

# Frugal Sensing: Wideband Power Spectrum Sensing From Few Bits

Omar Mehanna, *Student Member, IEEE*, and Nicholas D. Sidiropoulos, *Fellow, IEEE*

**Abstract**—Wideband spectrum sensing is a key requirement for cognitive radio access. It now appears increasingly likely that spectrum sensing will be performed using networks of sensors, or crowd-sourced to handheld mobile devices. Here, a network sensing scenario is considered, where scattered low-end sensors filter and measure the average signal power across a band of interest, and each sensor communicates a single bit (or coarsely quantized level) to a fusion center, depending on whether its measurement is above a certain threshold. The focus is on the underdetermined case, where relatively few bits are available at the fusion center. Exploiting non-negativity and the linear relationship between the power spectrum and the autocorrelation, it is shown that adequate power spectrum sensing is possible from few bits, even for dense spectra. The formulation can be viewed as generalizing classical nonparametric power spectrum estimation to the case where the data is in the form of inequalities, rather than equalities.

**Index Terms**—Cognitive radio, distributed spectrum compression, spectral analysis, spectrum sensing.

## I. INTRODUCTION

**E**FFICIENT utilization of the wireless spectrum has been a growing concern, due to the remarkable growth in the mobile Internet and the variety of emerging wireless devices and services competing for bandwidth. Actively seeking and exploiting transmission opportunities while respecting the ‘right of way’ of licensed users, cognitive radio is a promising *cohabitation* paradigm that is currently at the center stage of wireless communication and networking research.

Spectrum sensing is a core functionality for cognitive radio, as it forms the basis for adaptive spectrum sharing. The goal of spectrum sensing is to detect spectral occupancy, and perhaps coarsely estimate power levels, under sensing constraints that typically preclude explicitly scanning the full band. A variety of spectrum sensing methods have been developed in recent years, ranging from narrowband energy detection to wideband sensing, mostly based on isolated hypothesis testing per narrowband channel ‘bin’, without taking into account dependence

across frequency bins or exploiting any underlying parametrization. Reference [1] provides a good up-to-date review of spectrum sensing for cognitive radio.

The premise of cognitive radio is that most of the band is idle, most of the time, i.e., measured spectra are typically sparse. Building upon this premise, *compressive spectrum sensing* has been introduced to exploit frequency-domain sparsity to obtain accurate spectrum estimates at sub-Nyquist sampling rates, without frequency sweeping [2], [3]. A cooperative protocol for distributed compressive spectrum sensing has been developed in [4], enabling cognitive radio users to reach consensus on globally fused sensing outcomes.

Most work on spectrum sensing focuses on detecting activity in the *spectrum* versus the *power spectrum*, i.e., the Fourier transform of the signal, as opposed to the Fourier transform of its autocorrelation function. The power spectrum is an expectation that reflects long-term spectral activity patterns; short-term effects such as fading are integrated out. Power spectrum sensing has been explored very recently in [5]–[10], where it was shown that neither Nyquist-rate sampling nor full-band scanning is necessary when the goal is to estimate only a finite set of correlation lags, which is then Fourier transformed to yield an estimate of the power spectrum. This approach can decrease the sampling rate requirements by exploiting the ‘correlation parametrization’ (i.e., a low-order correlation model), without requiring spectrum sparsity. The key to this line of work is that power measurements are linear in the autocorrelation function, hence a finite number of autocorrelation lags can be estimated by collecting enough power measurements to build an over-determined system of linear equations. In [5], the power spectrum is estimated using sub-Nyquist rate sampling by exploiting the relationship between the autocorrelation function of the Nyquist-rate samples and that of the compressive measurements. The assumption that compressed measurements remain wide-sense stationary is relaxed in [6], where the under- and over-determined cases are considered. When over-determined, the power spectrum is estimated using linear least-squares, without recourse to additional signal properties. When under-determined, the problem is regularized by minimizing the  $\ell_1$  norm of the estimated power spectrum, thus relying on sparsity in this case.

A bank of periodic modulators is considered in [7], [8], where each branch is sampled at a fraction of the Nyquist rate, and cross-correlations of the branch outputs are used to build a system of linear equations in the unknown input correlation for a fixed number of lags. This approach has been generalized to the case of cyclostationary signals in [9]. In [10], multi-coset sampling is employed producing multi-resolution

Manuscript received August 04, 2012; revised January 13, 2013; accepted February 14, 2013. Date of publication March 15, 2013; date of current version April 26, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marius Pesavento. This work was supported in part by NSF CCF-0747332, and NSF ECCS-1231504.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: meha0006@umn.edu; nikos@umn.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2013.2252171

power spectral estimates at arbitrarily low average sampling rates. A different approach exploiting spectrum sparsity has been proposed in [11], where  $K$  wideband filters are used to detect occupancy in  $N$  channels with  $K < N$ , assuming that the number of occupied channels is up to  $O(K)$  (less than  $K/2$ ). Note that [11] does not exploit the autocorrelation parametrization.

References [5]–[11] assume analog amplitude samples (i.e., ignore quantization issues), which is reasonable for *lumped measurements* taken with relatively accurate A/D converters at a high number of bits per sample. The situation is very different in a network sensing setting using scattered low-end sensors with limited communication capabilities, which is the scenario considered here. Suppose that each sensor can only down-convert, filter, and measure average power at the output of its filter. Depending on the computed power level, the sensor may send a binary signal to the fusion center, or broadcast it to its peers. Is it possible to form a satisfactory estimate of the ambient power spectrum using just few such bits? This is the central question we set to address in this paper.

Power spectrum sensing from few bits has never been considered in the past, to the best of our knowledge—yet is a natural extension of classical spectral estimation to the case where the data is in the form of inequalities, rather than equalities. Exploiting linearity with respect to autocorrelation and important non-negativity properties in a novel optimization-based formulation, it is shown that the power spectrum sensing problem can be reduced to linear programming, and that adequate power spectrum sensing is possible from few bits, even for dense spectra. The tradeoffs that emerge in the selection of key parameters, such as filter length and power threshold, and how these affect spectrum sensing performance and complexity are studied. Also, relevant extensions, such as adaptive sensor polling and how to deal with inconsistent sensor readings, are discussed.

Our problem formulation may be reminiscent of *one-bit compressed sensing* [12]–[14]. In [12], [13], it has been shown that signals can be recovered with good accuracy from compressive sensing measurements quantized to just one bit per measurement. The reconstruction is performed by treating the 1-bit measurements as sign constraints, and further constraining the sparse signal on the unit sphere, such that it is recovered within a scaling factor (unavoidable, since 1-bit quantization eliminates all scaling information). The unit-sphere constraint is replaced by an  $\ell_1$ -norm equality constraint in [14] to obtain a linear programming formulation. The main differences between our work and the one-bit compressed sensing framework can be summarized as follows:

- We operate on the autocorrelation vector, instead of the signal per se, and for this reason we exploit positivity constraints that are not present in the one-bit compressed sensing framework.
- Our choice of (positive) thresholds mitigates the scaling problem, so we do not use a unit sphere constraint as in [12], [13], or the  $\ell_1$ -norm constraint as in [14].
- We do not need to assume sparsity of the unknown vector, and our method works even with few measurements due to the strong positivity constraints that we exploit.

It is also worth mentioning that 1-bit measurements were used to perform localization in a sensor network in [15].

The rest of the paper is organized as follows. Some preliminaries are presented in Section II. The proposed *frugal sensing* scheme is developed in Section III, followed by simulations and a discussion of the various design trade-offs in Section IV. Relevant extensions and variations are presented in Section V. Technical derivations and proofs are deferred to the Appendices. Conclusions are drawn in Section VI.

## II. PRELIMINARIES

Consider a discrete-time wide-sense stationary (WSS) signal  $x(n)$ , and let  $r(\ell) := \mathbb{E}[x(n)x^*(n - \ell)]$  denote its autocorrelation sequence, where  $r(\ell) = r^*(-\ell)$ ,  $\forall \ell$ , and  $r(0)$  is nonnegative, by definition. The power spectrum of  $x(n)$  is the discrete-time Fourier transform (DTFT) of  $r(\ell)$ ,  $S_x(\omega) = \sum_{\ell=-\infty}^{\infty} r(\ell)e^{-j\omega\ell}$ , where  $S_x(\omega)$  is real and nonnegative. If only a finite  $K$ -lag autocorrelation sequence is available, represented by the vector  $\mathbf{r}_x = [r(1 - K), \dots, r(-1), r(0), r(1), \dots, r(K - 1)]^T$ , then a windowed estimate of the power spectrum can be obtained as  $\hat{S}_x(\omega) = \sum_{\ell=-K+1}^{K-1} r(\ell)e^{-j\omega\ell}$ . Due to truncation to a finite number of lags, however, such an estimate is not guaranteed to be nonnegative at all frequencies. If we discretize the frequency axis, then an  $N_F$ -point estimate of the power spectrum can be obtained as  $\hat{\mathbf{s}}_x = \mathbf{F}\mathbf{r}_x$ , with  $\hat{s}_x(f) = \hat{S}_x\left(\frac{2\pi f}{N_F}\right)$ , for  $f = 0, \dots, N_F - 1$ , using the  $N_F \times (2K - 1)$  (phase-shifted) discrete Fourier transform (DFT) matrix:

$$\mathbf{F} = \begin{bmatrix} 1 & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}(K-1)} \\ e^{-j\frac{2\pi}{N_F}(1-K)} & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}2(K-1)} \\ e^{-j\frac{2\pi}{N_F}2(1-K)} & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}2(K-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e^{-j\frac{2\pi}{N_F}(N_F-1)(1-K)} & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}(N_F-1)(K-1)} \end{bmatrix}.$$

