

Low-Complexity Blind Parameter Estimation in Wireless Systems with Noisy Sparse Signals

Alexandra Gallyas-Sanhueza, *Graduate Student Member, IEEE*, and Christoph Studer, *Senior Member, IEEE*

Abstract—Baseband processing algorithms often require knowledge of the noise power, signal power, or signal-to-noise ratio (SNR). In practice, these parameters are typically unknown and must be estimated. Furthermore, the mean-square error (MSE) is a desirable metric to be minimized in a variety of estimation and signal recovery algorithms. However, the MSE cannot directly be used as it depends on the true signal that is generally unknown to the estimator. In this paper, we propose novel *blind* estimators for the average noise power, average receive signal power, SNR, and MSE. The proposed estimators can be computed at low complexity and solely rely on the large-dimensional and sparse nature of the processed data. Our estimators can be used (i) to quickly track some of the key system parameters while avoiding additional pilot overhead, (ii) to design low-complexity nonparametric algorithms that require such quantities, and (iii) to accelerate more sophisticated estimation or recovery algorithms. We conduct a theoretical analysis of the proposed estimators for a Bernoulli complex Gaussian (BCG) prior, and we demonstrate their efficacy via synthetic experiments. We also provide three application examples that deviate from the BCG prior in millimeter-wave multi-antenna and cell-free wireless systems for which we develop nonparametric denoising algorithms that improve channel-estimation accuracy with a performance comparable to denoisers that assume perfect knowledge of the system parameters.

I. INTRODUCTION

ACCURATE knowledge of system parameters, such as the average noise power, average signal power, and/or signal-to-noise ratio (SNR), is critical in wireless communication systems, as many baseband processing tasks rely on these quantities [3]. Virtually all existing wireless systems dedicate training phases to estimate such parameters. These training phases typically consist of sending *pilots*: signals that are known to the receiver and enable estimation of the desired parameters. As pilots do not convey information, minimizing the pilot overhead is desirable in practice. Furthermore, parameter

A. Gallyas-Sanhueza is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY; email: ag753@cornell.edu

C. Studer is with the Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland; email: studer@ethz.ch

The work of AGS and CS was supported in part by ComSenTer, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA, and in part by the US National Science Foundation (NSF) under grants CNS-1717559 and ECCS-1824379.

Part of this work was presented at the IEEE International Conference on Communications (ICC) 2021 [1]. This journal paper extends our work by (i) including a novel parametric noise power estimator with improved accuracy, (ii) evaluating the proposed blind estimators as an initializer for an expectation-maximization algorithm, and (iii) adding two applications examples.

MATLAB code to reproduce our simulations is available on GitHub: https://github.com/IIIP-Group/blind_and_nonparametric_estimators.

The authors thank Arian Maleki, Ramina Ghods, Charles Jeon, and Seyed Hadi Mirfarshbafan for discussions on signal recovery using SURE, and Haochuan Song for sharing the cell-free system simulator from [2].

estimation in wireless systems operating at millimeter-wave (mmWave) frequencies must be done frequently, since the propagation conditions can change at fast rates, e.g., blockers or interferers may appear or disappear quickly [4]. Thus, it is even more important to reduce the pilot overhead. In addition, such systems are expected to support several GHz of bandwidth and basestations will consist of a large number of antenna elements. It is therefore important to develop *low-complexity* solutions that quickly and accurately track such parameters for high-dimensional problems that must be processed at fast rates.

From a parameter estimation perspective, it is beneficial that many modern wireless communication systems often deal with high-dimensional data. For example, all-digital massive multiple-input multiple-output (MIMO) basestations are expected to be equipped with hundreds of antennas [5] or orthogonal frequency-division multiplexing (OFDM) systems will support thousands of subcarriers [6]. Since many of these high-dimensional signals arising in such systems exhibit structure (e.g., are sparse or are taken from a discrete set), one can design statistical methods that blindly estimate critical parameters without requiring a dedicated training phase.

In this paper, we focus on noisy observations of signal vectors that are *sparse*, i.e., only few entries carry most of the signals' energy. Examples of sparse vectors in wireless systems include (i) the beamspace-domain representation of all-digital mmWave multi-antenna channel vectors [7]–[9], (ii) the delay-domain representation of OFDM channel vectors [10], and (iii) the antenna-domain representation of channel vectors in cell-free MIMO wireless systems [11]. We will explain how sparsity can be exploited to estimate parameters and denoise noisy observations of sparse vectors. In Sections II, III, and IV, we decouple our results from wireless communication applications and study the general setting. In Section V, we apply our estimators and algorithms to three distinct applications in wireless systems.

In what follows, we will use the term “*blind*” for estimators that do not use any pilots or training sequences and instead rely only on the signal statistics; blind estimators may have tuning parameters. We will use the term “*nonparametric*” for estimators that do not need knowledge of system parameters and do not have parameters that need to be tuned manually; nonparametric estimators may use pilots or training sequences.

A. Prior Art in Blind and Nonparametric Estimation

Many of the existing blind noise power and SNR estimators exploit modulation-specific structure, such as the cyclic prefix

redundancy in OFDM [12], [13], or the periodicity of synchronization sequences [14]. Expectation-maximization (EM) has also been used for blind noise power or SNR estimation [15], and for joint sparse signal recovery and noise power estimation [16], [17]. However, the iterative nature of Bayesian algorithms and EM, and their relatively high per-iteration complexity renders such methods unsuitable for real-time estimation in wireless systems that operate with high-dimensional data at gigabit-per-second sampling rates. In contrast, we propose low-complexity blind estimators whose complexity only scales with $\mathcal{O}(D)$, where D is the dimension of the processed data. Our proposed low-complexity estimators can also be used as initialization methods that accelerate the convergence of existing EM algorithms.

Joint noise power estimation and sparse signal recovery was investigated in [18]; these methods require the choice of algorithm parameters, which affect the estimation accuracy and robustness. A parameter-free version of sparse signal recovery that combines approximate message passing (AMP) [19], [20] with Stein's unbiased risk estimate (SURE) [21], [22] was proposed in [23]. Similarly, the nonparametric equalizer (NOPE) [24] combines AMP with SURE to perform linear minimum mean-square error (MSE) equalization in massive MIMO systems without knowledge of the SNR. A drawback of such algorithms is the high per-iteration complexity, which prevents their use in wireless systems supporting large bandwidths and high-dimensional problems (see, e.g., [25], [26] for hardware results of sparse signal recovery). We therefore focus on low complexity, blind, and nonparametric algorithms for the fully-determined setting (in contrast to compressive sensing where one has fewer measurements than unknowns), which finds use in many practical situations. For example, all-digital massive MIMO architectures (which can be as energy efficient as hybrid analog-digital architectures [27]–[29]) and cell-free wireless systems can provide measurement vectors of the same dimension as the sparse signal. In OFDM systems, even though pilots are typically transmitted only on a subset of all subcarriers, interpolation and extrapolation algorithms can be used to extract channel state information on all subcarriers [30]; this also leads to the fully-determined setting that enables the use of our methods.

In this low-complexity setting, the concept of estimating tuning parameters directly from the noisy observations has been used recently for adaptive denoising of mmWave [7]–[9] or OFDM [10] channel vectors. Such denoising algorithms typically require a tuning parameter: the denoising threshold. While SURE can be used to automatically determine the MSE-optimal denoising threshold, it still requires knowledge of the noise power. In contrast to such results, we propose low-complexity blind estimators, which enable the design of nonparametric (i.e., parameter free) channel-vector denoising algorithms that deliver comparable performance to methods that assume perfect knowledge of the required parameters (e.g., the noise power).

Blind nonparametric algorithms have been proposed for denoising of real-valued signals. The authors in [31] have used power estimation methods based on the median absolute deviation (MAD) of real-valued signals for wavelet denoising.

The Python wavelet toolbox PyYAWT [32] includes MAD-based power estimation and adaptive wavelet denoising using SURE for real-valued signals. Our methods also build upon MAD and SURE, but are suitable for complex-valued signals. In addition, we provide a detailed derivation and a theoretical analysis, and extend the general concept to estimate other quantities that frequently arise in wireless systems. While some papers apply *real-valued* MAD for noise power estimation in the complex-valued setting (see, e.g., [33] for magnetic resonance imaging), there are non-negligible differences to the complex case. We therefore derive the complex-valued version, provide a theoretical accuracy analysis with a Bernoulli complex Gaussian (BCG) prior, and show application examples that deviate from this prior in order to highlight robustness and usefulness of our results.

B. Contributions

A variety of applications in communication systems deal with sparse and complex-valued signals whose observations are contaminated with noise. For such a model, we propose novel low-complexity blind estimators for the average noise power, average signal power, and SNR. In addition, we propose a blind estimator for the MSE of an estimation function that aims to recover the sparse signal. We use this blind MSE estimate to design a novel nonparametric channel-vector denoising algorithm. We conduct a theoretical analysis of our estimators for a BCG prior, and we showcase simulation results with synthetic data in order to demonstrate the efficacy and limits of our estimators in finite dimensions. In order to demonstrate the efficacy of our results in situations that deviate from a BCG prior, we provide three application examples of channel-vector denoising in mmWave and cell-free communication systems. We also show that our low-complexity estimators can be used to accelerate the convergence (and, hence, reduce the complexity) of existing estimators with a concrete example of an EM-based algorithm.

C. Notation

Lowercase and uppercase boldface letters denote column vectors and matrices, respectively. The d th entry of the vector $\mathbf{a} \in \mathbb{C}^D$ is a_d ; the real and imaginary parts are $\Re\{\mathbf{a}\}$ and $\Im\{\mathbf{a}\}$, respectively. We use $\mathbf{b} \triangleq |\mathbf{a}|^2$ to refer to $b_d = |a_d|^2$ for $d = 1, \dots, D$. For $\mathbf{a} \in \mathbb{C}^D$, the vector q -norm is defined as $\|\mathbf{a}\|_q \triangleq \left(\sum_{d=1}^D |a_d|^q \right)^{1/q}$ for $q \geq 1$ with $\|\mathbf{a}\|_\infty \triangleq \max_{d=1, \dots, D} |a_d|$ and the ℓ_0 -pseudo-norm $\|\mathbf{a}\|_0$ counts the number of nonzero entries in \mathbf{a} . The identity matrix is \mathbf{I} and the all-zeros vector is $\mathbf{0}$. The discrete Fourier transform matrix is denoted by \mathbf{F} and satisfies $\mathbf{F}^H \mathbf{F} = \mathbf{I}$, where the superscript H denotes the Hermitian (conjugate transpose) matrix. An i.i.d. circularly-symmetric complex Gaussian random vector $\mathbf{x} \in \mathbb{C}^D$ with variance E_x per complex dimension is denoted by $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, E_x \mathbf{I})$ and its probability density function (PDF) evaluated at \mathbf{x} is $f^{\mathcal{CN}}(\mathbf{x}; \mathbf{0}, E_x \mathbf{I})$. Sample estimates are denoted by a bar, e.g., the sample variance $\bar{E}_x \triangleq \frac{1}{D} \|\mathbf{x}\|_2^2$ of the random vector $\mathbf{x} \in \mathbb{C}^D$; statistical quantities are denoted by plain symbols, e.g., the

variance $E_x \triangleq \frac{1}{D} \mathbb{E}[\|\mathbf{x}\|_2^2]$, where $\mathbb{E}[\cdot]$ denotes expectation; blind estimators are denoted by a hat, e.g., \hat{E}_x . For $x \in \mathbb{R}$, rounding towards plus and minus infinity is denoted by $\lceil x \rceil$ and $\lfloor x \rfloor$, respectively, and $[x]_+ \triangleq \max\{x, 0\}$. Convergence in probability of a random sequence A_n to a random variable A is $A_n \xrightarrow{n \rightarrow \infty}^{\text{prob.}} A$ and almost sure convergence is $A_n \xrightarrow{n \rightarrow \infty}^{\text{a.s.}} A$.

II. PRACTICAL GUIDE TO LOW-COMPLEXITY BLIND ESTIMATORS

We now introduce two system models and propose low-complexity blind estimators for the average noise and signal powers, SNR, and MSE. The derivation of the proposed estimators and an analysis of the key properties are provided in Section III.

A. System Models

We say that a complex-valued vector $\mathbf{s} \in \mathbb{C}^D$ is *sparse* if the number of nonzero entries is smaller than the dimension D . As a sparsity measure, one can use, for example, the ℓ_0 -pseudonorm $\|\mathbf{s}\|_0$. This definition of sparsity allows us to derive theoretical results, but in practice, our algorithms also work for *approximately* sparse signals in which most entries are small compared to the noise (but not necessarily zero). We will focus on the following two system models.

System Model 1. Let $\mathbf{s} \in \mathbb{C}^D$ be a sparse signal with average power $E_s \triangleq \frac{1}{D} \mathbb{E}[\|\mathbf{s}\|_2^2]$. We model the input-output relation of a noisy observation of the sparse signal as

$$\mathbf{y} = \mathbf{s} + \mathbf{n}, \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^D$ is the noisy observation and $\mathbf{n} \in \mathbb{C}^D$ models noise with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I})$. We assume that the sparse signal vector \mathbf{s} and noise vector \mathbf{n} are statistically independent.

