

Compressive Wideband Power Spectrum Estimation

Dyonisius Dony Ariananda, *Student Member, IEEE*, and Geert Leus, *Fellow, IEEE*

Abstract—In several applications, such as wideband spectrum sensing for cognitive radio, only the power spectrum (a.k.a. the power spectral density) is of interest and there is no need to recover the original signal itself. In addition, high-rate analog-to-digital converters (ADCs) are too power hungry for direct wideband spectrum sensing. These two facts have motivated us to investigate compressive wideband power spectrum sensing, which consists of a compressive sampling procedure and a reconstruction method that is able to recover the unknown power spectrum of a wide-sense stationary signal from the obtained sub-Nyquist rate samples. It is different from spectrum blind sampling (SBS), which aims at reconstructing the original signal instead of the power spectrum. In this paper, a solution is first presented based on a periodic sampling procedure and a simple least-squares reconstruction method. We evaluate the reconstruction process both in the time and frequency domain. Then, we examine two possible implementations for the compressive sampling procedure, namely complex Gaussian sampling and multicorset sampling, although we mainly focus on the latter. A new type of multicorset sampling is introduced based on the so-called minimal sparse ruler problem. Next, we analyze the statistical properties of the estimated power spectrum. The computation of the mean and the covariance of the estimates allows us to calculate the analytical normalized mean squared error (NMSE) of the estimated power spectrum. Further, when the received signal is assumed to contain only circular complex zero-mean Gaussian i.i.d. noise, the computed mean and covariance can be used to derive a suitable detection threshold. Simulation results underline the promising performance of our proposed approach. Note that all benefits of our method arise without putting any sparsity constraints on the power spectrum.

Index Terms—Compressive sampling, multicorset sampling, power spectrum estimation, sparse ruler, wide-sense stationary signals.

I. INTRODUCTION

IN recent years, wideband spectrum estimation and sensing has become a popular topic in signal processing and telecommunications. A popular application is cognitive radio where unlicensed users have to sense a broad frequency range in order to locate the unoccupied licensed spectrum before establishing a communication link. One possible approach is to divide the entire wideband spectrum into a large number of narrowband channels followed by a channel-by-channel

sequential sensing. However, this approach might introduce a significant amount of delay in the spectrum sensing process. In [1], a filter bank structure is introduced to perform multichannel spectrum sensing in the wideband regime. Similarly, [2], [3] optimize a bank of multiple narrowband detectors to improve the aggregate opportunistic throughput of a cognitive radio system by introducing the so-called multiband joint detection and multiband sensing-time-adaptive joint detection, respectively. Again, these methods are not efficient due to the need for a large number of bandpass filters. Another approach is to directly scan the wideband spectrum using a high-rate analog-to-digital converter (ADC), such as in [4], where wavelets are used to detect the edges or boundaries of the occupied bands. However, such high-rate ADCs consume a large amount of power [5].

To reduce the burden on the ADCs, many researches have been performed to exploit specific features of the spectrum (such as sparsity in the spectrum or the edge spectrum [6]–[8]). These specific properties allow for a reduction of the sampling rate compared to the Nyquist rate while maintaining perfect signal reconstruction when no noise is present. In [9], the so-called multicorset sampling is examined and proposed to reduce the sampling rate when the considered multiband signals have a frequency support on a union of finite intervals. Given prior knowledge of the frequency support of the received signals, [9] has derived the condition for exact reconstruction as well as proposed an explicit reconstruction formula. Unfortunately, in many applications, such as cognitive radio, the frequency support is not known in advance and the multicorset sampling approach proposed in [9] is not suitable. In order to solve this problem, [7], [10] proposed solutions for signal reconstruction based on multicorset sampling without any prior knowledge about the frequency support of the original signal. Closely related ideas can also be found in [8], which discusses sub-Nyquist rate sampling for sparse multiband analog signals by means of a so-called modulated wideband converter, which consists of multiple branches, each of which employs a different periodic mixing function followed by low-pass filtering and low-rate uniform sampling. Since the objective of the methods discussed in [7], [8], and [10] is to sample a signal with unknown frequency support at minimal rate and reconstruct the spectrum from the samples by exploiting spectrum sparsity, these approaches fall in the class of spectrum blind sampling (SBS). In these works, it has been found that the minimum average sampling rate for most signals is given by the Landau lower bound (as studied in [9]), which is equal to the Nyquist rate multiplied with the frequency occupancy ratio. However, in the worst case scenario, the minimum average sampling rate increases and is given by the minimum of twice the Landau lower bound and the Nyquist rate. Note that all of the above approaches can be cast into a compressive sampling framework where the signal reconstruction can be carried out by using your favorite sparse recovery method such as the least

Manuscript received February 08, 2012; revised May 11, 2012; accepted May 11, 2012. Date of publication May 30, 2012; date of current version August 07, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Maria S. Greco. This work was supported by NWO-STW under the VICI program (project 10382). A portion of this work was presented at the IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2011), June 2011 and the International Conference on Digital Signal Processing (DSP 2011), July 2011.

The authors are with the Faculty of EEMCS, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: d.a.dyonisius@tudelft.nl; g.j.t.leus@tudelft.nl).

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.2012.2201153

absolute shrinkage and selection operator (LASSO) algorithm [11]. Also more classical methods can be adopted, such as the minimum variance distortionless response (MVDR) method [12], or multiple signal classification (MUSIC) [7].

All methods aforementioned concentrate on spectral estimation and aim at perfectly reconstructing the original signal. In fact, for spectrum sensing applications, only the power spectrum (a.k.a. the power spectral density), or equivalently, the autocorrelation function, needs to be recovered. Power spectrum estimation methods based on sub-Nyquist rate samples have been developed in [13], [14] by concentrating on the autocorrelation function instead of the original signal itself. In [13], the spectrum sensing approach proposed by [6], which exploits the embedded sparsity of the edge spectrum, is improved by taking advantage of the connection between the autocorrelation function of the compressive measurements and that of the Nyquist rate samples. Nevertheless, [13] assumes that the compressive measurements are wide-sense stationary, which is not true for most compressive sampling matrices. In [14], a compressive sampling framework is obtained by computing the output energy of a limited number of wideband filters to reconstruct the received energy in a large number of spectral bins. Unfortunately, [14] only exploits the output energy of each filter leading to an under-determined system of equations, while cross-correlations among the outputs of the different filters could also have been exploited. In [15], a power spectrum estimation method based on multicorset sampling is proposed by exploiting the fact that a wide-sense stationary signal corresponds to a diagonal covariance matrix of the frequency domain representation of the signal. This observation is used in [15] to build an overdetermined system of equations relating the frequency domain statistics of the compressive measurements with those of the signal, which is solvable by adopting a nonnegative least-squares algorithm. Another method labeled as coprime sampling is provided by [16]. This method aims at estimating the frequencies of sinusoids buried in noise by exploiting two uniform sub-Nyquist samplers with sampling periods that are coprime multiples of the Nyquist period.

This paper concentrates on efficient power spectrum reconstruction and aims at designing effective periodic sub-Nyquist sampling procedures for this, also labeled as power spectrum blind sampling (PSBS) in [17]. Theoretically, this approach is able to perfectly reconstruct the unknown power spectrum of a wide-sense stationary signal using least-squares by exploiting the cross-correlations between the different outputs of the periodic sampling device. The least-squares algorithm requires some rank conditions to be satisfied, which will guide the actual implementation of the sampling device. In this paper, sampling techniques based on random modulating waveforms can be adopted, as used in [8], but the main focus will be on multicorset approaches. A novel multicorset sampling implementation is designed based on the minimal sparse ruler problem. The theoretical statistical properties of the estimated power spectrum are also investigated leading to the mean and the covariance of the estimated power spectrum, which is useful for formulating the normalized mean squared error (NMSE) analytically. Moreover, the detection threshold used to evaluate the presence or absence of a signal at a specific frequency can also be de-

rived by assuming the received signal is merely circular complex zero-mean Gaussian i.i.d. noise. All the proposed schemes are compared via both analysis and simulations. In general, the developed sampling procedures can significantly decrease the sampling rate requirements by exploiting the spectral correlation properties without putting any sparsity constraints on the power spectrum.

II. SYSTEM MODEL AND PROBLEM STATEMENT

Let $x(t)$ be a wide-sense stationary analog signal, which is assumed to be complex-valued (e.g., the complex envelope of the observed real-valued signal) and bandlimited with bandwidth $\frac{1}{T}$ (which also indicates the Nyquist rate). We then consider a spectrum sensing application, where the task is to sense the power spectrum of $x(t)$. Fig. 1 depicts the employed sampling device, which can be regarded as one possible implementation of an analog to information converter (AIC) in a compressive sampling operation. However, note that this sampling device is capable of modeling any AIC implementation, such as those proposed in [18], [19]. The considered sampling device has M branches, where the i th branch modulates the signal $x(t)$ with a possibly complex-valued periodic waveform $p_i(t)$ of period NT followed by an integrate-and-dump device with period NT (thus with rate equal to $\frac{1}{N}$ times the Nyquist rate). The output of the i th branch at the k th sampling index can thus be expressed as

$$\begin{aligned} y_i[k] &= \frac{1}{NT} \int_{kNT}^{(k+1)NT} p_i(t)x(t)dt \\ &= \frac{1}{T} \int_{kNT}^{(k+1)NT} c_i(t - kNT)x(t)dt \end{aligned} \quad (1)$$

where $c_i(t)$ yields a single period of $\frac{1}{N}p_i(t)$, i.e., $c_i(t) = \frac{1}{N}p_i(t)$ for $0 \leq t < NT$ and $c_i(t) = 0$ elsewhere. Assume now that $c_i(t)$ is a piecewise constant function having constant values in every interval of length T , i.e., $c_i(t) = c_i[-n]$ for $nT \leq t < (n+1)T$, where $n = 0, 1, \dots, N-1$. Then, (1) can be rewritten as

$$\begin{aligned} y_i[k] &= \sum_{n=0}^{N-1} c_i[-n] \frac{1}{T} \int_{(kN+n)T}^{(kN+n+1)T} x(t)dt \\ &= \sum_{n=0}^{N-1} c_i[-n] x[kN+n] \\ &= \sum_{n=1-N}^0 c_i[n] x[kN-n] \end{aligned} \quad (2)$$

where $x[n]$ can be viewed as the output of an integrate-and-dump process with period T (thus with rate equal to the Nyquist rate) applied to $x(t)$, which is not explicitly computed due to its high complexity. Note that the average sampling rate of this periodic sampler is given by the Nyquist rate multiplied by $\frac{M}{N}$ and hence we will use $M < N$ to keep the complexity low. The presented sampling device is actually similar to the modulated wideband converter introduced in [8], where the values of $c_i[n]$ are randomly generated, e.g., adopting complex Gaussian sampling or random binary (from the set $\{\pm 1\}$) sampling. However, the sampler coefficients $c_i[n]$ can also be set to implement

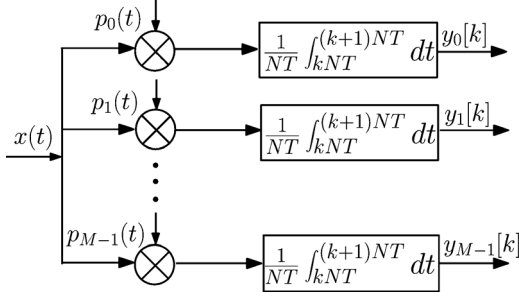


Fig. 1. Illustration of the sample acquisition scheme, which modulates the received analog signal with M different periodic waveforms followed by an integrate-and-dump process.

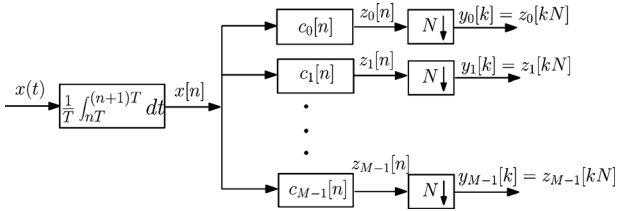


Fig. 2. Digital interpretation of the sampling device of Fig. 1, consisting of a high-rate integrate-and-dump process, followed by a bank of M branches, where each branch consists of a digital filtering operation followed by a down-sampling operation.

efficient multicoset sampling, which will be discussed in more detail in Section V.

Fig. 2 underlines the important fact that (2) can actually be perceived as a digital filtering operation of $x[n]$ by the filter $c_i[n]$ of length N followed by an N -fold decimation, i.e., $y_i[k] = z_i[kN]$, where

$$z_i[n] = c_i[n] \star x[n] = \sum_{m=1-N}^0 c_i[m]x[n-m]$$

with \star representing the convolution operator. This observation turns out to be useful for the reconstruction process.

The goal of this paper is to reconstruct the power spectrum of $x(t)$ based on the obtained samples $\{y_i[k]\}_{i,k}$. Since $x[n]$ is obtained from $x(t)$ by an integrate-and-dump device operating at Nyquist-rate, the spectrum of $x[n]$ is given by a periodic extension of a slightly changed version of the spectrum of $x(t)$ without aliasing. As a result, the power spectrum of $x[n]$ is uniquely determined by the power spectrum of $x(t)$ and vice versa, and thus we will concentrate on reconstructing the power spectrum of $x[n]$ in this paper.

