishali, Y. Eldar, and A. Elron, "Xampling: Signal acquisition and processing in union of subspaces," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4719–4734, 2011.

[7] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, "Signal processing with compressive measurements," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 445–460, 2010.

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

[9] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 520–544, 2010.

[10] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 3, 2006, pp. III–III.

[11] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 83–91, 2008.

[12] A. Petropulu, Y. Yu, and H. Poor, "Distributed mimo radar using compressive sampling," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, Oct 2008, pp. 203–207.

[13] Y. Yu, A. Petropulu, and H. Poor, "Measurement matrix design for compressive sensing x2013;based mimo radar," *Signal Processing, IEEE Transactions on*, vol. 59, no. 11, pp. 5338–5352, Nov 2011.

[14] ——, "Cssf mimo radar: Compressive-sensing and step-frequency based mimo radar," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 2, pp. 1490–1504, APRIL 2012.

[15] S. Gogineni and A. Nehorai, "Target estimation using sparse modeling for distributed mimo radar," *Signal Processing, IEEE Transactions on*, vol. 59, no. 11, pp. 5315–5325, Nov 2011.