Compressive Sampling in Array Processing

Ali Ahmed and Justin Romberg
Georgia Institute of Technology

***Abstract*—In this paper, we propose a sampling architecture for the efficient acquisition of multiple signals lying in a subspace. We show that without the knowledge of the signal subspace, the proposed sampling architecture acquires the signals at a sub-Nyquist rate. Prior to sampling at a sub-Nyquist rate, the analog signals are diversified using analog preprocessing. The preprocessing step is carried out using implementable components that inject "structured" randomness into the signals. We recast the signal reconstruction from fewer samples as a low-rank matrix recovery problem from generalized linear measurements. Our results also include a sampling theorem that provides the sufficient sampling rate for the exact reconstruction of the signals. We also discuss an application of this sampling architecture in the estimation of the covariance matrix, required for parameter estimation in several important array processing applications, from much fewer samples.**

I. INTRODUCTION

In this paper, we present a sampling scheme for the efficient acquisition of multiple signals lying in a subspace. In addition, we compliment the sampling architecture with a sampling theorem, which dictates the sampling rate required for the reconstruction of the signal ensemble. Multiple signals lying in a subspace often arise from the outputs of a sensor array in various signal processing applications, some of which are outlined in Section III. In such application, often the task is to estimate the signal parameters from their covariance matrix, e.g., the MUSIC algorithm in array processing use the covariance matrix to estimate signal parametrs, such as angle of arrival, and frequency offsets. In several wideband signal processing applications the sampling rate required to acquire the covariance matrix may be prohibitive; especially, in view of the increasing trend of using high frequency spectrum in some applications in array processing. We will show that our proposed sampling architecture can estimate the covariance matrix of the input signal ensemble with a much smaller sampling rate; hence, relieving the burden on the analog-to-digital converters (ADCs).

An ensemble of M signals, each of which is bandlimited to $W/2$ radians/sec can be captured completely at MW samples per second. This can be achieved using M ADCs, one for each signal and each taking samples at rate W. We will show that if the signals lie in a small subspace of dimension $R \ll M$, meaning that all the signals in the ensemble can be written as (or closely approximated by) distinct linear combinations of $R \ll M$ underlying signals, then the net sampling rate can be reduced considerably by using *analog diversification* [1], [2]. The signals will be diversified using implementable analog devices and the resultant signals will then be sampled at a smaller rate. In Section II-A, we will show that these samples can be expressed as linear measurements of a low-rank matrix. Over the course of one second, we want to acquire an $M \times W$ matrix comprised of samples of the ensemble taken at the Nyquist rate. The proposed sampling architecture produces a series of linear combinations of entries of this matrix. The conditions (on the signals and the acquisition system) under which this type of recovery is effective have undergone intensive study in the recent literature; see, for example, [3], [4]

The main contributions of this paper are as follows: first, the design of a *practical* measurement system constructed from components that can be implemented in hardware; second, the statement of Theorem 1, which proves the sub-Nyquist rate acquisition is possible without knowing the signal subspace in advance. We also discuss an application of this sampling scheme in the compressive estimation of the covariance matrices of the signals lying in a subspace.

A. Signal model

We will use notation $\boldsymbol{X_c}(t)$ to denote a signal ensemble of interest and $x_1(t), \ldots, x_M(t)$ to denote the individual signals within that ensemble. Conceptually, we may think of $\boldsymbol{X_c}(t)$ as a "matrix" with finite M number of rows, but each row contains a bandlimited signal. Our underlying assumption is that the signals in the ensemble lie in a subspace $\mathcal{S}$ of dimension R; that is, we can write

$$\boldsymbol{X}_c(t) \approx \boldsymbol{A}\boldsymbol{S}_c(t), \tag{1}$$

where $\boldsymbol{S}_c(t)$ is a smaller signal ensemble with R signals that lie in subspace $\mathcal{S}$ and $\boldsymbol{A}$ is a $M \times R$ matrix with entries $A[m, r]$. We will use the convention that fixed matrices operating to the left of the signal ensembles simply "mix" the signals point-by-point, and so (1) is equivalent to

$$x_m(t) \approx \sum_{r=1}^{R} A[m, r] s_r(t).$$

The only structure we will impose on individual signals is that they are real-valued, bandlimited, and periodic. Thus, signals live in a finite-dimensional linear subspace and provide a natural way of discretizing the problem; that is, what exists in $\boldsymbol{X_c}(t)$ for $t \in [0, 1]$ is all there is to know, and each signal can be captured exactly with W equally-spaced samples, which, for the most part, reduces the clutter in mathematics. In a detailed manuscript under preparation, we discuss how to adapt our results to more realistic signal models in which the (non-periodic) signal is windowed in time and overlapping

