Rapid Sensing of Underutilized, Wideband Spectrum Using the Random Demodulator

Andrew Harms*[1], Waheed U. Bajwa†[2], Robert Calderbank‡[3]
*Department of Electrical Engineering, Princeton University, Princeton, NJ, USA
†Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ, USA
‡Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA
[1]hharms@princeton.edu [2]waheed.bajwa@rutgers.edu [3]robert.calderbank@duke.edu

***Abstract*—Efficient spectrum sensing is an important problem given the large and increasing demand for wireless spectrum and the need to protect incumbent users. We can more efficiently use large swaths of underutilized spectrum by designing spectrum sensors that can quickly, and power-efficiently, find and opportunistically communicate over unused (or underutilized) pieces of spectrum, such as television bands. In this paper, we concentrate on a particular sensing architecture, the *Random Demodulator* (RD), and look at two aspects of the problem. First, we offer fundamental limits on how efficiently any algorithm can perform the sensing operation with the RD. Second, we analyze a very simple, low-complexity algorithm called *one-step thresholding* that has been shown to work near-optimally for certain measurement classes in a low SNR setting or when the non-zero input coefficients are nearly equal. We rigorously establish that the RD architecture is well-suited for near-optimal recovery of the locations of the non-zero frequency coefficients in similar settings using one-step thresholding and perform numerical experiments to offer some confirmation of our results.**

I. Introduction

Spectrum sensing has received wide attention recently because of the growing demand for wireless spectrum. The proliferation of mobile devices has increased the demand on the wireless medium for communication, and caused a push for more efficient use of spectrum. This has led to the investigation of methods to more efficiently use wireless spectrum. The so-called white-space devices are an example that utilize unused television spectrum for opportunistic communication transmissions. Efficiently utilizing the available spectrum has become an increasingly important topic because of a complicated regulatory framework and the need to protect legacy and primary users. Recently, the President's Council of Advisors on Science and Technology released a report calling for the federal government to identify 1GHz of spectrum for possible sharing with commercial and private users in the near-term future [1]. This report argues that the slow-moving regulatory process has created a need to share spectrum with secondary users and, in turn, a need for new technologies which more efficiently utilize large swaths of spectrum. For example, cognitive radio and white-space devices are being investigated to better utilize large swaths of spectrum containing only a few television and wireless microphone signals by actively sensing their environment and searching for the unused portions of the spectrum. Military applications that require the monitoring of a large swath of spectrum for the presence of enemy communications also have many of the same requirements. Specifically, the devices need to know which frequencies are currently accommodating other communication (e.g., the Fourier coefficients at those frequencies have non-zero magnitude) and which are unused.

When building spectrum sensing devices, we want to use the best method to perform this sensing, or detection, of active frequencies. Several criteria arise to describe what is meant by the 'best' method. One such criterion is to minimize the number of errors made when detecting active frequencies. The reason is twofold: first, we do not want to interfere with an existing user, and second, we do not want to miss the opportunity to utilize unused pieces of spectrum. Proper handling of these two situations is vital. A second criterion is to minimize the resources (e.g., time and power) used to conduct the sensing operation. The spectrum environment is likely to be highly dynamic (frequencies are used for a short period of time and then become silent again). We therefore want to spend as little time as possible performing the detection to avoid interference. The sensing device is also likely running from a battery if it is a mobile device, so computational resources cost power. We therefore want to use the quickest and lowest complexity method possible.

Perhaps the simplest sensing method is to sweep over a range of frequencies using a Fourier transform and compare the Fourier coefficients to a threshold value. The problem is that the complexity (i.e., the number of samples required) scales with the bandwidth (i.e., number of frequencies) of interest. Many times we may be interested in a very large bandwidth, but we know that only a small number of frequencies contain energy. Recently the Random Demodulator (RD) [2] was proposed that can accomplish sensing with much lower complexity in the case of wideband, underutilized spectrum. We consider the RD architecture and first analyze the fundamental limits on spectrum sensing for any recovery algorithm. Second, we know from [2] that several low-complexity algorithms can recover the sampled signal. In the sensing problem, however, we may only be interested in knowing which frequencies contain the signal energy. We therefore analyze how well an even lower-complexity algorithm can recover the locations of the non-zero frequencies of the signal. Lastly, we perform numerical experiments to confirm that this algorithm does work well in recovering the locations of the non-zero frequencies in certain cases. From [3], we know that these cases include when the SNR is not too large and when the size of the entries in β are nearly the same.

