Faster than Nyquist, Slower than Tropp

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*—The sampling rate of analog-to-digital converters is severely limited by underlying technological constraints. Recently, Tropp et al. proposed a new architecture, called a *random demodulator* (RD), that attempts to circumvent this limitation by sampling sparse, bandlimited signals at a rate much lower than the Nyquist rate. An integral part of this architecture is a random bi-polar modulating waveform (MW) that changes polarity at the Nyquist rate of the input signal. Technological constraints also limit how fast such a waveform can change polarity, so we propose an extension of the RD that uses a *run-length limited* MW which changes polarity at a slower rate. We call this extension a *constrained random demodulator* (CRD) and establish that it enjoys theoretical guarantees similar to the RD and that these guarantees are directly related to the *power spectrum* of the MW. Further, we put forth the notion of *knowledge-enhanced* CRD in the paper. Specifically, we show through simulations that matching the distribution of spectral energy of the input signal with the power spectrum of the MW results in the CRD performing better than the RD of Tropp et al.**

I. Introduction

Analog-to-digital converters (ADCs) are the foundation of modern signal processing. For example, acquiring a high fidelity representation of a guitar sound in the studio allows it to be processed by any number of digital devices in post-processing. Two of the defining characteristics of ADCs are *sampling rate* and *resolution*. In general these two are in contention with each other. Increase the rate and the resolution drops; increase the resolution and a drop in rate is required. This is captured by the rule-of-thumb that a doubling of the rate causes a one bit reduction in the resolution: $2^B \cdot f_s = P$ where B is the effective number of bits (ENOB) – a measure of ADC resolution, f_s is the sampling rate, and the constant P is a figure-of-merit determined by the choice of ADC architecture. Unfortunately, this figure-of-merit advances at a much slower pace than the pace of Moore's law for microprocessors [1], [2], and many current application are pushing the technology to the limit. For example, a software-defined radio performing spectrum sensing requires a sampling rate on the order of 1 GHz and can only sample with resolution of 10 ENOB using current technology [1].

One possible solution to this problem is to exploit prior information about the signal apart from the bandwidth. For example, if we have prior knowledge that the signal has a *sparse* representation in some basis, then we can use advanced techniques to sample the signal at a lower rate. It has long been known that bandlimited signals that are sparse in the frequency domain can be sampled at a much lower rate than the Nyquist rate [3]. Recent theoretical advances in Compressed Sensing (CS) [4] have reinvigorated the investigation of techniques which exploit signal sparsity to sample at a rate much lower than the Nyquist rate. Three architectures which seek to exploit these recent developments in CS are *Chirp Sampling* [5], *Xampling* [6], and the *Random Demodulator* [7]. In this paper we concentrate on signals which can be well-approximated by a small number of frequency tones and the Random Demodulator (RD) architecture seems to be an appropriate choice.

This work was supported, in part, by DARPA under the KECoM program and by AFOSR under grant FA9550-09-1-0422.

A. Our Contributions

In this paper, we emphasize two contributions to the RD architecture. Both relate to a white-noise-like bipolar modulating waveform (MW), a vital component of the RD. In the RD this waveform is generated from a random binary sequence where each entry of the sequence is generated independently. An important characteristic of this sequence is that it changes polarity at the Nyquist rate of the input signal. In [8], it is pointed out that hardware constraints can limit the rate at which this waveform can change polarity while retaining high fidelity. The proposed solution is to use *run-length limited* (RLL) codes to generate sequences with two run constraints. These run constraints limit the minimum, d, and maximum, k, fundamental time units between transitions in the waveform and allow for the use of waveforms that achieve higher fidelity for a given switching rate. The resulting system is called a *Constrained Random Demodulator* (CRD). Further, if there is a fundamental limit on the rate at which the waveform can change polarity (due to the physics of the hardware), then the CRD will have a higher operating bandwidth compared to the RD [8]. This advantage does, however, come at a cost. The dependence within the waveform causes a reduction in performance guarantees for a given bandwidth. Our first contribution in this regard is to expand on our earlier work on the CRD [8] and establish the relationship between the choice of RLL MW and the resulting performance guarantees.