It is well-known that the power spectrum or power spectral density (PSD) of $x[n]$ is given by

$$P_x(\omega) = \sum_{n=-\infty}^{\infty} r_x[n]e^{-jn\omega}, \quad 0 \leq \omega < 2\pi$$

where $r_x[n]$ represents the autocorrelation function of $x[n]$, defined as $r_x[n] = E(x[n]x^*[n-n])$. Therefore, estimating the power spectrum $P_x(\omega)$ amounts to estimating the autocorrelation function $r_x[n]$. The major contribution of this work is that we will take advantage of all the M^2 different cross-spectra of $y_i[k]$ with $y_j[k]$ for $i, j = 0, 1, \dots, M-1$, which will enable rate-compression without introducing any sparsity constraints

on $x(t)$. Note that the cross-spectrum or cross spectral density (CSD) of $y_i[k]$ with $y_j[k]$ is given by

$$P_{y_i, y_j}(\omega) = \sum_{k=-\infty}^{\infty} r_{y_i, y_j}[k]e^{-jk\omega}, \quad 0 \leq \omega < 2\pi$$

where $r_{y_i, y_j}[k] = E(y_i[l]y_j^*[l-k])$ is the cross-correlation function of $y_i[k]$ with $y_j[k]$. These ensemble quantities $\{r_{y_i, y_j}[k]\}_{i,j}$ can be estimated by their sample averages, which in turn result in estimates of $\{P_{y_i, y_j}(\omega)\}_{i,j}$. In the following sections, we will first describe a time-domain approach to reconstruct $r_x[n]$ given $r_{y_i, y_j}[k]$ for $i, j = 0, 1, \dots, M-1$. Next, a frequency-domain approach to estimate $P_x(\omega)$ given $P_{y_i, y_j}(\omega)$ for $i, j = 0, 1, \dots, M-1$ will be discussed.

III. TIME-DOMAIN RECONSTRUCTION APPROACH

A. Reconstruction Analysis

In this subsection, a method to reconstruct $r_x[n]$ given $r_{y_i, y_j}[k]$ for $i, j = 0, 1, \dots, M-1$ is presented. Since $y_i[k] = z_i[kN]$, the cross-correlation function of $y_i[k]$ with $y_j[k]$ can be expressed as the N -fold decimated version of the cross-correlation function of $z_i[n]$ with $z_j[n]$, as follows:

$$r_{y_i, y_j}[k] = E(y_i[l]y_j^*[l-k]) = E(z_i[lN]z_j^*[(l-k)N]) = r_{z_i, z_j}[kN]. \quad (3)$$

It is obvious that $r_{z_i, z_j}[n]$ can be written as

$$r_{z_i, z_j}[n] = r_{c_i, c_j}[n] \star r_x[n] = \sum_{m=-N+1}^{N-1} r_{c_i, c_j}[m]r_x[n-m] \quad (4)$$

where $r_{c_i, c_j}[n]$ is the “deterministic” cross-correlation function between $c_i[n]$ and $c_j[n]$

$$r_{c_i, c_j}[n] = c_i[n] \star c_j^*[-n] = \sum_{m=1-N}^0 c_i[m]c_j^*[m-n]. \quad (5)$$

From (3) and (4), we obtain

$$\begin{aligned} r_{y_i, y_j}[k] &= r_{z_i, z_j}[kN] = \sum_{m=-N+1}^{N-1} r_{c_i, c_j}[m]r_x[kN-m] \\ &= \sum_{l=0}^1 \mathbf{r}_{c_i, c_j}^T[l] \mathbf{r}_x[k-l] \end{aligned} \quad (6)$$

which is based on the following definitions:

$$\mathbf{r}_{c_i, c_j}[0] = [r_{c_i, c_j}[0], r_{c_i, c_j}[-1], \dots, r_{c_i, c_j}[-N+1]]^T \quad (7)$$

$$\mathbf{r}_{c_i, c_j}[1] = [r_{c_i, c_j}[N], r_{c_i, c_j}[N-1], \dots, r_{c_i, c_j}[1]]^T \quad (8)$$

$$\mathbf{r}_x[k] = [r_x[kN], r_x[kN+1], \dots, r_x[(k+1)N-1]]^T. \quad (9)$$

By cascading the M^2 different cross-correlation functions $r_{y_i, y_j}[k]$, we obtain the $M^2 \times 1$ vector $\mathbf{r}_y[k] = [\dots, r_{y_i, y_j}[k], \dots]^T$, for $i, j = 0, 1, \dots, M-1$, which can be derived from (6) as

$$\mathbf{r}_y[k] = \sum_{l=0}^1 \mathbf{R}_c[l] \mathbf{r}_x[k-l] \quad (10)$$

where $\mathbf{R}_c[0]$ and $\mathbf{R}_c[1]$ are the $M^2 \times N$ matrices given by $\mathbf{R}_c[0] = [\dots, \mathbf{r}_{c_i, c_j}[0], \dots]^T$ and $\mathbf{R}_c[1] = [\dots, \mathbf{r}_{c_i, c_j}[1], \dots]^T$, respectively, for $i, j = 0, 1, \dots, M-1$.

Due to the bandlimitedness of $x[n]$, $\mathbf{r}_y[k]$ basically has unlimited support. However, in many practical situations, $\mathbf{r}_y[k]$ only has significant values within a range $-L \leq k \leq L$ and negligible values outside this range, where L is a design parameter that can be chosen as large as required. Hence, let us relax the bandlimitedness condition and assume that the support of $\mathbf{r}_y[k]$ is strictly limited to $-L \leq k \leq L$, which is a rather standard approach when computing a cross-spectrum from a cross-correlation function. Since from (10), it is clear that $\mathbf{r}_y[k]$ depends on both $\mathbf{r}_x[k]$ and $\mathbf{r}_x[k-1]$, one could think that under the above assumption the support of $\mathbf{r}_x[k]$ is limited to $-L-1 \leq k \leq L$, but that would mean that also $\mathbf{r}_y[-L-1]$ is nonzero. As a consequence, the support of $\mathbf{r}_x[k]$ should also be limited to $-L \leq k \leq L$.

All these quantities can be collected into the following vectors:

$$\mathbf{r}_y = [\mathbf{r}_y^T[0], \mathbf{r}_y^T[1], \dots, \mathbf{r}_y^T[L], \mathbf{r}_y^T[-L], \dots, \mathbf{r}_y^T[-1]]^T \quad (11)$$

$$\mathbf{r}_x = [\mathbf{r}_x^T[0], \mathbf{r}_x^T[1], \dots, \mathbf{r}_x^T[L], \mathbf{r}_x^T[-L], \dots, \mathbf{r}_x^T[-1]]^T \quad (12)$$

where \mathbf{r}_y has size $(2L+1)M^2 \times 1$ and \mathbf{r}_x has size $(2L+1)N \times 1$. Let us further introduce two other important observations. First, based on the definition of $\mathbf{r}_x[k]$ in (9), the fact that the support of $\mathbf{r}_x[k]$ is limited to $-L \leq k \leq L$, and the complex conjugate symmetry in $r_x[n]$, it is clear that the support of $r_x[n]$ is limited to $-LN \leq n \leq LN$ and the last $N-1$ entries of $\mathbf{r}_x[L]$ are zero. Second, based on the definition of $\mathbf{r}_{c_i, c_j}[1]$ in (8) and the fact that the support of $r_{c_i, c_j}[n]$ is limited to $1-N \leq n \leq N-1$, it is clear that the first column of $\mathbf{R}_c[1]$ is zero. These two observations allow us to write the linear convolution in (10) as a circular convolution within $-L \leq k \leq L$, without any additional zero padding. Hence, we can eventually write the relation between \mathbf{r}_y and \mathbf{r}_x as

$$\mathbf{r}_y = \mathbf{R}_c \mathbf{r}_x \quad (13)$$

where \mathbf{R}_c is the $(2L+1)M^2 \times (2L+1)N$ matrix given by

$$\mathbf{R}_c = \begin{bmatrix} \mathbf{R}_c[0] & & & & \mathbf{R}_c[1] \\ \mathbf{R}_c[1] & \mathbf{R}_c[0] & & & \\ & \mathbf{R}_c[1] & \mathbf{R}_c[0] & & \\ & & \ddots & \ddots & \\ & & & \mathbf{R}_c[1] & \mathbf{R}_c[0] \end{bmatrix}. \quad (14)$$

Note that (13) is solvable using least-squares (LS) if \mathbf{R}_c has full column rank, which obviously requires $M^2 \geq N$.

The inverse problem of (13) can be further simplified by observing that \mathbf{R}_c is a block circulant matrix with blocks of size $M^2 \times N$, which can easily be converted into a block diagonal matrix \mathbf{Q}_c with blocks of size $M^2 \times N$. This can be carried out solely by using the $(2L+1)$ -point (inverse) discrete Fourier transform ((I)DFT)

$$\mathbf{R}_c = (\mathbf{F}_{2L+1}^{-1} \otimes \mathbf{I}_{M^2}) \mathbf{Q}_c (\mathbf{F}_{2L+1} \otimes \mathbf{I}_N)$$

where \mathbf{F}_{2L+1} is the $(2L+1) \times (2L+1)$ DFT matrix, and $\mathbf{Q}_c = \text{diag}\{\mathbf{Q}_c(0), \mathbf{Q}_c(2\pi \frac{1}{2L+1}), \dots, \mathbf{Q}_c(2\pi \frac{2L}{2L+1})\}$ with

$\mathbf{Q}_c(\omega)$ being the $M^2 \times N$ matrix spectrum of the $M^2 \times N$ matrix sequence $\{\mathbf{R}_c[k]\}_{k=0,1}$

$$\mathbf{Q}_c(\omega) = \sum_{k=0}^1 \mathbf{R}_c[k] e^{-jk\omega}. \quad (15)$$

Consequently, by defining the $(2L+1)N \times 1$ vector \mathbf{q}_x and the $(2L+1)M^2 \times 1$ vector \mathbf{q}_y as

$$\mathbf{q}_x = (\mathbf{F}_{2L+1} \otimes \mathbf{I}_N) \mathbf{r}_x \quad (16)$$

$$\mathbf{q}_y = (\mathbf{F}_{2L+1} \otimes \mathbf{I}_{M^2}) \mathbf{r}_y \quad (17)$$

we can re-express (13) as

$$\mathbf{q}_y = \mathbf{Q}_c \mathbf{q}_x. \quad (18)$$

From (11), (12), (16), and (17), we can also write \mathbf{q}_x and \mathbf{q}_y as

$$\mathbf{q}_x = \left[\mathbf{q}_x^T(0), \mathbf{q}_x^T\left(2\pi \frac{1}{2L+1}\right), \dots, \mathbf{q}_x^T\left(2\pi \frac{2L}{2L+1}\right) \right]^T$$

$$\mathbf{q}_y = \left[\mathbf{q}_y^T(0), \mathbf{q}_y^T\left(2\pi \frac{1}{2L+1}\right), \dots, \mathbf{q}_y^T\left(2\pi \frac{2L}{2L+1}\right) \right]^T$$

where $\mathbf{q}_x(\omega)$ and $\mathbf{q}_y(\omega)$ are, respectively, the $N \times 1$ and $M^2 \times 1$ vector spectra of the $N \times 1$ and $M^2 \times 1$ vector sequences $\mathbf{r}_x[k]$ and $\mathbf{r}_y[k]$

$$\mathbf{q}_y(\omega) = \sum_{k=-L}^L \mathbf{r}_y[k] e^{-jk\omega}$$

$$\mathbf{q}_x(\omega) = \sum_{k=-L}^L \mathbf{r}_x[k] e^{-jk\omega}. \quad (19)$$

Here, \mathbf{q}_y is obtained from the sample-averaged version of $\mathbf{r}_y[k]$, while \mathbf{q}_x , and thus $\mathbf{r}_x[k]$, is to be estimated. By using (19), we are able to rewrite (18) as a set of $2L+1$ matrix equations:

$$\mathbf{q}_y\left(2\pi \frac{l}{2L+1}\right) = \mathbf{Q}_c\left(2\pi \frac{l}{2L+1}\right) \mathbf{q}_x\left(2\pi \frac{l}{2L+1}\right). \quad (20)$$

If $\mathbf{Q}_c(2\pi \frac{l}{2L+1})$ has full column rank for $l = 0, 1, \dots, 2L$, we can compute $\mathbf{q}_x(2\pi \frac{l}{2L+1})$ using LS from (20) for $l = 0, 1, \dots, 2L$. Note that the above simplification unintentionally transformed the time-domain approach in some kind of frequency-domain approach. However, we can also directly start from the frequency domain, as indicated in Section IV. Having estimated \mathbf{q}_x , we can reconstruct \mathbf{r}_x using (16) and then compute the $(2L+1)N \times 1$ power spectrum vector \mathbf{s}_x as

$$\mathbf{s}_x = \mathbf{F}_{(2L+1)N} \mathbf{r}_x \quad (21)$$

where

$$\mathbf{s}_x = \left[P_x(0), P_x\left(2\pi \frac{1}{(2L+1)N}\right), \dots, P_x\left(2\pi \frac{(2L+1)N-1}{(2L+1)N}\right) \right]^T$$

and $\mathbf{F}_{(2L+1)N}$ is the $(2L+1)N \times (2L+1)N$ DFT matrix.