Define the  $K \times K$  Toeplitz-Hermitian autocorrelation matrix

$$\mathbf{R}_x = \text{Toeplitz}(\mathbf{r}_x) := \begin{bmatrix} r(0) & r(-1) & \dots & r(1-K) \\ r(1) & r(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r(-1) \\ r(K-1) & \dots & r(1) & r(0) \end{bmatrix}. \quad (1)$$

The construction of  $\mathbf{R}_x$  from  $\mathbf{r}_x$  can be explicitly parameterized as follows. Let  $\mathbf{E}_\ell$  denote the  $K \times K$  matrix with ones on the  $\ell$ -th lower diagonal and zeros elsewhere,  $\ell \in \{1, \dots, K - 1\}$ . Define the  $(K - 1) \times 1$  vectors  $\mathbf{r}_{\Re} := \text{Re}\{\mathbf{r}_x(K + 1 : 2K - 1)\}$  and  $\mathbf{r}_{\Im} := \text{Im}\{\mathbf{r}_x(K + 1 : 2K - 1)\}$ , such that  $r_{\Re}(\ell) + jr_{\Im}(\ell) = r(\ell)$ , for  $\ell \in \{1, \dots, K - 1\}$ , where  $(K + 1 : 2K - 1)$  denotes the range of indices from  $K + 1$  to  $2K - 1$ , and  $\text{Re}\{\cdot\}$ ,  $\text{Im}\{\cdot\}$  denote real and imaginary parts, respectively. Then

$$\begin{aligned} \mathbf{R}_x &= r(0)\mathbf{I} + \sum_{\ell=1}^{K-1} [(r_{\Re}(\ell) + jr_{\Im}(\ell))\mathbf{E}_\ell + (r_{\Re}(\ell) - jr_{\Im}(\ell))\mathbf{E}_\ell^T] \\ &= r(0)\mathbf{I} + \sum_{\ell=1}^{K-1} [r_{\Re}(\ell)\bar{\mathbf{E}}_\ell + r_{\Im}(\ell)\tilde{\mathbf{E}}_\ell] \end{aligned} \quad (2)$$

where  $\bar{\mathbf{E}}_\ell := \mathbf{E}_\ell + \mathbf{E}_\ell^T$  and  $\tilde{\mathbf{E}}_\ell := j(\mathbf{E}_\ell - \mathbf{E}_\ell^T)$ .

### III. POWER SPECTRUM SENSING FROM FEW BITS

Consider  $M$  scattered sensors measuring the ambient signal power and reporting to a fusion center—the measurement and reporting mechanisms will be specified shortly. We begin by assuming that all sensors sense a common signal, up to a sensor-specific constant modeling path loss and frequency-flat shadowing and fading, and that each sensor samples the signal at Nyquist rate. Both these assumptions will be lifted in the sequel, but they simplify exposition at this point. In Appendix A, it is shown that frequency-selective fading can be mitigated by averaging the measurements over a long period of time, and that the basic approach carries over without further modification. The Nyquist sampling requirement can be lifted by using an equivalent analog processing and integration chain—the details can be found in Appendix B, see also [7]. Note that we do not assume that the sensors are synchronized; sensing time offsets and phase shifts are allowed.

#### A. Sensor Measurement Chain

First, each sensor  $m \in \{1, \dots, M\}$  uses automatic gain control (AGC) to adjust the scaling of its received signal  $y_m(t) = \gamma_m x(t)$  to a common reference, where  $\gamma_m$  models the associated sensor-specific loss. Note that the power spectrum is invariant with respect to timing offset and phase shift, hence we may assume without loss of generality that every sensor processes the same signal,  $x(t)$ , after the AGC stage. Then,  $x(t)$  is sampled using an analog-to-digital converter operating at Nyquist rate, yielding the WSS sequence  $x(n)$ . Sensor  $m \in \{1, \dots, M\}$  then passes  $x(n)$  through a wideband FIR filter with impulse response  $g_m(n)$  of length  $K$  (i.e.,  $g_m(n) = 0$  for  $n < 0$  and  $n > K - 1$ ). In order to monitor a wide swath of spectrum with relatively few sensors, it is necessary to use broadband filters  $\{g_m(n)\}_{m=1}^M$ , which should somehow provide, loosely speaking, independent yet complementary views of the underlying power spectrum. We propose to use random complex pseudo-noise (PN) impulse responses, i.e.,  $g_m(n)$  is generated using a PN linear shift register, whose initial seed is unique for each sensor (e.g., its serial number) and known to the fusion center. This approach is simple, works well (as shown in the next section), and requires no coordination between sensors: A sensor may fail when its battery runs out, or new sensors may be added without re-programming the other ones. Using random PN filters can also be motivated from a random projections viewpoint, as for the compression matrix applied to sparse signals [2].

The filter's output sequence  $z_m(n)$  is the convolution of the signal  $x(n)$  with the impulse response  $g_m(n)$ , expressed as  $z_m(n) = \sum_{k=0}^{K-1} g_m(k)x(n-k)$ . Let  $\alpha_m := E[|z_m(n)|^2]$  denote the average power of the WSS signal  $z_m(n)$ . Each sensor estimates  $\alpha_m$  using a sample average:

$$\hat{\alpha}_m^{(N)} = \frac{1}{N} \sum_{n=0}^{N-1} |z_m(n)|^2$$

with  $\lim_{N \rightarrow \infty} \hat{\alpha}_m^{(N)} = \alpha_m$  under appropriate mixing conditions [16, p. 171]. Finally, each sensor compares the estimated  $\hat{\alpha}_m^{(N)}$  to a threshold (or set of thresholds). The simplest setup is to use a single threshold  $t$  and binary  $\{0, 1\}$  signaling. If  $\hat{\alpha}_m^{(N)} \geq t$ ,

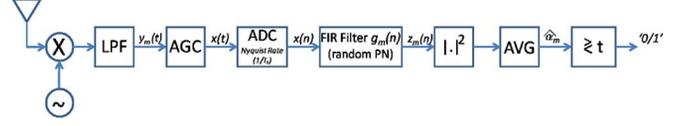


Fig. 1. Sensor measurement chain.

then sensor  $m$  sends ‘1’ to the fusion center, otherwise it sends ‘0’. This sensor measurement chain is shown in Fig. 1.

#### B. Fusion Center

Define the sets  $\mathcal{M}_a := \{m : \hat{\alpha}_m \geq t\}$  and  $\mathcal{M}_b := \{m : \hat{\alpha}_m < t\}$ , with  $M_a := |\mathcal{M}_a|$  and  $M_b := |\mathcal{M}_b|$  such that  $M_a + M_b = M$ . The superscript in  $\hat{\alpha}_m^{(N)}$  is dropped for brevity. Also, define the  $K \times 1$  vector  $\mathbf{g}_m := [g_m^*(K-1), g_m^*(K-2), \dots, g_m^*(0)]^T$  (conjugate reversal of  $g_m(n)$ ), and the  $K \times 1$  vector  $\mathbf{x}_{K,n} := [x(n), x(n+1), \dots, x(n+K-1)]^T$ . It can then be verified that the Toeplitz-Hermitian matrix  $\mathbf{R}_x$  defined in (1) is the autocorrelation matrix of  $\mathbf{x}_{K,n}$ , i.e.,  $\mathbf{R}_x = E[\mathbf{x}_{K,n} \mathbf{x}_{K,n}^H] \succeq 0$  (positive semi-definite), and that  $E[|z_m(n)|^2] = E[|\mathbf{x}_{K,n}^H \mathbf{g}_m|^2]$ . Hence  $\alpha_m = E[|\mathbf{x}_{K,n}^H \mathbf{g}_m|^2] = \mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m$ . It follows that, upon receipt of a ‘1’ (or ‘0’) from sensor  $m$ , the fusion center learns that  $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m \geq t$  (resp.  $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m < t$ ), assuming sufficient averaging such that sample averages converge to ensemble averages. Note that since we only need to ensure that the inequality is not reversed, sample averaging requirements are considerably relaxed relative to high-rate quantization.

The job of the fusion center is to estimate the ambient power spectrum based on the information it received from the sensors, represented by the partition  $\{\mathcal{M}_a, \mathcal{M}_b\}$ . This can be accomplished by reconstructing the  $K$ -lag autocorrelation function  $\hat{\mathbf{r}}_x$ , and then applying the DFT:  $\hat{\mathbf{s}}_x = \mathbf{F} \hat{\mathbf{r}}_x$ . Due to the truncation of the autocorrelation to  $K$  lags (as well as inaccurate estimation of  $\hat{\mathbf{r}}_x$ ), the corresponding  $\hat{\mathbf{s}}_x$  is no longer guaranteed to be nonnegative. In classical spectral analysis, non-negativity of the spectral estimate can be ensured by positive extension of the truncated correlation sequence [17]. There are infinitely many extensions that give rise to positive spectra, a popular one being Burg’s Maximum Entropy extension—this is a well-studied subject in spectral analysis.

