BEATING NYQUIST THROUGH CORRELATIONS: A CONSTRAINED RANDOM DEMODULATOR FOR SAMPLING OF SPARSE BANDLIMITED SIGNALS

Andrew Harms[†], *Waheed U. Bajwa*[‡], *and Robert Calderbank*[‡]

[†]Princeton University, Princeton, NJ 08544
[‡]Duke University, Durham, NC 27708

ABSTRACT

Technological constraints severely limit the rate at which analog-to-digital converters can reliably sample signals. Recently, Tropp et al. proposed an architecture, termed *random demodulator* (RD), that attempts to overcome this obstacle for sparse bandlimited signals. One integral component of the RD architecture is a white noise-like, bipolar modulating waveform that changes polarity at a rate equal to the signal bandwidth. Since there is a hardware limitation to how fast analog waveforms can change polarity without undergoing shape distortion, this leads to the RD also having a constraint on the maximum allowable bandwidth. In this paper, an extension of the RD, termed as *constrained* random demodulator (CRD), is proposed that bypasses this bottleneck by replacing the original modulating waveform with a *run-length limited* modulating waveform that changes polarity at a slower rate than the signal bandwidth. One of the main contributions of the paper is establishing that the CRD, despite employing a modulating waveform with correlations, enjoys theoretical guarantees that are quite similar to the original RD architecture. In addition, for a given sampling rate and rate of change in the modulating waveform polarity, numerical simulations confirm that the CRD can sample a signal with an even wider bandwidth without a significant loss in performance.

1. INTRODUCTION

One of the defining characteristics of analog-to-digital converters (ADCs) is the tradeoff between sampling rate and resolution. This tradeoff exists, in part, because the capacitors used to build ADC circuits take time to switch between charged and uncharged states, forcing designers to limit either the sampling rate or the resolution of an ADC [1, 2]. A rule of thumb for this rate–resolution tradeoff is that a doubling of the sampling rate causes a 1 bit reduction in the ADC resolution; in other words, $2^B \cdot f_s = P$, where B denotes the effective number of bits (ENOB)—a measure of ADC resolution, f_s denotes the sampling rate, and the constant P is determined by the state-of-the-art in ADC technology. Unfortunately, the constant P in ADC technology increases at a much slower pace than that dictated by Moore's law for microprocessors [1, 2]. This forces many applications to push the current ADC technology to the limit. For example, software-defined radios require sampling rate on the order of 1 GHz and therefore can only manage resolution of 10 ENOB using today's ADC technology [1].

Fortunately, the rate–resolution tradeoff of the ADC technology can be circumvented by exploiting prior knowledge of additional structure in signals. One such additional structure is *signal sparsity*; it has been known for quite some time now that bandlimited signals that are sparse in the frequency domain can be sampled at a rate that is much smaller than the Nyquist rate [3]. This old idea has been revisited in the past few years given the recent theoretical triumphs in the area of compressed sensing [4]. In particular, while several techniques have been put forward for sampling sparse bandlimited signals at sub-Nyquist rates, three candidate architectures that rely primarily on recent developments in compressed sensing are *chirp sampling* [5], *Xampling* [6], and *random demodulator* [7]. Our focus in this paper is on the sampling of bandlimited signals that can be well-approximated through a small number of frequency tones and the random demodulator (RD) architecture seems particularly well-suited for this specific problem, including near-optimal guarantees for robustness against noise. Therefore we concentrate on the RD architecture in this exposition, although some of the ideas presented also appear to be of relevance to the Xampling architecture.

1.1. Our Contributions

One integral component of the RD architecture is a white noise-like, bipolar modulating waveform that changes polarity at a rate equal to the signal bandwidth. Since there is a hardware limitation to how fast analog waveforms can change polarity without undergoing shape distortion, the RD also has a constraint on the maximum allowable signal bandwidth. This bottleneck is reminiscent of the challenges faced by researchers in the early days of magnetic recording systems. In magnetic disks, 0's and 1's are stored by magnetizing and demagnetizing the recording media and the reading head reports back the stored data as either positive or negative peaks in the read-back voltage. Increasing the recording density on a magnetic disk by packing more bits in a region causes the read-back voltage to rapidly change polarity, leading to significant distortions in the peak amplitudes, among other things, and causes a large number of read errors.