System Model 1 finds numerous applications in wireless communication systems. Prime examples are in describing estimated channel vectors (i) in multi-antenna mmWave systems, where the beampattern-domain representation of the channel vectors are typically sparse [7]–[9], (ii) in OFDM systems, where the delay-domain representation of the channel vectors are typically sparse [10], or (iii) in cell-free communication systems with centralized processing, where the antenna-domain representation of the channel vectors are typically sparse [11]. In what follows, we assume the sparse vector \mathbf{s} is unknown (in contrast to pilot-based estimation), which makes parameter estimation nontrivial in this blind scenario.

System Model 2. Let $\mathbf{y} \in \mathbb{C}^D$ be a noisy observation as in System Model 1. Fix a weakly differentiable function¹ $\eta : \mathbb{C} \rightarrow \mathbb{C}$ that operates entry-wise on vectors. We model the output after applying this function to the noisy observation as

$$\eta(\mathbf{y}) = \mathbf{s} + \mathbf{e}, \quad (2)$$

where $\mathbf{e} \in \mathbb{C}^D$ contains (likely non-Gaussian) residual distortion. We emphasize that the sparse signal vector \mathbf{s} and

¹A weakly differentiable function may be nondifferentiable only in zero-measure sets (e.g., for particular values), and has to be differentiable everywhere else.

the residual distortion vector \mathbf{e} are not necessarily statistically independent.

System Model 2 is relevant in the following scenarios: (i) Estimating a sparse signal \mathbf{s} from a noisy observation \mathbf{y} by applying an entry-wise denoising or estimation function, producing the signal estimate $\hat{\mathbf{s}} \triangleq \eta(\mathbf{y})$; this scenario finds use for channel-vector denoising [7]–[9]. (ii) Modeling nonlinearities caused by hardware impairments [34], in which case the distorted version of the noisy received signal can be expressed as $\mathbf{r} \triangleq \eta(\mathbf{y})$; this scenario finds use in signals sampled with low-resolution data converters [35], [36], for example.

B. Low-Complexity Blind Nonparametric Estimators

In what follows, we make use of the sample median, which we define as follows.

Definition 1 (Sample Median). Let $\mathbf{z} \in \mathbb{R}^D$ be a real-valued vector and $\mathbf{z}^{\text{sort}} \in \mathbb{R}^D$ be its sorted version (entries sorted in ascending order). Then, the sample median is defined as

$$\bar{m}(\mathbf{z}) \triangleq \frac{1}{2} \left(z_{\lfloor (D+1)/2 \rfloor}^{\text{sort}} + z_{\lceil (D+1)/2 \rceil}^{\text{sort}} \right). \quad (3)$$

The sample median is robust to outliers [37], [38], which makes it amenable to System Model 1, as the nonzero entries of the sparse vector \mathbf{s} can be considered to be outliers for the purpose of separating the sparse signal from noise. We emphasize that the sample median can be computed at a complexity of $\mathcal{O}(D)$ average time using quickselect [39] or of $\mathcal{O}(D)$ deterministic time using the MedianOfNinthers algorithm [40].

We now propose a range of low-complexity blind estimators (no pilots required) for complex-valued signals that require no parameters.

Estimator 1 (Average Noise Power). Consider System Model 1. We propose the following blind estimator

$$\hat{N}_0 \triangleq \frac{\bar{m}(|\mathbf{y}|^2)}{\log(2)} \quad (4)$$

for the average noise power defined as $N_0 \triangleq \mathbb{E}[\|\mathbf{n}\|_2^2] / D$.

Estimator 1 is blind as it only requires the absolute square entries of the noisy observation \mathbf{y} in equation (1). The estimate \hat{N}_0 can be computed efficiently in $\mathcal{O}(D)$ time, since the most complex operation is computing the median of a vector of dimension D . Estimator 1 exploits sparsity in the signal \mathbf{s} , but is independent of the signal sparsity, the signal power, or the statistical sparsity model. It is, however, important to understand that the accuracy of this estimator depends on all of these factors as it relies on the fact that the nonzero entries of the sparse vector \mathbf{s} can be treated as outliers for the purpose of estimating the average noise power. We note that this noise power estimator can be seen as a complex-

valued and squared version² of the median absolute deviation (MAD) estimator [37], [41], where we use the assumption that the noise in System Model 1 is zero mean. The intuition behind this estimator (and the $\log(2)$ factor) is the fact that the entries $|n_d|^2/(N_0/2)$, $d = 1, \dots, D$ are χ^2 distributed with two degrees of freedom, which have a median of $2\log(2)$, and that the median of $|\mathbf{y}|^2$ is not significantly “contaminated” by the sparse signal. Estimator 1 is used in the estimators proposed next.

Estimator 2 (Average Signal Power). *Consider System Model 1. We propose the following blind estimator*

$$\widehat{E}_s \triangleq \left[\frac{\|\mathbf{y}\|_2^2}{D} - \widehat{N}_0 \right]_+ \quad (5)$$

for the average signal power defined as $E_s \triangleq \mathbb{E}[\|\mathbf{s}\|_2^2]/D$.

Estimator 2 is blind as it only requires the sample estimate of the receive power $\overline{E}_y \triangleq \|\mathbf{y}\|_2^2/D$ and the blind noise estimate \widehat{N}_0 from Estimator 1. \widehat{E}_s can be computed efficiently in $\mathcal{O}(D)$ time, since the most complex operation is computing \widehat{N}_0 . The intuition behind this estimator comes from subtracting the estimated noise power from the total receive power, as done previously in [13] for an OFDM-specific estimator.

Estimator 3 (Signal-to-Noise Ratio). *Consider System Model 1. We propose the following blind estimator*

$$\widehat{\text{SNR}} \triangleq \left[\frac{\|\mathbf{y}\|_2^2}{D\widehat{N}_0} - 1 \right]_+ \quad (6)$$

for the SNR defined as $\text{SNR} \triangleq \mathbb{E}[\|\mathbf{s}\|_2^2]/\mathbb{E}[\|\mathbf{n}\|_2^2]$.

Estimator 3 is blind as it only requires the sample estimate of the receive power $\overline{E}_y \triangleq \|\mathbf{y}\|_2^2/D$ and the blind estimate \widehat{N}_0 from Estimator 1. $\widehat{\text{SNR}}$ can also be computed efficiently in $\mathcal{O}(D)$ time. The intuition behind this estimator comes from dividing the estimated signal and noise powers, as done previously in [13] for an OFDM-specific estimator.

Estimator 4 (Mean-Square Error). *Consider System Model 2 with a fixed function $\eta : \mathbb{C} \rightarrow \mathbb{C}$. We propose the following blind estimator*

$$\widehat{\text{MSE}} \triangleq \frac{1}{D} \|\eta(\mathbf{y}) - \mathbf{y}\|_2^2 - \widehat{N}_0 + \frac{\widehat{N}_0}{D} \sum_{d=1}^D \left(\frac{\partial \Re\{\eta(y_d)\}}{\partial \Re\{y_d\}} + \frac{\partial \Im\{\eta(y_d)\}}{\partial \Im\{y_d\}} \right) \quad (7)$$

for the MSE defined as $\text{MSE} \triangleq \mathbb{E}[\|\eta(\mathbf{y}) - \mathbf{s}\|_2^2]/D = \mathbb{E}[\|\mathbf{e}\|_2^2]/D$.

Estimator 4 is blind as it only requires the receive signal \mathbf{y} , the blind estimate \widehat{N}_0 from Estimator 1, and the function η . The complexity of the proposed MSE estimator depends on

²The squared median absolute deviation (MAD) estimator for real-valued signals provided in [38] corresponds to $\overline{m}(|\mathbf{y}|)^2$ whereas we propose to use $\overline{m}(|\mathbf{y}|^2)$. While $\overline{m}(|\mathbf{y}|^2) \leq \overline{m}(|\mathbf{y}|^2)$ if D is even, both estimators coincide if D is odd. What is more, our scaling factor $\log(2) \approx 0.6931$ differs considerably from the widely-used scaling factor of $(\Phi^{-1}(3/4))^2 \approx (0.6745)^2$ for real-valued signals [31]. We reiterate that the latter is derived for power estimation of *real-valued* Gaussians using the MAD estimator, while in our derivation we consider the case of *complex-valued* Gaussians.

the function η . For example, equation (7) can be computed efficiently in $\mathcal{O}(D)$ time for the soft-thresholding function with a given threshold. Even if the threshold is not given, searching for the best threshold and applying the soft-thresholding function can be done in $\mathcal{O}(D \log(D))$ time using the methods developed for the BEACHES algorithm in [7]. The MSE is a frequently used metric to evaluate the performance of estimation algorithms. Our blind MSE estimate, since it is independent of \mathbf{s} , can be used to automatically tune parameters in estimators. The intuition behind this estimator relies on SURE, and we refer the interested reader to [22] for an accessible derivation in the real-valued case and to [7], [8] for a derivation in the complex-valued case. Estimator 4 is used to obtain the nonparametric channel-vector denoising algorithm described in Section V.

C. Low-Complexity Blind Parametric Estimators

We now propose a low-complexity blind estimator (no pilots or training sequences required) that takes an estimate \hat{p} of the activity rate as a parameter. We then propose a family of parametric estimators for the activity rate.

Estimator 5 (Average Noise Power with Estimated SNR and Activity Rate Corrections). *Consider System Model 1, the low-complexity blind estimates $\widehat{\text{SNR}}$ from Estimator 3, and a parameter \hat{p} that is an estimate of the activity rate p . We propose the following blind parametric estimator*

$$\widehat{N}_0(\hat{p}) \triangleq \frac{1}{2} \widehat{N}_0 \left(\max \left\{ \frac{\log(2)}{\log\left(\frac{2-2\hat{p}}{1-2\hat{p}}\right)}, \frac{1}{1+\widehat{\text{SNR}}} \right\} + (1-\hat{p}) + \frac{\hat{p}^2}{\hat{p}+\widehat{\text{SNR}}} \right) \quad (8)$$

for the average noise power N_0 .

Estimator 5 is blind as it only requires the blind estimates \widehat{N}_0 , $\widehat{\text{SNR}}$, but is parametric as it depends on the activity rate estimate \hat{p} . $\widehat{N}_0(\hat{p})$ can be computed efficiently in $\mathcal{O}(D)$ time, since the most complex operations are computing \widehat{N}_0 and $\widehat{\text{SNR}}$, and eventually \hat{p} (but here we consider \hat{p} as a given parameter and ignore the complexity associated with obtaining it). The intuition behind this estimator will become clear after we present Theorem 1, as it is derived from averaging a lower and an upper bound on N_0 . As shown in Section IV, this parametric estimate $\widehat{N}_0(\hat{p})$ often yields better accuracy than the nonparametric estimate \widehat{N}_0 .

Since in some applications an estimate for \hat{p} may be unavailable, we next propose a family of estimators that attempt to extract the activity rate p directly from the noisy observation vector \mathbf{y} . Such estimators can, for example, be used to substitute \hat{p} in equation (8).