Unlike classical spectral analysis, the data here is in the form of linear inequalities involving the autocorrelation matrix. The setup is more heavily under-determined, and we need to employ all available structural properties and prior information to obtain a meaningful estimate of the power spectrum. Towards this end, we propose including both  $\mathbf{R}_x = \text{Toeplitz}(\mathbf{r}_x) \succeq 0$  and  $\mathbf{F} \mathbf{r}_x \geq 0$  as explicit constraints in an optimization-based formulation.

The remaining issue is to find an appropriate cost function. A reasonable choice is to minimize the total signal power, i.e.,  $r(0) = E[|x(n)|^2]$ , consistent with the premise of cognitive

<sup>1</sup>Nothing at all, when *censoring* is adopted. Censoring blends well with random access ‘uplink’ communication from the sensors to the fusion center, because it reduces contention. When fixed multiplexing (such as time/frequency- or code-division multiple-access) is used for sensor to fusion center communication, it is appealing to use ternary  $\{0, \pm 1\}$  signaling, corresponding to two power thresholds  $t_1$  and  $t_2$ , where  $t_2 > t_1$ . If  $\hat{\alpha}_m^{(N)} \geq t_2$ , then sensor  $m$  sends ‘−1’ to the fusion center, else if  $t_1 \leq \hat{\alpha}_m^{(N)} < t_2$  it sends ‘1’, else it sends ‘0’. We focus on binary signaling for simplicity and clarity of exposition.

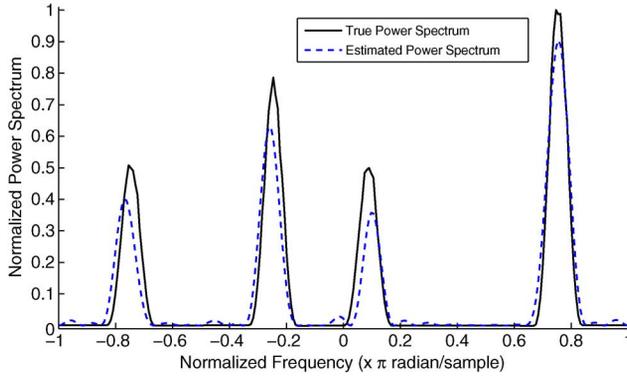


Fig. 2. Illustrative example for the proposed *frugal sensing* approach with sparse spectrum.

radio that most of the spectrum is unused in most places, most of the time. Interestingly, since we enforce  $\mathbf{s}_x = \mathbf{F}\mathbf{r}_x \geq 0$ , and since  $r(0) = \sum_{f=0}^{N_F-1} s_x(f)$ , it follows that  $r(0) = \|\mathbf{s}_x\|_1$ , i.e., minimizing the total signal power implicitly encourages sparsity in the reconstructed power spectrum. Putting everything together leads to the following problem formulation:

$$\begin{aligned}
 & \hat{\mathbf{r}}_x = \arg \min_{\mathbf{r}_x} r(0) \\
 \text{s.t. : } & \mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m \geq t, m \in \mathcal{M}_a, \\
 & \mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m < t, m \in \mathcal{M}_b, \\
 & \mathbf{r}_x := [r(1-K), \dots, r(-1), r(0), r(1), \dots, r(K-1)]^T, \\
 & r(\ell) = r^*(-\ell), \ell = 1, \dots, K-1, \\
 & \mathbf{R}_x = \text{Toeplitz}(\mathbf{r}_x), \mathbf{R}_x \succeq 0, \mathbf{F}\mathbf{r}_x \geq 0
 \end{aligned} \tag{3}$$

Note that the constraint  $\mathbf{R}_x = \text{Toeplitz}(\mathbf{r}_x)$  is a linear relation between  $\mathbf{R}_x$  and  $\mathbf{r}_x$  as expressed in (2). This implies that all the constraints in (3) are ordinary linear inequalities in the variables  $r(0) \in \mathbb{R}$  and  $\{r_{\Re}(\ell) \in \mathbb{R}, r_{\Im}(\ell) \in \mathbb{R}\}_{\ell=1}^{K-1}$ , except for the constraint  $\mathbf{R}_x \succeq 0$ , which is a linear matrix inequality (LMI). Hence, problem (3) is a semidefinite program (SDP) that can be optimally solved using efficient interior point methods. The following proposition, however, asserts that the constraint  $\mathbf{R}_x \succeq 0$  is redundant; it is in fact implied by the constraint  $\mathbf{F}\mathbf{r}_x \geq 0$ .

*Proposition 1:* For  $N_F \geq 2K-1$ ,  $\mathbf{F}\mathbf{r}_x \geq 0 \Rightarrow \mathbf{R}_x := \text{Toeplitz}(\mathbf{r}_x) \succeq 0$ . The converse is generally not true.

The proof can be found in Appendix C. Proposition 1 implies that problem (3) is not affected by removing the constraint  $\mathbf{R}_x \succeq 0$ . Thus, (3) can be expressed as the following linear program (LP):

$$\begin{aligned}
 & \min_{r(0) \in \mathbb{R}, \{r_{\Re}(\ell) \in \mathbb{R}, r_{\Im}(\ell) \in \mathbb{R}\}_{\ell=1}^{K-1}} r(0) \\
 \text{s.t. : } & \mathbf{g}_m^H \left[ r(0)\mathbf{I} + \sum_{\ell=1}^{K-1} (r_{\Re}(\ell)\tilde{\mathbf{E}}_{\ell} + r_{\Im}(\ell)\tilde{\mathbf{E}}_{\ell}) \right] \mathbf{g}_m \geq t, m \in \mathcal{M}_a, \\
 & \mathbf{g}_m^H \left[ r(0)\mathbf{I} + \sum_{\ell=1}^{K-1} (r_{\Re}(\ell)\tilde{\mathbf{E}}_{\ell} + r_{\Im}(\ell)\tilde{\mathbf{E}}_{\ell}) \right] \mathbf{g}_m < t, m \in \mathcal{M}_b, \\
 & r(\ell) = r^*(-\ell) = r_{\Re}(\ell) + jr_{\Im}(\ell), \quad \ell = 1, \dots, K-1 \\
 & \mathbf{r}_x := [r(1-K), \dots, r(-1), r(0), r(1), \dots, r(K-1)]^T, \\
 & \mathbf{F}\mathbf{r}_x \geq 0.
 \end{aligned} \tag{4}$$

The significance of this reduction from an SDP to an LP is that the latter is easier to solve using specialized algorithms. The LP problem (4) can be expressed in the standard form as follows. Define the two  $(2K-1) \times 1$  vectors:

$$\begin{aligned}
 \mathbf{d}_m := & \left[ g_m^*(0)g_m(K-1), \dots, \sum_{n=0}^{K-2} g_m^*(n)g_m(n+1), \right. \\
 & \left. \sum_{n=0}^{K-1} |g_m(n)|^2, \sum_{n=0}^{K-2} g_m(n)g_m^*(n+1), \dots, \right. \\
 & \left. g_m(0)g_m^*(K-1) \right]^T,
 \end{aligned}$$

$$\begin{aligned}
 \tilde{\mathbf{r}}_x := & \left[ r(0), \text{Re}\{r(1)\}, \dots, \text{Re}\{r(K-1)\}, \right. \\
 & \left. \text{Im}\{r(1)\}, \dots, \text{Im}\{r(K-1)\} \right]^T,
 \end{aligned}$$

where  $\mathbf{r}_x$  can be obtained from  $\tilde{\mathbf{r}}_x$  using a  $(2K-1) \times (2K-1)$  transformation matrix  $\mathbf{Q} : \mathbf{r}_x = \mathbf{Q}\tilde{\mathbf{r}}_x$ . For example, for  $K=3$ , the transformation matrix is:

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 1 & 0 & -j \\ 0 & 1 & 0 & -j & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & j & 0 \\ 0 & 0 & 1 & 0 & j \end{bmatrix}$$

Hence, it is easy to verify that  $\alpha_m = \mathbf{d}_m^T \mathbf{r}_x = \mathbf{d}_m^T \mathbf{Q}\tilde{\mathbf{r}}_x = \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x$ , where  $\tilde{\mathbf{d}}_m := \mathbf{Q}^T \mathbf{d}_m$ . Finally, defining  $\tilde{\mathbf{F}} := \mathbf{F}\mathbf{Q}$  and  $\mathbf{c} := [1, 0, \dots, 0]^T$ , problem (4) can be formulated in the standard LP form:

$$\begin{aligned}
 & \min_{\tilde{\mathbf{r}}_x \in \mathbb{R}^{(2K-1)}} \mathbf{c}^T \tilde{\mathbf{r}}_x \\
 \text{s.t. : } & \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x \geq t, \quad m \in \mathcal{M}_a, \\
 & \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x < t, \quad m \in \mathcal{M}_b, \\
 & \tilde{\mathbf{F}}\tilde{\mathbf{r}}_x \geq 0.
 \end{aligned} \tag{5}$$

#### IV. SIMULATIONS AND PARAMETER TUNING

In this section, we provide simulation results and discuss the effect of some design parameters on the quality of the power spectrum estimate. We begin with a simulation that illustrates what one can expect from the proposed approach. In Fig. 2 and Fig. 3, a scenario with  $M=100$  sensors was considered, and the estimated power spectrum (dashed line) has been obtained by solving the LP (5). For Fig. 2 the true power spectrum is sparse (solid line), filter length  $K=24$  was used, and the threshold  $t$  was set such that  $M_a=30$ ; whereas for Fig. 3 the true power spectrum is dense, filter length  $K=10$  was used, and  $t$  was set such that  $M_a=50$ . The plotted spectra have been normalized by the peak value of the true power spectrum. The quality of the estimates in Figs. 2, 3 is very satisfactory considering that only 100 bits have been used as input data—corresponding roughly to three single precision IEEE floats, or about what it would take to transmit three accurate power measurements, or  $r(0)$  and  $r(1)$  (note that  $r(1)$  is complex, requiring two floats).