In order to overcome this challenge in magnetic recording systems, Tang and Bahl [8] introduced the idea of *run-length limited* (RLL) sequences in which run-length constraints describe the minimum separation, d, and maximum separation, k, between transitions from one state to another. The idea in the case of magnetic recording being that one can use (d, k) RLL binary sequences to increase the number of bits written on the disk by a factor of $(d + 1)$ without affecting the read-back fidelity. Note that there is a rate loss associated with converting arbitrary binary sequences to (d, k) RLL binary sequences and the major breakthrough in magnetic recording was that the rate loss associated with (d, k) sequences is smaller than $d + 1$, leading to a net increase in recording density; we refer the reader to [9] for further details on this topic.

In this paper, we make use of the lessons learned from the research on magnetic recording systems and propose an extension of the RD architecture, termed as *constrained* random demodulator (CRD), that replaces the original RD modulating waveform with a (d, k) RLL modulating waveform. This is quite similar in spirit to the use of (d, k) sequences in magnetic recording systems and

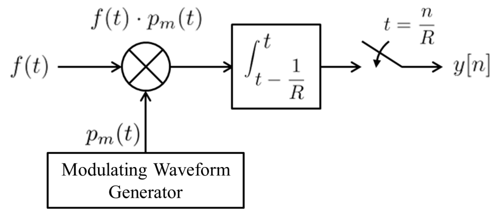


Fig. 1. Block diagram of the (constrained) random demodulator [7].

clearly leads to an increase in the operating bandwidth of the RD by a factor of $(d + 1)$ without any hardware modifications. This increase in the operating bandwidth however comes at the cost of introducing statistical dependence across the modulating waveform. One of our main contributions is establishing that the CRD, despite employing a modulating waveform with correlations, enjoys theoretical guarantees that are quite similar to the original RD architecture. In addition, one would expect an increase in the operating bandwidth to lead to an overall reduction in the allowable sparsity of the bandlimited signal. However, for a given sampling rate and rate of change in the modulating waveform polarity, we find through numerical simulations that the CRD can sample a signal with approximately $25\% \sim 35\%$ more bandwidth without a significant reduction in the signal sparsity.

2. BACKGROUND: THE RANDOM DEMODULATOR

In this section, we briefly review some of the 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 of this architecture.

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 $f(t)$. Now 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 the 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.

3. THE CONSTRAINED RANDOM DEMODULATOR

With the RD, it is possible to sample a sparse bandlimited signal at a significantly lower rate than the Nyquist rate. Still at issue though is the fact that the RD requires creation of a modulating waveform that changes polarity at the Nyquist rate. Given the nature of analog electronics, there is a hard bandwidth limit beyond which such waveforms cannot be generated without shape distortion. Stated differently, the RD makes use of a modulating waveform with an unconstrained $(d, k) = (0, \infty)$ RLL MS that in turn determines the maximum operating bandwidth of the RD architecture. On the other hand, the basic idea behind the CRD is to replace the unconstrained MS of the RD with an RLL MS with $d > 0$, thereby increasing the operating bandwidth of the architecture by a factor of $(d + 1)$ without any advances in the hardware technology. The only change to the system is the replacement of the entries of the matrix D.

There is of course a price to be paid by using RLL sequences to increase the addressable bandwidth. Specifically, recall that RLL sequences place constraints on separations between different states (transitions), which are characterized by the parameters d and k—the minimum and maximum separation, respectively. Therefore the price that we end up paying is that the elements of an RLL MS no longer remain statistically independent. However, the key insight here is that the dependence is only local and decays geometrically to zero as the elements get farther away from each other. In the following, we focus on one particular type of RLL MS for which the elements are independent whenever they are separated by a fixed number, say L. The following class of RLL MS have this particular property, and we believe that other examples can be found.