Estimator 6 (Activity Rate). *Consider System Model 1 and integers $1 \leq q < r$. We propose the following family of blind*

Algorithm 1 Baseline EM (adapted from [42, Alg. 8.1] to circularly-symmetric complex Gaussians)

```

1: input  $\mathbf{y}$ ,  $K^{\max}$ ,  $\xi$ ,  $N_0^{\text{init}}$ , and  $p^{\text{init}}$  // noisy signal, maximum iterations, tolerance, initial noise variance, initial activity rate
2: initialize // initialize weights ( $w$ ) and variances ( $v$ ) of distributions A and B as follows:
3:  $w_a \leftarrow 1 - p^{\text{init}}$ ,  $v_a \leftarrow N_0^{\text{init}}$  // initialize A as noise
4:  $w_b \leftarrow p^{\text{init}}$ ,  $v_b \leftarrow v_a + \frac{1}{p^{\text{init}}} \left[ \frac{\|\mathbf{y}\|_2^2}{D} - N_0^{\text{init}} \right]_+$  // initialize B as signal plus noise, ensuring  $w_a v_a + w_b v_b = \|\mathbf{y}\|_2^2 / D$ 
5:  $k \leftarrow 0$ ,  $v_a^{\text{old}} \leftarrow \infty$ ,  $v_b^{\text{old}} \leftarrow \infty$ 
6: while  $\xi < \frac{|v_a - v_a^{\text{old}}|}{v_a} + \frac{|v_b - v_b^{\text{old}}|}{v_b}$  and  $k < K^{\max}$  do // iterate until convergence or maximum iterations reached
7:  $k \leftarrow k + 1$ ,  $v_a^{\text{old}} \leftarrow v_a$ ,  $v_b^{\text{old}} \leftarrow v_b$ 
8: for  $d = 1$  to  $D$  do // likelihood that  $y_d$  comes from distribution A
9:    $a_d \leftarrow \frac{w_a f^{\mathcal{CN}}(y_d; 0, v_a)}{w_a f^{\mathcal{CN}}(y_d; 0, v_a) + w_b f^{\mathcal{CN}}(y_d; 0, v_b)}$ 
10:   $b_d \leftarrow \frac{w_b f^{\mathcal{CN}}(y_d; 0, v_b)}{w_a f^{\mathcal{CN}}(y_d; 0, v_a) + w_b f^{\mathcal{CN}}(y_d; 0, v_b)}$  // likelihood that  $y_d$  comes from distribution B
11: end for
12:  $w_a \leftarrow \frac{1}{D} \sum_{d=1}^D a_d$ ,  $w_b \leftarrow 1 - w_a = \frac{1}{D} \sum_{d=1}^D b_d$  // update weights based on updated likelihoods
13:  $v_a \leftarrow \frac{1}{w_a D} \sum_{d=1}^D a_d |y_d|^2$ ,  $v_b \leftarrow \frac{1}{w_b D} \sum_{d=1}^D b_d |y_d|^2$  // update variances based on updated likelihoods
14: end while
15: if  $v_a > v_b$  then // assign the smallest variance to noise, and the largest variance to signal plus noise
16:    $\hat{N}_0^{\text{EM}} \leftarrow v_b$ ,  $\hat{E}_s^{\text{EM}} \leftarrow w_a(v_a - v_b)$ ,  $\hat{p}^{\text{EM}} \leftarrow w_a$ 
17: else
18:    $\hat{N}_0^{\text{EM}} \leftarrow v_a$ ,  $\hat{E}_s^{\text{EM}} \leftarrow w_b(v_b - v_a)$ ,  $\hat{p}^{\text{EM}} \leftarrow w_b$ 
19: end if
20:  $K \leftarrow k$  // save number of iterations until convergence or stopping condition met
21: return  $\hat{N}_0^{\text{EM}}$ ,  $\hat{E}_s^{\text{EM}}$ , and  $\hat{p}^{\text{EM}}$  // estimated noise variance, estimated signal variance, estimated activity rate

```

parametric estimators³

$$\hat{p}(q, r) \triangleq \frac{1}{D} \left(\frac{\|\mathbf{y}\|_q}{\|\mathbf{y}\|_r} \right)^{\frac{1}{1/q - 1/r}} \quad (9)$$

for the activity rate⁴ defined as $p \triangleq \mathbb{E}[\|\mathbf{s}\|_0] / D$.

Estimator 6 is blind as it only requires the receive vector \mathbf{y} , but is parametric as it requires a choice for q and r . $\hat{p}(q, r)$ can be computed efficiently in $\mathcal{O}(D)$ time and, among others, the following choices for q and r require low complexity: $\hat{p}(1, 2) \triangleq \frac{1}{D} \left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_2} \right)^2$, $\hat{p}(1, \infty) \triangleq \frac{1}{D} \left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_\infty} \right)$, $\hat{p}(2, 4) \triangleq \frac{1}{D} \left(\frac{\|\mathbf{y}\|_2}{\|\mathbf{y}\|_4} \right)^4$, and $\hat{p}(2, \infty) \triangleq \frac{1}{D} \left(\frac{\|\mathbf{y}\|_2}{\|\mathbf{y}\|_\infty} \right)^2$. The parameters q and r must be chosen according to simulations, as we are unaware of a principled and reliable way to determine them. In our simulations, the choice $\hat{p}(1, \infty)$ performed best.

D. Blind Parametric Estimator Based on Expectation-Maximization (EM)

As a baseline, we also consider a blind EM estimator (no pilots required) that requires initialization values and algorithm parameters that determine the convergence criterion.

Estimator 7 (Noise Power, Signal Power, and Activity Rate). *Consider System Model 1. Algorithm 1, initialized with $N_0^{\text{init}} < \|\mathbf{y}\|_2^2 / D$ and $p^{\text{init}} < 0.5$, simultaneously estimates the noise power N_0 , the signal power E_s , and the activity rate p .*

Estimator 7 is blind as it only requires the noisy observation \mathbf{y} , but is parametric as it needs a choice for the maximum number of iterations K^{\max} , the tolerance ξ , and initialization values for the noise power N_0^{init} and activity rate p^{init} . The total number of EM iterations K is not fixed but depends on K^{\max} , ξ , N_0^{init} , p^{init} , and on the input \mathbf{y} itself. The complexity of Estimator 7 is $\mathcal{O}(KD)$. We note that this estimator is a variant of a classical EM algorithm for a two-component Gaussian mixture [42], where we use the assumption that the signal and the noise in System Model 1 are zero mean and complex valued. The intuition behind this estimator is the fact that each entry $|y_d|$, $d = 1, \dots, D$, of vector \mathbf{y} contains either noise or signal-plus-noise, and those two cases have Gaussian distribution with different variances.

We note that this baseline EM algorithm is only a minor variation of the method in [42, Alg. 8.1]. The iterative nature of such methods, however, results in (often significantly) higher complexity than our estimators. With this in mind, we propose an improved version that we call “accelerated EM,” which simply consists of initializing the baseline EM algorithm using our blind nonparametric noise variance estimator. As we will see in Section IV-B, this accelerated EM variant drastically reduces the number of iterations needed for convergence without degrading accuracy.

³In practice, we use $\min\{0.499, \hat{p}(q, r)\}$ in place of $\hat{p}(q, r)$, so that $\log\left(\frac{2-2\hat{p}(q, r)}{1-2\hat{p}(q, r)}\right)$ is always well-defined, as required by equation (14) on page 7.

⁴We define the activity rate as the fraction of nonzero entries of the vector \mathbf{s} . Values of p close to 0 indicate the vector is sparse and $p = 1$ indicates that all entries are nonzero.

TABLE I

COMPLEXITY AND ACCURACY SUMMARY. D IS THE SIGNAL DIMENSION AND $K^{\text{ACC}} \ll K^{\text{BL}}$ REFER TO THE NUMBER OF ITERATIONS IN THE ACCELERATED EM AND BASELINE EM ALGORITHMS, RESPECTIVELY.

	Power estimation	Complexity	Accuracy	
		Denoising	Synthetic data	Realistic channels
Baseline EM	$\mathcal{O}(K^{\text{BL}}D)$	$\mathcal{O}(K^{\text{BL}}D + D \log(D))$	(✓✓✓)	(✓✓)
Accelerated EM	$\mathcal{O}(K^{\text{ACC}}D)$	$\mathcal{O}(K^{\text{ACC}}D + D \log(D))$	(✓✓✓)	(✓✓)
Nonparametric	$\mathcal{O}(D)$	$\mathcal{O}(D \log(D))$	(✓✓)	(✓✓✓)
Parametric	$\mathcal{O}(D)$	$\mathcal{O}(D \log(D))$	(✓)	(✓✓✓)

E. Summary of Proposed Power Estimation and Denoising Algorithms

Table I summarizes the complexity and accuracy of the different estimators. ‘‘Baseline EM’’ refers to Estimator 7, ‘‘accelerated EM’’ to Estimator 7 initialized using Estimator 1, ‘‘nonparametric’’ to Estimator 1, and ‘‘parametric’’ to Estimator 5. The complexity for blind noise power estimation is mentioned below the definition of each of these algorithms. The complexity for denoising is the complexity of estimating the noise power plus the complexity of the BEACHES algorithm from [8]. Since BEACHES already sorts the magnitudes of the noisy signal, the nonparametric and parametric estimators that use the median require no additional complexity for estimating the noise power. Anticipating the results shown in Section IV and Section V, we illustrate (qualitatively) the accuracy of the estimators with synthetic data that perfectly matches the BCG prior, and with practical examples that deviate from this prior.

III. THEORY

We first show that the sample median approaches the median for $D \rightarrow \infty$ and introduce our statistical model for sparse vectors. We then derive and analyze Estimators 1 to 7. The observations made in this section are valid in the large-dimension limit and for the noisy BCG model to be introduced in Definition 4. We use simulations to demonstrate the accuracy of our estimators for finite (and small) dimensions D with the noisy BCG model in Section IV. To demonstrate the efficacy of our methods in practical scenarios with signals that deviate from the BCG model, we evaluate our denoising algorithms in three distinct scenarios in Section V.

A. Convergence of the Sample Median for $D \rightarrow \infty$

We will use the following definition of the median.

Definition 2 (Median). *Let X be an absolutely continuous random variable (RV) with cumulative distribution function (CDF) $F_X(x)$. Then, the median m_X of X is defined as*

$$F_X(m_X) = \frac{1}{2}. \quad (10)$$

While, analogously to the central limit theorem, the sample median is approximately Gaussian if D is large (see, e.g., [43]), we will only use the following result.

Lemma 1 (Lem. C.1 from [43]). *Let X be a RV whose PDF is differentiable in some neighborhood of the median m_X and*

vector \mathbf{x} contain i.i.d. samples of X . Then, for any $c > 0$ the sample median $\bar{m}(\mathbf{x})$ satisfies

$$\lim_{D \rightarrow \infty} \Pr[|\bar{m}(\mathbf{x}) - m_X| \geq c] = 0. \quad (11)$$

This result implies that in the *large-dimension limit* ($D \rightarrow \infty$), the sample median $\bar{m}(\mathbf{x})$ converges in probability to the median m_X . Hence, by observing a sufficiently large number of samples, which is possible in modern multi-antenna mmWave or OFDM systems, we can accurately estimate the median m_X .

B. Statistical Model for Complex-Valued Sparse Vectors

To derive and analyze the blind estimators proposed in Section II, we need a statistical model for the sparse signal \mathbf{s} . This model should (i) have as few parameters as possible while being able to model a large class of complex-valued sparse vectors typically arising in communication systems and (ii) facilitate a theoretical analysis. In what follows, we consider BCG random vectors [20], [44], which allow control over the signal sparsity and the signal power. We reiterate that the BCG model is instrumental *only* for our analysis. The provided simulation results in Section V will show that the proposed estimators exhibit robustness to model mismatch, e.g., for signals that are not necessarily i.i.d. Gaussian or circularly symmetric.

Definition 3 (BCG Random Vector). *A sparse vector $\mathbf{s} \in \mathbb{C}^D$ is BCG if each entry is nonzero with probability $p \in (0, 1]$, and the nonzero entries are i.i.d. circularly-symmetric complex Gaussian with variance E_s/p . The PDF of each entry s_d , $d = 1, \dots, D$, is therefore given by*

$$f_s(s_d) \triangleq (1-p)\delta(s_d) + p \frac{1}{\pi E_s/p} e^{-\frac{|s_d|^2}{E_s/p}}, \quad (12)$$

where $\delta(\cdot)$ is the Dirac delta distribution.

With this model, the activity rate is $p = \mathbb{E}[\|\mathbf{s}\|_0]/D$ (meaning the expected number of nonzero entries is $\mathbb{E}[\|\mathbf{s}\|_0] = pD$), and the average power of the sparse signal vector \mathbf{s} is $E_s = \frac{1}{D} \mathbb{E}[\|\mathbf{s}\|_2^2]$.

In System Model 1, we assumed that the noise vector \mathbf{n} is i.i.d. circularly-symmetric complex Gaussian with variance N_0 per complex entry. Hence, the PDF of each entry n_d , $d = 1, \dots, D$, is given by $f^{\mathcal{CN}}(n_d; 0, N_0) \triangleq \frac{1}{\pi N_0} e^{-|n_d|^2/N_0}$. Consequently, if \mathbf{s} is a BCG random vector, then the PDF of the noisy observation vector $\mathbf{y} = \mathbf{s} + \mathbf{n}$ is as follows.

Definition 4 (Noisy BCG Random Vector). *The PDF of the entries y_d , $d = 1, \dots, D$, of a BCG random vector per Definition 3 observed as in System Model 1 is given by*

$$f_Y(y_d) \triangleq (1-p) \frac{1}{\pi N_0} e^{-\frac{|y_d|^2}{N_0}} + p \frac{1}{\pi(N_0 + E_s/p)} e^{-\frac{|y_d|^2}{N_0 + E_s/p}}. \quad (13)$$

For this signal and observation model, we are now able to derive and analyze Estimators 1 to 7. We will make frequent use of the entry-wise square of vector \mathbf{y} that we will call $\mathbf{z} \triangleq |\mathbf{y}|^2$. We also define a random variable (RV) Z with the same distribution as any of the i.i.d entries of \mathbf{z} , and let m_Z be the median of Z .

C. Analysis of Estimator 1

We start with the blind noise power estimator defined in Estimator 1. We have the following key result. The proof is given in Appendix A.