Our second contribution is to show that the signal reconstruction performance of the CRD can be made significantly better than the RD through matching of the spectrum of the RLL MW with the probabilistic distribution of signal energy in the spectral domain. We call such a matched CRD a *knowledge-enhanced* CRD (or KECoRD), in the vein of the DARPA KECoM program [9]. Heuristically, it is easy to argue the superiority of the KECoRD over the RD. The spectrum of the MW in the RD is flat and therefore it sends energy from all frequencies to baseband with (on average) an equal weighting. On the other hand, the MW used in the CRD is generated from a statistically dependent sequence and has a characteristic, non-flat power spectrum associated with it [10]. The CRD therefore favors some frequencies over others during the modulation of the signal energy to baseband. Consequently, the CRD performs better than the RD whenever it samples signals with a probabilistic distribution of spectral energy which matches that of the RLL MW.

We conclude this discussion by pointing out a related, independent work carried out in [11], [12] to improve upon the performance of the RD. The biggest difference with our work is that we focus on both a rigorous mathematical understanding of the impact of the MW on the performance guarantees and the ability of the RLL MW to perform better for certain signal classes; [11], [12] limit themselves to performance improvement without any theoretical guarantees concerning the identifiability properties of input waveforms.

II. The Random Demodulator Architecture

We briefly review some key characteristics of the RD architecture as they pertain to the sampling of sparse, bandlimited signals. We refer the reader to [7] for a comprehensive overview.

The basic purpose of the RD is to take samples at a sub-Nyquist rate and still be able to reconstruct signals that are periodic, limited in bandwidth to W Hz, and are completely described by a total of $S \ll W$ tones. In other words, a signal $f(t)$ being fed as an input to the RD takes the parametric form

$$f(t) = \sum_{\omega \in \Omega} a_\omega e^{-2\pi i \omega t}, \ t \in [0,1) \tag{1}$$

where $\Omega \subset \{0, \pm 1, ..., \pm(W/2-1), W/2\}$ is a set of S integer-valued frequencies and $\{a_\omega : \omega \in \Omega\}$ is a set of complex-valued amplitudes. In order to acquire this sparse bandlimited signal $f(t)$, the RD performs three basic actions as illustrated in Fig. 1. First, it multiplies $f(t)$ with a modulating waveform $p_m(t)$ that is given by

$$p_m(t) = \sum_{n=0}^{W-1} \varepsilon_n 1_{\left[\frac{n}{W}, \frac{n+1}{W}\right)}(t) \tag{2}$$

where the discrete-time *modulating sequence* (MS) $\{\varepsilon_n\}$ independently takes values $+1$ or -1 with probability $1/2$ each. Next, it low-pass filters the continuous-time product $f(t) \cdot p_m(t)$. Finally, it takes samples at the output of the low-pass filter at a rate of $R \ll W$.

One of the major contributions of [7] is that it expresses the actions of the RD on a continuous-time, sparse bandlimited signal $f(t)$ in terms of the actions of an $R \times W$ matrix Φ_{RD} on a vector $\alpha \in \mathbb{C}^W$ that has only S nonzero entries. Specifically, let $\mathrm{x} \in \mathbb{C}^W$ denote a Nyquist-sampled version of the continuous-time input signal $f(t)$. Then it is easy to conclude from (1) that x can be written as $\mathrm{x} = F\alpha$, where the matrix $F = \frac{1}{\sqrt{W}}\left[e^{-2\pi i n \omega / W}\right]_{(n,\omega)}$ denotes a (normalized) discrete Fourier transform matrix and $\alpha \in \mathbb{C}^W$ has only S nonzero entries corresponding to the amplitudes of the nonzero frequencies in $f(t)$. Note that the effect of the modulating waveform on $f(t)$ in discrete-time is equivalent to multiplying a $W \times W$ diagonal matrix $D = \mathrm{diag}(\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_{W-1})$ with $\mathrm{x} = F\alpha$. Further, the effect of the low-pass filter on $f(t) \cdot p_m(t)$ in discrete-time is equivalent to multiplying an $R \times W$ matrix H, which has W/R consecutive ones starting at position $rW/R + 1$ in the r^{th} row of H, with $DF\alpha$.[1] Therefore, if one collects R samples at the output of the RD into a vector $\mathrm{y} \in \mathbb{C}^R$, then it follows from the preceding discussion that $\mathrm{y} = HDF\alpha = \Phi_{RD} \cdot \alpha$, where we have that the random demodulator matrix $\Phi_{RD} = HDF$.