978-1-4673-3146-3/13/$31.00 ©2013IEEE

sections are reconstructed jointly. Each bandlimited, periodic signal in the ensemble can be written as

$$x_m(t) = \sum_{f=-B}^{B} \alpha_m[f]\, e^{j2\pi f t},$$

where $\alpha_m[f]$ are complex but have symmetry $\alpha_m[-f] = \alpha_m[f]^*$ to ensure that $x_m(t)$ is real. We can capture $x_m(t)$ perfectly by taking $W = 2B+1$ equally spaced samples per row. We will call this the $M \times W$ matrix of samples $\boldsymbol{X}$; of course, knowing every entry in this matrix is the same as knowing the entire signal ensemble. We can write

$$\boldsymbol{X} = \boldsymbol{C}\boldsymbol{F}, \tag{2}$$

where $\boldsymbol{F}$ is a $W \times W$ normalized discrete Fourier matrix and $\boldsymbol{C}$ is a $M \times W$ matrix whose rows contain Fourier series coefficients for the signals in $\boldsymbol{X}_c(t)$. Matrix $\boldsymbol{F}$ is orthonormal, while $\boldsymbol{C}$ inherits the correlation structure of the original ensemble. An estimate of the covariance matrix of the ensemble $\boldsymbol{X}$ from a reasonably large number of samples is defined as

$$\boldsymbol{R}_{XX} = \frac{1}{W}\boldsymbol{X}\boldsymbol{X}^*.$$

We will be concerned with estimating $\boldsymbol{R}_{XX}$ from much fewer samples than dictated by Shannon-Nyquist framework.

II. The Random Demodulator for Multiple Signals Lying in a Subspace

To efficiently acquire the signal ensemble living in a subspace, our proposed sampling architecture, shown in Fig. 1, follows a two-step approach. In the first step, each of the M signals undergo analog preprocessing, which involves modulation, and low-pass filtering. The modulator takes an input signal $x_m(t)$ and multiplies it by a fixed and known $d_m(t)$. We will take $d_m(t)$ to be a binary ± 1 waveform that is constant over an interval of length $1/W$. Intuitively, the modulation results in the diversification of the signal *information* over the frequency band of width W. The diversified analog signals are then processed by an analog-low-pass filter; implemented using an integrator, see [5] for details. The low-pass filter only selects a frequency sub-band (or a subspace) of width roughly equal to Ω, and as will be shown in Theorem 1, this partial information is enough for the signal reconstruction. The partial information suffices as the signals are scrambled using modulators before low-pass filtering. Note that the low-pass filter in each channel in Fig. 1 can be replaced; in general, by a band-pass filter, i.e., the location of the band does not matter only its width does. This also explains why we don't need to know the subspace in which signals live in advance.

In the second step, the filtered signal is sampled by an ADC in each channel at a lower rate Ω. The result in Theorem 1 asserts that Ω is roughly of a factor of R/M smaller than the Nyquist rate W.

Compressive sampling architectures based on the random modulator have been analyzed previously in the literature [5], [6]. The principal finding is that if the input signal is spectrally sparse (meaning the total size of the support of its Fourier transform is a small percentage of the entire band), then the modulator can be followed by a low-pass filter and an ADC that takes samples at a rate comparable to the size of the active band. This architecture has been implemented in hardware in multiple applications [7], [8].