Theorem 1. *Let \mathbf{y} be a noisy BCG random vector with PDF as in Definition 4 and with activity rate satisfying*

$$p \leq p^{\max} \quad \text{with} \quad p^{\max} \triangleq \frac{e^2 - 2}{2e^2 - 2} \approx 0.421. \quad (14)$$

Let a lower bound LB and an upper bound UB be defined as follows:

$$LB \triangleq \frac{m_Z}{\min \left\{ \log \left(\frac{2-2p}{1-2p} \right), \log(2)(1+SNR) \right\}} \quad (15)$$

$$UB \triangleq \frac{m_Z}{\log(2)} \left((1-p) + \frac{p^2}{p+SNR} \right). \quad (16)$$

Then, the average noise power N_0 satisfies⁵

$$LB \leq N_0 \leq UB \leq \lim_{D \rightarrow \infty} \hat{N}_0. \quad (17)$$

Theorem 1 has the following key implications: (i) In the large-dimension limit, the proposed blind estimate \hat{N}_0 bounds the average noise power N_0 from above, i.e., we have developed a pessimistic estimator. (ii) If $SNR \rightarrow 0$ or $p \rightarrow 0$, then $LB = UB = m_Z / \log(2)$ in equation (17), and therefore $N_0 = m_Z / \log(2)$. Thus, either for $p \rightarrow 0$ or $SNR \rightarrow 0$, the proposed estimate is exact, i.e., $\hat{N}_0 \xrightarrow{D \rightarrow \infty} m_Z / \log(2) = N_0$. We summarize this important insight in the following remark.

Remark 1. *In the large-dimension limit ($D \rightarrow \infty$), the proposed blind nonparametric estimate \hat{N}_0 is pessimistic (i.e., overestimates the average noise power N_0), and becomes exact at low SNR or low activity rate p (i.e., for sparse vectors).*

Next, we present bounds on the relative error of Estimator 1. These bounds depend on the activity rate p and the SNR. The proof is given in Appendix B.

⁵Here we simplify the notation: $\lim_{D \rightarrow \infty} \hat{N}_0$ converges in probability to $m_Z / \log(2)$, and strictly speaking this latter expression is the upper bound.

Corollary 1. *For $p \leq p^{\max}$ as in equation (14), the relative error $\varepsilon \triangleq |\hat{N}_0 - N_0|/N_0$ of Estimator 1 in the large-dimension limit is bounded as follows:*

$$\frac{1}{1/SNR+1/p+1} \leq \lim_{D \rightarrow \infty} \varepsilon \leq \min \left\{ \log \left(\frac{1-p}{1-2p} \right), SNR \right\}. \quad (18)$$

An upper bound for the relative error ε can be obtained if (i) an upper bound on the SNR is known, or (ii) an upper bound on p is known, since $\log \left(\frac{1-p}{1-2p} \right)$ is nondecreasing for $p \in (0, 0.5)$. In addition, we confirm the second implication discussed below Theorem 1: Corollary 1 implies that if $p \rightarrow 0$ (irrespective of the SNR) or $SNR \rightarrow 0$ (irrespective of the sparsity), then the proposed estimator becomes exact, i.e., $\varepsilon = 0$ and therefore $\hat{N}_0 \xrightarrow{D \rightarrow \infty} N_0$.

D. Analysis of Estimator 2

For the blind estimate \hat{E}_s of the average signal power E_s , we use the following lemma, which is derived from the fact that the entries of the vector $\mathbf{z} \triangleq |\mathbf{y}|^2$ are i.i.d. with expected value of $\mathbb{E}[z_d] = \mathbb{E}[\|\mathbf{y}\|_2^2] / D = E_s + N_0$, $d = 1, \dots, D$.

Lemma 2. *Let \mathbf{y} be a noisy BCG random vector with PDF as in Definition 4. Then, according to the strong law of large numbers we have*

$$\frac{1}{D} \|\mathbf{y}\|_2^2 - N_0 \xrightarrow{D \rightarrow \infty} E_s. \quad (19)$$

To obtain Estimator 2 in equation (5) on page 4, we construct a blind estimator of E_s by taking the left side of equation (19) and replacing the average noise power N_0 with the blind estimate \hat{N}_0 from Estimator 1. To avoid negative values of E_s that have no physical meaning, we assign a value of zero to our estimate if $\|\mathbf{y}\|_2^2 / D - \hat{N}_0$ is negative. Since the estimate \hat{N}_0 overestimates the true average noise power N_0 , the blind estimate \hat{E}_s in equation (5) on page 4 tends to underestimate the signal power. From Theorem 1 it follows that for $p \rightarrow 0$ or $SNR \rightarrow 0$, the blind signal power estimate \hat{E}_s is exact.

E. Analysis of Estimator 3

The blind SNR estimator is obtained by simply taking the ratio of \hat{E}_s in equation (5) on page 4 and \hat{N}_0 in equation (4) on page 3. For $D \rightarrow \infty$, the blind signal power estimate underestimates the average signal power and the noise power estimate overestimates the average noise power, which means that the blind SNR estimate in equation (6) on page 4 underestimates the SNR. From Theorem 1 it follows that for $D \rightarrow \infty$ with either $p \rightarrow 0$ or $SNR \rightarrow 0$ the blind SNR estimate is exact.

F. Analysis of Estimator 4

In order to analyze Estimator 4, we first assume that the average noise power N_0 is known. For this scenario, we can borrow the following two theorems from [8].

Theorem 2 (Thm. 1 of [8]). *Consider System Model 2. Then, Stein's unbiased risk estimate given by*

$$\begin{aligned} SURE &\triangleq \frac{1}{D} \|\eta(\mathbf{y}) - \mathbf{y}\|_2^2 - N_0 \\ &+ \frac{N_0}{D} \sum_{d=1}^D \left(\frac{\partial \Re\{\eta(y_d)\}}{\partial \Re\{y_d\}} + \frac{\partial \Im\{\eta(y_d)\}}{\partial \Im\{y_d\}} \right) \end{aligned} \quad (20)$$

is an unbiased estimate of the MSE so that $\mathbb{E}[SURE] = MSE$.

Theorem 3 (Thm. 3 of [8]). *If η is pseudo-Lipschitz, then SURE in equation (20) converges to the MSE in the large-dimension limit, i.e., we have $\lim_{D \rightarrow \infty} SURE = MSE$.*

Theorem 3 implies that if N_0 were known perfectly, then one could perfectly estimate the MSE in the large-dimension limit without knowledge of the sparse signal vector \mathbf{s} . For smaller values of the dimension D , Theorem 2 only ensures equality in expectation (while the estimator remains MSE-optimal). Equality in expectation means that some realizations will underestimate and some realizations will overestimate the true MSE⁶.

Estimator 4 is a blind version of SURE, in which we have replaced the true average noise power N_0 by its estimate \hat{N}_0 . Consequently, for $D \rightarrow \infty$ and either $p \rightarrow 0$ or $SNR \rightarrow 0$, we have that: (i) Remark 1 states \hat{N}_0 will be exact, from which it follows that $\widehat{MSE} = SURE$, (ii) Theorem 3 ensures $SURE = MSE$, and therefore (iii) Estimator 4 will be exact ($\widehat{MSE} = MSE$) in this scenario. For higher values of p or SNR , we know that \hat{N}_0 tends to overestimate N_0 , but since this estimated quantity appears twice in equation (7) on page 4 with different signs, we cannot derive a simple rule that states whether Estimator 4 tends to underestimate or overestimate the MSE.

G. Analysis of Estimator 5

Estimator 5 is derived as the mean of the lower and upper bounds in equation (17) on the preceding page, utilizing the SNR estimate from Estimator 3 and an activity rate estimate \hat{p} of the user's choice. Estimator 5 often improves the performance (achieves lower bias) compared to Estimator 1, especially at high SNR. In contrast to Estimator 1, we no longer know if the noise power from Estimator 5 is being overestimated or underestimated. As this estimator takes \hat{p} as a parameter, it is especially useful in applications where p is known a priori or bounded (e.g., in OFDM systems the number of nonzero delay taps of the channel's impulse response should not exceed the cyclic prefix length).

H. Analysis of Estimator 6

To estimate the activity rate, we can use the equivalence of vector norms [45] that states $\|\mathbf{x}\|_q \leq L^{1/q-1/r} \|\mathbf{x}\|_r$ holds

⁶We have to keep in mind that we use the estimated MSE to determine parameters in the estimation function η that minimize the MSE for each given realization of \mathbf{y} . Therefore, offsets that depend on the realization of the noisy observation \mathbf{y} can be treated as a constant and thus be ignored, even if these offsets cause the MSE to take on negative values. In other words, we are not interested in the true value of the MSE, but rather in the *shape* of the MSE function with respect to the parameters in η .

for any vector $\mathbf{x} \in \mathbb{C}^L$ if $1 \leq q < r$. In particular, it holds for a vector $\mathbf{s}^{\text{nz}} \in \mathbb{C}^L$ of length $L \triangleq \|\mathbf{s}\|_0$ that contains only the nonzero entries of the sparse vector \mathbf{s} . For such vector, we have that $\|\mathbf{s}^{\text{nz}}\|_q \leq L^{1/q-1/r} \|\mathbf{s}^{\text{nz}}\|_r$. Since the entries of \mathbf{s} that are zero do not contribute to these norms, we note that $\|\mathbf{s}^{\text{nz}}\|_q = \|\mathbf{s}\|_q$ and $\|\mathbf{s}^{\text{nz}}\|_r = \|\mathbf{s}\|_r$, and therefore

$$\|\mathbf{s}\|_q \leq \|\mathbf{s}\|_0^{1/q-1/r} \|\mathbf{s}\|_r, \quad 1 \leq q < r. \quad (21)$$

Using equation (21), we can obtain a lower bound for the activity rate⁷:

$$\frac{1}{D} \left(\frac{\|\mathbf{s}\|_q}{\|\mathbf{s}\|_r} \right)^{\frac{1}{1/q-1/r}} \leq \frac{\|\mathbf{s}\|_0}{D} \approx p. \quad (22)$$

The inequality in equation (22) holds with equality if the nonzero entries of the signal are constant-modulus, i.e., if $|s_d^{\text{nz}}| = |s_{d'}^{\text{nz}}|, \forall d, d' \in \{1, \dots, L\}$. We obtain the blind estimator $\hat{p}(q, r)$ from the left side of equation (22), by replacing \mathbf{s} with its noisy version \mathbf{y} . With this substitution the inequality is not preserved (except if $N_0 = 0$), but we use that definition of $\hat{p}(q, r)$ as a rough activity rate estimate instead of picking an arbitrary value.

I. Analysis of Estimator 7

Estimator 7 is a specialized variant of a classical EM algorithm for a two-component Gaussian mixture [42], adapted to complex-valued and zero-mean variables. We consider signal and noise power estimation from a noisy BCG signal as in Definition 4. To understand it as a Gaussian source-separation problem, we consider that each entry of \mathbf{y} is a realization of either (i) just noise with distribution $\mathcal{CN}(0, N_0)$, or (ii) signal plus noise with distribution $\mathcal{CN}(0, N_0 + E_s/p)$. Just-noise realizations occur with probability $1-p$, while signal-plus-noise realizations occur with probability p . Using EM, we estimate the variances of the circularly-symmetric complex Gaussians N_0 and $(N_0 + E_s/p)$, and mixture weights $1-p$ and p . We use our previous knowledge to set the mean of the two distributions to zero, unlike classical EM algorithms that also estimate the means. We make the following observations: (i) This model allows any signal sparsity, as opposed to Estimator 1 which assumes a maximum activity rate p^{\max} . (ii) In the low SNR regime, EM may not be able to separate the noise and signal components as $N_0 + E_s/p \approx N_0$. (iii) The accuracy and the complexity of the algorithm will depend on the maximum number of iterations K^{\max} , the tolerance ξ , the variance and weight initializations, and the noisy realization \mathbf{y} .

To avoid EM converging to pathological solutions with arbitrary initialization, we initialize the algorithm with the following two minimum assumptions: (i) The signal is sparse, or equivalently $p^{\text{init}} \in (0, 0.5)$, and (ii) the power of the entries of \mathbf{y} that contain only noise is smaller than the power of the entries of \mathbf{y} that contain signal plus noise, or equivalently $N_0^{\text{init}} \leq \|\mathbf{y}\|_2^2/D$. This translates to initializing the Gaussian mixture variances v and weights w with $v_a \leq \|\mathbf{y}\|_2^2/D$, $w_b \in (0, 0.5)$, $w_a = 1 - w_b$, and $v_b = v_a + (\|\mathbf{y}\|_2^2/D - v_a)/w_b$. We verify that for this initialization, the average power of the mixture is $w_a v_a + w_b v_b = \|\mathbf{y}\|_2^2/D$, as expected.

⁷The activity rate is $p \triangleq \mathbb{E}[\|\mathbf{s}\|_0]/D = \lim_{D \rightarrow \infty} \|\mathbf{s}\|_0/D$. When D is finite, we have $\|\mathbf{s}\|_0/D \approx p$.

IV. SYNTHETIC RESULTS

We now characterize the accuracy of the estimators proposed in Section II. We use the sparse signal model in Definition 4. Without loss of generality, we fix the noise power to $N_0 = 1$, while varying the signal power E_s , the activity rate p , and the dimension D of the vectors. For different sets of parameters, we perform Monte-Carlo simulations with 10,000 trials. In the plots, the thicker line with markers shows the average performance of an estimator, while the shaded area shows the region closer than one standard deviation away from the mean performance, a measure of the *precision* of the estimator.