B. Alternative Time-domain Approach

In this subsection, we present an alternative yet different version of the time-domain reconstruction approach presented in

Section III-A. We start by rewriting (2) in matrix-vector notation

$$\mathbf{y}[k] = \mathbf{C}\mathbf{x}[k] \quad (22)$$

where the $M \times 1$ measurement vectors $\mathbf{y}[k]$ and the $N \times 1$ vector sequence $\mathbf{x}[k]$ are, respectively, defined as

$$\mathbf{y}[k] = [y_0[k], y_1[k], \dots, y_{M-1}[k]]^T \quad (23)$$

$$\mathbf{x}[k] = [x[kN], x[kN+1], \dots, x[kN+N-1]]^T \quad (24)$$

while the $M \times N$ compressive sampling matrix \mathbf{C} is given by

$$\mathbf{C} = [\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{M-1}]^T$$

with $\mathbf{c}_i = [c_i[0], c_i[-1], \dots, c_i[1-N]]^T$.

Next, we compute the $M \times M$ autocorrelation matrix of $\mathbf{y}[k]$ in (23), which is given by $\mathbf{R}_y[0] = E(\mathbf{y}[k]\mathbf{y}^H[k])$. If we also construct the $N \times N$ autocorrelation matrix of $\mathbf{x}[k]$ in (24) as $\mathbf{R}_x^a = E(\mathbf{x}[k]\mathbf{x}^H[k])$, the relationship between \mathbf{R}_x^a and $\mathbf{R}_y[0]$ can be expressed as

$$\mathbf{R}_y[0] = \mathbf{C}\mathbf{R}_x^a\mathbf{C}^H \quad (25)$$

where the elements of \mathbf{R}_x^a are given by $[\mathbf{R}_x^a]_{ij} = r_x[i-j] = r_x^*[j-i]$ due to the wide-sense stationary property of $x[n]$. While \mathbf{R}_x^a has a Toeplitz structure, this is not the case for $\mathbf{R}_y[0]$ because the elements of $\mathbf{y}[k]$ in (23) are generally not wide-sense stationary due to the nature of the compressive sampling matrix \mathbf{C} . Therefore, it is theoretically possible to exploit all columns of $\mathbf{R}_y[0]$ to estimate one of the columns of \mathbf{R}_x^a by first stacking all columns of $\mathbf{R}_y[0]$ into the $M^2 \times 1$ vector $\text{vec}(\mathbf{R}_y[0])$, which is nothing else than $\mathbf{r}_y[0]$ in (11). Here, $\text{vec}(\cdot)$ is the operator that stacks all columns of a matrix in a large column vector. Based on (25), it is evident that $\mathbf{r}_y[0]$ in (11) can be written as

$$\mathbf{r}_y[0] = \text{vec}(\mathbf{R}_y[0]) = (\mathbf{C}^* \otimes \mathbf{C})\text{vec}(\mathbf{R}_x^a) \quad (26)$$

where \otimes represents the Kronecker product operation. Since all columns of \mathbf{R}_x^a contain the same information, \mathbf{R}_x^a can be condensed into the $2N-1$ vector $\mathbf{r}_x^a = [r_x[0], r_x[1], \dots, r_x[N-1], r_x[1-N], \dots, r_x[-1]]^T$, and we can write

$$\text{vec}(\mathbf{R}_x^a) = \mathbf{T}\mathbf{r}_x^a \quad (27)$$

where \mathbf{T} is a special $N^2 \times (2N-1)$ repetition matrix with the i th row of \mathbf{T} given by the $((i-1+(N-2)\lfloor \frac{i-1}{N} \rfloor) \bmod (2N-1) + 1)$ th row of the identity matrix \mathbf{I}_{2N-1} . By combining (26) and (27), we obtain

$$\mathbf{r}_y[0] = (\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}\mathbf{r}_x^a = \mathbf{R}_c^a\mathbf{r}_x^a \quad (28)$$

where the $M^2 \times (2N-1)$ matrix $\mathbf{R}_c^a = (\mathbf{C}^* \otimes \mathbf{C})\mathbf{T}$ is actually given by

$$\mathbf{R}_c^a = [\mathbf{R}_c^a[0], \mathbf{R}_c^a[1]] \quad (29)$$

with $\mathbf{R}_c^a[0]$ equal to $\mathbf{R}_c[0]$ and $\mathbf{R}_c^a[1]$ obtained by removing the first column of $\mathbf{R}_c[1]$ (which actually has zero entries). If \mathbf{R}_c^a has full column rank, it is possible to reconstruct the autocorrelation vector \mathbf{r}_x^a from (28) using LS. Then, we can compute the power

spectrum vector \mathbf{s}_x as in (21) by replacing \mathbf{r}_x with a zero-padded version of \mathbf{r}_x^a denoted as $\tilde{\mathbf{r}}_x^a$, i.e., $\mathbf{s}_x = \mathbf{F}_{(2L+1)N}\tilde{\mathbf{r}}_x^a$. Note that as the time-domain approach only gives a valid power spectrum estimate when $r_x[n]$ has negligible correlation values above lag LN , where L can be freely selected, the alternative time-domain approach should only be preferred if $r_x[n]$ has negligible correlation values above lag $N-1$. Hence, for a fixed N , e.g., when the sampler is fixed, this is a clear disadvantage of the alternative approach.

Comparing the alternative approach to (13), we further observe that it is a different method that can not merely be obtained from the time-domain approach by setting $L=0$ in (11) and (12). If we set $L=0$ in (11), the support of $\mathbf{r}_y[k]$ is limited to $k=0$. However, as we explained in Section III-A, the support of $\mathbf{r}_x[k]$ is then also limited to $k=0$. Further, from the complex conjugate symmetry in $r_x[n]$, we can then conclude that setting $L=0$ is only possible when $x[n]$ is assumed to be a white sequence, i.e., $r_x[n] = r_x[0]\delta[n]$. Hence, the minimum possible value of L in (11) and (12) is $L=1$.

IV. FREQUENCY-DOMAIN RECONSTRUCTION APPROACH

In this section, we develop a frequency-domain reconstruction approach. The reason for presenting this frequency-domain method is its tight connection to spectrum blind sampling (SBS) presented in [7], [8], [10]. SBS also starts from a frequency-domain viewpoint but focuses on spectrum reconstruction instead of power spectrum reconstruction. Since $y_i[k] = z_i[kN]$, we can also write the cross-spectrum of $y_i[k]$ with $y_j[k]$ as an N -fold aliased version of the cross-spectrum of $z_i[n]$ and $z_j[n]$

$$P_{y_i, y_j}(\omega) = \frac{1}{N} \sum_{n=0}^{N-1} P_{z_i, z_j} \left(\omega \frac{1}{N} + 2\pi \frac{n}{N} \right), \quad 0 \leq \omega < 2\pi. \quad (30)$$

It is well known that $P_{z_i, z_j}(\omega)$ can be written as

$$P_{z_i, z_j}(\omega) = P_{c_i, c_j}(\omega)P_x(\omega), \quad 0 \leq \omega < 2\pi \quad (31)$$

where $P_{c_i, c_j}(\omega)$ is the ‘‘deterministic’’ cross-spectrum between $c_i[n]$ and $c_j[n]$

$$P_{c_i, c_j}(\omega) = C_i(\omega)C_j^*(\omega), \quad 0 \leq \omega < 2\pi.$$

From (30) and (31), we can thus write

$$\begin{aligned} P_{y_i, y_j}(\omega) &= \frac{1}{N} \sum_{n=0}^{N-1} P_{c_i, c_j} \left(\omega \frac{1}{N} + 2\pi \frac{n}{N} \right) P_x \left(\omega \frac{1}{N} + 2\pi \frac{n}{N} \right) \\ &= \mathbf{p}_{c_i, c_j}^T(\omega) \mathbf{p}_x(\omega), \quad 0 \leq \omega < 2\pi \end{aligned}$$

where we have that

$$\begin{aligned} \mathbf{p}_{c_i, c_j}(\omega) &= \frac{1}{N} \left[P_{c_i, c_j} \left(\omega \frac{1}{N} \right), \dots, P_{c_i, c_j} \left(\omega \frac{1}{N} + 2\pi \frac{N-1}{N} \right) \right]^T \\ \mathbf{p}_x(\omega) &= \left[P_x \left(\omega \frac{1}{N} \right), \dots, P_x \left(\omega \frac{1}{N} + 2\pi \frac{N-1}{N} \right) \right]^T. \quad (32) \end{aligned}$$

Note that reconstructing $\mathbf{p}_x(\omega)$ for $0 \leq \omega < 2\pi$ is equivalent to reconstructing $P_x(\omega)$ for $0 \leq \omega < 2\pi$.

Stacking the M^2 different cross-spectra $P_{y_i, y_j}(\omega)$ in the $M^2 \times 1$ vector $\mathbf{p}_y(\omega) = [\dots, P_{y_i, y_j}(\omega), \dots]^T$, for $i, j = 0, 1, \dots, M-1$, we finally obtain

$$\mathbf{p}_y(\omega) = \mathbf{P}_c(\omega)\mathbf{p}_x(\omega), \quad 0 \leq \omega < 2\pi \quad (33)$$

where $\mathbf{P}_c(\omega)$ is the $M^2 \times N$ matrix given by $\mathbf{P}_c(\omega) = [\dots, \mathbf{P}_{c_i, c_j}(\omega), \dots]^T$, for $i, j = 0, 1, \dots, M-1$. Assuming that $\mathbf{P}_c(\omega)$ has full column rank for $0 \leq \omega < 2\pi$, we can solve (33) for $0 \leq \omega < 2\pi$ using LS.

The above approach allows us to estimate the power spectrum at any specific frequency, and as such does not require any limits on the support of $\mathbf{r}_y[k]$ and thus on the support of $\mathbf{r}_x[n]$. However, to compute $\mathbf{p}_y(\omega)$ from $\mathbf{r}_y[k]$ in practice, the support of $\mathbf{r}_y[k]$ has to be truncated. So we could again, as before, relax the bandlimitness condition and assume that the support of $\mathbf{r}_y[k]$ and thus $\mathbf{r}_x[k]$ is limited to $-L \leq k \leq L$. In that case, to reconstruct the overall power spectrum $P_x(\omega)$, it suffices to compute $\mathbf{p}_x(\omega)$ for $\omega = 0, 2\pi \frac{1}{2L+1}, \dots, 2\pi \frac{2L}{2L+1}$, which from (33) is completely determined by $\mathbf{p}_y(\omega)$ for $\omega = 0, 2\pi \frac{1}{2L+1}, \dots, 2\pi \frac{2L}{2L+1}$. It can be shown that such an approach would be exactly equivalent to the special form of the time-domain approach presented in (20).

V. MINIMAL SPARSE RULER SAMPLING

To ensure the uniqueness of the LS solution of (13) and (20), many different implementations of the considered sampling procedure can be investigated. Although many types of random modulating waveforms can be studied [8], such as complex Gaussian sampling or random binary (from the set $\{\pm 1\}$) sampling, we mainly focus on multicoset sampling in this paper. More specifically, we propose some new multicoset implementations based on the so-called minimal sparse ruler problem, which we will label as minimal sparse ruler sampling.

Observing (2), multicoset sampling can be implemented by simply setting for every branch i , one different entry of $c_i[n]$ to one and the others to zero, i.e., $c_i[n] = 1$ if $-n = n_i$ and $c_i[n] = 0$ if $-n \neq n_i$, where $n_i \neq n_j$ whenever $i \neq j$. Concisely, $c_i[n] = \delta[-n - n_i]$, where $n_i \neq n_j, \forall i \neq j$. This is actually identical to selecting M different rows from the identity matrix \mathbf{I}_N . However, note that this row selection cannot be random, because we need to deterministically guarantee the full column rank of \mathbf{R}_c in (13) or equivalently of $\{\mathbf{Q}_c(2\pi \frac{l}{2L+1})\}_{l=0}^{2L}$ in (20). Observe that every row of \mathbf{R}_c only contains a single one, which means that the full rank conditions can be fulfilled by ensuring that \mathbf{R}_c has at least a single one in each of its columns. We can find from (5), (7) and (8) that when $\mathbf{R}_c[0]$ has a one in the column corresponding to lag $-n$, $\mathbf{R}_c[1]$ has a one in the column corresponding to lag n . As a result, if the first $\lfloor \frac{N}{2} \rfloor + 1$ columns of $\mathbf{R}_c[0]$ all have at least a single one, then also the last $\lfloor \frac{N}{2} \rfloor$ columns of $\mathbf{R}_c[1]$ all have at least a single one, where $\lfloor x \rfloor$ represents the largest integer not greater than x . Hence, a sufficient condition to guarantee that all columns of \mathbf{R}_c have at least a single one can be achieved by ensuring that the first $\lfloor \frac{N}{2} \rfloor + 1$ columns of $\mathbf{R}_c[0]$ all have at least a single one. Therefore, the problem we now like to solve is how to choose a proper combination of rows of \mathbf{I}_N to generate the coefficients of $c_i[n]$ for $i = 0, 1, \dots, M-1$, such that $\mathbf{R}_c[0]$ has at least a single one in

each of its first $\lfloor \frac{N}{2} \rfloor + 1$ columns. Further, note that the aim is to keep the number of selected rows minimal, in order to minimize the number of branches M and thus to minimize the compression rate $\frac{M}{N}$.