A. Fundamental Limits on Spectrum Sensing

The first main contribution we make in this paper is an analysis on the fundamental limits afforded when using the RD for sensing of wideband, underutilized spectrum. We build off of work by Reeves and Gastpar [4] and extend it to the case of the RD. Here, we concentrate on the scaling regime where the number of (noisy) samples acquired and the number of active frequencies both scale linearly with the bandwidth. In the analysis, a small number of errors are allowed in the recovery and the bandwidth is allowed to grow to infinity. We quantify the ratio of samples to bandwidth beyond which recovery of the locations of the non-zero frequencies with a ratio of errors smaller than some parameter α is not possible.

B. Spectrum Sensing using Thresholding

Our second main contribution is to show that a simple thresholding algorithm can recover the non-zero frequencies with high probability if the number of samples scales logarithmically with the bandwidth and linearly with the sparsity. In [2], signals that are noisily sampled with the RD can be recovered if the number of samples m satisfies $m \asymp k \log^6 n$ where n is the total number of tones and k is the number of non-zero tones. The recovery can be accomplished through several low-complexity algorithms (e.g., CoSaMP [5], LASSO [6]). In the wideband sensing problem, however, we are really only interested in finding which frequencies contain energy in the signal, not the values of the coefficients. To this end, a very simple, low-complexity algorithm called *one-step thresholding* (OST) (see Algorithm 1 in [3] or the description in Section IV-B) is a potential choice. We show that OST can recover the locations of the non-zero tones with high probability if the number of samples scales as $m \asymp \max\{k \log n, \log^2 n\}$. In particular, the OST algorithm performs near-optimally when either the non-zero entries have nearly the same magnitude or the signal-to-noise ratio (SNR) is not too large.

II. Problem Formulation

In this paper, we are concerned with the measurement model

$$y = \mathrm{X}\beta + w \tag{1}$$

where X is an $m \times n$ (complex) RD matrix, β is a length n k-sparse complex signal vector ($||\beta||_0 \leq k$[1]), and w is a length m complex white Gaussian noise vector. We denote the sparsity pattern, or locations of the non-zero entries, of the vector β with $S = \{1 \leq i \leq n : \beta_i \neq 0\}$. The RD architecture is shown in Fig. 1, and a key aspect of [2] is showing that the architecture can be represented by the RD matrix $\mathrm{X} = \mathrm{HDF}$ where F is a unitary Fourier matrix, $\mathrm{D} = \mathrm{diag}(\varepsilon_i)$ is a diagonal matrix with a Rademacher sequence ε_i for $i = 1, ..., n$ on the diagonal, and H is a matrix with n/m consecutive 1's in each row with each row orthogonal. An example for H is

$$\mathrm{H} = \begin{bmatrix} 1 & 1 & 1 & & & & & & \\ & & & 1 & 1 & 1 & & & \\ & & & & & & 1 & 1 & 1 \end{bmatrix}.$$

The Rademacher sequence is a random sequence with independent entries equally likely $+1$ and -1.

[1] $||\cdot||_0$ counts the number of non-zero coefficients

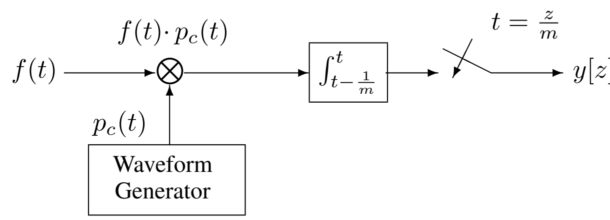

Fig. 1. Block diagram of the random demodulator [2]: The input signal is multiplied by a random waveform generated from a Rademacher chipping sequence, then low-pass filtered, and finally sampled at a sub-Nyquist rate $m \ll n$ where n is the bandwidth of $f(t)$.

A key result from [2] that we will use tells us that the columns of X are nearly unit-norm with high probability provided m is large enough.

Theorem 1. [2, Theorem 8] *Suppose* $m \geq c_1 \delta^{-2} \log n$. *Then an* $m \times n$ *RD matrix satisfies*

$$\mathbb{P}\left[\max_i \left| ||x_i||_2^2 - 1 \right| \geq \delta \right] \leq n^{-1}$$

where x_i *is the* i^{th} *column of* X *and* c_1 *is a constant.*

In the next two sections, we first present our results on the limits of recovery using an information-theoretic argument. We then present thresholding as an achievable scheme and discuss when this scheme is nearly optimal compared to the information-theoretic limit.