A. Evaluation of the Noise Power, Signal Power, SNR, and Activity Rate Estimators

Figure 1 shows the effect of the SNR on the performance of the proposed blind nonparametric estimator \widehat{N}_0 from equation (4) on page 3 and the proposed blind parametric estimate $\widehat{N}_0(\hat{p})$ from equation (8) on page 4, for which we only include results using $\hat{p}(q, r)$ with $q = 1$ and $r = \infty$ for the activity rate estimate, as these parameters showed the best performance in our simulations, outperforming other values of p and q , and a fixed-value of $\hat{p} = 0.25$ which is the center of the simulated range $p \in (0, 0.5)$. We also simulate the baseline EM estimate $\widehat{N}_0^{\text{EM}}$ described in Estimator 7, initialized with $N_0^{\text{init}} = 0.4\|\mathbf{y}\|_2^2/D$ and $p^{\text{init}} = \hat{p}(1, \infty)$, a maximum of $K^{\text{max}} = 30$ iterations and early stopping if the total parameter change is below $\xi = 0.1\%$. As a baseline, we plot the genie-aided estimator $\overline{N}_0 \triangleq \frac{1}{D}\|\mathbf{n}\|_2^2$ that has separate knowledge of \mathbf{n} and the reference parameter N_0 .

Figure 2 shows the effect of the SNR on the performance of the proposed signal power and SNR estimators for an activity rate of $p = 0.1$ and a dimension of $D = 64$. In this case, $\text{SNR}^{\text{EM}} \triangleq \widehat{E}_s^{\text{EM}}/\widehat{N}_0^{\text{EM}}$, the genie-aided estimators that have separate knowledge of \mathbf{s} and \mathbf{n} are $\overline{E}_s \triangleq \frac{1}{D}\|\mathbf{s}\|_2^2$ and $\overline{\text{SNR}} \triangleq \overline{E}_s/\overline{N}_0$, and the reference parameters are E_s and $\text{SNR} \triangleq E_s/N_0$.

From Figures 1 and 2, we observe the following facts about the blind nonparametric estimators: (i) For sparse vectors ($p = 0.1$), our estimators have a precision comparable to that of the genie-aided estimators even for a small sample size of $D = 64$. (ii) The precision of all considered estimators decreases as D increases. (iii) As predicted by our theory, the average noise power is overestimated while the signal power and SNR are underestimated. (iv) At low SNR, the median-based estimators for these three quantities become exact. We also observe that the proposed blind parametric estimate $\widehat{N}_0(\hat{p})$ with $\hat{p} = \hat{p}(1, \infty)$ is more accurate than the blind nonparametric estimate \widehat{N}_0 at high SNR. However, $\widehat{N}_0(\hat{p})$ has fewer theoretical guarantees and is not an upper bound on N_0 .

Figure 3 shows the accuracy of the blind, parametric activity rate Estimator 6. We see that at low and high SNR, $\hat{p}(1, 2)$ tends to overestimate p while $\hat{p}(2, \infty)$ tends to underestimate it. Overall, $\hat{p}(1, \infty)$ results in the best performance when combined with Estimator 5. Admittedly, this is only a rough estimator and we include it as an example of what could be plugged into Estimator 5 or Estimator 7. Nonetheless, we

emphasize that side information about the signal's sparsity should be utilized whenever available.

In comparison with EM (cf. Figures 1 and 2), our methods provide a less-accurate estimate at higher SNR, but require significantly lower complexity. The complexity of the baseline EM algorithm (in terms of the number of operations such as real-valued additions, real-valued multiplications, and exponentials) is more than $K(16D + 12) + 3D$ operations—with early stopping, the average number of iterations observed in our simulations ranges from $K = 8$ to $K = 28$ depending on the SNR. In contrast, our proposed median-based noise estimator has an average complexity of no more than $7.7D + 9$ operations if the median is computed using quickselect [39], and avoids the evaluation of operations such as exponentials and divisions. Hence, our proposed blind estimator is more than $17\times$ less complex than the baseline EM algorithm, which renders our method suitable (i) for low-complexity parameter estimation and (ii) as a potential initializer for EM-based estimators.

B. Accelerated Convergence of EM Using Median-Based Initialization

Figure 4 shows the effect of initialization on the EM algorithm. To study the rate of convergence, we disable early stopping by setting $\xi = 0$ so that the number of iterations is always $K = K^{\text{max}}$, and plot the relative error $\varepsilon^{\text{EM}} \triangleq |\widehat{N}_0^{\text{EM}} - N_0|/N_0$ as we vary K^{max} . We compare the convergence of (i) the accelerated EM algorithm (diamond markers) which is initialized with the blind nonparametric estimate \widehat{N}_0 , and (ii) the baseline EM algorithm (circular markers) initialized with a fixed initialization of $\text{SNR} = 5$, which corresponds to setting the noise power to $1/6$ of the received power $\|\mathbf{y}\|_2^2/D$. We simulated various values of p and SNR , and picked four examples that are representative. At low SNR or high sparsity (low p), the accelerated EM algorithm converges already in the first iteration. In contrast, the baseline EM algorithm converges in more than 16 iterations in some cases. The only case we observe the baseline to outperform the accelerated EM algorithm is in Figure 4(d), in which (i) the baseline has advantage since $\frac{1}{6}\|\mathbf{y}\|_2^2/D$ coincides exactly with the true value of N_0 , and (ii) the SNR is high and the sparsity is low, making it the worst case for the \widehat{N}_0 estimate used by the accelerated EM. We also examine the effect of initializing the activity rate with (i) a fixed value of 0.25, versus (ii) the blind parametric estimate $\hat{p}(1, \infty)$, and we observe no significant difference, especially for the preferred accelerated EM algorithm; however, as $\hat{p}(1, \infty)$ showed superior performance than a fixed value when used in the parametric noise power estimator $\widehat{N}_0(\hat{p})$, we prefer $\hat{p}(q, r)$ when no side information about the signal's sparsity is available.

C. Evaluation of the MSE Estimator

To evaluate the performance of the MSE estimator, we consider System Model 2 with η being the soft-thresholding function defined as

$$\eta(x; \tau) \triangleq \begin{cases} \frac{x}{|x|} \max\{|x| - \tau, 0\} & x \neq 0 \\ 0 & x = 0, \end{cases} \quad (23)$$

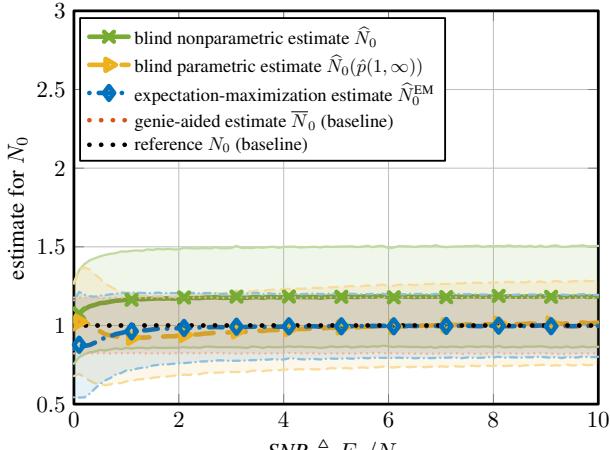
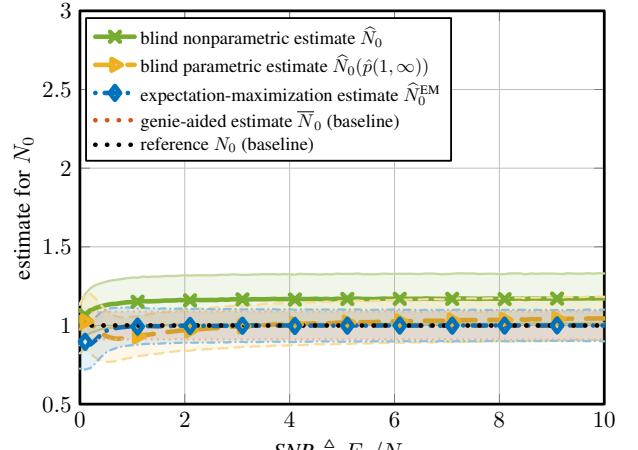
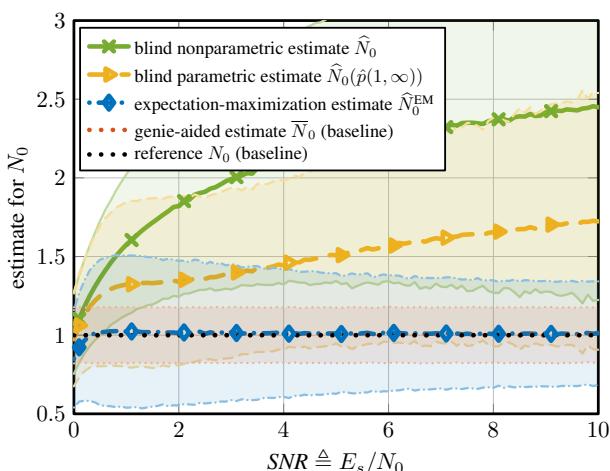
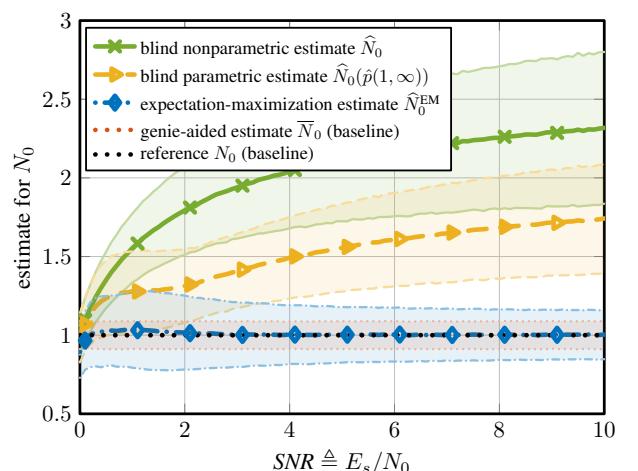
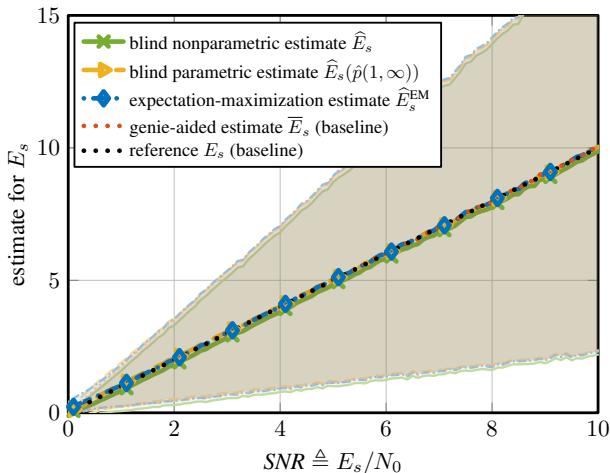
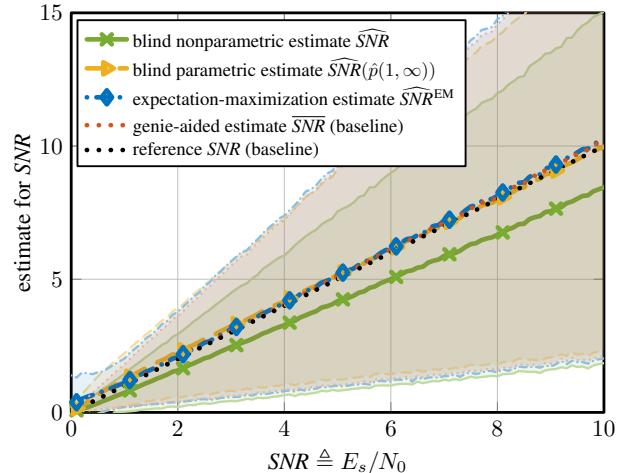
(a) $p = 0.1, D = 64$ (b) $p = 0.1, D = 256$ (c) $p = 0.4, D = 64$ (d) $p = 0.4, D = 256$

Fig. 1. Effect of varying the SNR on the proposed low-complexity blind noise power estimators, for different values of the activity rate p and the dimension D . Blind estimators are more accurate for smaller values of p , and the precision of all estimators is higher for larger values of D .



(a) Average signal power



(b) Signal-to-noise ratio

Fig. 2. Effect of varying the SNR on the proposed low-complexity blind signal power and SNR estimators, for a signal of dimension $D = 64$ with 10% nonzero entries.