Since $c_i[n] = \delta[-n - n_i]$, it is obvious from (5) that

$$r_{c_i, c_j}[n] = \delta[n - n_i + n_j] \quad (34)$$

which depends on the differences $n_i - n_j$. By introducing S as a set of M indices selected from $\{0, 1, \dots, N-1\}$, representing the row indices of \mathbf{I}_N selected by the multicoset sampler, and Ω as the set of related index-differences, given by $\Omega = \{|n_i - n_j| \mid \forall n_i, n_j \in S\}$, the problem of constructing the sampler coefficients $\{c_i[n]\}_{i=0}^{M-1}$ becomes

$$\min_S |S| \quad \text{s.t.} \quad \left\{0, 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor\right\} \subset \Omega \quad (35)$$

where $|S|$ represents the cardinality of the set S . While the solution of (35) can be found by exhaustive or greedy search procedures, one possible way to find a suboptimal solution of (35) is by reformulating the problem as a so-called minimal length- $\lfloor \frac{N}{2} \rfloor$ sparse ruler problem, which has been well-studied. This is done by introducing S' as a set of M indices selected from $\{0, 1, \dots, \lfloor \frac{N}{2} \rfloor\}$ and Ω' as the set of related index-differences, given by $\Omega' = \{|n_i - n_j| \mid \forall n_i, n_j \in S'\}$, and by solving

$$\min_{S'} |S'| \quad \text{s.t.} \quad \Omega' = \left\{0, 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor\right\}. \quad (36)$$

A sparse ruler with length $\lfloor \frac{N}{2} \rfloor$ can be regarded as a ruler having $k < \lfloor \frac{N}{2} \rfloor + 1$ distance marks $0 = n_0 < n_1 < \dots < n_{k-1} = \lfloor \frac{N}{2} \rfloor$, but is still able to measure all integer distances from 0 up to $\lfloor \frac{N}{2} \rfloor$. The length- $\lfloor \frac{N}{2} \rfloor$ sparse ruler having k distance marks is called minimal if there is no length- $\lfloor \frac{N}{2} \rfloor$ sparse ruler having $k-1$ marks. The minimal sparse ruler problem has for instance been investigated in [20]. Many exact and approximate solutions for the sparse ruler problem have been precomputed and tabulated. By making the connection between the sparse ruler problem and our multicoset design problem, the sampler coefficients $\{c_i[n]\}_{i=0}^{M-1}$ can be constructed using any known sparse ruler, which guarantees the full rank property of \mathbf{R}_c and thus the uniqueness of the simple LS solution to power spectrum reconstruction.

For the alternative time-domain approach, the aim is to ensure the uniqueness of the LS solution of (28), which can be achieved if \mathbf{R}_c^a in (28) has full column rank. In the case of multicoset sampling, the full rank condition of \mathbf{R}_c^a can be achieved if each column of \mathbf{R}_c^a has at least a single one because we know that every row of \mathbf{R}_c^a will only contain a single one by considering (5), (7), (8), and (29). Following the same analysis as in the previous paragraphs, the problem of constructing $\{c_i[n]\}_{i=0}^{M-1}$ while ensuring that \mathbf{R}_c^a has at least a single one in each of its columns boils down to solving a minimal length- $(N-1)$ sparse ruler problem. For the same N , this obviously leads to a worse compression rate $\frac{M}{N}$ than for the time-domain approach. However, while the minimal length- $\lfloor \frac{N}{2} \rfloor$ sparse ruler only provides a suboptimal solution for the time-domain approach, the minimal length- $(N-1)$ sparse ruler offers the minimum possible compression rate $\frac{M}{N}$ for the alternative time-domain approach. This can easily be verified.

TABLE I
 EXAMPLES OF MINIMAL SPARSE RULERS (TD = TIME DOMAIN APPROACH, ATD = ALTERNATIVE TIME DOMAIN APPROACH)

N	length- $\frac{N}{2}$	M for length- $\frac{N}{2}$	M/N for TD	length- $(N-1)$	M for length- $(N-1)$	M/N for ATD
2	1	2	1	1	2	1
11	5	4	0.3636	10	6	0.5455
18	9	5	0.2778	17	7	0.3889
39	19	8	0.2051	38	11	0.2821
78	39	11	0.1410	77	15	0.1923
84	42	11	0.1310	83	16	0.1905
128	64	14	0.1094	127	20	0.1563

Table I shows some examples of minimal sparse ruler samplers for the time-domain (TD) and alternative time-domain (ATD) approaches.

VI. ESTIMATION AND DETECTION PERFORMANCE

In this section, we evaluate the estimation and detection performance of the proposed power spectrum estimators. We first derive the mean and the covariance of the estimated power spectrum in (21). Based on the derived mean and covariance, we then formulate the analytical normalized mean squared error (NMSE) of the estimate. To evaluate the detection performance, we assume the received sequence $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise. Based on this assumption, we simplify the earlier derived mean and covariance of the estimator and we show that asymptotically, for a sufficient number of measurements, the estimated power spectrum is Gaussian as well. Based on these results, we can finally formulate the decision threshold for a constant false alarm rate.

A. Estimation Performance

Given K measurement vectors $\mathbf{y}[k]$, the unbiased estimate for $r_{y_i, y_j}[k]$ in (6) can be written as

$$\hat{r}_{y_i, y_j}[k] = \frac{1}{K - |k|} \sum_{l=\max(0, k)}^{K-1+\min(0, k)} y_i[l] y_j^*[l - k] \quad (37)$$

where $\max(a, b)$ and $\min(a, b)$ gives the largest and smallest value of a and b , respectively.

Obviously, $E(\hat{r}_{y_i, y_j}[k]) = r_{y_i, y_j}[k]$ since $\hat{r}_{y_i, y_j}[k]$ is an unbiased estimate. Following Section III-A, we can now compute the covariance matrix of the estimate $\hat{\mathbf{r}}_y$ for \mathbf{r}_y in (11), which is given by the $(2L+1)M^2 \times (2L+1)M^2$ matrix $\mathbf{C}_{\hat{\mathbf{r}}_y} = E(\hat{\mathbf{r}}_y \hat{\mathbf{r}}_y^H) - E(\hat{\mathbf{r}}_y) E(\hat{\mathbf{r}}_y^H)$. The elements of $\mathbf{C}_{\hat{\mathbf{r}}_y}$ are given by

$$\begin{aligned} & \text{Cov}(\hat{r}_{y_i, y_j}[k], \hat{r}_{y_w, y_v}[q]) \\ &= E(\hat{r}_{y_i, y_j}[k] \hat{r}_{y_w, y_v}^*[q]) - E(\hat{r}_{y_i, y_j}[k]) E(\hat{r}_{y_w, y_v}^*[q]) \\ &= E(\hat{r}_{y_i, y_j}[k] \hat{r}_{y_w, y_v}^*[q]) - r_{y_i, y_j}[k] r_{y_w, y_v}^*[q] \end{aligned} \quad (38)$$

where $0 \leq i, j, w, v \leq M-1$, $-L \leq k, q \leq L$, and $E(\hat{r}_{y_i, y_j}[k] \hat{r}_{y_w, y_v}^*[q])$ can be expressed as

$$\begin{aligned} & E(\hat{r}_{y_i, y_j}[k] \hat{r}_{y_w, y_v}^*[q]) \\ &= E\left(\left(\frac{1}{K - |k|} \sum_{l=\max(0, k)}^{K-1+\min(0, k)} y_i[l] y_j^*[l - k]\right)\right) \end{aligned}$$

$$\begin{aligned} & \times \left(\frac{1}{K - |q|} \sum_{p=\max(0, q)}^{K-1+\min(0, q)} y_w^*[p] y_v[p - q] \right) \\ &= \frac{1}{(K - |k|)(K - |q|)} \sum_{l=\max(0, k)}^{K-1+\min(0, k)} \sum_{p=\max(0, q)}^{K-1+\min(0, q)} \\ & E(y_i[l] y_j^*[l - k] y_w^*[p] y_v[p - q]) \end{aligned} \quad (39)$$

Observe that the computation of $\mathbf{C}_{\hat{\mathbf{r}}_y}$ is not trivial since it involves the computation of fourth order moments. Knowledge about the distribution of the received sequence $x[n]$ is thus required. For example, if $x[n]$ is Gaussian distributed, then the sequences $y_i[k]$ are also jointly Gaussian and the fourth order moments in (39) can be simplified as the sum of products of second order moments [22].

Based on the statistical properties of $\hat{\mathbf{r}}_y$, we can now compute the expected value and the covariance of the recovered power spectrum. Denote the estimated power spectrum for \mathbf{s}_x in (21) by $\hat{\mathbf{s}}_x$. From (13) and (21), assuming that \mathbf{R}_c has full column rank, the relationship between $\hat{\mathbf{s}}_x$ and $\hat{\mathbf{r}}_y$ is given by

$$\hat{\mathbf{s}}_x = \mathbf{F}_{(2L+1)N} \mathbf{R}_c^\dagger \hat{\mathbf{r}}_y \quad (40)$$

where $\mathbf{R}_c^\dagger = (\mathbf{R}_c^H \mathbf{R}_c)^{-1} \mathbf{R}_c^H$. The expected value of $\hat{\mathbf{s}}_x$ is thus given by

$$E(\hat{\mathbf{s}}_x) = \mathbf{F}_{(2L+1)N} \mathbf{R}_c^\dagger E(\hat{\mathbf{r}}_y) = \mathbf{F}_{(2L+1)N} \mathbf{R}_c^\dagger \mathbf{r}_y \quad (41)$$

Correspondingly, we denote the $(2L+1)N \times (2L+1)N$ covariance matrix of $\hat{\mathbf{s}}_x$ by $\mathbf{C}_{\hat{\mathbf{s}}_x}$, which can be written as

$$\mathbf{C}_{\hat{\mathbf{s}}_x} = \mathbf{F}_{(2L+1)N} \mathbf{R}_c^\dagger \mathbf{C}_{\hat{\mathbf{r}}_y} (\mathbf{R}_c^\dagger)^H \mathbf{F}_{(2L+1)N}^H \quad (42)$$

where the elements of $\mathbf{C}_{\hat{\mathbf{r}}_y}$ are given by (38). Note that the variance of the elements of $\hat{\mathbf{s}}_x$ can be found on the diagonal of $\mathbf{C}_{\hat{\mathbf{s}}_x}$.

It is well known that the NMSE of the estimated power spectrum $\hat{\mathbf{s}}_x$ is then given by

$$\text{NMSE} = \frac{E(\|\hat{\mathbf{s}}_x - \mathbf{s}_x\|_2^2)}{\|\mathbf{s}_x\|_2^2} = \frac{\text{tr}(\mathbf{C}_{\hat{\mathbf{s}}_x}) + \|E(\hat{\mathbf{s}}_x) - \mathbf{s}_x\|_2^2}{\|\mathbf{s}_x\|_2^2} \quad (43)$$

where $\text{tr}(\cdot)$ is the trace operator. Since $\hat{\mathbf{s}}_x$ is a linear function of $\hat{\mathbf{r}}_y$ [see (40)] and $\hat{\mathbf{r}}_y$ is an unbiased estimate of \mathbf{r}_y [see (37)], $\hat{\mathbf{s}}_x$ is an unbiased estimate of \mathbf{s}_x as long as the support of the autocorrelation of the received signal $r_x[n]$ is limited to $-LN \leq n \leq LN$, which is the assumption we adopt in (12). When this is the case, the NMSE of $\hat{\mathbf{s}}_x$ is equal to $\frac{\text{tr}(\mathbf{C}_{\hat{\mathbf{s}}_x})}{\|\mathbf{s}_x\|_2^2}$.

Note that in (37), we have used an unbiased estimate $\hat{r}_{y_i, y_j}[k]$ of the cross-correlation $r_{y_i, y_j}[k]$ instead of a biased one. The

reason for this can be explained as follows. Using an unbiased estimate instead of a biased one, the realness of the resulting power spectrum estimate \hat{s}_x is not jeopardized, but the positiveness of \hat{s}_x can be affected. However, even when using a biased estimate, i.e., using a normalization factor $\frac{1}{K}$ instead of $\frac{1}{(K-|k|)}$ in (37), the positiveness of \hat{s}_x can generally not be guaranteed, in contrast to the Nyquist-rate sampling case. This is due to the fact that the pseudo-inverse of \mathbf{R}_c is used in (40), which does not introduce an additional bias when transforming $\hat{\mathbf{r}}_y$ into \hat{s}_x . That is the main reason why we started from an unbiased estimate from the beginning. We will come back to this issue in Section VII.