Start by defining an RLL Bernoulli sequence, $\mathrm{c} = \{c_i\}$ with $c_i \in \{0, 1\}$, as a sequence in which consecutive 1's are separated by at least d, and at most k, consecutive 0's. Next, use 1's in the RLL Bernoulli sequence to specify transitions in the RLL MS, $\mathrm{v} = \{v_i\}$

[1] Here, and throughout the rest of this paper, it is assumed without loss of generality that R divides W.

with $v_i \in \{+1, -1\}$, generated from c as $v_i = (-1)^{c_i} v_{i-1}$. Now define a block code, C, of all RLL MS exponentiated from RLL Bernoulli sequences of length L that also begin and end with $(d+1)$ 0's. Notice that concatenation of any two codes in C produces a (d, k') RLL MS with

$$k' = \begin{cases} k + 2d, & L > k + 2(d+1) \\ \infty, & L \leq k + 2(d+1) \end{cases}.$$

In order to minimize L, we require that $k = \infty$ and that L satisfies $L \geq 2(d+1) + 1 = 2d + 3$. We can now create RLL MS of arbitrary length if we concatenate sequences drawn independently from C. Denote such a RLL MS as $\beta = \{\beta_i\}$, and it has the following property: $\forall i, \beta_i$ independent of β_{i+t} if $|t| \geq L$.

As an example, let us construct a code with $L = 5$, $d = 1$, and $k = \infty$. C contains four sequences: $(+1, +1, -1, -1, -1)$, $(-1, -1, +1, +1, +1)$, $(+1, +1, +1, +1, +1)$, and $(-1, -1, -1, -1, -1)$. Sequences are chosen equally likely from C and concatenated together to form a MS, β. If β is chosen in such a way, then any two elements of the sequence, β_i and β_{i+t}, are independent if $|t| \geq 5$.

The structure of the RD matrix Φ_{RD} is analyzed in [7]. Crucial to the reconstruction of signals sampled by the RD is the satisfaction of the Restricted Isometry Property (RIP) by Φ_{RD}. The independence of the MS and the independence of the rows of Φ_{RD} are used to show the satisfaction of the RIP. Our use of RLL sequences in the CRD matrix, Φ_{CRD}, introduces a dependent MS and dependence among the rows of Φ_{CRD}, and we must show that Φ_{CRD} satisfies the RIP. To accomplish this, we apply an argument similar to that used in [10] for Toeplitz matrices and present results that show the RIP is satisfied if elements of the MS are independent if sufficiently separated. We can use this new analysis to show that Φ_{CRD} satisfies the RIP because RLL sequences exhibit this type of behavior. To this end, we give the following theorem for which a proof will be given in a future journal paper.[2]

Theorem 1 (Recovery of General Bandlimited Signals)**.** *Suppose that the sampling rate, R, satisfies*

$$R \geq \ell^3 \cdot C \cdot S \log^6(W)$$

and that R divides W and $\ell = 2d + 3$, $d > 0$. Draw a $R \times W$ CRD matrix Φ_{CRD} using an RLL MS as described in Section 3 with parameters $(d, k = \infty)$. The following statement holds, except with probability $O(W^{-1})$.

Suppose that γ is an arbitrary amplitude vector and ν is a noise vector with $||\nu||_2 \leq \eta$. Let $\mathrm{y} = \Phi_{CRD}\gamma + \nu$ be noisy samples collected by the CRD. Then every solution $\hat{\gamma}$ to the convex program $\hat{\gamma} = arg \min ||\mathrm{v}||_1$ subject to $||\Phi_{CRD}\mathrm{v} - \mathrm{y}||_2 \leq \eta$ approximates the target vector γ

$$||\hat{\gamma} - \gamma||_2 \leq C \max\left\{\eta, \frac{1}{\sqrt{S}}||\gamma - \gamma_S||_1\right\}$$

where γ_S is a best S-sparse approximation to γ with respect to the ℓ_1 norm.

4. IMPLEMENTATION ISSUES

In practice, the MS would be generated from a finite-state machine (i.e. a pseudo-random binary sequence). The sequences considered up to now (the Rademacher sequence and RLL sequences with $k = \infty$) require an infinite number of states. For this reason, we also consider general RLL sequences generated from a Markov chain with $k < \infty$, of which the generation and correlation characteristics has been analyzed [9, 11]. The Markov chain is stationary and has $2k + 2$ states described by a transition matrix $P = \{p_{ij}\}$ [9]. While current techniques do not prove that the dependence in such a sequence is local, the correlation decreases geometrically to zero as the separation within the sequence grows. This leads us to believe that these sequences will perform well as a MS in the CRD.