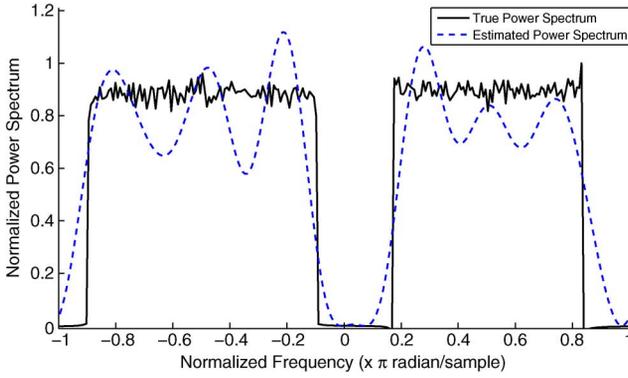


Fig. 3. Illustrative example for the proposed *frugal sensing* approach with dense spectrum.

In the rest of the paper, we use the normalized mean square error (NMSE) to measure the quality of the power spectrum estimate. The NMSE is defined as

$$\text{NMSE} := \mathbb{E} \left[ \frac{\|\mathbf{s}_x - \hat{\mathbf{s}}_x\|^2}{\|\mathbf{s}_x\|^2} \right] \quad (6)$$

where the expectation is taken with respect to the random signal and the random impulse responses of the FIR filters, obtained via Monte-Carlo simulations. Note that using  $\mathbb{E}[\|\mathbf{s}_x - \hat{\mathbf{s}}_x\|^2]/\mathbb{E}[\|\mathbf{s}_x\|^2]$  instead of (6) to define NMSE made very little difference in our experiments—the results were almost identical.

#### A. Threshold Selection

In this subsection, we show that, from an estimation performance point of view, the threshold  $t$  should be selected according to the sparsity level of the power spectrum (assuming prior sparsity knowledge is available). Let  $\eta$  denote the *sparsity ratio*, defined as the ratio of the nonzero<sup>2</sup> entries to the total length of the power spectrum, and define  $\nu$  as the ratio of the number of sensors with measurements above  $t$  to the total number of reporting sensors (i.e.,  $\nu := M_a/M$ ).

In Fig. 4, we plot the NMSE versus the ratio  $\nu$ , for signals with different sparsity ratios  $\eta$ . The sparse signal was fixed for each  $\eta$ , and 1000 Monte-Carlo simulations for each  $\nu$  were used to obtain the corresponding NMSE (here the expectation was taken with respect to the random FIR filters only). The setup included  $M = 60$  sensors and the filter length was set to  $K = 8$ . Two main points can be deduced from Fig. 4. First, we see that as the sparsity ratio  $\eta$  increases, the NMSE is minimized at a higher ratio  $\nu$ . This means that the threshold  $t$  should be tuned such that number of sensors reporting measurements above  $t$  decreases as the power spectrum becomes more sparse. Historical data can be used to get an expectation for  $\eta$ , and to identify the distribution of  $\alpha_m$ . Exploiting such prior statistical information, the threshold  $t$  can be selected such that  $\nu$  minimizes the NMSE for the corresponding  $\eta$ . The second point that can be drawn from Fig. 4 is that the minimum NMSE increases as the power spectrum becomes less sparse. This implies that the quality of the estimated power spectrum using the proposed approach is relatively better for sparser signals. It is worth

<sup>2</sup>Or above a small quantity  $\epsilon$ .

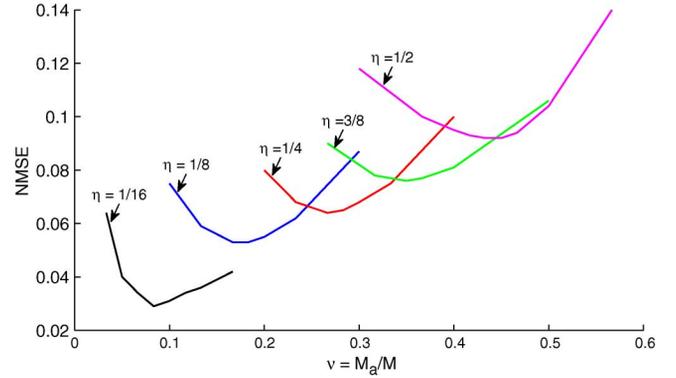


Fig. 4. The optimum  $\nu$  that yields the minimum NMSE depending on  $\eta$  of the signal being estimated.

mentioning that an adaptive threshold selection algorithm for the one-bit compressed sensing framework has been introduced in [18], assuming a signal with a separable distribution that is known *a priori*.

#### B. Filter Type and Length $K$

Next, we look at how the filter length  $K$  affects the quality of the power spectrum estimate, and also discuss two candidate classes of random filters. Note that the number of filter taps  $K$  is also the number of estimated autocorrelation lags. Truncation of the autocorrelation sequence smears the estimated power spectrum [17], and the smaller  $K$  is, the more pronounced this smearing will be. This is the reason why  $K = 24$  has been used in Fig. 2, where the spectrum is a sparse superposition of narrowband spectra, whereas  $K = 10$  has been used in Fig. 3 which features two main lobes occupying more than half the bandwidth. On the other hand,  $K$  is also the number of unknowns, and the larger  $K$  is, relative to the number of inequality constraints in (5), the more under-determined the problem becomes, which counteracts the reduced smearing. The choice of  $K$  thus determines the trade-off between smearing and inequalities-versus-unknowns considerations. In addition, the complexity of solving (5) is roughly  $O(K^{3.5})$ , which is another reason why  $K$  should be kept moderate.

Fig. 5 illustrates this tradeoff, showing the NMSE as a function of  $K$  for various  $M$ . In Fig. 5, two types of random impulse responses were used for the filters: (a) complex binary antipodal ( $\pm 1 \pm j$ )-valued random PN, and (b) normalized white complex Gaussian random variables. Random sparse signals with  $\eta = 0.25$  were generated and the reported NMSE for each  $K$  is the result of averaging across more than 1000 Monte-Carlo simulations (with respect to the random signals and filters). Three scenarios were considered with  $M = 50, 100$  and  $200$  sensors, where  $t$  was selected such that  $M_a = 12, 25$  and  $50$ , respectively<sup>3</sup>. Fig. 5 confirms our intuition about the trade-off in the choice of  $K$ . Fig. 5 also shows that the optimal  $K$  is an increasing function of  $M$ , which can be understood by noting that as  $M$  increases, the number of inequalities increases, hence

<sup>3</sup>The results in the figure were obtained by varying the threshold with each simulation run to sustain the required  $M_a$  in each run. Very similar results were obtained when the threshold was fixed across all simulation runs, which was selected as the average of the different thresholds that sustain the required  $M_a$  in each run.

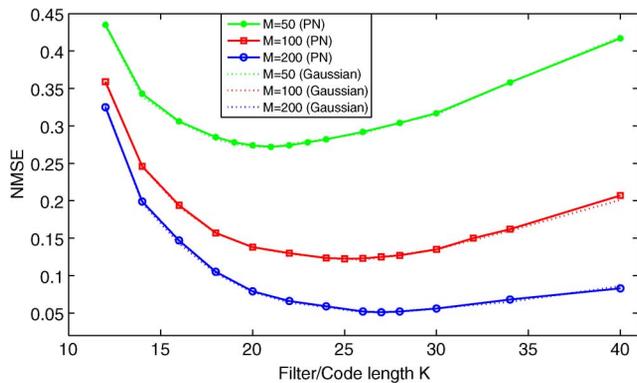


Fig. 5. Tradeoff between the NMSE and the filter length  $K$ .

one can afford more unknowns. Another point worth noting is that the performance of Gaussian filters (dotted lines) is almost identical to that of binary PN filters. However, binary PN filters are much simpler to implement via cheap linear shift registers, hence preferable to Gaussian filters.

## V. RELEVANT EXTENSIONS

In this section we discuss some extensions and variations to the proposed frugal sensing scheme.

### A. Another Reconstruction Method

So far, we have considered minimizing the total signal power as our objective function in (5), which implicitly encourages sparsity in the reconstructed power spectrum. In this subsection, we consider a different formulation of the reconstruction problem. First, note that the feasible region:

$$C := \{\tilde{\mathbf{r}}_x | \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x \geq t, m \in \mathcal{M}_a, \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x < t, m \in \mathcal{M}_b, \tilde{\mathbf{F}} \tilde{\mathbf{r}}_x \geq 0\}$$

is a convex polyhedron, whose volume is a measure of the uncertainty in  $\tilde{\mathbf{r}}_x$  associated with the constraint set  $C$ ; however, finding the volume of a convex polyhedron is NP-hard [19]. The optimal solution of the LP (5) will always be on the boundary of  $C$ —in fact, without loss of optimality, can be taken to be a vertex of  $C$ . Thus the boundary of  $C$  is associated with sparse feasible spectra. If the sought spectrum is known to be non-sparse, then it makes sense to steer away from the boundary of  $C$ , and a good way to enforce this is to use the “center” of  $C$  to estimate  $\tilde{\mathbf{r}}_x$ . There are different ways to define the center of  $C$ , and we use the center of the maximum volume inscribed ellipsoid.

