Shaping the Power Spectra of Bipolar Sequences with Application to Sub-Nyquist Sampling

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*—In the design of line codes for data transmission and modulation codes for data storage it is important to match the spectral properties of bipolar sequences (containing the symbols ± 1) to the spectral properties of the transmission or storage medium. This paper demonstrates that spectral shaping is possible with bipolar sequences, and explores the advantage of bipolar sequences with shaped spectra in Analog-to-Digital Converters (ADCs) where they are used to modulate a sparse input waveform that is described by a small number of parameters. Performance is improved by matching the spectrum of the modulating waveform to prior information about the spectrum of the sparse input. The focus is on the Random Demodulator where sparse signals are reconstructed from sub-Nyquist samples.**

I. Introduction

The characteristics of a recording (or communication) medium define a channel; for example, readback voltage responds to changes in magnetization, defining a one-dimensional channel. This "partial-response" (or differential) channel leads to inter-symbol interference (ISI) between closely spaced symbols. It is important to write sequences to the medium that minimize ISI or that have small spectral energy at the nulls of the corresponding partial-response channel. Tang and Bahl [1] introduced run-length limited (RLL) (d, k) codes in 1970 to limit inter-symbol interference in bipolar sequences. These codes limit the minimum, d, and maximum, k, length runs allowed in the sequence without changing symbol and are analyzed further in [2], [3].

A recent application of bipolar sequences is to the design of Analog-to-Digital Converters (ADCs) that overcome the Nyquist limit linking the sampling rate of ADCs to the bandwidth of acquired signals. Spectrum shaping of sequences is very important in this application because matching the spectrum of the sequence to the distribution of spectral energy in the input signal significantly improves performance [4]. This paper focuses exclusively on the application to sub-Nyquist sampling. Two defining characteristics of ADCs are *sampling rate* and *resolution*, and they are generally in contention. An increase in resolution requires a drop in rate, and vice-versa. The rule-of-thumb is that doubling 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 resolution, f_s is the sampling rate, and the constant P is a figure-of-merit determined by the ADC architecture. This figure-of-merit advances at a much slower pace than Moore's law for microprocessors [5]–[7], and many current applications push 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 [5]. An ADC sampling below the Nyquist rate of the acquired signal offers higher resolution.

The Nyquist limit is overcome by exploiting prior information about the signal other than the bandwidth, including prior knowledge that the signal has a *sparse* representation in some basis. It has long been known that bandlimited signals that are sparse in the frequency domain can be sampled at rates much lower than the Nyquist rate [8]. Recent advances in Compressed Sensing (CS) [9] have breathed new life into the investigation of exploiting signal sparsity to sample at much lower than Nyquist rates. Two architectures which exploit these recent developments in CS are *Xampling* [10] and the *Random Demodulator* [11], and both use modulation with bipolar waveforms as a major piece of the architecture. In this paper we concentrate on signals that are well-approximated by a small number of frequency tones and on the Random Demodulator (RD) architecture.

The modulating sequences in the RD are random bipolar sequences where each entry is generated independently, and transitions in the polarity of the waveform occur at the Nyquist rate of the input signal. In [4], RLL sequences are introduced into the RD to overcome hardware constraints that can limit the rate at which the waveform changes polarity while retaining high fidelity. The resulting system is coined a *Constrained Random Demodulator* (CRD). The RLL sequences also introduce a correlation structure and a characteristic power spectrum. The primary benefit is realized when the spectrum of the modulating sequence is *tuned* to the spectrum of the input signal, a so-called *Knowledge-Enhanced* CRD (KECoRD). If the spectrum of the random waveform is proportional to the probabilistic distribution on the input tones, then the spectrum is considered *matched* to the input signal. In this case, more energy is *passed through* the CRD to the measurements, and signal reconstruction performance is significantly improved [4]. The CRD favors some frequencies over others and, consequently, performs better than the RD when sampling signals with a *matched* distribution of spectral energy.

A. Our Contributions

The analysis in [4] concentrates on RLL sequences that limit transitions in the modulating waveform (MW), i.e., $d > 0$. It does not address how to shape spectra to match a given input signal, nor does it offer a thorough analysis of the advantages of matching the MW to input signal. This paper begins the analysis of shaping power spectra. Our primary contribution is an analysis of the spectra of a family of bipolar sequences and highlighting their utility in ADC architectures. The MW is generated from a statistically dependent sequence and has a characteristic, non-flat power spectrum associated with it [2], [3]. The design of sequences with arbitrary spectra is vital to matching sequences with input signals of arbitrary spectral energy distribution. However, it is also complex, and we focus on families of sequences that are generated by tuning a particular Markov chain introduced in [3] for analysis of RLL sequences. We simplify the analysis of these sequences by limiting the analysis to a family of Markov chains described by a single parameter. Despite the simplification, the family of sequences generated from these simple Markov chains can be tuned to cover the entire range of frequencies.