Calculation of the auto-correlation function $R_x(m)$ for such a sequence requires the vector a : $a_i = \sum_{u=1}^{2k+2} \pi_u p_{ui} y_{ui}$ containing the weighted sum of symbols transmitted on arriving at each state and the vector b : $b_j = \sum_{v=1}^{2k+2} p_{jv} y_{jv}$ containing the weighted sum of symbols transmitted on departing each state. Here, π_u is the u^{th} entry of the stationary distribution for P and y_{ij} is the symbol transmitted on departing state i and arriving at state j. Now $R_x(m) = \mathrm{b}^T P^{(m-1)} \mathrm{a}$, where by stationarity, the auto-correlation is a function of only the separation between elements [11].

Fig. 2 shows a plot of the auto-correlation function for several (d, k) sequences. The correlation decays geometrically as m, the separation in time, increases.

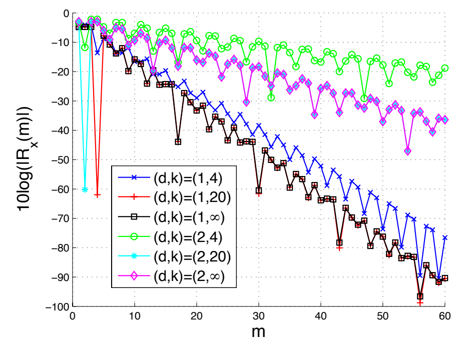


Fig. 2. Log-magnitude plot of the auto-correlation of an RLL code as a function of the time separation. Larger values of d and smaller values of k exhibit stronger correlation. The function for $k = 20$ is nearly identical to the function for $k = \infty$. For reference, an independent sequence is given by $d = 0$ and $k = \infty$.

5. NUMERICAL RESULTS AND DISCUSSION

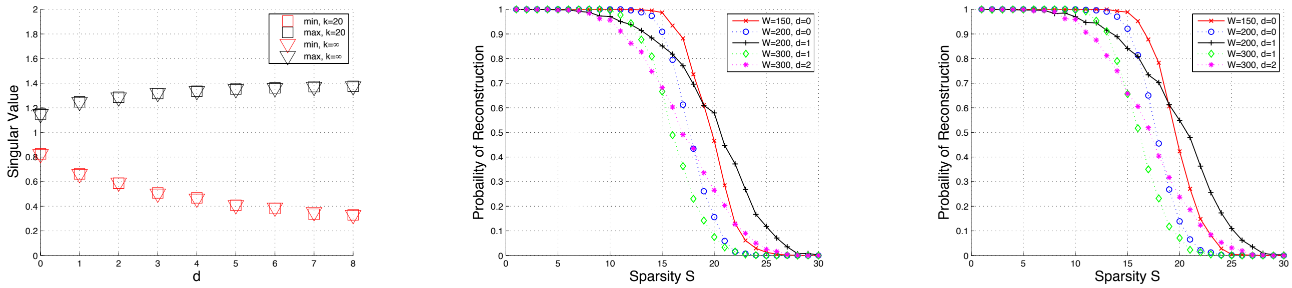
The RIP of a matrix Φ is important for the recovery of signals using the techniques of compressive sensing. The RIP of order N with restricted isometry constant δ_N is satisfied for Φ if

$$\left|\frac{||\Phi \mathrm{x}||_2^2 - ||\mathrm{x}||_2^2}{||\mathrm{x}||_2^2}\right| \leq \delta_N \quad (3)$$

with $\delta_N \in (0, 1)$ and $||\mathrm{x}||_0 \leq N$. Stated alternatively, singular values of $W \times N$ sub-matrices of Φ satisfy $\sqrt{1 - \delta_N} < \sigma_N < \sqrt{1 + \delta_N}$.

The RIP is shown to hold for the RD [7] and for the CRD with a particular RLL MS in Theorem 1. We do not have theoretical results to show that it holds for a CRD using general RLL MS, but numerical results suggest that it does. Fig. 3(a) shows the minimum and maximum singular values as a function of d obtained by evaluating several sub-matrices of a CRD matrix. The plot for $k = \infty$ was obtained using RLL MS described in Section 3 and for which Theorem 1 gives theoretical guarantees. The plot for $k = 20$ was obtained using general RLL MS described in Section 4. While we cannot give

[2]For $d = 0$, $\ell = 1$ and the result of [7] is obtained.