B. Constant False Alarm Rate (CFAR) Detection Performance

For the detection performance evaluation, let us assume that the received sequence $x[n]$ in Fig. 2 only contains circular complex zero-mean Gaussian i.i.d. noise with variance σ^2 , i.e., $E(x[n]x^*[m]) = \sigma^2\delta[n-m]$ and $E(x[n]x[m]) = 0$ for all m and n . When this is the case, $E(\hat{r}_{y_i, y_j}[k])$ can be simplified to

$$E(\hat{r}_{y_i, y_j}[k]) = r_{y_i, y_j}[k] = r_{c_i, c_j}[0]\sigma^2\delta[k]. \quad (44)$$

The elements of $\mathbf{C}_{\hat{\mathbf{r}}_y}$ in (42) can then also be simplified according to Appendix A and they are given by

$$\text{Cov}(\hat{r}_{y_i, y_j}[k], \hat{r}_{y_w, y_v}[q]) = \frac{\sigma^4 r_{c_i, c_w}[0] r_{c_j, c_v}^*[0] \delta[q-k]}{(K-|k|)}. \quad (45)$$

The expected value of \hat{s}_x is then given by (41) and the covariance matrix of \hat{s}_x by (42) where the elements of $E(\hat{\mathbf{r}}_y)$ are given by (44) and those of $\mathbf{C}_{\hat{\mathbf{r}}_y}$ by (45).

Asymptotically, when the number of measurement vectors K is sufficiently large compared to L in (11), i.e., $L \ll K$, it can be shown that the Gaussian approximation is applicable for the distribution of each element of \hat{s}_x , when $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise. This evaluation on the asymptotic statistical distribution of \hat{s}_x is provided in Appendix B.

Combining (42), (45), and the analysis in Appendix B, we are now ready to evaluate the detection problem at frequencies $\omega = 0, 2\pi\frac{1}{(2L+1)N}, \dots, 2\pi\frac{(2L+1)N-1}{(2L+1)N}$ under the asymptotic Gaussian behavior of $\hat{P}_x(\omega)$. Note that it is always possible to increase the number of grid points in the frequency domain by simply padding zeros to $\hat{\mathbf{r}}_x$, and thus it is always possible to evaluate the detection problem at frequencies other than the above specified frequencies. The detection problem is modeled as a selection between hypothesis $H_{1,\omega}$, which represents the occupancy of frequency ω , and hypothesis $H_{0,\omega}$, which represents the absence of a signal at frequency ω . Here, we want to maximize the detection probability given the false alarm probability and thus we adopt the Neyman-Pearson theorem [23]. The decision rule is given by

$$\hat{P}_x(\omega) \underset{H_{0,\omega}}{\overset{H_{1,\omega}}{\geq}} \gamma_\omega$$

where γ_ω is the decision threshold for frequency ω . Let us denote the mean and variance of the noise power at frequency ω as $\mu_{n,\omega}$ and $\sigma_{n,\omega}^2$, respectively. Note that the value of $\mu_{n,\omega}$ and

$\sigma_{n,\omega}^2$ for $\omega = 0, 2\pi\frac{1}{(2L+1)N}, \dots, 2\pi\frac{(2L+1)N-1}{(2L+1)N}$ can be found as the elements of $E(\hat{s}_x)$ in (41) and those of the diagonal of $\mathbf{C}_{\hat{s}_x}$ in (42) where the elements of $E(\hat{\mathbf{r}}_y)$ in (41) and $\mathbf{C}_{\hat{\mathbf{r}}_y}$ in (42) are, respectively, given by (44) and (45). Due to the asymptotic Gaussian behavior of $\hat{P}_x(\omega)$, the false alarm probability at frequency ω , $p_{fa,\omega}$, given the threshold value γ_ω , is given by

$$p_{fa,\omega} = Q\left(\frac{\gamma_\omega - \mu_{n,\omega}}{\sigma_{n,\omega}}\right)$$

where $Q(\cdot)$ is the tail probability of the standard Gaussian distribution

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du.$$

As a result, for a given false alarm probability $p_{fa,\omega}$, the decision threshold at frequency ω can be computed as

$$\gamma_\omega = Q^{-1}(p_{fa,\omega})\sigma_{n,\omega} + \mu_{n,\omega}. \quad (46)$$

C. Alternative Time-Domain Approach Case

In this subsection, we again evaluate the estimation and detection performance of the proposed power spectrum estimators but now for the alternative time-domain approach of Section III-B. Note that this approach is not a special case of the time-domain approach and requires a separate analysis. We start the analysis on the estimation performance by considering $\hat{\mathbf{r}}_y[0]$ as an unbiased estimate of $\mathbf{r}_y[0]$ in (26). Note that the elements of $\hat{\mathbf{r}}_y[0]$ are simply given by (37) with $k = 0$ and it is evident that $E(\hat{\mathbf{r}}_y[0]) = \mathbf{r}_y[0]$. Next, we derive the covariance matrix of $\hat{\mathbf{r}}_y[0]$, which is given by the $M^2 \times M^2$ matrix $\mathbf{C}_{\hat{\mathbf{r}}_y[0]} = E(\hat{\mathbf{r}}_y[0]\hat{\mathbf{r}}_y[0]^H) - E(\hat{\mathbf{r}}_y[0])E(\hat{\mathbf{r}}_y[0]^H)$ whose elements are simply given by (38) with $k = q = 0$.

Based on the statistical properties of $\hat{\mathbf{r}}_y[0]$, we can compute the expected value and the covariance of $\hat{\mathbf{r}}_x^a$, which is the estimate of \mathbf{r}_x^a in (28). Assuming that \mathbf{R}_c^a in (28) has full column rank, the expected value of $\hat{\mathbf{r}}_x^a$ can be written as

$$E(\hat{\mathbf{r}}_x^a) = \mathbf{R}_c^{a\dagger} E(\hat{\mathbf{r}}_y[0]) = \mathbf{R}_c^{a\dagger} \mathbf{r}_y[0]. \quad (47)$$

Correspondingly, we denote the $(2N-1) \times (2N-1)$ covariance matrix of $\hat{\mathbf{r}}_x^a$ by $\mathbf{C}_{\hat{\mathbf{r}}_x^a}$, which can be expressed as

$$\mathbf{C}_{\hat{\mathbf{r}}_x^a} = \mathbf{R}_c^{a\dagger} \mathbf{C}_{\hat{\mathbf{r}}_y[0]} (\mathbf{R}_c^{a\dagger})^H \quad (48)$$

where $\mathbf{R}_c^{a\dagger} = (\mathbf{R}_c^{aH} \mathbf{R}_c^a)^{-1} \mathbf{R}_c^{aH}$. Let us denote the zero padded version of $\hat{\mathbf{r}}_x^a$ by $\hat{\mathbf{r}}_x^a$ and replace $E(\hat{\mathbf{r}}_x^a)$ and $\mathbf{C}_{\hat{\mathbf{r}}_x^a}$ by the $(2L+1)N \times 1$ vector $E(\hat{\mathbf{r}}_x^a)$ and the $(2L+1)N \times (2L+1)N$ matrix $\mathbf{C}_{\hat{\mathbf{r}}_x^a}$. The mean and the covariance matrix of \hat{s}_x are then given by

$$E(\hat{s}_x) = \mathbf{F}_{(2L+1)N} E(\hat{\mathbf{r}}_x^a) \quad (49)$$

and

$$\mathbf{C}_{\hat{s}_x} = \mathbf{F}_{(2L+1)N} \mathbf{C}_{\hat{\mathbf{r}}_x^a} \mathbf{F}_{(2L+1)N}^H \quad (50)$$

respectively. Similar to the time-domain approach case, the NMSE of \hat{s}_x depends on its variance and bias. Since \hat{s}_x is a linear function of $\hat{\mathbf{r}}_y[0]$, which is an unbiased estimate of $\mathbf{r}_y[0]$, \hat{s}_x is also an unbiased estimate of \mathbf{s}_x as long as the support of the autocorrelation of the received signal $r_x[n]$ is limited to

$-N + 1 \leq n \leq N - 1$. When this is the case, the NMSE of $\hat{\mathbf{s}}_x$ is again equal to $\frac{\text{tr}(\mathbf{C}_{\hat{\mathbf{s}}_x})}{\|\mathbf{s}_x\|_2^2}$.

For the detection performance evaluation, we again assume that the received sequence $x[n]$ in Fig. 2 only contains circular complex zero-mean Gaussian i.i.d. noise. When this is the case, the elements of $E(\hat{\mathbf{r}}_y[0])$ in (47) are given by (44) with $k = 0$, and the elements of $\mathbf{C}_{\hat{\mathbf{r}}_y[0]}$ in (48) are given by (45) with $k = q = 0$. Similar to the time-domain approach case, the Gaussian approximation is also applicable for the distribution of each element of $\hat{\mathbf{s}}_x$ under the alternative time-domain approach when $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise and as long as K is sufficiently large. This is shown in Appendix C. The derivation of the detection threshold given a fixed false alarm probability for this alternative time-domain approach case follows Section VI-B.

VII. ADDITIONAL CONSTRAINTS

So far, we have assumed that the power spectrum \mathbf{s}_x can be estimated without any additional constraints relying on the assumption that there are enough equations available. As we explained in Section VI-A, this estimated spectrum is real but not necessarily positive. Hence, we could think about adding a positivity constraint to our reconstruction problem. To do so, let us first combine (16), (18), and (21) to produce

$$\mathbf{q}_y = \Phi \mathbf{s}_x \quad (51)$$

where $\Phi = \mathbf{Q}_c(\mathbf{F}_{2L+1} \otimes \mathbf{I}_N) \mathbf{F}_{(2L+1)N}^{-1}$ is of size $(2L+1)M^2 \times (2L+1)N$. The power spectrum estimate is then given by the solution of the following positivity-constrained least-squares problem:

$$\min_{\mathbf{s}_x} \|\hat{\mathbf{q}}_y - \Phi \mathbf{s}_x\|_2^2 \text{ s.t. } \mathbf{s}_x \geq \mathbf{0} \quad (52)$$

where $\mathbf{0}$ is the $(2L+1)N \times 1$ vector containing only zeros and \geq is a component-wise inequality.

Similarly, if we know that the power spectrum is sparse, we could think about adding a sparsity constraint to our reconstruction problem. The power spectrum estimate is then given by the solution of the following sparsity-constrained least-squares problem:

$$\min_{\mathbf{s}_x} \|\hat{\mathbf{q}}_y - \Phi \mathbf{s}_x\|_2^2 + \lambda \|\mathbf{s}_x\|_1 \quad (53)$$

where the weight $\lambda \geq 0$ balances the sparsity-bias tradeoff. And naturally, it is also possible to combine the positivity constraint with the sparsity constraint.

While this positivity and/or sparsity constraint might lead to more accurate power spectrum estimates, they also allow to solve the underdetermined case, i.e., when $M^2 < N$, and as such allow to further reduce the sampling rate requirements. However, we decided not to include these constraints from the beginning, because of two reasons. First of all, the constrained least-squares problems are harder to solve than the unconstrained one, and thus lead to a higher computational complexity. And second, the unconstrained solution allows for an analytical performance evaluation (as carried out in Section VI), while this is not trivial for the constrained solutions, since we lose the linear relationship between the

estimated power spectrum $\hat{\mathbf{s}}_x$ and the cross-correlation vector $\hat{\mathbf{r}}_y$.

Finally, note that a positive solution can always be obtained from the unconstrained solution, by simply setting the negative entries to zero. In that case, the NMSE derived in Section VI-A can be viewed as an upper bound on the true NMSE while the optimal detection threshold (as well as the related detection and false alarm rate) at a specific frequency derived in Section VI-B will not change at all.

VIII. SIMULATION RESULTS

In this section, we present some simulation results describing the effectiveness of our proposed methods. In the first part, we examine the estimation performance of both the time-domain and alternative time-domain reconstruction approaches while the detection performance of both approaches is presented in the second part. We consider minimal sparse ruler sampling and complex Gaussian sampling, where the latter is merely considered to show that the proposed techniques also work for other samplers than multicoset samplers. In this section, minimal sparse ruler sampling refers to the multicoset sampling technique for which we define the sampler coefficients $c_i[n]$ by selecting the rows of an $N \times N$ identity matrix according to Section V, and for which we acquire larger compression rates by randomly adding extra rows of the identity matrix to the already selected rows.

A. Estimation Performance

First of all, we evaluate the performance of our time-domain approach presented in Section III-A. We consider a complex baseband signal $x[n]$ spanning the frequency bands $[-0.9\pi, -0.65\pi]$, $[0.1\pi, 0.35\pi]$, and $[0.55\pi, 0.8\pi]$. To generate this signal, we pass circular complex zero-mean Gaussian i.i.d. noise with variance σ^2 through a digital filter $h[n]$ of length $LN + 1$. As a result, the support of the true autocorrelation sequence $r_x[n]$ is limited to $-LN \leq n \leq LN$, as required by our theory, and it is given by

$$r_x[n] = h[n] \star h^*[-n] \star \sigma^2 \delta[n]. \quad (54)$$

In this subsection, we take $L = 2$ and $N = 84$, and we vary the compression rate $\frac{M}{N}$. The motivation to fix N at $N = 84$ is computational complexity since a higher N will generally result in a higher complexity.