III. Information-Theoretic Bound on Reliable Sensing

To examine some fundamental limits of the RD in the linear scaling regime, we analyze the RD matrix in the setting of [4]. We begin with the measurement model (1) and also impose the relation $m = \lceil \rho n \rceil$ and $k = \lfloor \Omega n \rfloor$ so that both m and k are linear in the problem size n. In other words, $\rho < \infty$ represents the sampling rate in the problem, and $0 \leq \Omega \leq 1/2$ represents the sparsity ratio. We also impose a distribution on the non-zero entries of β denoted as F with mean μ_F and variance σ_F^2. We then define the variance of the vector β as $V(\Omega, F) = \Omega(1-\Omega)\mu_F^2 + \Omega\sigma_F^2$. Further, each of the $\binom{n}{k}$ realizations of S is equally likely. Finally, the distortion for an estimate $\hat{S}$ of S is (as defined in [4])

$$d(S, \hat{S}) = \max(\mathrm{MDR}(S, \hat{S}), \mathrm{FAR}(S, \hat{S}))$$

with missed detection rate

$$\mathrm{MDR}(S, \hat{S}) = \frac{1}{|S|} \sum_{i=1}^{n} 1(i \in S, i \notin \hat{S})$$

and false alarm rate

$$\text{FAR}(S,\hat{S}) = \frac{1}{|\hat{S}|}\sum_{i=1}^{n} 1(i \notin S, i \in \hat{S}).$$

We then require that $d(S,\hat{S}) \leq \alpha$. Finally, we must rescale the RD matrix so that the assumptions of [4] are satisfied, namely that the rows of X are unit norm in expectation, i.e., $\mathbb{E}[\text{trace}(\text{XX}^H)] = m$ where X^H denotes Hermitian transpose. For the RD matrix,

$$\text{XX}^H = \text{HDF}(\text{HDF})^H = \text{HDFF}^H\text{D}^H\text{H}^H = \text{HH}^H = \frac{n}{m}\text{I}_m,$$

and we therefore must scale the RD matrix by $\sqrt{m/n}$ to satisfy this condition.

With this in mind, one of our main results is as follows.

Theorem 2. *If* X *is a (scaled)* $m\times n$ *RD matrix and* $m/n \to \rho$ *as* $n \to \infty$*, then in order to satisfy a distortion* α *for the estimated sparsity pattern* $\hat{S}$*,* ρ *must satisfy*

$$\rho \geq \frac{2R(\Omega,\alpha)}{\log\left(1+V(\Omega,F)\right)}$$

where

$$R(\Omega,\alpha) = \begin{cases} H(\Omega) - \Omega H(\alpha) \\ \quad -(1-\Omega)H(\frac{\Omega\alpha}{1-\Omega}), & \alpha < 1-\Omega \\ 0, & \alpha \geq 1-\Omega \end{cases} \tag{2}$$

and $H(\cdot)$ *is the binary entropy function.*

Proof: The important result that we use is Lemma 2 from [4]. It states that for a randomly generated vector β, as described above, the pair (ρ,α) is not achievable for a sequence of $m\times n$ sampling matrices X if

$$\limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_{\text{X}} I(S;y) < R(\Omega,\alpha)$$

where $I(S;y)$ is the mutual information between the sparsity pattern S and the samples y. We bound this mutual information with the mutual information between the noiseless samples and the noisy samples using (1):

$$I(S;y) \leq I(\text{X}\beta;y).$$

Conditioned on the matrix X, we can bound

$$I(\text{X}\beta;y) \leq \max_z I(z;z+w)$$

where z is a random vector satisfying $\mathbb{E}[zz^H] = V(\Omega,F)\text{XX}^H$. This maximization occurs for z a Gaussian vector with the required covariance, so that

$$I(\text{X}\beta;y) \leq \frac{1}{2}\log\left|\text{I}_m + V(\Omega,F)\text{XX}^H\right|. \tag{3}$$

Recall that for the scaled RD matrix, we have $\text{XX}^H = \text{I}_m$, and as a result:

$$I(\text{X}\beta;y) \leq \frac{m}{2}\log\left(1+V(\Omega,F)\right)$$

for any RD matrix X.