Given the discrete-time representation $\mathrm{y} = \Phi_{RD} \cdot \alpha$, recovering the continuous-time signal $f(t)$ described in (1) is equivalent to recovering the S-sparse vector α from y. In this regard, the primary objective of the RD is to guarantee that α can be recovered from y even when the sampling rate R is far below the Nyquist rate W. Fortunately, recent theoretical developments in the area of compressed sensing have provided us with numerous greedy as well as convex optimization based methods that are guaranteed to recover α (or a good approximation of α) from y as long as the *sensing matrix* Φ_{RD} can be shown to satisfy certain geometrical properties [4]. The highlight of [7] in this regard is that the RD matrix is explicitly shown to satisfy the requisite geometrical properties as long as the sampling rate R scales linearly with the number of frequency tones S in the signal and (poly)logarithmically with the signal bandwidth W.

[1]Throughout this paper, we assume that R divides W.

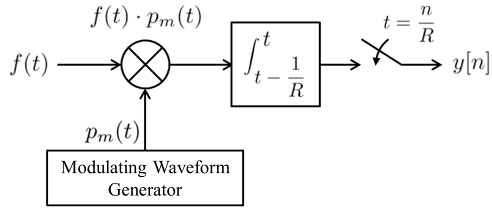

Fig. 1. Block diagram of the (constrained) random demodulator [7].

III. Constrained Random Demodulator

Motivated by the desire to use a MS that switches at a rate lower than the Nyquist rate, we introduce the CRD which uses an RLL sequence with statistically dependent entries. The only change to the RD architecture is a replacement of the independent MS by a dependent MS. For the matrix representation, this means changing the diagonal entries of the matrix D from an independent sequence to an RLL sequence. We must, however, be careful when doing so. The dependence in the RLL sequence causes a change in the structure of the measurement matrix Φ_{RD}. It turns out that if the correlation structure of the sequence is sufficiently well-behaved (defined below) then we enjoy guarantees that take only a slight hit over the RD.

A very important tool for studying the behavior of our measurement matrix is the *Restricted Isometry Property* (RIP). The RIP of order S with restricted isometry constant δ_S is satisfied for a matrix Φ if

$$(1-\delta_S)||x||_2^2 \leq ||\Phi x||_2^2 \leq (1+\delta_S)||x||_2^2 \tag{3}$$

or equivalently

$$\left| \frac{\|\Phi \mathrm{x}\|_2^2 - \|\mathrm{x}\|_2^2}{\|\mathrm{x}\|_2^2} \right| \leq \delta_S \tag{4}$$

with $\delta_S \in (0,1)$ and $\|\mathrm{x}\|_0 \leq S$. It is convenient to use the "triple-bar" norm of [7] to describe the RIP condition. Given a matrix A, this norm captures the largest magnitude eigenvalue of any $S \times S$ principal submatrix of A: $|||A||| = \sup_{|\Omega| \leq S} \|A|_{\Omega \times \Omega}\|$. Therefore, (3) is satisfied if and only if $|||\Phi^*\Phi - I||| \leq \delta_S$.

We have the following theorem describing the RIP for a CRD.

Theorem 1 (RIP for the CRD)**.** *Let* Φ_{CRD} *be a* $R \times W$ *CRD matrix using a MS with maximum dependence distance* ℓ *and matrix* Δ *such that* $|||\Delta||| < \delta$ *for a fixed* $\delta \in (0,1)$*. Next, suppose that* R *satisfies*

$$R \geq \ell^3 (\delta - |||\Delta|||)^{-2} \cdot \mathrm{C} \cdot S \log^6(W) \tag{5}$$

where C *is a positive constant. Then with probability* $\mathcal{O}(1 - W^{-1})$ *the CRD matrix* Φ_{CRD} *satisfies the RIP of order* S *with constant* $\delta_S \leq \delta$*.*

The proof and a more detailed discussion will be presented in a journal paper. The maximum dependence distance ℓ is the smallest ℓ such that any two entries in the MS separated by this distance are independent; Δ is a matrix that is determined by the correlation properties of the MS. It has entries

$$\Delta_{\alpha\omega} = \sum_{j \neq k} \eta_{jk} f_{j\alpha}^* f_{k\omega} \mathbb{E}[\varepsilon_j \varepsilon_k]$$

where $\eta_{jk} = \langle h_j, h_k \rangle$, h_j is the jth column of H, $f_{k\omega}$ is the (k,ω)-th entry of F, and $\varepsilon = \{\varepsilon_j\}$ is the MS. We can further relate $|||\Delta|||$ to the *power spectrum* of the MS. We can show that

$$|||\Delta||| \approx \max_\omega |\tilde{F}_\varepsilon(\omega)| \tag{6}$$

where $\tilde{F}_\varepsilon(\omega)$ is referred to as the *reduced spectrum* of the MS and is trivially related to the power spectrum $F_\varepsilon(\omega) = \sum_m R_\varepsilon(m) e^{-\frac{2\pi}{W} m\omega}$