Define:

$$\begin{cases} \mathbf{a}_m = -\tilde{\mathbf{g}}_m, b_m = -t & \text{if } m \in \mathcal{M}_a \\ \mathbf{a}_m = \tilde{\mathbf{g}}_m, b_m = t & \text{if } m \in \mathcal{M}_b \end{cases}$$

and let the vector  $\mathbf{v}_n$  correspond to the negative of the  $n$ -th row of  $\tilde{\mathbf{F}}$ , where  $n \in \{1, \dots, N_F\}$  (i.e.,  $\mathbf{F} \mathbf{r}_x \geq 0 \Leftrightarrow \mathbf{v}_n^T \tilde{\mathbf{r}}_x \leq 0, \forall n$ ). Finding the ellipsoid  $\varepsilon := \{\mathbf{B} \mathbf{u} + \mathbf{p} \mid \|\mathbf{u}\|_2 \leq 1\}$  of maximum volume that lies inside the convex polyhedron  $C$  can be used to lower bound the actual volume of  $C$ . This can be

expressed as a convex optimization problem in the variables  $\mathbf{B}$  and  $\mathbf{p}$  [20, Sec. 8.4.2]:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{p}} \quad & \log \det \mathbf{B}^{-1} \\ \text{s.t. :} \quad & \|\mathbf{B} \mathbf{a}_m\|_2 + \mathbf{a}_m^T \mathbf{p} \leq b_m, \quad m = 1, \dots, M \\ & \|\mathbf{B} \mathbf{v}_n\|_2 + \mathbf{v}_n^T \mathbf{p} \leq 0, \quad n = 1, \dots, N_F. \end{aligned} \quad (7)$$

The volume of the ellipsoid  $\varepsilon$  is proportional to  $\det \mathbf{B}$ , and  $\mathbf{p}$  is the center of  $\varepsilon$  [20, Sec. 8.5.2]. Now, instead of minimizing the total signal power  $r(0) = \|\mathbf{s}_x\|_1$  as in (5), we propose setting the estimate of  $\tilde{\mathbf{r}}_x$  to  $\mathbf{p}$ , i.e., the estimated autocorrelation  $\tilde{\mathbf{r}}_x$  is the center of the maximal inscribable ellipsoid.

Clearly, this approach does not promote sparsity, however it can yield better estimates, as compared to (5), when the spectrum is non-sparse. This was numerically verified for the following setup. The setup included  $M = 60$  sensors, the filter length was set to  $K = 6$ , and the threshold was selected such that  $M_a = 30$ . A non-sparse spectrum was randomly generated, and the NMSE was obtained using 500 Monte-Carlo simulations. Using the LP reconstruction method (5), the NMSE was found to be 0.2544, whereas using (7) gave an NMSE of 0.2228, showing a slight advantage for (7) over (5). The real reason for introducing the ellipsoid approximation though is discussed in the next subsection.

### B. Sensor Polling—Adaptive Sensing

So far, we have assumed that sensors are active and the fusion center is passive; each sensor sends a bit based on its own measurement, while the fusion center collects the sensor reports and estimates  $\mathbf{r}_x$ . A more intelligent strategy is to allow the fusion center to selectively poll sensors on the basis of previously received sensor reports. The idea here is that, given partial information about the sought spectrum, certain sensors are more valuable than others. Polling also makes sense from an energy conservation point of view for battery-operated sensors, which can be put to sleep until polled by the fusion center. Thus, the question we are addressing here is:

Assuming that the fusion center has already obtained measurements from  $M$  sensors, which are the best sensors to poll next among the  $J$  remaining ones, and in what order?

We propose the following *greedy* approach. Since finding the exact volume of the feasible region  $C$  is NP-hard [19], we use the volume of the maximal inscribable ellipsoid, which is obtained by solving (7), as an uncertainty measure for the estimated power spectrum. The volume of this ellipsoid  $V$  is proportional to  $\det \mathbf{B}$ , i.e.,  $V = c \det \mathbf{B}$ , where  $c$  is a constant. Polling sensor  $j$  will result in either adding  $\tilde{\mathbf{d}}_j^T \tilde{\mathbf{r}}_x \geq t$  or  $\tilde{\mathbf{d}}_j^T \tilde{\mathbf{r}}_x < t$  to the set of constraints. Let  $V_j'$  denote the new volume of the maximal inscribable ellipsoid corresponding to the addition of the first inequality, and  $V_j''$  the volume corresponding to the addition of the second inequality. The proposed approach is to poll the sensor  $j$  that yields the minimum worst-case volume after its corresponding inequality is included in the constraint set, i.e.,  $j^* = \arg \min_{j \in \{1, \dots, J\}} \max(V_j', V_j'')$ , where  $j^*$  is the selected sensor. This approach requires that

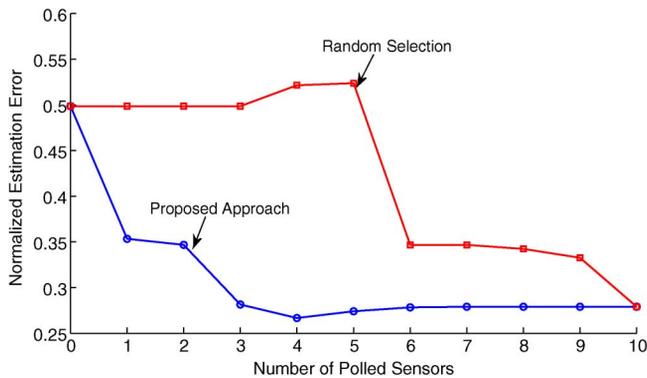


Fig. 6. The decrease of the estimation error as more sensors are polled, for a typical scenario.

the fusion center searches through all remaining non-polled  $J$  sensors and solves  $2J$  problems of type (7) before deciding on which sensor to poll at each step. This can be a heavy computational burden, but note that for modest sensor populations all required computations can be performed once off-line, and the results stored for on-line use.

In Fig. 6, we illustrate the performance of the proposed sensor polling scheme as compared to randomly selecting any sensor, for a typical scenario. A dense power spectrum is considered, and a short filter length  $K = 6$  is used. It is assumed that the fusion center has already received the 1-bit measurements from  $M = 10$  sensors, and  $J = 10$  sensors remain to be polled. The normalized error in the power spectrum estimate ( $= \|\mathbf{s}_x - \hat{\mathbf{s}}_x\|^2 / \|\mathbf{s}_x\|^2$ ), as each of the remaining  $J$  sensors is polled by the fusion center, is plotted in the figure. The figure shows that using the proposed sensor polling scheme, the error significantly decreases after polling each of the first 3 sensors due to the *good* choice of sensors to be polled; whereas randomly selecting the sensor to poll does not give the same performance. Note that both curves meet at the end when all  $J = 10$  sensors are polled, as expected. Also note that polling some sensors may have no effect on the feasible region, and consequently no effect on the estimated power spectrum. That is why the error does not change for the proposed scheme when polling each of the last 5 sensors, as shown in the figure.

In Fig. 7, we report the average performance considering a similar setup as in Fig. 6, but with  $J = 15$ . A total of 5 sensors are polled in each run, and we plot the NMSE, obtained using 20 Monte-Carlo simulations, when each one of them is polled using the proposed sensor polling scheme and with random sensor selection. The figure shows the better performance of the proposed scheme due to the *good* choice of sensors to be polled.

### C. Higher-Resolution Quantization

It is clear that finer-grained quantization of  $\hat{\alpha}_m$  will improve the quality of the power spectrum estimate, but at the cost of higher signaling rate and sensor hardware complexity. Using multi-bit quantization should be considered *vis-a-vis* the alternative of employing more single-bit sensors while holding hardware, energy, and signaling costs fixed. Another factor that must be taken into account in deciding the right number of quantization levels is that coarse quantization is naturally more robust

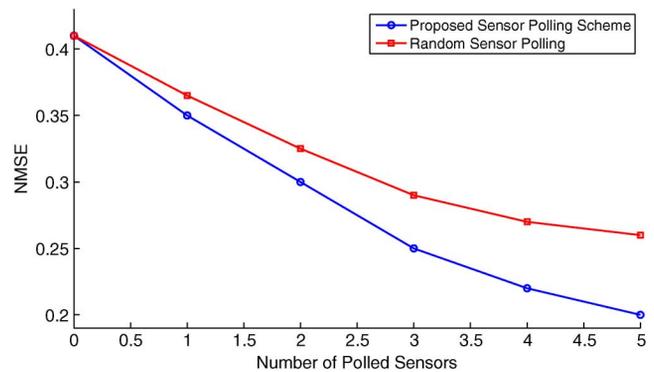


Fig. 7. The decrease of the NMSE as more sensors are polled.

to sample averaging errors in estimating output power. In the limit, if the analog  $\{\hat{\alpha}_m\}_{m=1}^M$  are communicated to the fusion center (e.g., using analog modulation), the power spectrum can be estimated by solving the following weighted least squares minimization

$$\begin{aligned} \min_{\tilde{\mathbf{r}}_x \in \mathbb{R}^{(2K-1)}} \quad & \mathbf{c}^T \tilde{\mathbf{r}}_x + \sum_{m=1}^M \mu_m \left\| \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x - \hat{\alpha}_m \right\|^2 \\ \text{s.t.} : \quad & \tilde{\mathbf{F}} \tilde{\mathbf{r}}_x \geq 0. \end{aligned} \quad (8)$$

where the weights  $\{\mu_m\}$  reflect the relative accuracy of  $\hat{\alpha}_m$  and  $\mu = \sum_{m=1}^M \mu_m$  trades off the data term versus prior information on the total power (and sparsity) of the measured power spectrum.