978-1-4673-3146-3/13/$31.00 ©2013IEEE

II. Spectra of Bipolar Sequences

Consider a bipolar sequence $\varepsilon_i \in \{+1, -1\}$ for $i = 0, ..., W-1$ generated by the Markov chain in Fig. 1. Nodes in the top row produce the symbol $+1$, while those in the bottom row produce the symbol -1. The number below each edge indicates the transition probability between connected nodes. These transition probabilities are denoted by p_{ij}. The symmetry of the block diagram ensures both symbols are equally likely and the entries are not independent.

The transition probabilities can be collected into a matrix $\mathrm{P} = [p_{ij}]$, where P is a stochastic matrix. The stationary distribution of the Markov chain is denoted as π and satisfies the relation $\pi^T = \pi^T \mathrm{P}$. Additionally, the symbols output from each state are collected into the vector $\mathrm{b} = [+1 \cdots +1 \; -1 \cdots -1]$ where k entries of $+1$ are followed by k entries of -1. With these in hand, the autocorrelation of a sequence generated from such a Markov chain is

$$R_\varepsilon(m) = \mathrm{a}^T \mathrm{P}^m \mathrm{b} \tag{1}$$

where $\mathrm{a}^T = \mathrm{b}^T \cdot \mathrm{diag}(\pi)$, $\mathrm{diag}(\mathrm{x})$ is a diagonal matrix with x on the diagonal, and $R_\varepsilon(-m) = R_\varepsilon(m)$ [12]. We can understand the behavior of $R_\varepsilon(m)$ from the eigenvalues of P. Denote the eigenvalues of P as λ_i for $i = 1, ..., 2k$ and label them in descending order of magnitude

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{2k}|.$$

Since P is a stochastic matrix, $\lambda_1 = 1$ (with a corresponding all ones eigenvector), and $|\lambda_2| < 1$. We can bound the autocorrelation function [13, Theorem 8.5.1]

$$|R_\varepsilon(m)| \leq \lambda_2^m$$

and see that the autocorrelation experiences geometric decay.

The autocorrelation is written explicitly in terms of the eigenvalues through an eigen-decomposition of P as

$$R_\varepsilon(m) = \mathrm{a}^T \mathrm{P}^m \mathrm{b} = \mathrm{a}^T \mathrm{V} \Lambda^m \mathrm{V}^{-1} \mathrm{b} \tag{2}$$

where Λ is a diagonal matrix with the eigenvalues of P and V has columns containing the corresponding eigenvectors of P. Calculating P^m is much easier with this formulation because Λ is diagonal. The power spectrum of the sequence is the Fourier transform of the autocorrelation function $F_\varepsilon(\omega) = \sum_m R_\varepsilon(m) e^{-\frac{2\pi}{W} m\omega}$.

A. Single Parameter Markov Chains

Analysis of (2) for P corresponding to an arbitrary Markov chain of the form in Fig. 1 is complex. We focus the discussion on a Markov chain described by a single parameter p, which occurs when $k = d + 1$. This family of Markov chains shows that varying p and d allows us to place the spectral peak across the entire range of frequencies.

We start with the Markov chain resulting from $d = 0$ and $k = 1$, which has a four state diagram with transition probability matrix

$$\mathrm{P} = \begin{bmatrix} 0 & p & 1-p & 0 \\ 0 & 0 & 1 & 0 \\ 1-p & 0 & 0 & p \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{3}$$

where we have used the substitution $p_1 = p$. The eigenvalues are given by $\lambda_1 = 1$, $\lambda_2 = -p$, $\lambda_3 = -\frac{1}{2}(1-p) + \sqrt{\frac{1}{4}(1-p)^2 - p}$, and $\lambda_4 = -\frac{1}{2}(1-p) - \sqrt{\frac{1}{4}(1-p)^2 - p}$.