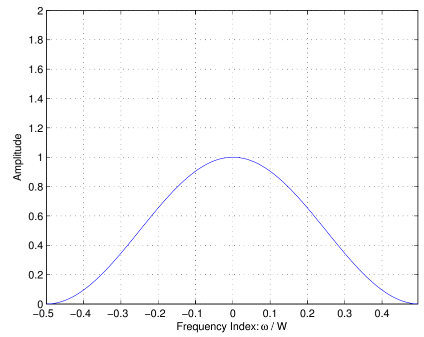

Fig. 2. Spectrum of a rate-1/2 repetition-coded sequence. The spectrum rolls off to zero at high frequencies.

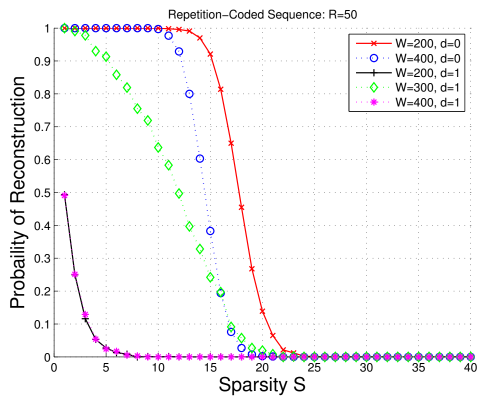

Fig. 3. Reconstruction comparison for a rate-$1/2$ RCS with $R = 50$. We can see that for $W = 200$ and $W = 400$ the results are particularly bad. This tells us that our theory is fairly tight given that the RCS does not satisfy our criteria for the RIP.

by $\tilde{F}_\varepsilon(\omega) = F_\varepsilon(\omega) - 1$. This means the RIP constant in Theorem 1 is minimal for a flat spectrum ($F_\varepsilon(\omega) = 1$, produced by an independent sequence) and is approximately equal to the largest deviation of the power spectrum from a flat spectrum.

We now consider two example sequences to compare against the RD. The first example is a *repetition-coded sequence* (RCS) – a simple and naive choice. A RCS is generated by repeating each element of an independent MS d times. The resulting sequence is not stationary but cycle-stationary. The power spectrum is not defined for such sequences; however, averaging over the period of the sequence provides an adequate *average* power spectrum for our purposes. The average power spectrum is shown in Fig. 2. It goes to zero at high frequencies, and from (6) we see that $|||\Delta||| \approx 1$ and Theorem 1 is not in force. To show that the theory is correct to exclude such sequences, we perform numerical experiments reconstructing vectors sampled with the CRD measurement matrix; the results are shown in Fig. 3. For these experiments, and all subsequent, we generate input vectors with S non-zero entries at random locations and containing random complex numbers from the unit circle. For each of $1,000$ randomly generated realizations of the measurement matrix, we sample a new randomly generated input vector and reconstruct the input from the samples with the Basis Pursuit algorithm. The *probability of success* corresponds to the ratio of successful trials. The results are poor for the CRD with the RCS ($d = 1$ curve) especially for $W = 200$ and $W = 400$. The dependence on W

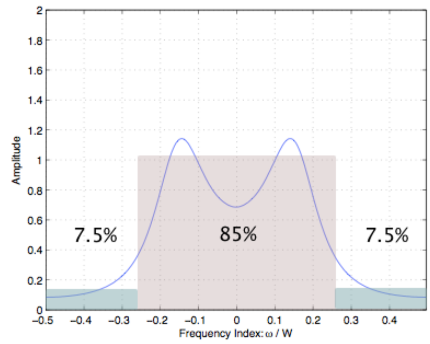

Fig. 4. Spectrum of a $(1, 20)$ RLL sequence. The spectrum gets smaller at high frequencies, but never reaches zero. The two shaded regions show the proportion of energy contained in the lower and higher frequencies.