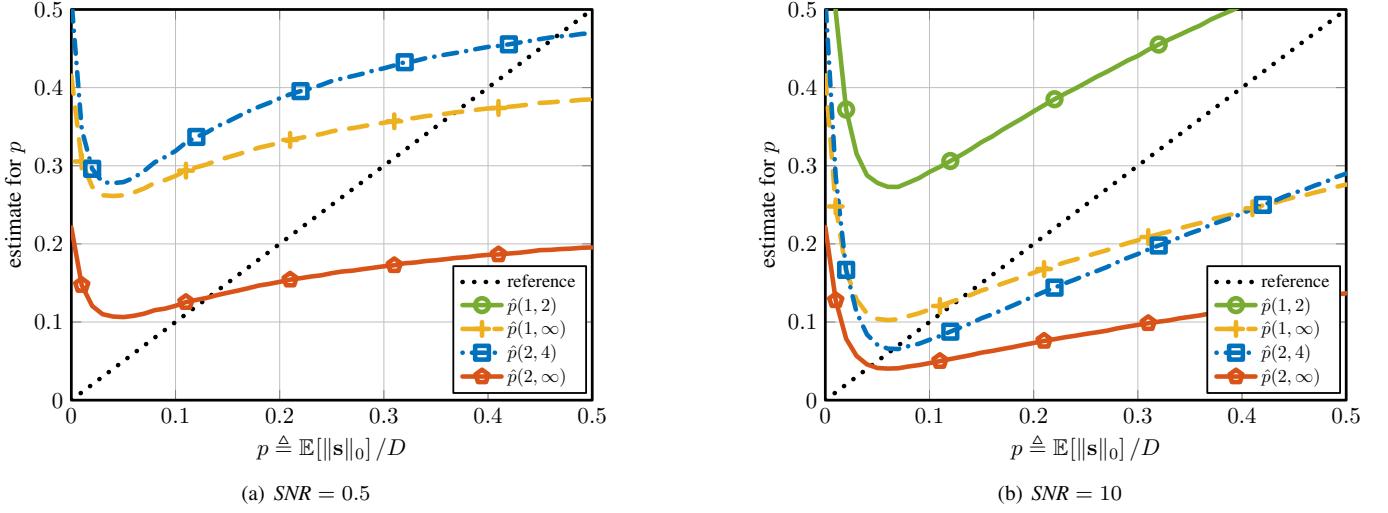


Fig. 3. Average accuracy of the blind parametric activity rate estimator $\hat{p}(q, r)$ for different combinations of q and r with vectors of length $D = 64$. For $SNR = 0.5$, the results for $\hat{p}(1, 2)$ are not visible as the values exceed 0.5. While the overall accuracy of this family of estimators appears to be low, the estimate obtained by $\hat{p}(1, \infty)$ combined with Estimator 5, for example, shows superior performance compared to using the mean activity rate in the considered interval (i.e., 0.25).

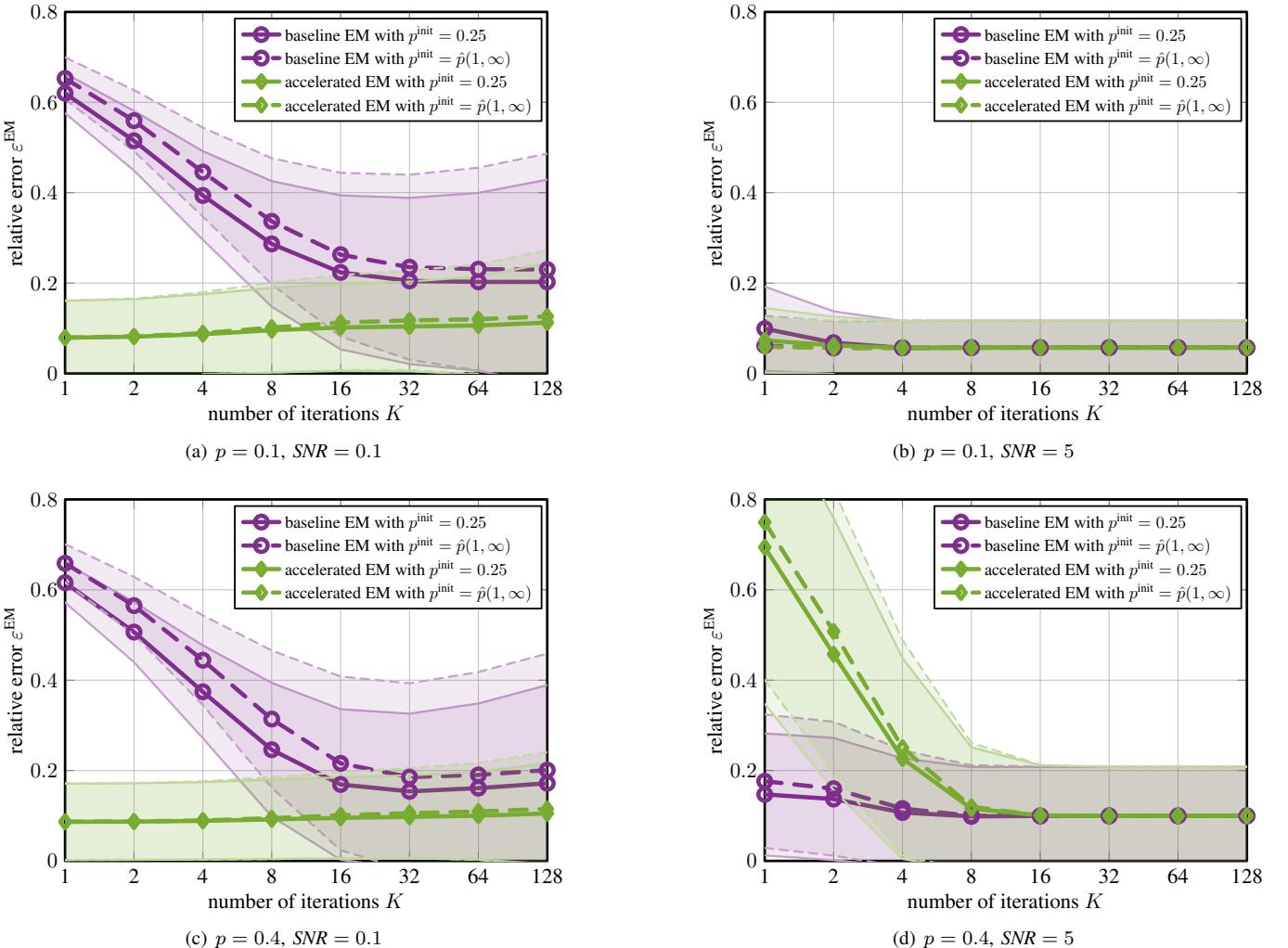


Fig. 4. Relative error $\varepsilon_{EM} \triangleq |\hat{N}_0^{EM} - N_0|/N_0$ of the baseline and accelerated EM noise power estimators vs. the number of iterations K , for different values of SNR and activity rate p . Initializing with the low-complexity estimator \hat{N}_0 accelerates convergence.

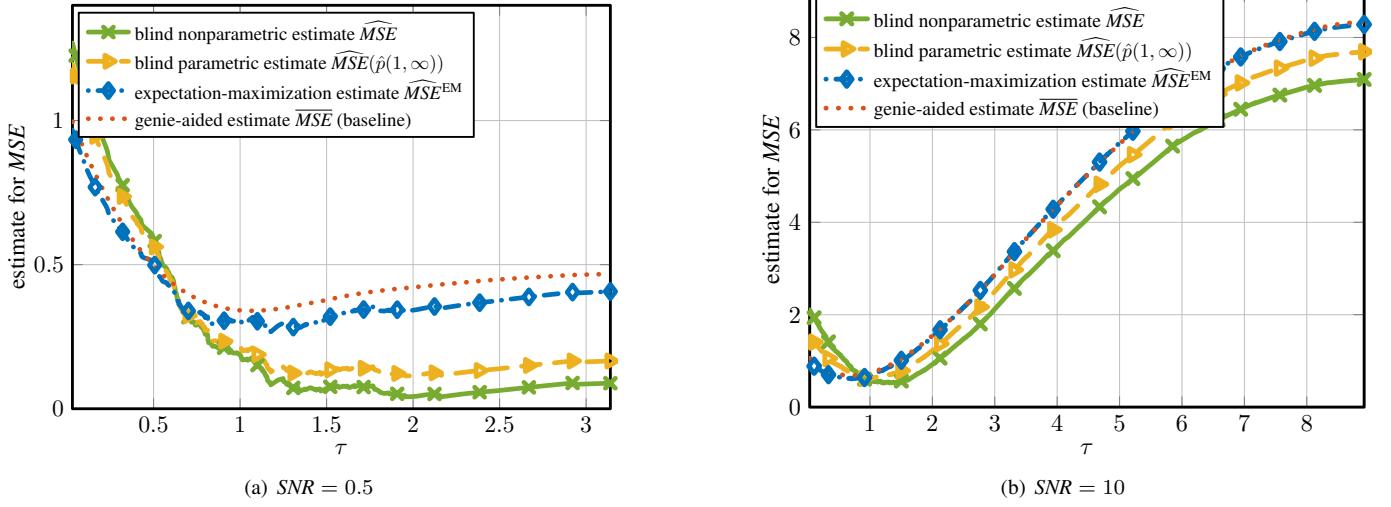


Fig. 5. Examples of the estimated MSE versus τ given one realization of \mathbf{y} at (a) low SNR and (b) high SNR.

where the denoising threshold is a real number $\tau \geq 0$.

Figure 5 shows two realizations of the estimated MSE as a function of the tuning parameter τ . The only reference in this case is the genie-aided estimator $\widehat{MSE} \triangleq \frac{1}{D} \|\eta(\mathbf{y}; \tau) - \mathbf{s}\|_2^2$. We picked two examples that are representative of what we have observed through multiple experiments with different system parameters to illustrate the following observations: (i) If the MSE function has a pronounced minimum as in Figure 5(b), then the value of τ that minimizes the blind estimate tends to be very close to the value that minimizes the genie-aided MSE function. (ii) If the MSE function has a less pronounced minimum as in Figure 5(a), then the value of τ that minimizes the blind estimate may be far from the value that minimizes the genie-aided MSE function. In spite of that, because the MSE function is flat near the minimum, the genie-aided MSE function evaluated at these two values of τ returns values that are similar. In other words, (i) and (ii) summarize our observations that our algorithm finds a near-optimal (sub-optimal) denoising threshold τ when the MSE of the denoised channel is (not) sensitive to τ . Note that here we have only picked two representative realizations; in Section V, we validate our estimator with quantitative results by showing the denoising performance averaged over many realizations.

V. APPLICATIONS TO NONPARAMETRIC CHANNEL-VECTOR DENOISING

We show three applications in wireless systems, in which the quality of channel estimates is essential for data detection. Concretely, we show that our algorithms can be applied to adaptively denoise pilot-based channel estimates, resulting in a reduced (improved) bit-error-rate (BER).

A. Infinite-Resolution Massive Multiuser MIMO System

We start with an application of Estimator 4 for beamspace channel estimation. As in [7], we simulate an uplink massive multiuser (MU) MIMO system in which $U = 8$ single-antenna user equipments (UEs) transmit channel-estimation pilots and

data to a basestation (BS) equipped with a uniform linear array of $D = 128$ antenna elements. The UEs are randomly placed with a uniform distribution in a 120° circular sector around the BS, with a minimum distance of 10 m and maximum distance of 110 m from the BS. A minimum angular separation of 4° between UEs is enforced. We assume UE-side perfect power control (UEs adjust their transmit power so that the received power at the BS is equal for all UEs), and we ignore quantization at transmitter and receiver sides, assuming infinite-resolution signals.