(a) Minimum and maximum singular values of sub-matrices of a CRD matrix as a function of d with k both infinite and finite ($d = 0$, $k = \infty$ gives the RD of [7]). The matrix is constructed with $W = 512$, $R = 256$, and $S = 10$ while 100 randomly chosen sub-matrices are evaluated.

(b) $k = \infty$: Probability of successful reconstruction vs. sparsity using the Basis Pursuit algorithm for 1000 instances of Φ for each set of parameters. The (x) and (o) marks are the RD with a 150Hz and 200Hz signal. The (+) and ($\Diamond$) marks are the CRD with (d, k)=$(1, \infty)$ and a 200Hz and 300Hz signal. The (*) marks are the CRD with (d, k)=$(2, \infty)$ and a 300Hz signal. Also, $R = 50$ and the transition width is fixed for each.

(c) $k = 20$: Probability of successful reconstruction vs. sparsity using the Basis Pursuit algorithm for 1000 instances of Φ for each set of parameters. The (x) and (o) marks are the RD with a 150Hz and 200Hz signal. The (+) and ($\Diamond$) marks are the CRD with (d, k)=$(1, 20)$ and a 200Hz and 300Hz signal. The (*) marks are the CRD with (d, k)=$(2, 20)$ and a 300Hz signal. Also, $R = 50$ and the transition width is fixed for each.

Fig. 3. Numerical Results

theoretical guarantees for these MS, we see that the singular values are bounded close to 1 and nearly identical for both classes of MS. This leads us to believe that the RIP will be satisfied, and we will not see a significant performance hit if we use the more practical RLL MS with $k < \infty$. Again, for reference the RD uses (d, k)=$(0, \infty)$.

Fig. 3(b) and Fig. 3(c) plot the probability of successful reconstruction as a function of the input signal sparsity, S. The curves were obtained with different (d, k) sequences and signal bandwidth W. Fig. 3(b) uses RLL MS with $k = \infty$ and for which we provide a theoretical background of its performance. Fig. 3(c) uses RLL MS with $k = 20$. Again, these MS do not provably become independent and so do not exactly fit into our theoretical framework. The correlation, however, decays quickly, and the performance is only slightly degraded compared to the $k = \infty$ RLL MS. The $d = 0$ curve is again the RD of [7] and provides the baseline for our comparison, and all curves were created with a random waveform that switches at the same rate. The CRD with $d = 1$ offers comparable performance to the unconstrained RD but with the benefit of acquiring an input signal with more bandwidth. Depending on the tolerance in reconstruction probability, the CRD can provide up to a 33% increase in the acquirable bandwidth. Even at 50% greater bandwidth, the CRD only reduces the sparsity by $\sim$4 (25%). This shows numerically that observable bandwidth can be increased with a slight to no drop in the sparsity.

In summary, the RD shows that a sparse bandlimited signal can be sampled not only based on the bandwidth of the signal, but also the sparsity of the signal. The underlying hardware also gives a minimum transition width of the random waveform, so when at this limit we are limited to viewing a bandwidth W with the RD. However, by using RLL MS and creating a CRD, we can increase the bandwidth up to $W' \leq (d+1)W$ if we are willing to incur a penalty in the sparsity of the signal. We also give a theoretical bound on this penalty but believe that limit to be loose. Numerical simulations show that only a small penalty is encountered for bandwidth increases of even 50%. Despite the sparsity penalty, we still gain a great advantage; the RD is limited by the hardware to viewing signals of a particular bandwidth while the CRD can look beyond this bandwidth.

6. 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, Sept. 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, Sept. 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] D. Tang and L. Bahl, "Block codes for a class of constrained noiseless channels," *Inform. Cont.*, pp. 436–461, Dec 1970.

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

[10] W. U. Bajwa, *New Information Processing Theory and Methods for Exploiting Sparsity in Wireless Systems*, Ph.D. thesis, University of Wisconsin - Madison, 2009.

[11] G. Bilardi, R. Padovani, and G. Pierbon, "Spectral analysis of functions of Markov chains with applications," *IEEE Trans. Commun.*, pp. 853–861, July 1983.