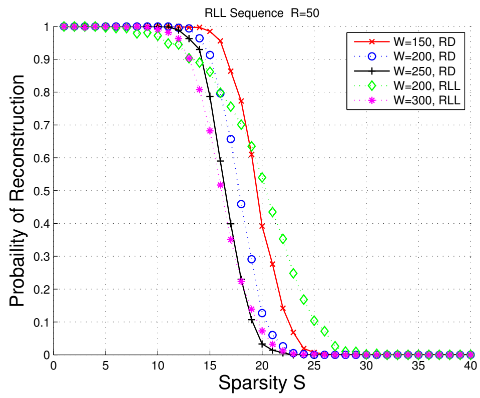

Fig. 5. Reconstruction comparison of the RD and a $(1, 20)$ RLL sequence with $R = 50$ held constant.

can be attributed to the divisibility of W/R by $(d + 1)$.

The second example we consider is an RLL sequence generated from a Markov chain and for which the correlation properties have been well-studied [13]. Fig. 4 shows the power spectrum of a RLL sequence with $(d, k) = (1, 20)$. We see that while it decreases at higher frequencies, it does not approach zero at any point; (6) tells us that $|||\Delta||| \approx 0.9$, and Theorem 1 is in effect. To verify this, we again perform numerical experiments reconstructing signals sampled with the CRD measurement matrix generated with an RLL sequence, shown in Fig. 5. The CRD using the RLL MS results are much better than the CRD with the RCS and comparable to the RD with only a slight decrease in sparsity for the CRD at the same probability.

IV. Knowledge-Enhanced CRD

Notice that (6) is a maximum over all frequencies of the spectrum and tells us that the worst case in the spectrum determines the worst-case performance for *any* sparse input signal. This is essentially confirming uniqueness of each tone's signature within the baseband. In practice, however, we are almost always concerned with the average-case, rather than worst-case, reconstruction performance. Indeed, it is infeasible to numerically evaluate the worst-case performance; even the earlier experiments report only the average-case performance.

In this section, we argue that in the average case if we impose additional constraints on the distribution of signal energy across different frequencies then we can perform better sampling and reconstruction of these signals using a carefully chosen RLL MS. This insight is motivated by the workings of the general architecture of Fig. 1;

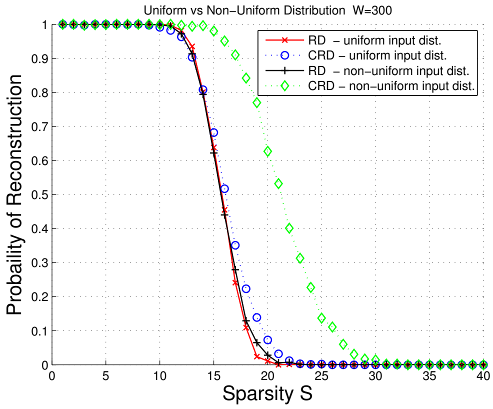

Fig. 6. The advantage of matching the spectrum of the MS with the statistics of the input signal. These reconstructions were performed with $R = 50$ and $W = 300$.

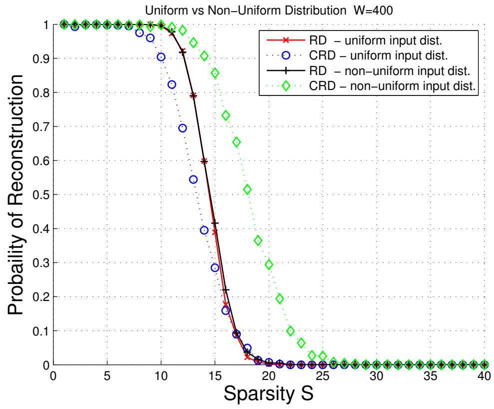

Fig. 7. These reconstructions were performed with $R = 50$ and $W = 400$. The CRD seems to have larger dependence on W/R.

the RD/CRD modulates the input signal with the MW and low-pass filters the resulting signal. Ignoring the identifiability aspects of the problem, our intuition suggests that the reconstruction performance will improve if more energy from the input signal can be modulated to baseband. The RD of course is ill-suited for this since it sends every spectral region to the baseband with (on average) equal weighting. However, the CRD, because of the non-uniform spectrum of the associated MW, is well-suited for *knowledge-enhanced* sampling of signals. To show this, we impose a matching of the spectrum of the RLL MW with the probabilistic distribution of signal energy in the spectral domain. Doing so ensures that the modulated signal, on average, contains a large amount of energy at baseband. Note that recent investigations have looked at similar ideas [11], [12], but they fail to rigorously analyze the associated identifiability aspects of the problem that we have reported in Section III of this paper.