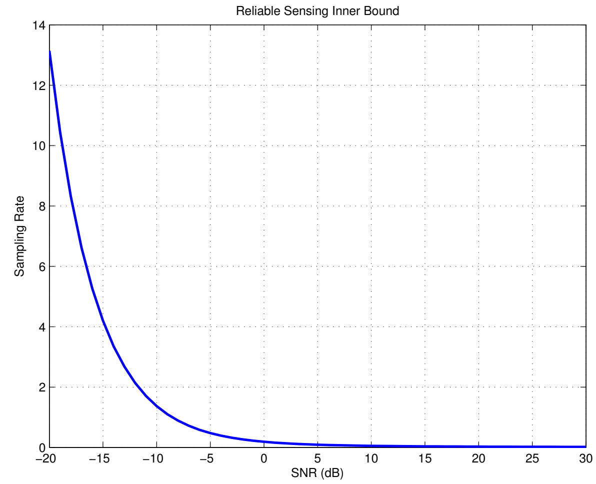

Fig. 2. Plot of the lower bound (Theorem 2) for sparsity ratio $\Omega = 0.01$ and distortion $\alpha = 0.1$. For small SNR, starting around -10 dB, the sampling rate $\rho > 1$ meaning that the number of measurements must be larger than the size of β if too much noise is present.

Consequently, we also have that:

$$I(S;y) \leq \frac{m}{2}\log\left(1+V(\Omega,F)\right) \tag{4}$$

and

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}I(S;y) &= \frac{m}{2n}\log\left(1+V(\Omega,F)\right)\\ &= \frac{\rho}{2}\log\left(1+V(\Omega,F)\right)\end{aligned}$$

because we have $\frac{m}{n} \to \rho$ as $n \to \infty$. Theorem 2 thus follows. ■

To illustrate Theorem 2, Fig. 2 shows a plot of the lower bound on the sampling rate as a function of SNR in dB for a sparsity ratio $\Omega = 0.01$ and distortion $\alpha = 0.1$. Here, SNR is given by the power of the distribution of β, defined as $P(\Omega,F) = \Omega(\mu_F^2 + \sigma_F^2)$, because of the row normalization of the scaled RD matrix X. The sampling rate grows very large at small values of SNR. In particular, $\rho > 1$ for SNR smaller than about -10dB, meaning that the number of measurements needed is larger than the size of β if too much noise is present. We also expect this bound to be loose at high SNR based on the results of [4]. The apparent vanishing sampling rate at large SNR seems to make this expectation reasonable. We also show that this expectation holds true at least for the thresholding algorithm considered in the next section.

IV. Sufficient Conditions for Thresholding

We now show that we can recover the locations of the non-zero tones of a sparse, wideband signal using the RD and the thresholding algorithms in [3]. First, we must show that the RD matrix satisfies the Coherence Property (CP). Let X be an (unscaled) $m\times n$ RD matrix and and let x_i denote the i^{th} column of X. Theorem 1 tells us that X has nearly unit-norm columns. In [3], the *worst-case coherence* is defined as

$$\mu = \max_{i,j:i\neq j} |\langle x_i, x_j\rangle|, \tag{5}$$

the *average coherence* is defined as

$$\nu = \frac{1}{n-1} \max_i \left| \sum_{j:j\neq i} \langle x_i, x_j\rangle \right|, \tag{6}$$

and X is said to satisfy the CP if

$$\mu \leq \frac{0.1}{\sqrt{2\log n}}$$

and

$$\nu \leq \frac{\mu}{\sqrt{m}}.$$

A. Worst-case and Average Coherence of the Random Demodulator

To show that thresholding can recover the sparsity pattern of signals sampled with the RD, we need a result from [2].

Theorem 3. [2, Theorem 9] *Suppose $m \geq 2\log n$. Then*

$$\mathbb{P}\left[\mu \geq c_2\sqrt{\frac{\log n}{m}}\right] \leq n^{-1}$$

where c_2 is a constant.

From this result, the requirement on μ in the CP is satisfied with high probability if $m \geq c_3 \log^2 n$ and c_3 is chosen appropriately.