Here, we consider a fixed bit-budget  $\mathcal{B} = MB$  setup, where  $B$  is the number of quantization bits used to describe the estimated  $\alpha_m$  at each sensor (i.e.,  $2^B$  quantization levels), and compare the performance of the different quantization schemes. We assume that the measurements  $\{\hat{\alpha}_m\}$  are mapped to discrete levels via a uniform quantizer. In Fig. 8, we plot the NMSE as a function of  $B$  for different bit-budgets. Random sparse signals with  $\eta = 0.25$  were generated and the reported NMSE for each point was averaged over more than 1000 Monte-Carlo simulations (with respect to the random signal and filters). The filter length was set to  $K = 24$ . Selecting the threshold for the one-bit quantization problem (5) as the average threshold that yields  $\nu = M_a/M = 0.25$  results in the NMSE point that is connected to the  $B = 2$  point via a dashed line, whereas the NMSE points that correspond to the uniform quantizer are connected via a solid line.

Fig. 8 shows that the NMSE can be significantly decreased by properly selecting the threshold in the one-bit quantization scenario (compared to uniform one-bit quantization). It can also be seen that if the bit-budget  $\mathcal{B}$  is small relative to  $K$ , then it is better to have a larger number of sensors with coarsely-quantized power measurements (i.e., small  $B$ ), whereas for a larger  $\mathcal{B}$  relative to  $K$ , increasing  $B$  gives better performance. More specifically, we can see that the one-bit quantization with the adapted threshold yields the minimum NMSE for  $\mathcal{B} = 24$  and  $\mathcal{B} = 48$ , while it is very close to the minimum NMSE for  $\mathcal{B} = 96$  and  $\mathcal{B} = 144$ . Therefore, considering the implementation and complexity advantages of 1-bit quantizers, these results motivate the usage of 1-bit sensors.

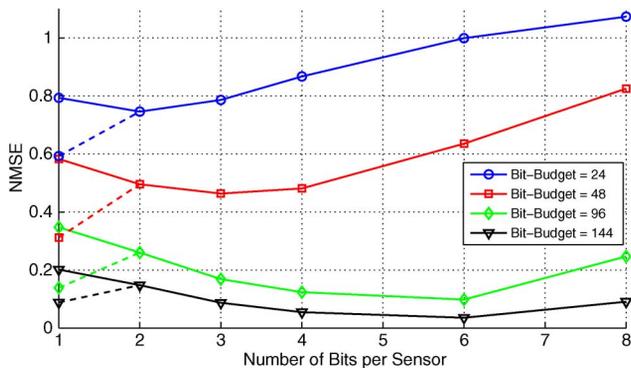


Fig. 8. NMSE for different quantization schemes at different bit-budgets.

It is worth mentioning that a similar trade off in performance between the number of measurements and the number of bits per measurement has been studied in [21] for the compressed sensing setting. In addition to the autocorrelation-specific positivity constraints that are imposed in our formulation as opposed to [21], reference [21] considers the tradeoff in presence of errors due to both signal noise and quantization, whereas we do not consider any errors. Interestingly, [21] also concludes that it is better to acquire as few as 1 bit per measurement in many practical applications.

#### D. Robust Estimation: Inconsistent Sensor Measurements

Due to insufficient sample averaging in the estimation of  $\alpha_m$ , and/or decoding errors in the sensor to fusion center communication links, it is possible that the set of correlation matrices satisfying the constraints in (4) can be empty. In such cases, it makes sense to find  $\mathbf{R}_x$  that is consistent with as many inequalities as possible. This can be formulated as follows. Add a slack variable  $q(m) \geq 0$ , that represents the possible error in the measurement or reporting of  $\alpha_m$ , to the constraints of type  $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m \geq t$  ( $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m \leq t$ ), such that they become  $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m + q(m) \geq t$  (resp.  $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m - q(m) \leq t$ ). Then, add a sparsity-inducing penalty  $\lambda \|\mathbf{q}\|_1 = \sum_{m=1}^M q(m)$  to the cost function, where  $\mathbf{q} := [q(1), \dots, q(M)]^T$ , to promote sparsity among the slack variables, in order to (approximately) minimize the number of inconsistent inequalities. In this way, problem (5) is modified to the following robust LP:

$$\begin{aligned} \min_{\tilde{\mathbf{r}}_x \in \mathbb{R}^{(2K-1)}, \mathbf{q} \in \mathbb{R}^M} \quad & \mathbf{c}^T \tilde{\mathbf{r}}_x + \lambda \mathbf{1}^T \mathbf{q} \\ \text{s.t. :} \quad & \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x + q(m) \geq t, \quad m \in \mathcal{M}_a, \\ & \tilde{\mathbf{d}}_m^T \tilde{\mathbf{r}}_x - q(m) < t, \quad m \in \mathcal{M}_b, \\ & \tilde{\mathbf{F}} \tilde{\mathbf{r}}_x \geq 0, \quad \mathbf{q} \geq 0. \end{aligned} \quad (9)$$

where  $\mathbf{1}$  is the vector of all ones, and  $\lambda \geq 0$  is a tuning parameter that controls the level of sparsity. It is worth mentioning that using the  $\ell_1$ -norm for robust estimation was introduced in [22], see also [23].

In Fig. 9, we consider a similar setup to that used for Fig. 2, assuming a sparse power spectrum (solid line),  $M = 100$ , and  $K = 24$ . The plotted spectra have been normalized by the peak value of the true power spectrum. To model for inconsistencies and errors in the reported measurement bits, an independent uniform random variable is added to each  $\hat{\alpha}_m$ . As a result, the fusion center received 20 *wrong* bits from the sensors

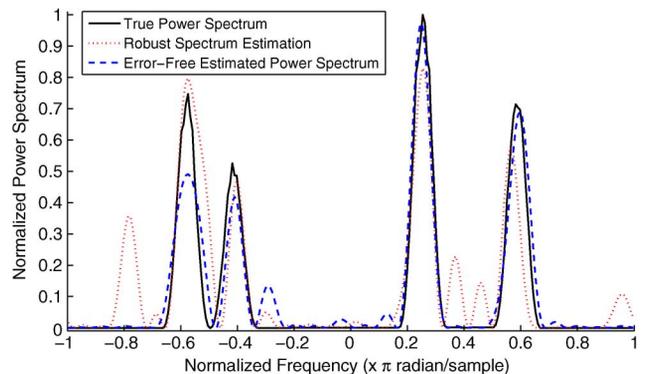


Fig. 9. Example showing the performance of the proposed robust frugal sensing scheme.

(i.e., 20 reversed inequalities); 14 ‘0’ bits are received as ‘1’, and 6 ‘1’ bits are received as ‘0’. This resulted in an infeasible problem (5). The estimated power spectrum that has been obtained by solving the robust LP (9) is plotted as the dotted line, where the tuning parameter  $\lambda$  was set to 1. It is worth noting that the resulting sparse  $\mathbf{q}$  after solving (9) included only 16 nonzero entries (representing the inconsistencies). If the true measurement bits are received by the fusion center such that the inequality constraints are consistent, the estimated power spectrum obtained by solving (9) is given as the dashed line. Note that in this case problem (9) is equivalent to problem (5), since the added sparsity-inducing penalty  $\lambda \mathbf{1}^T \mathbf{q}$  in the objective of (9) gives  $\mathbf{q} = 0$ , for  $\lambda$  sufficiently large. The quality of the power spectrum estimate using the robust LP (9) is very satisfactory, considering that 20% of the received measurement bits were flipped.

## VI. CONCLUSIONS

A network sensing scenario was considered, where scattered low-end sensors pass the received signal through a random filter, measure average power at the output of the filter, and send out a bit or coarsely quantized power level to a fusion center. The fusion center obtains an estimate of the power spectrum by solving an under-determined linear program comprising inequality constraints derived from the sensor data, plus prior information in the form of the cost function and non-negativity constraints. It was shown that adequate power spectrum sensing is possible from relatively few bits, even for dense spectra. The selection of some key design parameters was considered, and important trade-offs were revealed and illustrated in pertinent simulations. It was demonstrated that judicious choice of the filter length is needed to balance smearing effects against inequalities-versus-unknowns considerations, and the detection threshold at the sensors should be tuned such that number of sensors reporting measurements above it decreases as the power spectrum becomes more sparse. Some extensions and variations were also considered, notably an active sensor polling/adaptive sensing scheme that minimizes an estimate of the worst case uncertainty *after* sensor selection. This polling strategy performs considerably better than passive listening or random selection.

The formulation here can be viewed as generalizing classical nonparametric power spectrum estimation to the case where the data is in the form of inequalities, rather than equalities. A key

challenge is that estimation relies on solving appropriate optimization problems, and cannot be put in closed form. This makes performance analysis challenging as of this writing, however we hope to pursue new directions and tackle some of these issues in future work.