The spectra are plotted in Figs. 2-4 for a range of values of p. Consideration of the extreme values of $p = 0$ and $p = 1$ provides intuition into these plots. In the case of $p = 0$, the

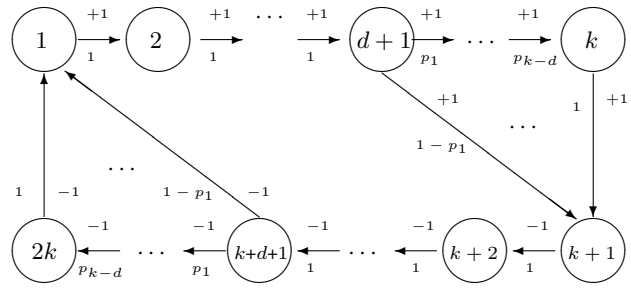

Fig. 1. State diagram of the Markov chain generating an RLL sequence. The transition probabilities are symmetric in the sense that $p_{(i+k)(j+k)} = p_{ij}$ where the sum is taken modulo $2k$. The top half outputs the symbol +1 while the bottom half outputs -1. The symbols are shown above the edges connecting nodes, while the transition probabilities are shown below the connecting edges.

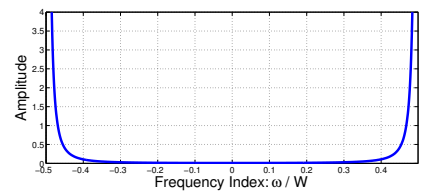

Fig. 2. Spectrum of a sequence with $d = 0$, $k = 1$, and $p = 0.01$.

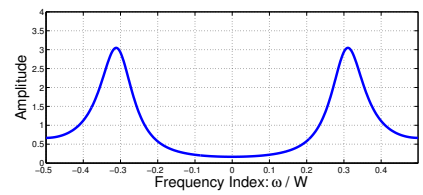

Fig. 3. Spectrum of a sequence with $d = 0$, $k = 1$, and $p = 0.5$.

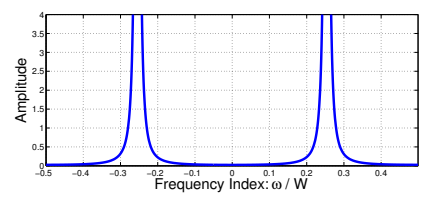

Fig. 4. Spectrum of a sequence with $d = 0$, $k = 1$, and $p = 0.95$.

Markov chain only visits two states, state 1 and state 3. The resulting sequence consecutively alternates between $+1$ and -1. The spectrum for this sequence is a dirac-delta function at frequency ± 0.5, and the spectrum in Fig. 2 is approaching a dirac-delta at ± 0.5 for a small value of p. On the other hand, if $p = 1$ the Markov chain cycles through all four states in order. The resulting sequence always contains symbols in alternating pairs, two symbols $+1$ followed by two symbols -1. The spectrum is therefore a dirac-delta at frequency ± 0.25, half the frequency of the previous sequence, and the spectrum in Fig. 4 is approaching a dirac-delta at ± 0.25 for a large value of p. The spectrum for $p = 0.5$ contains a peak, with non-zero width, at a frequency between ± 0.25 and ± 0.5. Such a sequence contains some repeated symbols, but never more than a single repeated symbol, and as a result low-frequency content is suppressed.

We can produce spectra with more low-frequency energy by increasing d but keeping $k - d = 1$. We retain a single parameter Markov chain, and as p ranges between $p = 0$ and $p = 1$,

the spectrum contains a peak that moves between two limiting frequencies. The upper limit is given by $\frac{1}{2(d+1)}$, and the lower limit is given by $\frac{1}{2(k+1)}$.

III. Application: Sub-Nyquist Sampling of Frequency-Sparse Signals

Armed with the family of sequences generated from single parameter Markov chains, we build a tunable ADC that can be matched to *a priori* knowledge about the input signal. We show the benefits in the context of the RD architecture, and briefly review its key aspects while referring the reader to [11] for a comprehensive overview.

A. Review of the Constrained Random Demodulator

The basic purpose of the RD is sampling at a sub-Nyquist rate while retaining the ability to reconstruct signals that are periodic, limited in bandwidth to W Hz, and completely described by a total of $S \ll W$ tones. In other words, a signal $f(t)$ input to the RD takes the parametric form

$$f(t) = \sum_{\omega \in \Omega} a_\omega e^{-2\pi i \omega t}, \ t \in [0,1) \tag{4}$$

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. To acquire $f(t)$, the RD performs three basic actions as illustrated in Fig. 5. First, it multiplies $f(t)$ with a modulating waveform $p_m(t)$ consisting of a train of shifted square pulses with amplitudes determined by a *modulating sequence* (MS) $\{\varepsilon_n\}$ independently taking values $+1$ or -1 each with probability $1/2$. Next, it low-pass filters the continuous-time product $f(t) \cdot p_m(t)$. Finally, it takes samples at the output of the filter at a rate of $R \ll W$.