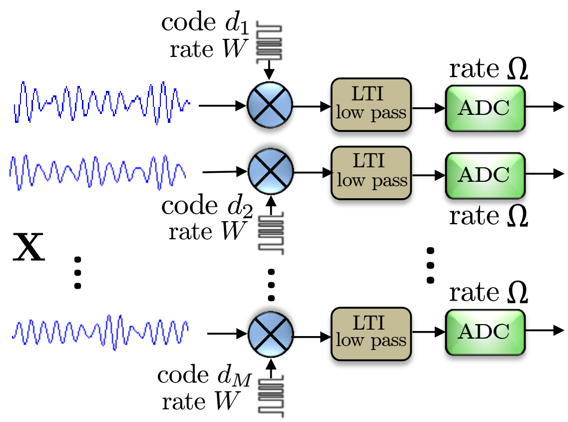

Fig. 1. The random demodulator for multiple signals lying in a subspace: M signals lying in a subspace are preprocessed in analog using a bank of independent modulators, and low-pass filters. The resultant signal is then sampled uniformly by an ADC in each channel operating at rate Ω samples per second. The net sampling rate is $\Delta = \Omega M$ samples per second.

A. System in Matrix Form

Each of the M input signals $x_m(t), 1 \le m \le M$ is multiplied by an independently generated random binary waveform $d_m(t), 1 \le m \le M$ alternating at rate W. That is, the output after the modulation in the mth channel is

$$y_m(t) = x_m(t) \cdot d_m(t), \quad m = 1, \cdots, M, \text{ and } t \in [0, 1)$$

The $y_m(t)$ are then low-pass filtered using an integrator, which integrates $y_m(t)$ over an interval of width $1/\Omega$ and the result is then sampled at rate Ω using an ADC. The samples taken by the ADC in the mth channel are

$$y_m[n] = \int_{(n-1)/\Omega}^{n/\Omega} y_m(t)dt, \quad n = 1, \cdots, \Omega.$$

The integration operation commutes with the modulation process; hence, we can equivalently integrate the signals $x_m(t), 1 \le m \le M$ over the interval of width $1/W$, and treat them as samples $\boldsymbol{X}_0 \in \mathbb{R}^{M\times W}$ of the ensemble $\boldsymbol{X}_c(t)$. The entries $X_0[m,n]$ of the matrix $\boldsymbol{X}_0$ are

$$\begin{aligned} X_0[m,n] &= \int_{(n-1)/W}^{n/W} x_m(t)dt, \\ &= \sum_{|\omega|\le W/2} C[m,\omega]\left[\frac{e^{\iota 2\pi\omega/W}-1}{\iota 2\pi\omega}\right] e^{-\iota 2\pi\omega n/W}, \end{aligned}$$

where the bracketed term representing the low-pass filter

$$\tilde{L}[\omega] = \left[\frac{e^{\iota 2\pi\omega/W}-1}{\iota 2\pi\omega}\right]$$

is evaluated in the window $\omega = 0, \pm 1, \cdots, \pm(W/2-1), W/2$. We will use an equivalent evaluation $L[\omega]$ of $\tilde{L}[\omega]$ in the window $\omega = 1, \cdots, W$. The Fourier coefficients of $C[m,\omega]$ of $\boldsymbol{X}$ defined in (2) are related to the Fourier coefficients $C_0[m,\omega]$ of $\boldsymbol{X}_0$

$$C_0[m,\omega] = C[m,\omega]L[\omega] \quad \omega = 1, \cdots, W, \tag{3}$$

and in time domain

$$\boldsymbol{X}_0 = \boldsymbol{C}_0 \boldsymbol{L} \boldsymbol{F}, \tag{4}$$

where $\boldsymbol{L}$ is a $W \times W$ diagonal matrix containing $L[\omega]$ as its diagonal entries, $\boldsymbol{F}$ is the $W \times W$ DFT matrix, and $\boldsymbol{C}_0$ is the coefficients matrix with entries defined in (3). Since $\boldsymbol{C}_0$ inherits its low-rank structure from $\boldsymbol{C}$; therefore, $\boldsymbol{X}_0$ is also a low-rank matrix of rank R. In the rest of this write up, we will consider recovering the rank R matrix $\boldsymbol{X}_0$. Since $\boldsymbol{L}$ is well-conditioned, the recovery of $\boldsymbol{X}_0$ implies the recovery of $\boldsymbol{X}$ in (2).