## APPENDIX

### A. Fading Considerations

First note that if the discrete signal  $y_m(n)$  is received in presence of frequency-flat fading, then the difference in the received power spectrum across sensors can be compensated for using AGC. Consider now a more general frequency-selective fading scenario. The received signal  $y_m(n)$  is the convolution of the transmitted discrete-time WSS signal  $x(n)$  with the linear (possibly time-varying) finite-impulse response fading channel  $\{h_m(n; \ell)\}_{\ell=0}^{L-1}$ , expressed as  $y_m(n) = \sum_{\ell=0}^{L-1} h_m(n; \ell)x(n - \ell)$ . Assuming that  $x(n)$  is independent of  $\{h_m(n; \ell)\}$ , the received autocorrelation is thus given as

$$\begin{aligned} & \mathbb{E}[y_m(n)y_m^*(n-k)] \\ &= \mathbb{E}\left[\sum_{\ell_1=0}^{L-1} h_m(n; \ell_1)x(n-\ell_1)\sum_{\ell_2=0}^{L-1} h_m^*(n-k; \ell_2)x^*(n-k-\ell_2)\right] \\ &= \sum_{\ell_1=0}^{L-1} \sum_{\ell_2=0}^{L-1} \mathbb{E}[h_m(n; \ell_1)h_m^*(n-k; \ell_2)]r_x(k+\ell_2-\ell_1). \quad (10) \end{aligned}$$

Next, we consider two scenarios for the fading channel.

*Scenario 1:*  $\{h_m(n; \ell)\}$  is random, time-invariant, and the correlation between two filter taps is only a function of the ordinal distance between them. This implies that

$$\begin{aligned} \mathbb{E}[h_m(n; \ell_1)h_m^*(n-k; \ell_2)] &= \mathbb{E}[h_m(\ell_1)h_m^*(\ell_2)] \\ &= r_{h_m}(\ell_1 - \ell_2). \end{aligned}$$

Then, from (10):

$$\begin{aligned} & \mathbb{E}[y_m(n)y_m^*(n-k)] \\ &= \sum_{\ell_1=0}^{L-1} \sum_{\ell_2=0}^{L-1} r_{h_m}(\ell_1 - \ell_2)r_x(k + \ell_2 - \ell_1) \\ &= \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{h_m}(\ell)r_x(k - \ell) \\ &= r_{y_m}(k) \end{aligned}$$

and thus  $y_m(n)$  is WSS, and the received power spectrum is expressed as

$$\begin{aligned} S_{y_m}(\omega) &= \sum_{k=-\infty}^{\infty} r_{y_m}(k)e^{-j\omega k} \\ &= \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{h_m}(\ell)\sum_{k=-\infty}^{\infty} r_x(k-\ell)e^{-j\omega k} \\ &= \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{h_m}(\ell)e^{-j\omega\ell}\sum_{m=-\infty}^{\infty} r_x(m)e^{-j\omega m} \\ &= S_{h_m}(\omega)S_x(\omega) \end{aligned}$$

where  $S_{h_m}(\omega) := \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{h_m}(\ell)e^{-j\omega\ell}$ . Note that since the channel frequency response is given as  $H_m(\omega) = \sum_{\ell=0}^{L-1} h_m(\ell)e^{-j\omega\ell}$ , then

$$\begin{aligned} \mathbb{E}[|H_m(\omega)|^2] &= \sum_{\ell_1=0}^{L-1} \sum_{\ell_2=0}^{L-1} \mathbb{E}[h_m(\ell_1)h_m^*(\ell_2)]e^{-j\omega(\ell_1-\ell_2)} \\ &= \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{h_m}(\ell)e^{-j\omega\ell} \\ &= S_{h_m}(\omega) \end{aligned}$$

Assuming that  $\mathbb{E}[|H_m(\omega)|^2]$  is the same across all sensors, and that sensors acquire sufficient samples with different channel realizations such that the sample average converges to the expectation, then all sensors will be reporting consistent power spectrum measurements. This effectively assumes that the channel remains constant over a relatively long period of time, then jumps to a new realization, dwells there for another measurement epoch, and so on. This is a reasonable model if each sensor only spends a small part of its time to sense the spectrum, while it does other things most of the time. Every time it returns to the spectrum sensing task, it will encounter a new channel realization, not only because of drift but also due to acquiring a new carrier/phase lock. If the reported measurements reflect averaging over many such epochs, then the proposed model is well-motivated.

*Scenario 2:* The *Wide Sense Stationary Uncorrelated Scattering* (WSSUS) channel model [24, Sec. 3.3], first introduced by Bello [25], where  $h_m(n; \ell)$  is WSS with respect to the time variable  $n$  and uncorrelated across the lag variable  $\ell$ . This implies that  $\mathbb{E}[h_m(n; \ell_1)h_m^*(n-k; \ell_2)] = r_{h_m}(k; \ell_1)\delta(\ell_1 - \ell_2)$ . Hence, substituting in (10) yields:

$$\begin{aligned} \mathbb{E}[y_m(n)y_m^*(n-k)] &= \sum_{\ell=0}^{L-1} \mathbb{E}[h_m(n; \ell)h_m^*(n-k; \ell)]r_x(k) \\ &= \phi_m(k)r_x(k) = r_{y_m}(k) \end{aligned}$$

where  $\phi_m(k) := \sum_{\ell=0}^{L-1} \mathbb{E}[h_m(n; \ell)h_m^*(n-k; \ell)]$ . For slowly varying channels,  $h_m(n; \ell) \approx h_m(n-k; \ell)$  for the (small) range of autocorrelation lags considered here, which implies that  $\phi_m(k) \approx \sum_{\ell=0}^{L-1} \mathbb{E}[|h_m(n; \ell)|^2]$  is approximately constant (not a function of  $k$ ). Hence, all sensors will be reporting consistent power spectrum measurements, assuming that sensors acquire sufficient samples such that the sample average converges to the expectation.

### B. Analog Sensor Measurement Chain

Assume that the complex-valued analog signal  $x(t)$  is band limited with two-sided bandwidth  $1/T$  (i.e. Nyquist rate =  $1/T$ ). Let  $\tilde{g}_m(t)$  be the impulse response of the analog filter of duration  $KT$  that corresponds to the FIR filter  $g_m(n)$ , satisfying  $\tilde{g}_m(t) = g_m(n)$  for  $nT < t \leq (n+1)T$ , where  $n = 0, \dots, K-1$ , and  $\tilde{g}_m(t) = 0$  for  $t > KT$  and  $t < 0$ . Let the discrete-time signal  $x(n)$  be the output samples from passing  $x(t)$  through an integrate and dump device operating at Nyquist rate:

$$x(n) = \int_{(n-1)T}^{nT} x(t)dt.$$

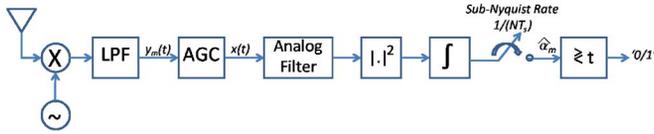


Fig. 10. Sensor measurement chain: Analog processing.

Passing the signal  $x(t)$  through the filter  $\tilde{g}_m(t)$  yields

$$\begin{aligned} \tilde{z}_m(t) &= \int_0^{KT} \tilde{g}_m(\tau) x(t - \tau) d\tau \\ &= \int_0^T \tilde{g}_m(\tau) x(t - \tau) d\tau + \int_T^{2T} \tilde{g}_m(\tau) x(t - \tau) d\tau \\ &\quad + \dots + \int_{(K-1)T}^{KT} \tilde{g}_m(\tau) x(t - \tau) d\tau \\ &= g_m(0) \int_0^T x(t - \tau) d\tau + g_m(1) \int_T^{2T} x(t - \tau) d\tau \\ &\quad + \dots + g_m(K-1) \int_{(K-1)T}^{KT} x(t - \tau) d\tau \\ &= \sum_{\ell=0}^{K-1} g_m(\ell) \int_{\ell T}^{(\ell+1)T} x(t - \tau) d\tau \end{aligned}$$

Now, consider the Nyquist-rate samples of  $\tilde{z}_m(t)$  at  $t = nT$ ,

$$\begin{aligned} \tilde{z}_m(nT) &= \sum_{\ell=0}^{K-1} g_m(\ell) \int_{\ell T}^{(\ell+1)T} x(nT - \tau) d\tau \\ &= \sum_{\ell=0}^{K-1} g_m(\ell) \int_{(n-\ell)T}^{(n-\ell+1)T} x(\tilde{\tau}) d\tilde{\tau} \\ &= \sum_{\ell=0}^{K-1} g_m(\ell) x(n - \ell) \end{aligned}$$

which is the discrete-time convolution of  $x(n)$  and  $g_m(n)$ . This shows that

$$\frac{1}{T} \int_0^{NT} |\tilde{z}_m(t)|^2 dt \approx \sum_{n=0}^N |z_m(n)|^2$$

The modified analog measurement chain is depicted in Fig. 10.