A major contribution of [11] is showing that the actions of the RD on $f(t)$ are expressed 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. Let $\mathrm{x} \in \mathbb{C}^W$ denote a Nyquist-sampled version of the continuous-time input signal $f(t)$. Then x is 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 $f(t)$. 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 $\mathrm{x} = F\alpha$. 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 rth row of H, with $DF\alpha$.[1] If R samples are collected at the output of the RD into a vector $\mathrm{y} \in \mathbb{C}^R$, then it follows that $\mathrm{y} = HDF\alpha = \Phi_{RD} \cdot \alpha$, where we have the random demodulator matrix $\Phi_{RD} = HDF$.

Given the discrete representation $\mathrm{y} = \Phi_{RD} \cdot \alpha$, recovering $f(t)$ described in (4) 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. Recent theoretical developments in CS provide a rich array of greedy and 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} is shown to satisfy certain geometrical properties [9]. One such property is the *Restricted Isometry Property* (RIP). The RIP of order S with restricted isometry constant δ_S is satisfied for a matrix Φ with unit-norm columns if

$$(1-\delta_S)||x||_2^2 \le ||\Phi x||_2^2 \le (1+\delta_S)||x||_2^2 \tag{5}$$

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

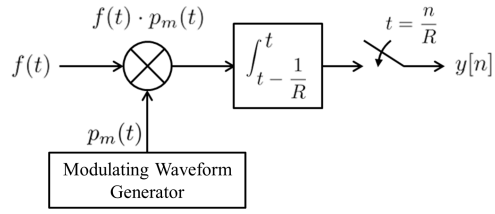

Fig. 5. Block diagram of the (constrained) random demodulator [11].

with $\delta_S \in (0,1)$ and $\|\mathrm{x}\|_0 \le S$. The notation of [11] compactly describes the RIP. For a matrix A, the norm $|||A||| = \sup_{|\Omega| \le S} \|A|_{\Omega\times\Omega}\|$ captures the largest singular value of any $S \times S$ principal submatrix of A, and (5) is satisfied if and only if $|||\Phi^*\Phi - I||| \le \delta_S$. A highlight of [11] is that the RD matrix is explicitly shown to satisfy the RIP 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.

Motivated by the desire to use a MS that switches at a rate lower than the Nyquist rate, [4] introduced the CRD using an RLL sequence with statistically dependent entries. The only change to the RD architecture is a replacement of the independent MS by a dependent one. For the matrix representation, the diagonal entries of the matrix D are changed from an independent sequence to an RLL sequence. If the correlation structure of the sequence is sufficiently well-behaved, then the CRD enjoys RIP guarantees that take only a slight hit over the RD. An $R \times W$ CRD satisfies RIP with high probability if

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

where C is a positive constant [4, Theorem 1]. The maximum dependence distance ℓ is the smallest integer such that any two entries in the MS separated by ℓ are independent. The matrix Δ is determined by the correlation properties of the MS and has entries

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

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. The quantity $|||\Delta|||$ is related to the *power spectrum* of the MS by

$$|||\Delta||| \approx \max_\omega |\tilde{F}_\varepsilon(\omega)| \tag{7}$$

where $\tilde{F}_\varepsilon(\omega) = F_\varepsilon(\omega) - 1$ and $F_\varepsilon(\omega) = \sum_m R_\varepsilon(m) e^{-\frac{2\pi}{W} m\omega}$ is the power spectrum. This means the RIP constant δ is minimal for a flat spectrum (produced by an independent sequence) and is approximately the largest deviation of the power spectrum from a flat spectrum. For the RIP, it is important that $0 < F_\varepsilon(\omega) < 2$.

Fig. 6 shows the power spectrum of an RLL sequence with $(d,k) = (1,20)$. We see that while it decreases at higher frequencies, it does not approach zero at any point; $|||\Delta||| \approx 0.9$, and [4, Theorem 1] is applicable.

B. Knowledge-Enhanced CRD and Spectra of Bipolar Sequences

The RD relies on two principles to recover sparse input signals: 1) identifiability of each tone and 2) sufficient energy capture of each tone. Notice that (7) is a maximum over all frequencies 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.

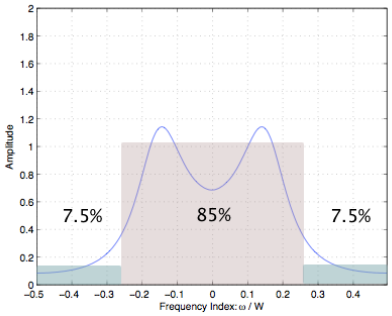

Fig. 6. Spectrum of a $(1, 20)$ RLL sequence. The spectrum gets smaller at high frequencies, but never reaches zero. The two shaded regions show the proportion of energy contained in the low and high frequency regions.