Turning to ν, we have that:

$$\langle x_i, x_j\rangle = \delta_{ij} + \sum_{r,s:r\neq s} \varepsilon_r\varepsilon_s\eta_{rs} f_{ri}^* f_{sj}$$

where δ_{ij} is the Kronecker delta, f_{ri} is an entry of the (unitary) Fourier matrix F, η_{rs} is the inner product between columns of the matrix H, and ε_i is the i^{th} entry of the Rademacher sequence. To calculate (6), we start with

$$\begin{aligned} \left|\sum_{j:j\neq i} \langle x_i, x_j\rangle\right| &= \left|\sum_{j:j\neq i}\sum_{r,s:r\neq s} \varepsilon_r\varepsilon_s\eta_{rs} f_{ri}^* f_{sj}\right| \\ &= \left|\sum_{r,s:r\neq s} \varepsilon_r\varepsilon_s\eta_{rs} f_{ri}^* \left(\sum_{j:j\neq i} f_{sj}\right)\right| \\ &= \left|\sum_{r,s:r\neq s} \varepsilon_r\varepsilon_s\eta_{rs} f_{ri}^* A_{si}\right| \end{aligned} \tag{7}$$

where

$$\begin{aligned} A_{si} &= \sum_{j:j\neq i} f_{sj} \\ &= \begin{cases} \sqrt{n} - f_{si}, & s = 0 \\ -f_{si}, & s \neq 0 \end{cases} \\ &= -f_{si} + \delta_s\sqrt{n} \end{aligned}$$

and

$$\delta_s = \begin{cases} 1, & s = 0 \\ 0, & s \neq 0. \end{cases}$$

Substituting this back into (7):

$$\begin{aligned} &\left|\sum_{j:j\neq i} \langle x_i, x_j\rangle\right| \\ &= \left|\sum_{r,s:r\neq s} \varepsilon_r\varepsilon_s\eta_{rs} f_{ri}^* \left(-f_{si}\right) + \sum_{r\neq 0} \varepsilon_r\varepsilon_0\eta_{r0} f_{ri}^*\sqrt{n}\right| \\ &\leq \left|\sum_{r,s:r\neq s} \varepsilon_r\varepsilon_s\eta_{rs} f_{ri}^* f_{si}\right| + \sqrt{n}\left|\sum_{r\neq 0} \varepsilon_r\varepsilon_0\eta_{r0} f_{ri}^*\right|. \end{aligned}$$

Finally, using this result in (6), we have:

$$\begin{aligned} &\frac{1}{n-1} \max_i \left|\sum_{j:j\neq i} \langle x_i, x_j\rangle\right| \leq \\ &\frac{1}{n-1} \max_i \left|\sum_{r,s:r\neq s} \varepsilon_r\varepsilon_s\eta_{rs} f_{ri}^* f_{si}\right| + \frac{\sqrt{n}}{n-1} \max_i \left|\sum_{r\neq 0} \varepsilon_r\varepsilon_0\eta_{r0} f_{ri}^*\right| \\ &= \frac{1}{n-1} \max_i B_i^{(1)} + \frac{\sqrt{n}}{n-1} \max_i B_i^{(2)}. \end{aligned}$$

Lemma 6 of [2] tells us that if $m \geq 2\log n$, then

$$\mathbb{P}\left[\max_i B_i^{(1)} \geq c_4\sqrt{\frac{\log n}{m}}\right] \leq n^{-1}.$$

Further, Lemma 5 of [2] also tells us that

$$\mathbb{P}\left[\max_i B_i^{(2)} \geq \sqrt{\frac{10\log n}{m}}\right] \leq n^{-1}.$$

Again, if we require that $m \geq c_3 \log^2 n$ for an appropriately chosen c_3 then we can ensure that ν is sufficiently small to satisfy the CP. Combining these results, we can say that as long as $n - 1 \geq m$, then

$$\nu \leq \frac{\bar{\mu}}{\sqrt{m}}$$

with probability $1 - 2n^{-1}$ where $\bar{\mu}$ is the upper bound on μ.

Finally taking a union bound over both of these conditions holding, we have that the CP is satisfied with probably at least $1 - 3n^{-1}$. Because the CP is satisfied, we can use (for example) Theorem 1 of [3] to show that OST can be successful in recovering the locations of the non-zero frequencies as long as $m > \max\left\{c_5 k\log n, c_6\log^2 n\right\}$ for appropriately chosen c_5 and c_6.