### C. Proof of Proposition 1

We show that enforcing nonnegativity of the discretized  $N_F$ -point power spectrum estimate, i.e.,  $\mathbf{s}_x = \mathbf{F}\mathbf{r}_x \geq 0$ , where  $s_x(f) = S_x\left(\frac{2\pi f}{N_F}\right)$ ,  $f \in \{0, \dots, N_F - 1\}$ , and  $\mathbf{F}$  is the  $N_F \times (2K - 1)$  (phase-shifted) DFT matrix, implies a positive

semidefinite  $K \times K$  autocorrelation matrix  $\mathbf{R}_x$ . We consider  $N_F \geq (2K - 1)$  and assume that  $N_F$  is odd (extending the proof to even  $N_F$  follows along the same lines). Define the  $N_F \times 1$  vector  $\bar{\mathbf{r}}_x$  as the zero-padded extension of  $\mathbf{r}_x$ ,

$$\bar{\mathbf{r}}_x := \begin{bmatrix} \underbrace{0 \dots 0}_{N_F - 2K + 1} & \mathbf{r}_x^T & \underbrace{0 \dots 0}_{N_F - 2K + 1} \end{bmatrix}^T.$$

Also, define  $n_F := \frac{N_F - 1}{2}$  and let  $\bar{\mathbf{F}}$  be the square  $N_F \times N_F$  phase-shifted DFT matrix:

$$\bar{\mathbf{F}} = \begin{bmatrix} 1 & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}(n_F)} \\ e^{-j\frac{2\pi}{N_F}(-n_F)} & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}(n_F)} \\ e^{-j\frac{2\pi}{N_F}2(-n_F)} & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}2(n_F)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e^{-j\frac{2\pi}{N_F}(N_F-1)(-n_F)} & \dots & 1 & \dots & e^{-j\frac{2\pi}{N_F}(N_F-1)(n_F)} \end{bmatrix}.$$

It is easy to verify that  $\bar{\mathbf{F}}\bar{\mathbf{r}}_x = \mathbf{F}\mathbf{r}_x = \mathbf{s}_x$ . Let matrix  $\mathbf{W}$  be the original (non-phase-shifted)  $N_F$ -point DFT matrix, vector  $\mathbf{v}$  be the first column of  $\bar{\mathbf{F}}$ , and define the diagonal matrix  $\mathbf{D} := \text{diag}(\mathbf{v})$  with elements of  $\mathbf{v}$  on the main diagonal, such that  $\bar{\mathbf{F}} = \mathbf{D}\mathbf{W}$  (and  $\mathbf{W} = \mathbf{D}^H\bar{\mathbf{F}}$ ).

Let  $\check{\mathbf{r}}_x^{(j)}$  be the  $j$ -th circular shift of  $\bar{\mathbf{r}}_x$  obtained by removing the last  $j$  entries of  $\bar{\mathbf{r}}_x$  and putting them as the first  $j$  entries (with  $\check{\mathbf{r}}_x^{(0)} = \bar{\mathbf{r}}_x$ ). A negative  $j$  signifies a shift in the reverse direction. Define the  $N_F \times N_F$  circulant matrix  $\mathbf{R}_c := [\check{\mathbf{r}}_x^{(-n_F)}, \dots, \check{\mathbf{r}}_x^{(0)}, \dots, \check{\mathbf{r}}_x^{(n_F)}]$ . For example, for  $K = 2$  and  $N_F = 5$ ,

$$\mathbf{R}_c = \begin{bmatrix} r(0) & r(-1) & 0 & 0 & r(1) \\ r(1) & r(0) & r(-1) & 0 & 0 \\ 0 & r(1) & r(0) & r(-1) & 0 \\ 0 & 0 & r(1) & r(0) & r(-1) \\ r(-1) & 0 & 0 & r(1) & r(0) \end{bmatrix}.$$

Circulant matrices are diagonalized by a DFT:  $\mathbf{R}_c = \frac{1}{N_F} \mathbf{W}^H \mathbf{\Lambda} \mathbf{W}$ , where  $\mathbf{\Lambda} = \text{diag}(\mathbf{W}\check{\mathbf{r}}_x^{(-n_F)})$  holds the eigenvalues of  $\mathbf{R}_c$  [26, p. 107]. Note that  $\mathbf{W}\check{\mathbf{r}}_x^{(-n_F)} = \bar{\mathbf{F}}\check{\mathbf{r}}_x^{(0)}$ . Since we enforce  $\bar{\mathbf{F}}\bar{\mathbf{r}}_x = \mathbf{s}_x \geq 0$ , this directly implies that  $\mathbf{R}_c$  is positive semidefinite. Next, it is easy to see that the  $K \times K$  autocorrelation matrix  $\mathbf{R}_x = \text{Toeplitz}(\mathbf{r}_x)$  can be obtained by deleting the last  $N_F - K$  rows and the last  $N_F - K$  columns of  $\mathbf{R}_c$ , i.e.,  $\mathbf{R}_x$  is the  $K$ -th order leading principal submatrix of  $\mathbf{R}_c$ . Sylvester's criterion states that a matrix is positive semidefinite if and only if the determinant of every principal submatrix is nonnegative [26, p. 160]. This implies that if  $\mathbf{R}_c \succeq 0$ , then the principal submatrix  $\mathbf{R}_x \succeq 0$ . Hence, we showed that enforcing  $\mathbf{F}\mathbf{r}_x \geq 0$  implies that  $\mathbf{R}_x \succeq 0$ . The converse is not true since  $\mathbf{R}_x \succeq 0$  does not necessarily imply that  $\mathbf{R}_c \succeq 0$ .

### REFERENCES

- [1] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.
- [2] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [3] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, vol. 4, pp. 1357–1360.

- [4] F. Zeng, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb. 2011.
- [5] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wideband spectrum sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2337–2340.
- [6] D. D. Ariananda and G. Leus, "Wideband power spectrum sensing using sub-Nyquist sampling," in *Proc. IEEE SPAWC*, San Francisco, CA, USA, Jun. 26–29, 2011, pp. 101–105.
- [7] D. D. Ariananda and G. Leus, "Power spectrum blind sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 443–446, Aug. 2011.
- [8] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.
- [9] G. Leus and Z. Tian, "Recovering second-order statistics from compressive measurements," in *Proc. IEEE CAMSAP*, San Juan, PR, Dec. 13–16, 2011, pp. 337–340.
- [10] M. Lexa, M. Davies, J. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3884–3887.
- [11] V. Havary-Nassab, S. Hassan, and S. Valaee, "Compressive detection for wideband spectrum sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 3094–3097.
- [12] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," presented at the Conf. Inf. Sci., Syst. (CISS), Princeton, NJ, USA, Mar. 2008.
- [13] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-Bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, Apr. 2013.
- [14] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Commun. Pure Appl. Math.* [Online]. Available: arxiv.org/abs/1109.4299, doi: 10.1002/cpa.21442
- [15] V. Cevher, P. Boufounos, R. Baraniuk, A. Gilbert, and M. Strauss, "Nearoptimal bayesian localization via incoherence and sparsity," presented at the Int. Conf. Inf. Process. Sensor Netw. (IPSN), San Francisco, CA, USA, Apr. 137–16, 2009.
- [16] R. Gray and L. Davidson, *Random Processes: A Mathematical Approach for Engineers*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1986.
- [17] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.
- [18] U. Kamilov, A. Bourquard, A. Amini, and M. Unser, "One-bit measurements with adaptive thresholds," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 607–610, Oct. 2012.
- [19] M. Dyer and A. Frieze, "On the complexity of computing the volume of a polyhedron," *SIAM J. Comput.*, vol. 17, no. 5, pp. 967–974, 1988.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [21] J. Laska and R. Baraniuk, "Regime change: Bit-depth versus measurement-rate in compressive sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3496–3505, Jul. 2012.
- [22] J. J. Fuchs, "A new approach to robust linear regression," in *Proc. 14th IFAC World Congr.*, Beijing, China, Jul. 1999, pp. 427–432.
- [23] S. Farahmand, D. Angelosante, and G. B. Giannakis, "Doubly robust Kalman smoothing by exploiting outlier sparsity," presented at the Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 7–10, 2010.
- [24] A. J. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.
- [25] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.
- [26] F. Zhang, *Matrix Theory: Basic Results and Techniques*. New York, NY, USA: Springer-Verlag, 1999.



**Omar Mehanna** (S'05) received the B.Sc. degree in electrical engineering from Alexandria University, Egypt, in 2006 and the M.Sc. degree in electrical engineering from Nile University, Egypt, in 2009.

Since 2009, he has been working towards the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA. His current research focuses on signal processing for communications, ad-hoc networks, and cognitive radio.



**Nicholas D. Sidiropoulos** (F'09) received the Diploma degree in electrical engineering from the Aristotelian University of Thessaloniki, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland—College Park, MD, USA, in 1988, 1990, and 1992, respectively.

He served as Assistant Professor at the University of Virginia (1997–1999); Associate Professor at the University of Minnesota—Minneapolis, MN, USA (2000–2002); Professor at the Technical University of Crete, Greece (2002–2011); and Professor at the

University of Minnesota—Minneapolis (2011 to present). His current research focuses primarily on signal and tensor analytics, with applications in cognitive radio, big data, and preference measurement.

Dr. Sidiropoulos received the NSF/CAREER award (1998), the IEEE Signal Processing Society (SPS) Best Paper Award (2001, 2007, and 2011), and the IEEE SPS Meritorious Service Award (2010). He has served as IEEE SPS Distinguished Lecturer (2008–2009) and Chair of the IEEE Signal Processing for Communications and Networking Technical Committee (2007–2008).