1) Action of the Modulator: As explained before, the modulator in the mth channel takes input signal $x_m(t)$ and returns $d_m(t)x_m(t)$. If we take W equally-spaced samples of $d_m(t)x_m(t)$ in a time window $t \in [0,1)$, then we can write the vector of samples $\boldsymbol{y}_m$ as $\boldsymbol{D}_m \boldsymbol{x}_m$, where $\boldsymbol{x}_m$ is the W-vector containing the W uniformly-spaced samples of $x_m(t)$, and $\boldsymbol{D}_m$ is a $W \times W$ diagonal matrix whose entries are samples $\boldsymbol{d}_m \in \mathbb{R}^W$ of $d_m(t)$. We will choose a binary sequence that randomly generates $d_m(t)$, which amounts to $\boldsymbol{D}_m$ being a random matrix of the following form:

$$\boldsymbol{D}_m = \begin{bmatrix} d_m[1] & & & \\ & d_m[2] & & \\ & & \ddots & \\ & & & d_m[W] \end{bmatrix}, \tag{5}$$

where $d_m[n] = \pm 1$ with probability 1/2,

and the $d_m[n]$ are independent. Conceptually, the modulator disperses the information in the entire band of $x_m(t)$ — this allows us to acquire the information at a smaller rate by filtering a sub-band.

2) Action of the Low-pass Filter: The samples $\boldsymbol{y}_m \in \mathbb{R}^{\Omega}$ taken by the ADC in the mth branch can be written in matrix form as

$$\boldsymbol{y}_m = \boldsymbol{P} \boldsymbol{D}_m \boldsymbol{x}_m,$$

where $\boldsymbol{x}_m \in \mathbb{R}^W$ are the rows of $\boldsymbol{X}_0$ defined in (4); $\boldsymbol{D}_m$ is the $W \times W$ random diagonal matrix defined in (5), which corresponds to the modulator in the kth branch; and $\boldsymbol{P} : \Omega \times W$ is the matrix for the integrator (used as low-pass filter; for more details, see [5]) that contains ones in locations $(\alpha, \beta) \in (j, \mathcal{B}_j)$, for $j = 1, \cdots, \Omega$, where

$$\mathcal{B}_j = \{(j-1)W/\Omega + 1 : jW/\Omega\} \quad 1 \le j \le \Omega,$$

where we are assuming for simplicity that Ω is a factor of W. Since the action of the integrator commutes with the action of the modulator, the operation of the integrator can be simply represented as a block-diagonal matrix $\boldsymbol{P}$ operating on the modulated entries of the rows of $\boldsymbol{X}_0$, which contains the samples of the integrated signals. Putting it all together, the samples acquired by the ADCs can be written as a random block diagonal matrix times the vector $\mathrm{vec}(\boldsymbol{X}_0)$, formed by stacking the rows of low-rank $\boldsymbol{X}_0$ as

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{P}\boldsymbol{D}_1 & & \\ & \ddots & \\ & & \boldsymbol{P}\boldsymbol{D}_M \end{bmatrix} \cdot \mathrm{vec}(\boldsymbol{X}_0), \tag{6}$$

where $\boldsymbol{y} \in \mathbb{R}^{\Omega M}$ is the vector containing the samples acquired by all the ADCs. We will denote by Δ, the total number of samples per second $M\Omega$ taken by all the ADCs.

B. Sampling Theorem: Exact Recovery

Clearly form the last section, we are observing the low-rank matrix $\boldsymbol{X}_0$ through an underdetermined linear operator $\mathcal{A} : \mathbb{R}^{M \times W} \to \mathbb{R}^{\Delta}$, i.e., the observations $\boldsymbol{y}$ can be equivalently expressed as

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0).$$

To solve for $\boldsymbol{X}_0$, we use the nuclear-norm minimization program subject to affine constraints as below:

$$\begin{aligned} &\text{minimize} \quad \|\boldsymbol{Z}\|_* \\ &\text{subject to} \quad \boldsymbol{y} = \mathcal{A}(\boldsymbol{Z}). \end{aligned} \tag{7}$$