We now examine the proposed minimal sparse ruler sampling discussed in Section V and complex Gaussian sampling. Both estimates are computed using LS. In the first method, the coefficients of $c_i[n]$ for $i = 0, 1, \dots, M - 1$ are generated according to the length-42 minimal sparse ruler having $M = 11$ distance marks. This is equivalent to selecting the corresponding $M = 11$ rows from the first 43 rows of the identity matrix \mathbf{I}_{84} leading to $M = 11$ branches in our sampling device. As a result, we have matrices $\mathbf{R}_c[0]$ and $\mathbf{R}_c[1]$ of size 121×84 in (10). We then implement the larger $\frac{M}{N}$ cases by randomly adding additional rows of \mathbf{I}_{84} to the already selected 11 rows. In the complex Gaussian sampling case, we simply vary the compression rate $\frac{M}{N}$ from $\frac{11}{84}$ to 0.5. The coefficients of $c_i[n]$ for $i = 0, 1, \dots, M - 1$ are randomly generated according to a circular complex Gaussian distribution with zero mean and variance $\frac{1}{M}$. Note that we keep these coefficients fixed over the

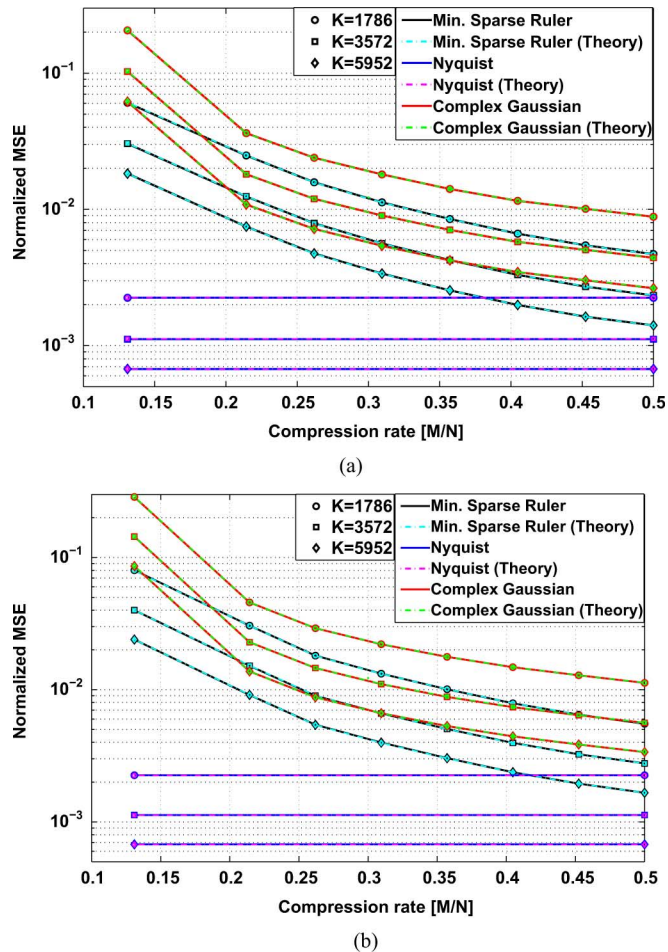


Fig. 3. The normalized MSE between the estimated power spectrum (minimal sparse ruler and complex Gaussian sampling) and the true one for various numbers of MVs (K). (a) Noise-free. (b) Noisy (SNR = 10 dB in active bands).

different simulation runs. In Fig. 3, the NMSE between the estimated power spectrum \hat{s}_x and the true one s_x is calculated for both the minimal sparse ruler sampling and complex Gaussian sampling. While no noise is considered in Fig. 3(a), random circular complex zero-mean Gaussian i.i.d. noise is introduced in Fig. 3(b), in such a way that the signal-to-noise ratio (SNR) in the active bands is given by 10 dB. Both the simulated and analytical NMSE are calculated for a varying number of measurement vectors (MVs) K (which is analogous to KN Nyquist rate samples) as an attempt to represent different sensing times. We also provide the simulated and analytical NMSE between the estimated power spectrum produced by Nyquist rate sampling and the true one for different sensing times as a benchmark. Here, the Nyquist rate estimate is obtained from the proposed multiset approach in Section V by setting $M = N$. From the figures, it is obvious that the quality of the estimation improves with $\frac{M}{N}$ and it slowly converges towards that of the Nyquist rate. We can also see how the NMSE improves as the sensing time increases, which is to be expected. Observe that the proposed minimal sparse ruler sampling generally performs better than complex Gaussian sampling. If we compare the structure of \mathbf{R}_c in (14) for complex Gaussian sampling with the one for minimal sparse ruler sampling, we can easily see that, for the minimal sparse ruler case, the columns of \mathbf{R}_c are not only independent but also orthogonal. As a result, the condition number

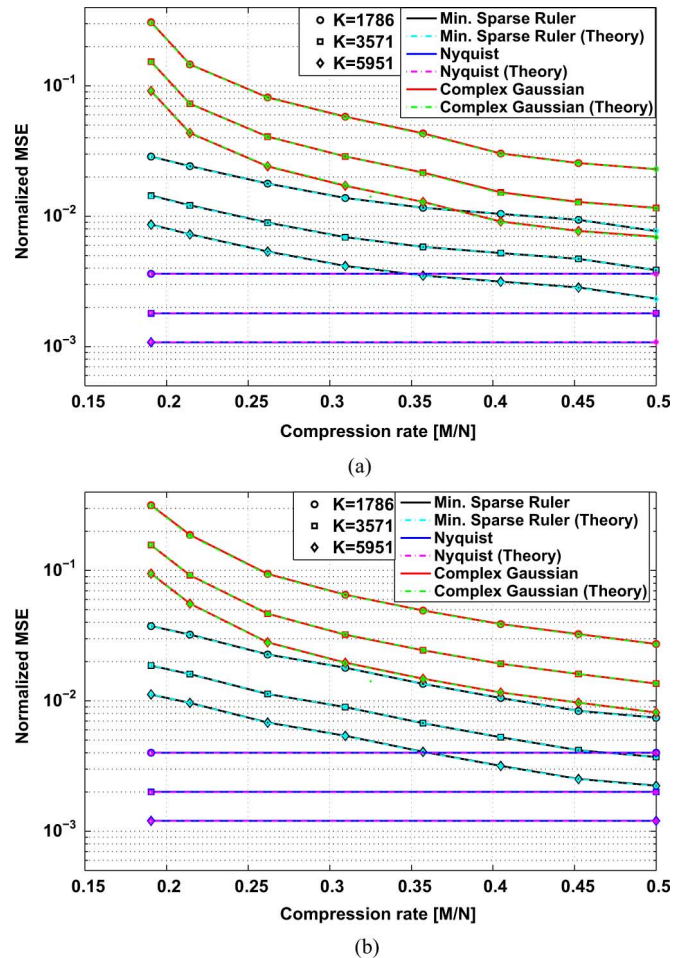


Fig. 4. The normalized MSE between the estimated power spectrum (minimal sparse ruler and complex Gaussian sampling) based on the alternative time-domain approach and the true one for various numbers of MVs (K). (a) Noise-free. (b) Noisy (SNR = 10 dB in active bands).

of \mathbf{R}_c for minimal sparse ruler sampling is smaller than the one for complex Gaussian sampling. This issue might explain why the performance of minimal sparse ruler sampling is better than that of complex Gaussian sampling. Note also how the simulated NMSE is on top of the analytical NMSE for the minimal sparse ruler, complex Gaussian, and Nyquist rate sampling.

Next, we consider the performance of the alternative time-domain approach in Section III-B for both minimal sparse ruler and complex Gaussian sampling. The signal model that is used here has the same characteristics as the one used in the time-domain approach except for the fact that the filter used to generate the signal now has length N . Hence, the support of the true autocorrelation sequence $r_x[n]$ is now limited to $-N + 1 \leq n \leq N - 1$, as required by our theory. We again select $N = 84$. For the minimal sparse ruler sampling case, we first generate the coefficients $c_i[n]$ for $i = 0, 1, \dots, M - 1$ according to the length-83 minimal sparse ruler leading to $M = 16$ branches in our sampling device. This will result in a matrix \mathbf{R}_c^c of size 256×167 in (28). We again implement the larger $\frac{M}{N}$ cases by randomly adding additional rows of \mathbf{I}_{84} to the already selected 16 rows. For complex Gaussian sampling, we simply vary the compression rate $\frac{M}{N}$ from $\frac{16}{84}$ to 0.5. Note however that complex Gaussian sampling is theoretically able to offer a smaller compression rate than the minimal sparse

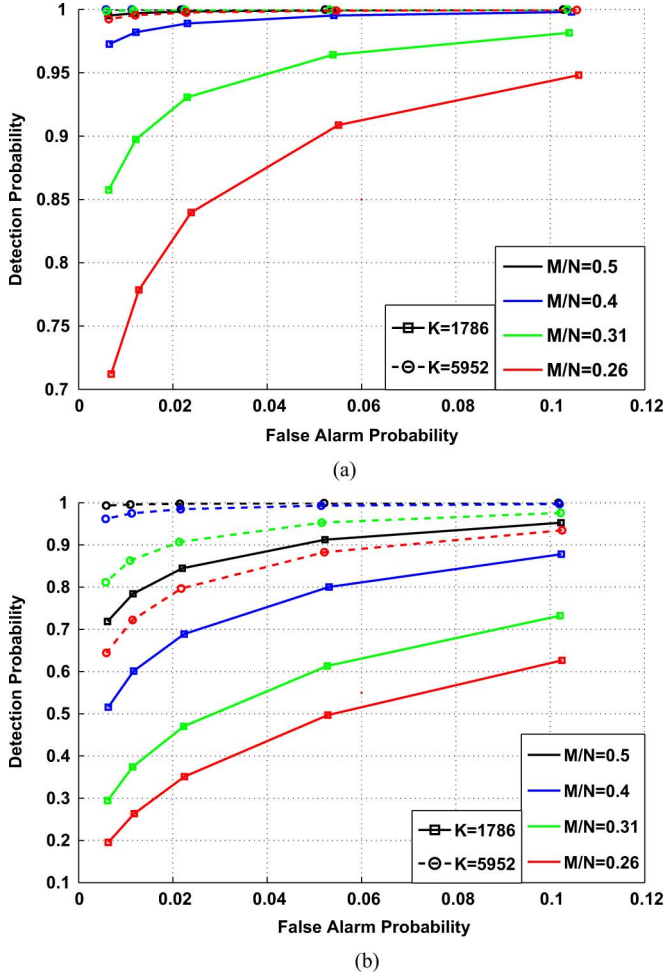


Fig. 5. The detection performance of the proposed time-domain approach (minimal sparse ruler sampling) for various numbers of MVs (K) and $\frac{M}{N}$. (a) $\text{SNR} = -2$ dB in active band. (b) $\text{SNR} = -5$ dB in active band.

ruler counterpart while maintaining the full column rank property of \mathbf{R}_c^a in (28) although it might not result in an acceptable performance. Similar to the time-domain reconstruction approach, we randomly generate the coefficients $c_i[n]$ according to a circular complex Gaussian distribution with zero mean and variance $\frac{1}{M}$ and keep these coefficients fixed over the different simulation runs. Both the simulated and analytical NMSE between the estimated power spectrum and the true one for the alternative time-domain approach are depicted in Fig. 4. The Nyquist rate based estimates in Fig. 4 are obtained from the multicaset implementation of the alternative time-domain approach in Section III-B with $M = N$. In general, we find similar trends as the ones observed for the time-domain reconstruction approach (Fig. 3) with respect to the impact of the compression rate and sensing time, as well as the relative performance between the minimal sparse ruler and complex Gaussian sampling.

B. Detection Performance

In this subsection, we consider a complex baseband signal spanning a frequency band between 0.15π and 0.25π . Again, this signal is generated by passing circular complex zero-mean Gaussian i.i.d. noise through a digital filter $h[n]$ of length $LN+1$ where L and N are set to $L = 2$ and $N = 84$, respectively.