To examine the idea of KECoRD, we perform numerical experiments with the RD and the CRD using a $(d, k) = (1, 20)$ RLL sequence. When generating the random input vectors, we use two distributions over the input signal frequencies. The first distribution is uniform over all frequencies; the second is non-uniform and places an emphasis on lower frequencies as shown in Fig. 4. Here, each tone has an 85% likelihood of falling in the lower half of the spectrum and a 15% likelihood of falling in the upper half. For the experiments we fixed $R = 50$ and show results for $W = 300$ in Fig. 6 and for $W = 400$ in Fig. 7. The curves marked as non-uniform distribution use an input signal generated as described previously. The performance of the RD, which has a flat spectrum, depends little on the class of input signal. However, the performance of the CRD depends highly on the class of input signal. For both values of W, the CRD performs much better if the *non-uniform* input signal is sampled. The CRD even performs better than the RD! The reconstruction does depend on W/R and understanding this relationship needs to be investigated. We conclude that the power spectrum of the RLL sequence holds the key to understanding sampling and reconstruction performance of the CRD; if the distribution of spectral energy of the input signal matches the spectrum of the MS, then reconstruction performance is improved. Providing a more detailed mathematical understanding for this phenomena is currently underway.

V. Conclusion

In conclusion, we have introduced statistically dependent sequences into the RD architecture and have shown insight into how the statistics of the sequence affect the reconstruction performance of the system. The first insight relates the power spectrum of the MS to the RIP of the CRD measurement matrix. The "flatter" the power spectrum, the smaller the RIP constant. Specifically, the RIP constant depends on the largest deviation of the power spectrum from a flat spectrum. The second insight is how to tailor the sampling system if additional knowledge about the input is available. Here, we relate the power spectrum of the MS to the spectrum of the input signal. If the distribution of energy across frequencies in the input signal coincides with the power spectrum of the MS, then signals sampled by such a CRD can be reconstructed very well. Providing a predictive ability to tailor the CRD for specific classes of input signals is currently under investigation.

References

[1] R. H. Walden, "Analog-to-digital converters and associated IC technologies," in *Proc. IEEE CSICS*, Oct. 2008, pp. 1–2.

[2] B. Le, T. W. Rondeau, J. H. Reed, and C. W. Bostian, "Analog-to-digital converters: A review of the past, present, and future," *IEEE Signal Proc. Mag.*, pp. 69–77, Nov. 2005.

[3] D. Rife and R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inf. Th.*, pp. 591–598, 1974.

[4] *IEEE Signal Processing Mag., Special Issue on Compressive Sampling*, vol. 25, no. 2, Mar. 2008.

[5] L. Applebaum, S. Howard, S. Searle, and R. Calderbank, "Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery," *Appl. Comp. Harmonic Anal.*, pp. 283–290, Sep. 2008.

[6] M. Mishali, Y. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," submitted to *IET J. Circuits, Devices and Systems*, Dec. 2009.

[7] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Th.*, pp. 520–544, Jan. 2010.

[8] A. Harms, W. U. Bajwa, and R. Calderbank, "Beating nyquist through correlations: A constrained random demodulator for sampling of sparse bandlimited signals," in *ICASSP 2011*.

[9] DARPA-BAA-10-38. http://www.darpa.mil/Our_Work/DSO/Programs/Knowledge_Enhanced_Compressive_Measurement_(KECoM).aspx

[10] A. Gallopoulos, C. Heegard, and P. Siegel, "The power spectrum of run-length-limited codes," *IEEE Trans on Comm*, vol. 37, no. 9, pp. 906–917, Sep 1989.

[11] J. Ranieri, R. Rovatti, and G. Setti, "Compressive sensing of localized signals: Application to analog-to-information conversion," *ISCAS 2010*.

[12] M. Mangia, R. Rovatti, and G. Setti, "Analog-to-information conversion of sparse and non-white signals: Statistical design of sensing waveforms," in *ISCAS 2011*.

[13] K. Immink, P. Siegel, and J. Wolf, "Codes for digital recorders," *IEEE Trans. Inf. Th.*, pp. 2260–2299, Oct 1998.