We simulate a noiseless channel matrix $\mathbf{H} \in \mathbb{C}^{D \times U}$ using line-of-sight (LoS) realizations from the mmMAGIC QuaDRiGa model [46] with a carrier frequency of $f_c = 60$ GHz. Each complex-valued entry $\mathbf{H}_{d,u}$ of the channel matrix contains the attenuation and phase between the u th UE and the d th BS antenna. For the channel estimation step, the UEs transmit orthogonal pilots. The maximum likelihood (ML) estimate of the channel matrix is obtained by right-multiplying the (noisy) received pilot sequence with the inverse of the orthogonal pilot matrix, resulting in

$$\mathbf{H}^{\text{ML}} = \mathbf{H} + \mathbf{N}^{\text{CE}}, \quad (24)$$

where $\mathbf{H} \in \mathbb{C}^{D \times U}$ is the antenna-domain channel matrix, $\mathbf{N}^{\text{CE}} \in \mathbb{C}^{D \times U}$ is complex Gaussian channel estimation noise with power N_0^{CE} per complex entry, and $\mathbf{H}^{\text{ML}} \in \mathbb{C}^{D \times U}$ is the ML channel estimate, which is a noisy observation of \mathbf{H} . The beamspace representation of the ML estimate is obtained by taking a spatial Fourier transform across the antenna array resulting in

$$\tilde{\mathbf{H}}^{\text{ML}} = \tilde{\mathbf{H}} + \tilde{\mathbf{N}}^{\text{CE}}. \quad (25)$$

Here, beamspace-domain quantities are designated by a tilde. Then, $\tilde{\mathbf{H}} = \mathbf{F}\mathbf{H}$ is the beamspace channel matrix, $\tilde{\mathbf{N}}^{\text{CE}} = \mathbf{F}\mathbf{N}^{\text{CE}}$ has the same distribution as \mathbf{N}^{CE} as the discrete Fourier transform matrix \mathbf{F} is unitary, and $\tilde{\mathbf{H}}^{\text{ML}}$ is the beamspace ML channel estimate, which is a noisy observation of $\tilde{\mathbf{H}}$. Column indices of $\tilde{\mathbf{H}}$ correspond to UEs, while row indices correspond to different angles-of-arrival to the BS. Since electromagnetic

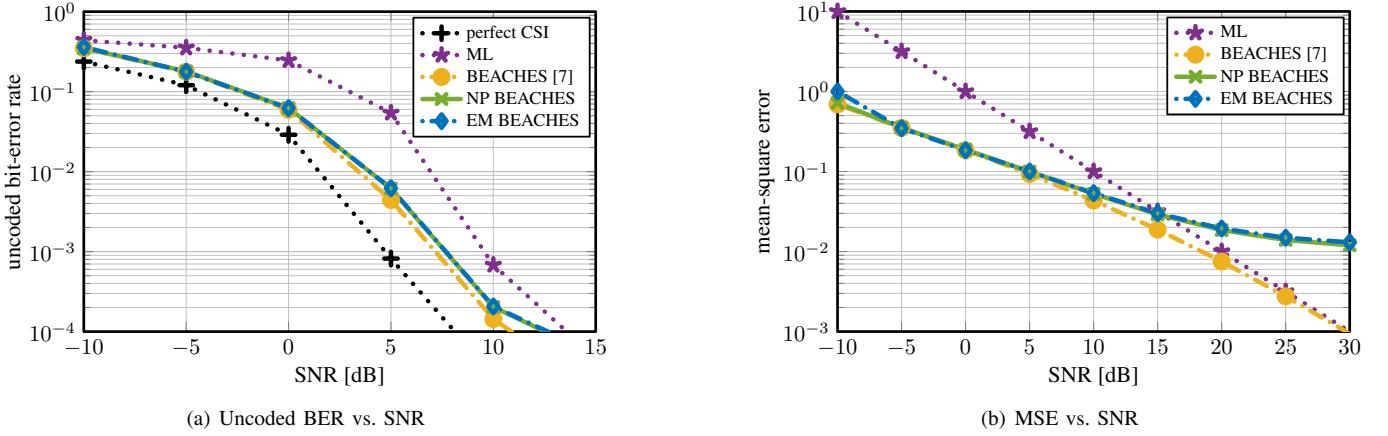


Fig. 6. Uncoded BER (a) and MSE (b) of mmWave channel estimation methods, including the nonparametric (NP) BEACHES variant which estimates the noise power and denoising parameter directly from the receive vector.

waves at high carrier frequencies experience strong attenuation, typical mmWave channels consist only of a small number of dominant propagation paths arriving at the BS. Thus, each column of $\tilde{\mathbf{H}}$ (which is the beamspace *channel vector* of one UE) will be approximately sparse, with many entries being close to zero.

By writing each column of equation (25) on the preceding page as an independent equation, we can express the channel estimation problem in the form of System Model 1, that is, each beamspace channel vector (that contains only few nonzero entries) corresponds to a sparse signal \mathbf{s} . The sparsity property implies that we can perform denoising to improve the ML channel estimate. After channel estimation, all UEs transmit data simultaneously using uncoded 16-QAM symbols and the BS performs data detection using the estimated channel vectors and linear minimum MSE equalization.

Figure 6 shows simulation results for 10,000 Monte–Carlo trials. For different channel estimation methods, we compute the MSE of the channel estimates and the resulting BER. We simulate beamspace channel estimation (BEACHES) as in [7], which denoises the columns of $\tilde{\mathbf{H}}^{\text{ML}}$ in equation (25) on the previous page by applying the soft-thresholding function in equation (23) on page 9; the thresholding parameter τ is adaptively selected for each noisy observation by minimizing SURE using an $\mathcal{O}(D \log(D))$ algorithm that assumes perfect knowledge of the average noise power N_0^{CE} . We compare this to NP BEACHES, a new nonparametric BEACHES variant which also applies soft-thresholding to the columns of $\tilde{\mathbf{H}}^{\text{ML}}$, but uses the (nonparametric) threshold τ that minimizes \widehat{MSE} as in Estimator 4; since \widehat{MSE} is a nonparametric version of SURE, NP BEACHES does not require knowledge of N_0^{CE} . In addition, we include a variant that we call EM BEACHES, which uses a version of \widehat{MSE} in which \widehat{N}_0 in equation (7) on page 4 is replaced by $\widehat{N}_0^{\text{EM}}$ from Estimator 7; for $\widehat{N}_0^{\text{EM}}$, we use $N_0^{\text{init}} = 0.4\|\mathbf{y}\|_2^2/D$ and $p^{\text{init}} = \hat{p}(1, \infty)$, a maximum of $K^{\text{max}} = 30$ iterations and early stopping if the total parameter change is below $\xi = 0.1\%$. The three versions of BEACHES as described above, after denoising the beamspace channel vectors, use the inverse Fourier transform to obtain an antenna-domain

channel estimate to be used for data detection. As a reference, we show the performance of perfect channel state information (CSI) that uses the ground truth (noiseless) channel matrix \mathbf{H} , and ML estimation that simply takes the noisy observation \mathbf{H}^{ML} in equation (24) on the preceding page as the estimate.

From Figure 6, we observe that NP BEACHES achieves virtually the same performance as the original BEACHES algorithm (which requires knowledge of N_0^{CE}), except at high SNR where Estimator 1 tends to overestimate N_0^{CE} . We reiterate that NP BEACHES requires no parameters and exhibits the same low complexity of $\mathcal{O}(D \log(D))$ as the original BEACHES algorithm, because the latter already sorts the entries of $|\mathbf{y}|^2$, which we can reuse to compute the median in Estimator 1. We observe that EM BEACHES achieves higher (worse) MSE at low SNR and does not outperform NP BEACHES at higher SNR.

In summary, denoising methods can significantly improve the ML channel estimate. All three BEACHES variants achieve similar BER performance. However, BEACHES needs knowledge of the noise power and EM BEACHES exhibits higher complexity than our nonparametric estimate, which renders NP BEACHES the preferable denoising method in this application scenario.

B. Low-Resolution Massive Multiuser MIMO System

Next, we consider the same uplink massive MU-MIMO system as Section V-A, but in this case each radio-frequency (RF) chain at the BS is equipped with a pair of 1-bit analog-to-digital converters (ADCs) to quantize the in-phase and quadrature baseband signals. Each RF chain applies a quantization function $Q(x) \triangleq \text{sign}(\Re\{x\}) + j \text{sign}(\Im\{x\})$ to the baseband signal, where $j^2 = -1$. For simplicity, we assume that the pilot matrix is an identity, i.e., each UE has a dedicated time slot to transmit one pilot while all other UEs are silent. The receive pilots then correspond to the 1-bit version of the ML channel estimate, which we call 1-bit ML⁸

$$\mathbf{H}^{\text{1-bit ML}} = Q(\mathbf{H} + \mathbf{N}^{\text{CE}}). \quad (26)$$

⁸ $\mathbf{H}^{\text{1-bit ML}}$ is simply the 1-bit version of \mathbf{H}^{ML} , not to be confused with the maximum likelihood channel estimate given a one-bit observation.

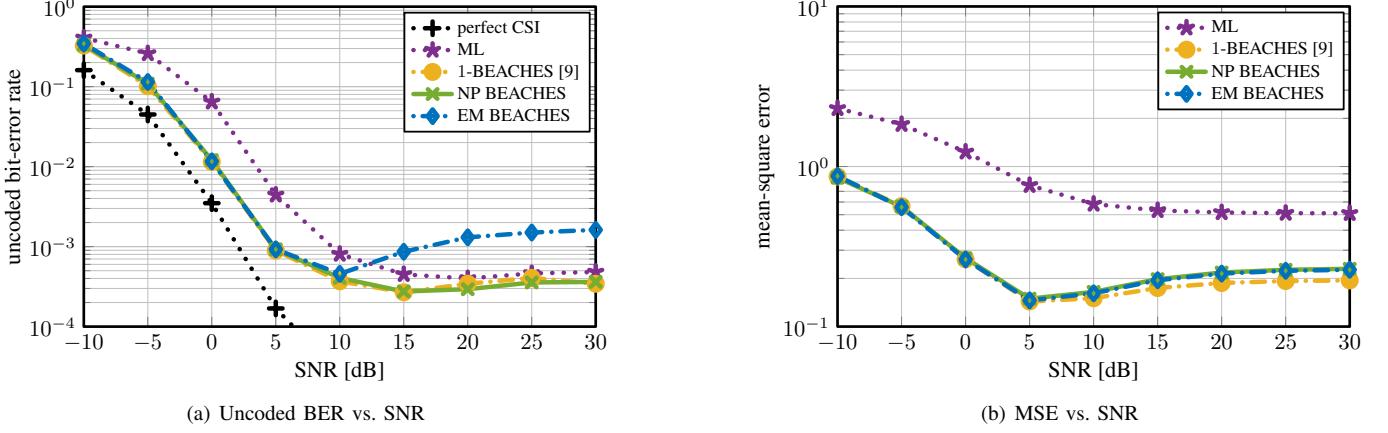


Fig. 7. Uncoded BER (a) and MSE (b) of channel estimation methods in a mmWave system with 1-bit quantization.

Here, quantization happens in the antenna domain and yet, when the quantized noisy channel is converted to beamspace, the sparse structure that is present in infinite-resolution beamspace channel vectors is also present in the coarsely quantized beamspace channel vectors. Thus,

$$\tilde{\mathbf{H}}^{1\text{-bit ML}} = \mathbf{F} \mathbf{Q} (\mathbf{H} + \mathbf{N}^{\text{CE}}) \quad (27)$$

has sparse columns that can be denoised. For more details on the validity of this statement, see [9] where $\tilde{\mathbf{H}}^{1\text{-bit ML}}$ was decomposed in a linear combination of $\tilde{\mathbf{H}}$ plus a residual.

Figure 7 shows simulation results for 10,000 Monte–Carlo trials. For different channel estimation methods, we compute the MSE and BER. All UEs simultaneously transmit uncoded QPSK symbol and the BS is using the estimated channels and performs the 1-bit Bussgang linear minimum MSE equalizer from [47].

We simulate 1-BEACHES as in [9]. This denoising algorithm decomposes equation (27) as $\tilde{\mathbf{H}}^{1\text{-bit ML}} = \tilde{\mathbf{H}} + \tilde{\mathbf{Q}}$, where $\tilde{\mathbf{Q}}$ represents the equivalent noise-plus-quantization error with average power $Q_0 = 2E_s - 4E_s/\sqrt{\pi(E_s + N_0)}$ per entry [9]. The 1-BEACHES algorithm denoises the columns of $\tilde{\mathbf{H}}^{1\text{-bit ML}}$ with the threshold τ that minimizes SURE, assuming perfect knowledge of Q_0 . We also use the nonparametric algorithms NP BEACHES and EM BEACHES (described in Section V-A) to denoise the columns of $\tilde{\mathbf{H}}^{1\text{-bit ML}}$. After denoising the beamspace channel vectors, these three BEACHES variants use the inverse Fourier transform to obtain an antenna-domain channel estimate. We compare these estimators with $\mathbf{H}^{1\text{-bit ML}}$ from equation (26) on the preceding page, and with the perfect CSI estimate that uses the ground truth \mathbf{H} as the channel estimate.

Since NP BEACHES uses the median-based noise estimate (which in this case estimates the effective “noise” floor that includes quantization errors), it is robust to outliers and is able to achieve MSE and BER performance very close to 1-BEACHES that has perfect knowledge of the noise-plus-quantization power. The EM estimator, however, strongly relies on the distribution of the noise and signal being Gaussian. Here, the signal is a realistic channel vector which is not Gaussian; more importantly, $\tilde{\mathbf{Q}}$ contains the effect of noise but

also quantization error, which means the equivalent noise also deviates from a Gaussian distribution. We attribute the higher (worse) BER of EM BEACHES to these two factors. We note that 1-BEACHES is designed specifically for 1-bit quantization and that the expression for Q_0 (which requires knowledge of the noise power and the signal power) would be different if the ADCs use a different number of bits. In contrast, our nonparametric denoiser is agnostic to the quantizer’s resolution and automatically determines the power of the noise plus quantization, as long as the signal is approximately sparse and the noise is approximately Gaussian.

C. Cell-Free Communication System

We simulate an uplink cell-free communication system with $U = 16$ single-antenna UEs and $D = 256$ single-antenna BSs. The UEs and BSs are randomly placed with a uniform distribution in a square of side length 1 km. The UEs transmit orthogonal pilots followed by QPSK data. All of the UEs transmit simultaneously and the received signal at all the BSs is processed at a central processing unit (CPU) that performs channel estimation and linear minimum MSE detection.