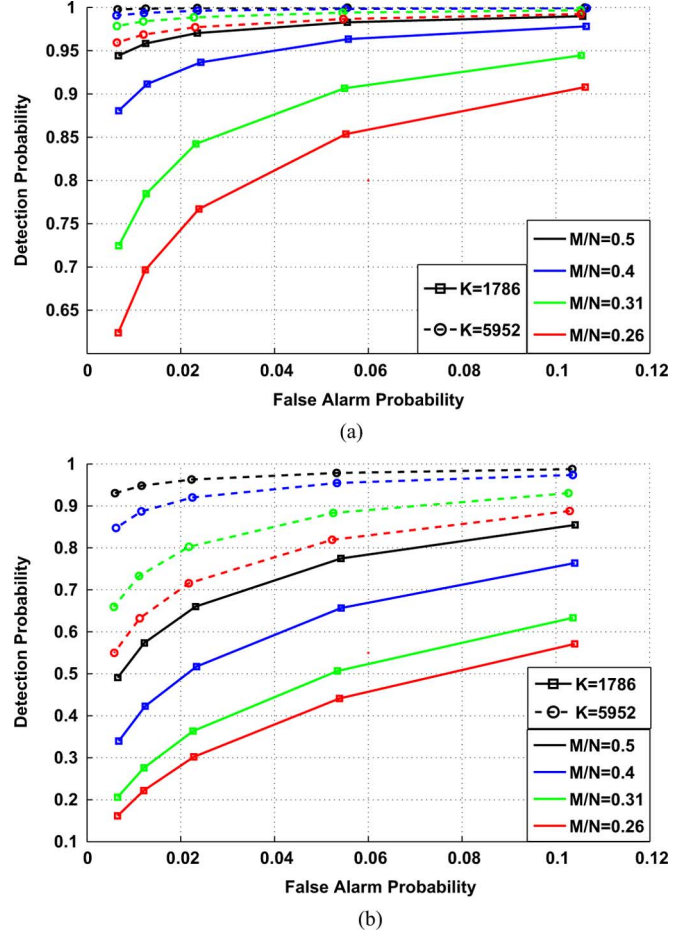
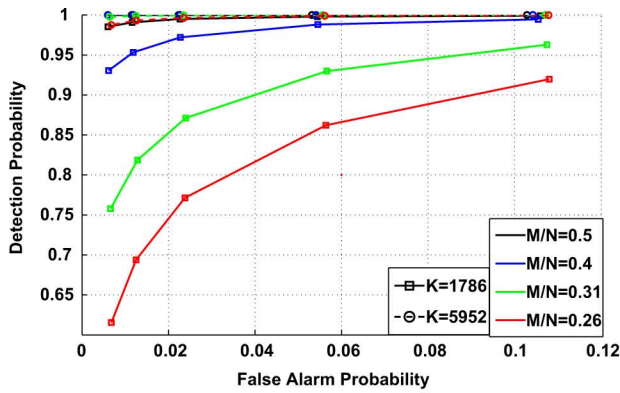


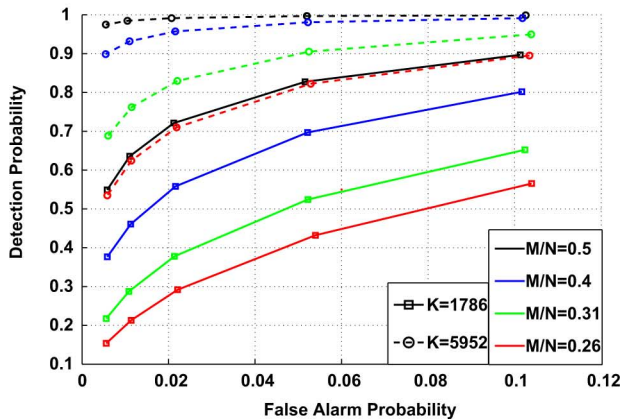
Fig. 6. The detection performance of the proposed time-domain approach (complex Gaussian sampling) for various numbers of MVs (K) and $\frac{M}{N}$. (a) $\text{SNR} = -2$ dB in active band. (b) $\text{SNR} = -5$ dB in active band.

On top of this spectrum, we add circular complex zero-mean Gaussian i.i.d. noise such that a specific SNR is obtained in the active band. In order to simplify the analysis in the simulation study, the same filter $h[n]$ is used for both the time-domain and alternative time-domain approaches. This certainly results in a bias for the estimate \hat{s}_x under the alternative time-domain approach. The detection probability should be evaluated in the occupied band. Note however that the active band is not perfectly rectangular and there are two transition bands around the edges of the occupied band. Therefore, we decide to leave small guard bands around the two edges of the active band and evaluate the detection performance at the points in the band from 0.154π to 0.243π . Meanwhile, the false alarm probability is based on a band that is significantly far from the occupied band, namely from -0.92π to -0.68π . The SNR in the active band is varied from -2 dB to -5 dB while the compression rate is varied between $\frac{M}{N} = 0.26$ and $\frac{M}{N} = 0.5$. The false alarm probability is set by determining the detection threshold at each frequency, which is computed according to (46). For simulation purposes, we try to vary the false alarm probability between $p_{fa} = 0.005$ and $p_{fa} = 0.1$, which is smaller than the value suggested by the IEEE 802.22 standard [24].

We first consider the detection performance of the time-domain approach for both the minimal sparse ruler and complex Gaussian sampling implementation, as shown in Figs. 5 and 6, respectively. Here, we calculate the false alarm and detection



(a)

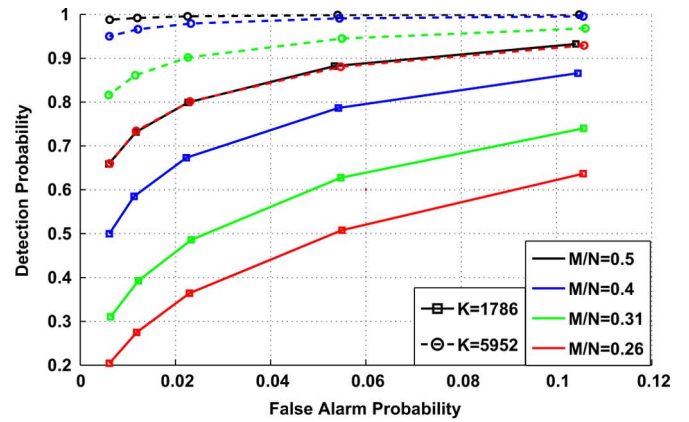


(b)

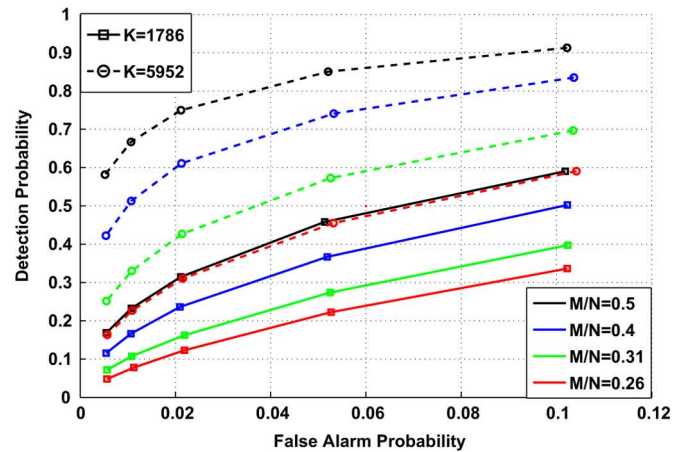
Fig. 7. The detection performance of the proposed alternative time-domain approach (minimal sparse ruler sampling) for various numbers of MVs (K) and $\frac{M}{N}$. (a) SNR = -2 dB in active band. (b) SNR = -5 dB in active band.

occurrences at each frequency point separately based on the estimated power at that point. For a given band, we then combine the amount of false alarm and detection occurrences at all frequency points within that band to calculate the false alarm and detection probability. The sampler coefficients for the minimal sparse ruler and complex Gaussian sampling are generated in the same way as in Section VIII-A. Two different numbers of measurement vectors K are simulated, i.e., $K = 1786$ and $K = 5952$. As it is obvious from the figures, minimal sparse ruler sampling generally has a better performance than complex Gaussian sampling. This somehow confirms the relative estimation performance between both approaches in the previous subsection. Both methods, however, have a good performance for all simulated $\frac{M}{N}$ when $K = 5952$ and SNR = -2 dB. For SNR = -5 dB, the performance of minimal sparse ruler sampling still reaches an acceptable level as long as $K = 5952$ and $\frac{M}{N} \geq 0.4$.

Next, we evaluate the performance of the alternative time-domain approach for the two sampling techniques, depicted in Figs. 7 and 8. Again, the sampler coefficients for the minimal sparse ruler and complex Gaussian sampling are generated in the same way as in Section VIII-A. We observe that the alternative time-domain approach suffers from a performance degradation compared to its time-domain approach counterpart and a poorer performance is found for the complex Gaussian sampling case. For $K = 5952$ and SNR = -2 dB, however, the performance of the minimal sparse ruler sampling is still acceptable.



(a)



(b)

Fig. 8. The detection performance of the proposed alternative time-domain approach (complex Gaussian sampling) for various numbers of MVs (K) and $\frac{M}{N}$. (a) SNR = -2 dB in active band. (b) SNR = -5 dB in active band.

IX. CONCLUSION

In this paper, we have developed a new approach for power spectrum estimation of wide-sense stationary signals based on samples produced by a sub-Nyquist sampling device. No sparsity constraints are required for this method. In general, the solution can be derived by solving simple LS problems, which are solvable as long as the rank condition of the corresponding system equations are satisfied. We have focused on multicoset sampling where we cast the design of the sampling device as a minimal sparse ruler problem. We have shown that any sparse ruler can produce a multicoset sampling design that ensures the full rank condition of the formulated sampling problem, and thereby guarantees the uniqueness of the power spectrum estimate as the solution to a set of simple LS problems. Moreover, when minimal sparse rulers are employed, the resulting samplers approach the minimum sampling rate resulting in a strong compression. Finally, we have derived the mean and the variance of the estimated power spectrum. Based on these results, we are able to derive the analytical MSE of our power spectrum estimates. Moreover, by assuming that the received signal only contains circular complex zero-mean Gaussian i.i.d. noise, we are able to derive the detection threshold which is advantageous when we intend to adopt the Neyman-Pearson theorem for evaluating the detection performance. The simulation study shows that the performance of our proposed approach is quite

acceptable in terms of both estimation and detection, therefore making it a promising candidate for power spectrum estimation and sensing of wide band signals. Cognitive radio is for instance one possible interesting application.

APPENDIX A
DERIVATION OF $\text{Cov}(\hat{r}_{y_i, y_j}[k], \hat{r}_{y_w, y_v}[q])$ FOR
CFAR DETECTION PERFORMANCE EVALUATION
(TIME-DOMAIN APPROACH)

We start by considering $\text{Cov}(\hat{r}_{y_i, y_j}[k], \hat{r}_{y_w, y_v}[q])$ in (38) and $E(\hat{r}_{y_i, y_j}[k]\hat{r}_{y_w, y_v}^*[q])$ in (39), which are valid for a general received signal $x[n]$. When the received signal $x[n]$ has a Gaussian distribution, $y_i[k]$ also has a Gaussian distribution. We can then adopt the following result for Gaussian random variables, which is proven in [21], [22]: If x_1, x_2, x_3 , and x_4 are jointly complex or real Gaussian random variables then $E(x_1x_2x_3x_4) = E(x_1x_2)E(x_3x_4) + E(x_1x_3)E(x_2x_4) + E(x_1x_4)E(x_2x_3) - 2E(x_1)E(x_2)E(x_3)E(x_4)$. This allows us to rewrite $E(\hat{r}_{y_i, y_j}[k]\hat{r}_{y_w, y_v}^*[q])$ in (39) as (55), shown at the bottom of the page.

For CFAR detection performance evaluation, we assume that the received signal $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise. When this is the case, the third and last terms in (55) are zero since $E(y_i[l]) = \sum_{n=1-N}^0 c_i[n]E(x[lN-n]) = 0$, and

$$E(y_i[l]y_v[p-q]) = \sum_{n=1-N}^0 \sum_{m=1-N}^0 c_i[n]c_v[m] \times E(x[lN-n]x[pN-qN-m]) = 0.$$

As a result, (55) can be rewritten as

$$E(\hat{r}_{y_i, y_j}[k]\hat{r}_{y_w, y_v}^*[q]) = r_{y_i, y_j}[k]r_{y_w, y_v}^*[q] + \frac{1}{(K-|k|)(K-|q|)} \times \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} \left\{ r_{y_i, y_w}[l-p]r_{y_j, y_v}^*[l+q-k-p] \right\}. \quad (56)$$

By inserting (56) into (38), $\text{Cov}(\hat{r}_{y_i, y_j}[k], \hat{r}_{y_w, y_v}[q])$ can be expressed as

$$\text{Cov}(\hat{r}_{y_i, y_j}[k], \hat{r}_{y_w, y_v}[q]) = \frac{1}{(K-|k|)(K-|q|)} \times \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} \left\{ r_{y_i, y_w}[l-p]r_{y_j, y_v}^*[l+q-k-p] \right\}. \quad (57)$$

If we then consider (44), we can observe in (57) that only the terms with $l=p$ and $k=q$ have a nonzero value. Hence, we have

$$\text{Cov}(\hat{r}_{y_i, y_j}[k], \hat{r}_{y_w, y_v}[q]) = \frac{r_{y_i, y_w}[0]r_{y_j, y_v}^*[0]\delta[q-k]}{(K-|k|)} = \frac{\sigma^4 r_{c_i, c_w}[0]r_{c_j, c_v}^*[0]\delta[q-k]}{(K-|k|)} \quad (58)$$

which is the result provided in (45).