B. Numerical Experiments

To verify the results of Section IV-A, we performed numerical experiments using the RD and the OST algorithm. The OST algorithm only requires two steps. First, a signal proxy g is formed

$$g = \mathrm{X}^H y.$$

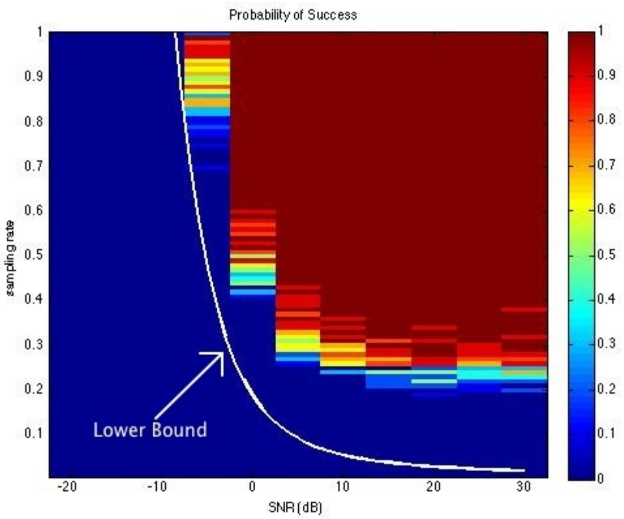

Fig. 3. One-step thresholding results using the RD. The empirical success rate is plotted against sampling rate and SNR in dB. The lower bound from Theorem 2 is also shown as a reference and to highlight that OST is nearly optimal for small values of SNR. For all these experiments, the sparsity ratio $\Omega = 0.01$.

Next, the locations of the non-zero coefficients are selected via thresholding as follows:

$$\hat{S} = \{1 \leq i \leq n : |g_i| > \lambda\}$$

where g_i is the i^{th} entry of g and λ is an appropriately chosen threshold (see [3]). Here, we used $\lambda = \max\{20\mu\sqrt{n\text{SNR}}, 2\sqrt{2}\}\sqrt{2\sigma^2 \log n}$. We also note here that for the RD, the first step is even less computationally intensive than a matrix multiplication because of the special structure of X. The results are presented in Fig. 3 along with the lower bound from Theorem 2 (with $\Omega = 0.01$ and $\alpha = 0.1$). The plot shows the empirical success rate as a function of the sampling rate ρ and the SNR in dB. For each experiment a new RD matrix was drawn as well as a new signal vector. The signal vector had a constant sparsity ratio $\Omega = 0.01$ (to match the lower bound) and all the coefficients had the same value. For each sampling rate and SNR pair 100 experiments were performed.

The results confirm that the OST algorithm is nearly optimal, relative to the lower bound, for small values of SNR (see e.g., [3] or [7] for further details on this point). The gap grows larger at higher SNR values.

V. Conclusions and Future Work

In this paper, we have proposed using the RD architecture [2] for the sensing of wideband, underutilized spectrum. We first analyzed the best performance we can get without considering a specific algorithm. If the number of measurements and level of sparsity both scale linearly with the bandwidth, then in the asymptotic regime recovery is not possible if the number of measurements does not grow fast enough as the bandwidth increases. Second, we analyzed a specific algorithm, OST, that is very low-complexity and which works near-optimally if the SNR is not too large and if the non-zero frequency coefficients are all nearly the same. In the case of OST, recovery of the locations of the non-zero coefficients succeeds if the number of measurements m scales as $m \asymp \max\{k \log n, \log^2 n\}$. We also performed numerical experiments to confirm that OST can recover the sparsity pattern and confirmed that OST is indeed near optimal if the SNR is small.

Some interesting directions for future work include improving the lower bound, especially in the large SNR regime, and finding upper bounds for the RD. An upper bound could include analyzing thresholding in the asymptotic regime or other achievable schemes that might perform better at larger SNR values. Additionally, considering other a priori information, such as a prior (non-uniform) distribution on β similar to that explored in [8], we feel would lead to some useful and interesting results.

References

[1] P. C. of Advisors on Science and Technology, "Realizing the full potential of government-held spectrum to spur economic growth," Executive Office of the President, Tech. Rep., July 2012.

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

[3] W. U. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," *J. Commun. Netw.*, vol. 12, no. 4, pp. 289–307, Aug 2010.

[4] G. Reeves and M. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," *arXiv:1002.4458*, 2012.

[5] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, vol. 26, pp. 301–321, 2008.

[6] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[7] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec 2009.

[8] A. Harms, W. U. Bajwa, and R. Calderbank, "A constrained random demodulator for sub-Nyquist sampling," *to appear in IEEE Trans. Signal Proc.*