Let $\boldsymbol{X}_0 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ be the reduced form svd of $\boldsymbol{X}_0$ with $\boldsymbol{U} : M \times R$, $\boldsymbol{V} : W \times R$ being the matrices of left and right singular vectors, respectively, and $\boldsymbol{\Sigma} : R \times R$ being a diagonal matrix containing singular values of $\boldsymbol{X}_0$. The coherence of $\boldsymbol{X}_0$ is defined as

$$\mu_1^2 = \frac{M}{R} \max_{1 \le k \le M} \|\boldsymbol{U}^{\mathrm{T}} \boldsymbol{e}_k\|_2^2, \tag{8}$$

$$\mu_2^2 = \frac{W}{R} \max_{1 \le k \le W} \|\boldsymbol{V}^{\mathrm{T}} \boldsymbol{e}_k\|_2^2, \tag{9}$$

and

$$\mu_3^2 = \frac{M\Omega}{R} \max_{\substack{1 \le i \le M \\ 1 \le j \le \Omega}} \sum_{k \sim \mathcal{B}_j} \langle \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}, \boldsymbol{e}_i \boldsymbol{e}_k^{\mathrm{T}} \rangle^2. \tag{10}$$

Note that $1 \le \mu_1^2 \le M/R$, and $1 \le \mu_2^2 \le W/R$. The coherences take smallest values for equally dispersed singular vectors and largest values for sparse singular vectors [4]

Theorem 1. *Suppose $\Delta = M\Omega$ measurements of the ensemble $\boldsymbol{X}_0$ are taken using* (6)*. If*

$$\Omega \ge C_\beta \mu_3^2 R \max((W/M)\mu_1^2, \mu_2^2) \log^3(WM) \tag{11}$$

for a constant C_β (that depends on $\beta > 1$), the minimizer $\hat{\boldsymbol{X}}$ to the problem (7) *is unique and equal to $\boldsymbol{X}_0$ with probability at least $1 - O(WM)^{-\beta}$.*

The result indicates that each ADC operates at a rate Ω that is smaller than the Nyquist rate W by a factor of R/M. The net sampling rate Δ scales with the number R of independent signals rather than with the total number M of signals in the ensemble. The coherence terms suggest that the sampling architecture is more effective for sampling signals with energy dispersed across channels and time.

III. Compressive Parameter Estimation

In many signal processing applications, the goal is to estimate from the measurements, the parameters of multiple signals generated at an antenna array by incident wavefronts. The parameters of interest include, for example, the angle of arrival of wavefronts impinging on antenna arrays in several applications, such as radars, sonars, seismic exploration, and surveillance. Other examples are the estimation of frequency offsets in OFDMA-based wireless communications, and the estimation of frequency in myriad of applications.

As an illustration, we will discuss the angle-of-arrival estimation. Many practical applications are concerned with the detection of the location of R point sources radiating energy. A reasonable assumption is that the energy arrives at the sensors as a sum of plane waves and the signals are narrow-band centered around frequency ω_c. The rth signal can be written in the complex from as

$$s_r(t) = g(t)e^{-j\omega_c t},$$

where the narrow-band assumption implies that the envelop $g(t)$ is slowly varying, i.e., for small time delays τ, we have $g(t-\tau) \approx g(t)$. For this reason, the time delay only induces a phase shift on $s_r(t)$. This is to say,

$$s_r(t-\tau) \approx s_r(t)e^{-j\omega_c \tau}.$$

As a result, the signal $x_m(t)$ at the mth antenna element is

$$x_m(t) = \sum_{r=1}^{R} a_m(\theta_r) s_r(t - \tau_m(\theta_r)),$$

where $\tau_m(\theta_r)$ is the propagation delay at the mth antenna with respect to a reference point, and $a_m(\theta_r)$ is the mth array element response to the plane wave incident at an angle θ_r. By arranging the signals $x_m(t), 1 \leq m \leq M$ as the rows of $\boldsymbol{X}_c(t)$, we can write the signal ensemble received at the antenna array as

$$\boldsymbol{X}_c(t) = \boldsymbol{A}(\theta)\boldsymbol{S}_c(t), \tag{12}$$

where $\boldsymbol{A}(\theta)$ is an $M \times R$ matrix containing as its rth column, the array steering vector

$$\boldsymbol{a}(\theta_r) = [a_1(\theta_r)e^{-j\omega_c \tau_1(\theta_r)}, \cdots, a_M(\theta_r)e^{-j\omega_c \tau_M(\theta_r)}],$$