We argue that, in the average case, imposing additional constraints on the signal energy distribution across frequencies improves the sampling and reconstruction performance for these signals if carefully crafted modulating sequences are used in the CRD. This insight is motivated by the workings of the architecture of Fig. 5; the RD modulates the input signal with the MW and low-pass filters the resulting signal. Setting aside the identifiability aspects of the problem, intuition suggests that the reconstruction performance will improve if more energy from the input signal is modulated to baseband. The RD sends every spectral region to baseband with (on average) equal weighting, and is ideal if no additional knowledge about the signal is available. However, a correlated sequence with a non-flat spectrum will *favor* some tones over others in the input signal. To show this, we match the spectrum of the RLL sequence with the probabilistic distribution of signal energy in the spectral domain. This ensures that the modulated signal, on average, contains a large amount of energy at baseband. The family of sequences described in Section II-A allows the tuning of modulating sequences to match the input signals with a single peak in the spectrum.

To verify the idea of KECoRD, we perform numerical experiments with the RD and CRD. When generating the random input vectors, the occurrence of tones has a distribution proportional to the spectrum in Fig. 3. If the spectrum of the random sequence matches the distribution of spectral energy in the input signal, as in Fig. 7, then reconstruction performance is improved over the RD. The performance of the RD, which has a flat spectrum, is shown in Fig. 8. The 50% recovery line is shown in blue. The number of tones at this 50% line is improved by roughly 5% for the CRD over the RD.

IV. Conclusions

We have motivated the need for sub-Nyquist sampling in spectrum sensing applications and the usefulness of shaping the power spectrum of bipolar sequences used in sub-Nyquist architectures. We have shown that we can shape the spectra of sequences and place energy in a peak at frequencies across the entire band of interest. Additionally, we show the utility of spectrum shaping by showing improved performance when spectra are *matched* to prior knowledge about the class of input signals. An interesting direction for future work is the systematic analysis of a more complicated Markov chain with more than one degree of freedom to allow the shaping of more complicated multimodal spectra.

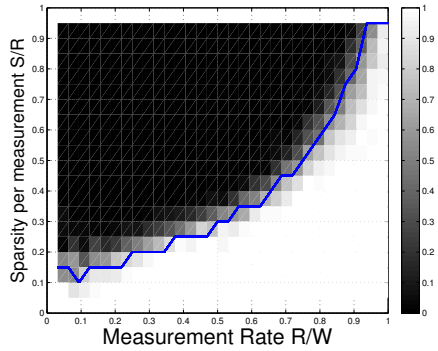

Fig. 7. Phase transition plot of the CRD with a $(d, k) = (0, 1)$ sequence with $p = 0.5$. The 50% empirical recovery line is plotted in blue and is shifted up and left relative to the RD.

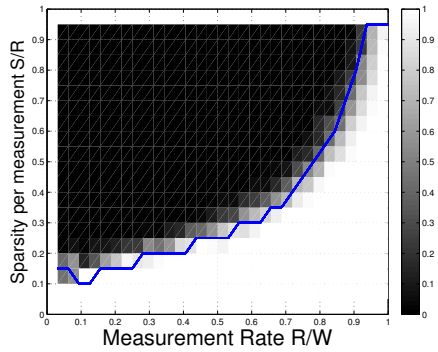

Fig. 8. Phase transition plot of the RD with a $(d, k) = (0, \infty)$ sequence with $p = 0.5$. The 50% empirical recovery line is plotted in blue and is shifted down and right relative to the CRD.

References

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

[2] 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.

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

[4] A. Harms, W. U. Bajwa, and R. Calderbank, "A constrained random demodulator for sub-Nyquist sampling," *IEEE Trans. Signal Proc.*, vol. 61, no. 3, pp. 707–723, Feb 2013.

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

[6] 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 Sig. Proc. Mag.*, pp. 69–77, Nov. 2005.

[7] B. Murmann. ADC performance survey 1997-2012. [Online]. Available: http://www.stanford.edu/ murmann/adcsurvey.html

[8] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. 20, no. 5, pp. 591–598, Sep. 1974.

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

[10] M. Mishali, Y. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET J. Circ., Dev., and Sys.*, vol. 5, no. 1, pp. 8–20, Jan 2011.

[11] 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.

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

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.