APPENDIX B
EVALUATION OF THE STATISTICAL DISTRIBUTION OF
 $\hat{\mathbf{s}}_x$ FOR CIRCULAR COMPLEX ZERO-MEAN GAUSSIAN
I.I.D. NOISE $x[n]$ (TIME-DOMAIN APPROACH)

In order to evaluate the statistical distribution of $\hat{\mathbf{s}}_x$ for circular complex zero-mean Gaussian i.i.d. noise $x[n]$, we first rewrite (37) as

$$\hat{r}_{y_i, y_j}[k] = \frac{1}{K-|k|} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} y_i[l]y_j^*[l-k] = \frac{1}{K-|k|} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \left\{ \sum_{n=1-N}^0 c_i[n]x[lN-n] \right\} \left\{ \sum_{m=1-N}^0 c_j^*[m]x^*[lN-kN-m] \right\} = \frac{1}{K-|k|} \sum_{n=1-N}^0 \sum_{m=1-N}^0 c_i[n]c_j^*[m] \sum_{l=\max(0,k)}^{K-1+\min(0,k)} x[lN-n]x^*[lN-kN-m]. \quad (59)$$

$$E(\hat{r}_{y_i, y_j}[k]\hat{r}_{y_w, y_v}^*[q]) = \frac{1}{(K-|k|)(K-|q|)} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} \left\{ E(y_i[l]y_j^*[l-k])E(y_w^*[p]y_v[p-q]) + E(y_i[l]y_w^*[p]) \times E(y_j^*[l-k]y_v[p-q]) + E(y_i[l]y_v[p-q])E(y_j^*[l-k]y_w^*[p]) - 2E(y_i[l])E(y_w^*[p])E(y_j^*[l-k])E(y_v[p-q]) \right\} = \frac{1}{(K-|k|)(K-|q|)} \sum_{l=\max(0,k)}^{K-1+\min(0,k)} \sum_{p=\max(0,q)}^{K-1+\min(0,q)} \left\{ r_{y_i, y_j}[k]r_{y_w, y_v}^*[q] + r_{y_i, y_w}[l-p]r_{y_j, y_v}^*[l+q-k-p] + E(y_i[l]y_v[p-q])E(y_j^*[l-k]y_w^*[p]) - 2E(y_i[l])E(y_w^*[p]) \times E(y_j^*[l-k])E(y_v[p-q]) \right\}. \quad (55)$$

For now, we only concentrate on the most inner summation in (59). Without loss of generality, we only consider $k \geq 0$ in the following analysis to simplify the writing. Note that the results of the analysis also hold for $k < 0$. We have three possible cases for $z = \sum_{l=k}^{K-1} x[lN - n]x^*[lN - kN - m]$

- 1) $n \neq m$: In this case, z is the sum of the product of two i.i.d. Gaussian distributions. Note also that the terms in the summation have identical distributions and they are also independent to one another. Hence, we can exploit the central limit theorem to assume that in this case, z has a Gaussian distribution for sufficiently large $K - k$.
- 2) $k = 0$ and $n = m$: In this case, z will have a chi-square distribution with $K - k$ degrees of freedom. The chi-square distribution will converge to a Gaussian distribution for a sufficiently large value of $K - k$.
- 3) $k \neq 0$ and $n = m$: Similar to the first case, z is now again the sum of the product of two i.i.d. Gaussian distributions. Even though the terms in the summation are identically distributed, they are generally dependent. As an example, the summation for $k = 2$ and $K = 20$ can be written as

$$\begin{aligned} & \sum_{l=2}^{19} x[lN - n]x^*[(l-2)N - n] \\ &= x[2N - n]x^*[-n] + x[3N - n]x^*[N - n] \\ & \quad + x[4N - n]x^*[2N - n] + \dots \\ & \quad + x[19N - n]x^*[17N - n]. \end{aligned} \quad (60)$$

Note that the same sample $x[2N - n]$ has a contribution in the first and third terms in (60). In general, for all possible values of k , every single term has a statistical dependency on at most two other terms in the summation. In order to simplify the analysis, let us split z into two separate summations, that is

$$\begin{aligned} & \sum_{l=k}^{K-1} x[lN - n]x^*[lN - kN - n] \\ &= \left\{ \sum_{s=1}^{\lfloor \frac{K-k}{2k} \rfloor} \sum_{l=0}^{k-1} x[((2s-1)k+l)N - n] \right. \\ & \quad \times x^*[((2s-2)k+l)N - n] \\ & \quad \left. + \sum_{l=(2\lfloor \frac{K-k}{2k} \rfloor + 2)k-1}^{\min((2\lfloor \frac{K-k}{2k} \rfloor + 2)k-1, K-1)} x[lN - n]x^*[lN - kN - n] \right\} \\ & \quad + \left\{ \sum_{s=1}^{\lfloor \frac{K-k}{2k} \rfloor} \sum_{l=0}^{k-1} x[(2sk+l)N - n]x^*[((2s-1)k+l)N - n] \right. \\ & \quad \left. + \sum_{l=(2\lfloor \frac{K-k}{2k} \rfloor + 2)k}^{K-1} x[lN - n]x^*[lN - kN - n] \right\}. \end{aligned} \quad (61)$$

Note that we basically put every k consecutive terms in z together into one group where every term is a product of two i.i.d. Gaussian random variables. The first pair of brackets in (61) contains the odd groups whereas the second pair of brackets contains the even groups. Observe that the summation in each

pair of brackets is the sum of independent and identically distributed terms. For instance, if we consider (61) for $k = 2$ and $K = 20$, we obtain

$$\begin{aligned} & \sum_{l=2}^{19} x[lN - n]x^*[(l-2)N - n] \\ &= \{x[2N - n]x^*[-n] + x[3N - n]x^*[N - n] + \dots \\ & \quad + x[18N - n]x^*[16N - n] + x[19N - n]x^*[17N - n]\} \\ & \quad + \{x[4N - n]x^*[2N - n] + x[5N - n]x^*[3N - n] + \dots \\ & \quad + x[16N - n]x^*[14N - n] + x[17N - n]x^*[15N - n]\}. \end{aligned} \quad (62)$$

Note that, within each pair of brackets in (62), there is no sample that contributes to more than one term in the summation. This characteristic can also be found if we apply different values of k and K to (61). As a result, we can also exploit the central limit theorem as long as $\lfloor \frac{K-k}{2k} \rfloor k$ is sufficiently large. When this is the case, we have the summation of two terms, each of which has a Gaussian distribution. As a result, z can again be approximated by a Gaussian distribution for large $\lfloor \frac{K-k}{2k} \rfloor k$.

By covering all cases, we can conclude that the Gaussian approximation is applicable for the distribution of $\{\hat{r}_{y_i, y_j}[k]\}_{i,j=0}^{M-1}$ as long as $\lfloor \frac{K-k}{2k} \rfloor k$ is sufficiently large. From (11), we know that $\{\hat{r}_{y_i, y_j}[k]\}_{i,j=0}^{M-1}$ has a support limited to $-L \leq k \leq L$. As a result, the Gaussian assumption for $\{\hat{r}_{y_i, y_j}[k]\}_{i,j=0}^{M-1}$ is valid if $L \ll K$. By taking (11) and (40) into account, we can find that the distribution of each element of $\hat{\mathbf{s}}_x$ for the case of circular complex zero-mean Gaussian i.i.d. noise $x[n]$, is asymptotically Gaussian for a large K .

APPENDIX C

EVALUATION OF THE STATISTICAL DISTRIBUTION OF $\hat{\mathbf{s}}_x$ FOR CIRCULAR COMPLEX ZERO-MEAN GAUSSIAN I.I.D. NOISE $x[n]$ (ALTERNATIVE TIME-DOMAIN APPROACH)

We investigate the statistical distribution of $\hat{\mathbf{s}}_x$ under the alternative time-domain approach by first considering the statistics of $\hat{\mathbf{r}}_y[0]$ in (47). Recall that the elements of $\hat{\mathbf{r}}_y[0]$ are $\hat{r}_{y_i, y_j}[0]$, which are given by (59) with $k = 0$:

$$\hat{r}_{y_i, y_j}[0] = \frac{1}{K} \sum_{n=1-N}^0 \sum_{m=1-N}^0 c_i[n]c_j^*[m] \sum_{l=0}^{K-1} x[lN - n]x^*[lN - m]. \quad (63)$$

We only pay attention to the most inner summation in (63). When $x[n]$ only contains circular complex zero-mean Gaussian i.i.d. noise, we can observe that $z = \sum_{l=0}^{K-1} x[lN - n]x^*[lN - m]$ is the sum of i.i.d. random variables for both $m = n$ and $m \neq n$. As a result, for sufficiently large K , we are again able to exploit the central limit theorem to assume that z has a Gaussian distribution. By combining (47)–(50) and (63), we can conclude that for the case of circular complex zero-mean Gaussian i.i.d. noise $x[n]$, each element of $\hat{\mathbf{s}}_x$ is asymptotically Gaussian distributed for a large K under the alternative time-domain approach.

REFERENCES

- [1] M. Kim and J. Takada, "Efficient multichannel wideband spectrum sensing technique using filter bank," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC 2009)*, Tokyo, Japan, Sep. 2009, pp. 1014–1018.

- [2] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.
- [3] P. Pysarvi-Hoseini and N. C. Beaulieu, "Optimal wideband spectrum sensing framework for cognitive radio systems," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1170–1182, Mar. 2011.
- [4] Z. Tian and G. B. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," in *Proc. IEEE Int. Conf. Cogn. Radio Oriented Wireless Netw. Commun.*, Greece, Jun. 2006.
- [5] B. Le, T. Rondeau, J. Reed, and C. Bostian, "Analog-to-digital converters," *IEEE Signal Process. Mag.*, vol. 22, no. 6, Nov. 2005.
- [6] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2007)*, Honolulu, HI, Apr. 2007, vol. 4, pp. IV/1357–IV/1360.
- [7] Y. Bresler, "Spectrum-blind sampling and compressive sensing for continuous-index signals," in *Proc. Inf. Theory Appl. Workshop (ITA 2008)*, Jan. 2008, pp. 547–554.
- [8] M. Mishali and Y. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.
- [9] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bound on aliasing error in sub-Nyquist nonuniform sampling of multi-band signals," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2173–2183, Sep. 2000.
- [10] M. Mishali and Y. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
- [11] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Royal Stat. Soc. B*, vol. 58, pp. 267–288, 1996.
- [12] Y. Wang, A. Pandharipande, and G. Leus, "Compressive sampling based MVDR spectrum sensing," in *Proc. Int. Workshop on Cogn. Inf. Process. (CIP 2010)*, Elba Island, Italy, Jun. 2010, pp. 333–337.
- [13] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP2009)*, Taipei, Apr. 2009, pp. 2337–2340.
- [14] V. Havary-Nassab, S. Hassan, and S. Valaee, "Compressive detection for wide-band spectrum sensing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2010)*, Dallas, TX, Mar. 2010, pp. 3094–3097.
- [15] M. A. Lexa, M. E. Davies, J. S. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2011)*, Prague, May 2011, pp. 3884–3887.
- [16] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *Proc. IEEE Digital Signal Process. Signal Process. Educ. Workshop*, Sedona, AZ, Jan. 2011, pp. 289–294.
- [17] G. Leus and D. D. Ariananda, "Power spectrum blind sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 443–446, Aug. 2011.
- [18] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, New Orleans, LA, May 2007, pp. 1959–1962.
- [19] S. Kirolos, T. Ragheb, J. Laska, M. F. Duarte, Y. Massoud, and R. G. Baraniuk, "Practical issues in implementing analog-to-information converters," in *Proc. 6th Int. Workshop on Syst.-on-Chip for Real-Time Appl.*, Cairo, Egypt, Dec. 2006, pp. 141–146.
- [20] J. Leech, "On the representation of $1, 2, \dots, n$ by differences," *J. London Math. Soc.*, vol. 31, pp. 160–169, Apr. 1956.
- [21] P. H. M. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Trans. Autom. Control*, vol. 33, no. 9, pp. 867–870, Sep. 1988.
- [22] W. Bar and F. Ditttrich, "Useful formula for moment computation of normal random variables with nonzero means," *IEEE Trans. Autom. Control*, vol. 16, no. 3, pp. 263–265, Jun. 1971.
- [23] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory, Upper Saddle River*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
- [24] *Functional Requirements for the 802.22 WRAN Standard*, IEEE 802.22-05/0007r46, 2005.



Dyonisius Dony Ariananda (S'11) was born in Semarang, Central Java, Indonesia. He received the Bachelor degree (*cum laude*) from Gadjah Mada University, Yogyakarta, Indonesia, in 2000 and the M.Sc degree (*cum laude*) from Delft University of Technology, the Netherlands, in 2009, both in electrical engineering. Apart from that, he was also granted the Australian Development Scholarship (ADS) by the Australian government to complete the M.S. degree (cumulative GPA: high distinction) in internetworking at the University of Technology Sydney, Australia, in 2005.

He has been a member of the junior education staff at the Informatics Engineering Department, Sanata Dharma University, Yogyakarta, from October 2000 to December 2003, and with the Electrical Engineering Department, Atma Jaya University, Jakarta, Indonesia, from January 2006 to July 2007. His research is mainly focused on signal processing for communication. He is currently pursuing the Ph.D. degree with the Circuit and System (CAS) group, Delft University of Technology, where he works on spectral analysis and spectrum estimation for cognitive radio.



Geert Leus (M'00–SM'05–F'12) was born in Leuven, Belgium, in 1973. He received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in June 1996 and May 2000, respectively.

He has been a Research Assistant and a Post-doctoral Fellow of the Fund for Scientific Research—Flanders, Belgium, from October 1996 to September 2003. During that period, he was affiliated with the Electrical Engineering Department, Katholieke Universiteit Leuven. Currently, he is an Associate Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands. During summer 1998, he visited Stanford University, Stanford, CA, and from March 2001 to May 2002, he was a Visiting Researcher and Lecturer at the University of Minnesota. His research interests are in the area of signal processing for communications.

Dr. Leus received the 2002 IEEE Signal Processing Society Young Author Best Paper Award and the 2005 IEEE Signal Processing Society Best Paper Award. He was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE SIGNAL PROCESSING LETTERS. Currently, he serves on the Editorial Board of the *EURASIP Journal on Advances in Signal Processing*.