and $\boldsymbol{S}_c(t)$ can be thought of as a matrix containing R independent analog signals $s_r(t), 1 \leq r \leq R$ as its rows. The model in (12) is more general and is applicable to a wide variety of problems involving estimation of other parameters like frequency, or the estimation of location in an azimuth/elevation/range system, where the location of sources is specified by three angles θ, ϕ, and γ. In general, the number R of point sources is much smaller than the number M of antenna arrays; that is, the signal lives in a smaller subspace. Multiple signal classification algorithm such as MUSIC [9] estimate the signal subspace based on the estimate of the signal covariance matrix, and then find the intersection of this subspace with the array manifold, which is a set composed of all steering vectors $\boldsymbol{a}(\theta_r)$ for the entire range of the parameter θ_r [10]. This procedure reveals the estimates of the unknown parameter, which in this case is the direction of arrival. The central role in this computation is the estimation of the covariance matrix $\boldsymbol{R}_{XX}$, which requires sampling the signal ensemble $\boldsymbol{X}_c(t)$ to obtain the corresponding matrix of samples $\boldsymbol{X}$ given as

$$\boldsymbol{X} = \boldsymbol{A}(\theta)\boldsymbol{S}, \tag{13}$$

where $\boldsymbol{X} : M \times W$, and $\boldsymbol{S} : R \times W$ are the matrix of samples. The transformation from (12) to (13) involves sampling analog signals $x_m(t)$ using ADCs. With an ever increasing trend of radars operating at high frequencies in the range of 35-40 GHz, the sampling burden on the ADCs keeps escalating. The advances in the sampling rate of the ADCs are not up to pace with the advances in signal processing. Therefore, it is important to design the systems in a way that reduces the sampling burden on the ADCs. Since $\boldsymbol{X}_c(t)$ is by construction an ensemble consisting of multiple signals lying in a subspace with $R \ll M$, the random demodulator presented can be used to acquire the ensemble $\boldsymbol{X}_c(t)$ efficiently at a lower sampling rate. The benefits are two-fold: first, the covariance matrix $\boldsymbol{R}_{XX} := \lim_{W\to\infty} 1/W\boldsymbol{X}\boldsymbol{X}^*$ of input signal ensemble $\boldsymbol{X}_c(t)$ can be estimated accurately from fewer samples; second, in some cases the effective frequency range at which radar can operate can increase as the ADCs are sampling at sub-Nyquist rate.

In summary, the random demodulator can be employed to efficiently acquire multiple signals lying in a subspace, which may be useful in compressively estimating several parameters of interest in various signal processing applications. The reduction in sampling rate is achieved by combining the acquisition and compression step together by making use of the fact that the signals live in a (a priori unknown) subspace.

References

[1] A. Ahmed and J. Romberg, “Compressive sampling of correlated signals,” in *IEEE Conf. Signals, Syst. and Comput.*, 2011, pp. 1188–1192.

[2] ——, “Compressive multiplexing of correlated signals,” *arXiv preprint arXiv:1308.5146*, 2013.

[3] M. Fazel, “Matrix rank minimization with applications,” Ph.D. dissertation, Stanford University, March 2002.

[4] E. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Comput. Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[5] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, “Beyond Nyquist: Efficient sampling of sparse bandlimited signals,” *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, 2010.

[6] M. Mishali and Y. Eldar, “Blind multiband signal reconstruction: Compressed sensing for analog signals,” *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, 2009.

[7] J. Yoo, S. Becker, M. Loh, M. Monge, and E. Candès, “A 100mhz-2ghz 12.5x sub-Nyquist rate receiver in 90nm CMOS,” in *Submitted to Proc. IEEE Radio Freq. Integrated Cir. Conf.*, 2012.

[8] J. Yoo, C. Turnes, E. Nakamura, C. Le, S. Becker, E. Sovero, M. Wakin, M. Grant, J. Romberg, A. Emami-Neyestanak, and E. Candès, “A compressed sensing parameter extraction platform for radar pulse signal acquisition,” *Submitted to IEEE J. Emerging Topics Cir. and Sys.*, February 2012.

[9] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas and Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[10] R. Roy and T. Kailath, “Esprit-estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.