We simulate a cell-free channel matrix \mathbf{H} using the model proposed by [48], with parameters as in [2] but without power control and with a transmit power of 12.5 mW per UE. As in equation (24) on page 12, the ML estimate of the channel matrix is obtained by right-multiplying the pilot sequence received at the CPU with the inverse of the orthogonal pilot matrix (we used a Hadamard pilot matrix), resulting in

$$\mathbf{H}^{\text{ML}} = \mathbf{H} + \mathbf{N}^{\text{CE}}. \quad (28)$$

The columns of \mathbf{H} (or *channel vectors*) contain the attenuations and phases between one UE and all BSs. For each UE, the BSs that have LoS or are closer to this UE will receive significantly higher power than the other BSs that are not nearby. This means that in the cell-free system, the channel vectors are approximately sparse [11] and the ML estimate can be denoised. Although the thermal noise variance at different basestations may differ, we assume i.i.d. noise in this paper.

Figure 8 shows the results of 10,000 Monte–Carlo trials. On the left, we plot the CDF of the MSE of the channel estimates,

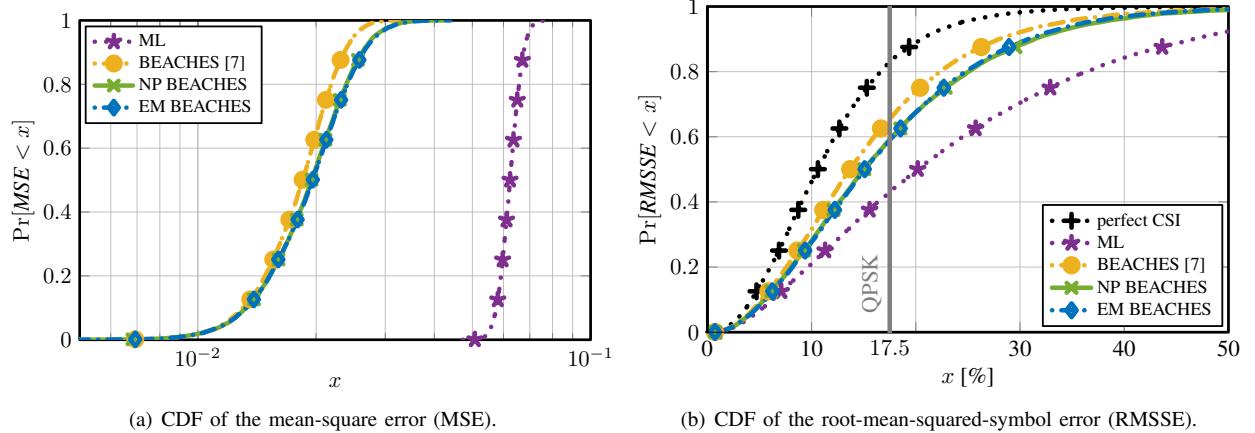


Fig. 8. Cumulative distribution function (CDF) of the (a) mean-square error (MSE) and (b) root-mean-squared-symbol error (RMSSE) of channel estimation methods in a cell-free wireless system.

and on the right, the CDF of the root-mean-squared-symbol-error (RMSSE). The RMSSE is a measure of how far the expected QPSK symbol is from the received data symbol after equalization with the channel estimates, and can be seen as equivalent to the error-vector-magnitude (EVM) for one UE.

In Figure 8, we observe a clear MSE improvement of the three denoising algorithms over the ML estimate: For a given value x , there are more realizations of channel estimates whose MSE is smaller than x for denoised channels than for ML. The fact that denoising improves the channel estimates is reflected in the RMSSE, since equalization is more effective and the obtained symbols are closer to the expected constellation points. We consider the RMSSE requirement of 17.5 % for QPSK from [49, Table 6.5.2.2-1]. The probability that a UE meets the requirement grows from 0.43 with ML channel estimation, to 0.59 with NP BEACHES or EM BEACHES denoising, an increase of 0.16. BEACHES with perfect knowledge of the noise power has a slight additional advantage, with a probability of meeting the requirement of 0.66.

VI. CONCLUSIONS

We have proposed blind estimators for the average noise power, signal power, SNR, and MSE. Our estimators can be calculated at low complexity and only require the noisy observation vector, avoiding the need for additional pilot signals entirely. We have analyzed our estimators for a Bernoulli complex Gaussian sparsity model and evaluated their accuracy via simulations. Using three channel-vector denoising tasks in (i) a multi-antenna mmWave system, (ii) a 1-bit quantized multi-antenna mmWave system, and (iii) a cell-free system, we have demonstrated that our blind estimators can be used to develop a novel nonparametric denoiser that achieves comparable performance and the same complexity as BEACHES in [7], [8] which requires knowledge of average noise power. We believe that the proposed blind estimators find potential use in a large number of other wireless communication applications that contain sparse complex-valued signals.

There are many avenues for future work. For signals that are less sparse (i.e., $p > 0.421$), one may want to replace the median by a higher quantile and the scaling factor $\log(2)$ needs

to be adjusted accordingly—a derivation of such estimators would follow immediately from our results in Section III-C. Huber M-estimators [50] combine the idea of mean and median, and they may also prove useful for blind noise power estimation in the presence of sparse signals. In the case of non-Gaussian, non-circularly-symmetric, or non-i.i.d sparse signals, new estimators can be tailored to exploit specific statistical properties (e.g., structured sparsity). Extending the statistical model, e.g., to signals with correlation or structured sparsity, can lead to improved estimators and is left for future work. In the case of colored noise (e.g., stemming from interference or large variations in radio-frequency circuitry), noise whitening techniques could be considered.

APPENDIX A PROOF OF THEOREM 1

A. Prerequisites

In what follows, we will need the distribution of $\mathbf{z} \triangleq |\mathbf{y}|^2$, where we assume \mathbf{y} is distributed according to Definition 4. Given a circularly-symmetric complex Gaussian RV A with variance E_a , the RV $B = |A|^2$ is exponentially distributed with CDF $F_B(b) \triangleq 1 - e^{-\frac{b}{E_a}}$, $b \geq 0$. Then, the CDF of each entry of the absolute-square noisy observation is as follows.

Definition 5 (Noisy BCG Power RV). *Let \mathbf{y} be as in Definition 4 and let $\mathbf{z} \triangleq |\mathbf{y}|^2$. Then, for $z_d \geq 0$, the CDF of each entry of \mathbf{z} is given by*

$$F_Z(z_d) \triangleq (1 - p) \left(1 - e^{-\frac{z_d}{N_0}} \right) + p \left(1 - e^{-\frac{z_d}{N_0 + E_s/p}} \right). \quad (29)$$

B. Upper Bounds on the Median

We start with the following two upper bounds on the median m_Z of a noisy BCG power RV Z with CDF given in equation (29).

Lemma 3. *For a noisy BCG power RV in Definition 5 with $p < 0.5$, the median is bounded from above by*

$$m_Z \leq N_0 \log \left(\frac{2 - 2p}{1 - 2p} \right). \quad (30)$$

Proof. We start from the definition of the median in equation (10) on page 6 for the RV Z with CDF as in equation (29) on the preceding page:

$$(1-p) \left(1 - e^{-\frac{m_Z}{N_0}}\right) + p \left(1 - e^{-\frac{m_Z}{N_0 + E_s/p}}\right) = \frac{1}{2}. \quad (31)$$

Since the second term is nonnegative, we can omit it to obtain the following inequality:

$$(1-p) \left(1 - e^{-\frac{m_Z}{N_0}}\right) \leq \frac{1}{2}. \quad (32)$$

Note that this bound will be useful for vectors s that are sparse, i.e., where p is small. We can simplify equation (32) as

$$1 - e^{-\frac{m_Z}{N_0}} \leq \frac{1}{2(1-p)} \quad (33)$$

$$\log \left(\frac{1-2p}{2-2p} \right) \leq -\frac{m_Z}{N_0}, \quad (34)$$

which leads to the upper bound on the median m_Z . In order to take the logarithm in equation (34), we require $p \in (0, 0.5)$. ■

Lemma 4. For a noisy BCG power RV Z in Definition 5 with $p \leq \frac{1/2 - e^{-2}}{1 - e^{-2}} \approx 0.421$, the median is bounded from above by

$$m_Z \leq \log(2)(N_0 + E_s). \quad (35)$$

Proof. We start from the definition of the median as in equation (31). Let us define the function $g(r) \triangleq e^{-1/r}$ with $r > 0$. We can now rewrite equation (31) as follows:

$$\frac{1}{2} = (1-p)g\left(\frac{N_0}{m_Z}\right) + pg\left(\frac{N_0 + E_s/p}{m_Z}\right). \quad (36)$$

The function $g(r)$ is concave for $r \geq 1/2$. Therefore, to ensure concavity of $g(r)$ in equation (36), we need

$$\frac{N_0}{m_Z} \geq \frac{1}{2} \quad \text{and} \quad \frac{N_0 + E_s/p}{m_Z} \geq \frac{1}{2}. \quad (37)$$

The two conditions in equation (37) are guaranteed as long as $2N_0 \geq m_Z$. Because CDFs are nondecreasing functions, requiring $2N_0 \geq m_Z$ is equivalent to requiring $F_Z(2N_0) \geq F_Z(m_Z) = 1/2$, which we can simplify as

$$\underbrace{(1-p) \left(1 - e^{-\frac{2N_0}{N_0}}\right) + p \left(1 - e^{-\frac{2N_0}{N_0 + E_s/p}}\right)}_{F_Z(2N_0)} \geq \underbrace{\frac{1}{2}}_{F_Z(m_Z)} \quad (38)$$

$$\frac{1}{2} - e^{-2} \geq p \left(e^{-\frac{2}{1+E_s/(pN_0)}} - e^{-2} \right). \quad (39)$$

Finally, to ensure equation (39) holds for all values of E_s and N_0 , we require

$$\frac{1}{2} - e^{-2} \geq p(1 - e^{-2}), \quad (40)$$

which implies that the condition $p \leq p^{\max}$ in equation (14) on page 7 ensures concavity of $g(r)$. Then, assuming $p \leq p^{\max}$, we can now use Jensen's inequality on the expression in equation (36) to get

$$\frac{1}{2} \leq g \left((1-p) \frac{N_0}{m_Z} + p \frac{N_0 + E_s/p}{m_Z} \right) = g \left(\frac{N_0 + E_s}{m_Z} \right). \quad (41)$$

We can now simplify this expression to

$$m_Z \leq \log(2)(N_0 + E_s), \quad (42)$$

which is the inequality in Lemma 4. ■

C. Lower Bound on the Median

We now establish the following lower bound on the median.

Lemma 5. For a noisy BCG power RV Z in Definition 5 with $p \in (0, 1]$, the median is bounded from below by

$$\frac{\log(2)N_0}{(1-p) + \frac{p^2}{p+SNR}} \leq m_Z. \quad (43)$$

Proof. We start from the definition of the median as in equation (31). Since the exponential CDF $F_B(b) \triangleq 1 - e^{-\frac{b}{E_a}}$ for $E_a \geq 0$ is concave in b , Jensen's inequality leads to

$$1 - e^{-(1-p)\frac{m_Z}{N_0} - p\frac{m_Z}{N_0 + E_s/p}} \geq \frac{1}{2}. \quad (44)$$

We can simplify this expression to obtain the following bound

$$\frac{1}{2} \geq e^{-(1-p)\frac{m_Z}{N_0} - p\frac{m_Z}{N_0 + E_s/p}} \quad (45)$$

$$\log(1/2) \geq -(1-p)\frac{m_Z}{N_0} - p\frac{m_Z}{N_0 + E_s/p}, \quad (46)$$

which leads to the inequality in Lemma 5 we wanted to prove. ■

D. Combining the Results

For all values of $p \in (0, 1]$ and $SNR \geq 0$, we have that

$$\frac{m_Z}{\log(2)} \left((1-p) + \frac{p^2}{p+SNR} \right) \leq \frac{m_Z}{\log(2)}, \quad (47)$$

and we defined \widehat{N}_0 such that $\widehat{N}_0 \xrightarrow[D \rightarrow \infty]{\text{prob.}} m_Z/\log(2)$ according to Lemma 1.

Finally, we can combine equation (47) with Lemma 3, Lemma 4 and Lemma 5 to obtain equation (17) on page 7.

APPENDIX B PROOF OF COROLLARY 1

Proof. Let the relative error of Estimator 1 be $\varepsilon \triangleq |\widehat{N}_0 - N_0|/N_0$. Using the inequalities from Theorem 1 and the quantities LB and UB defined there, we can bound ε as follows:

$$\lim_{D \rightarrow \infty} \frac{\widehat{N}_0 - UB}{UB} \leq \lim_{D \rightarrow \infty} \varepsilon \leq \lim_{D \rightarrow \infty} \frac{\widehat{N}_0 - LB}{LB}, \quad (48)$$

By using $\widehat{N}_0 \xrightarrow[D \rightarrow \infty]{\text{prob.}} m_Z/\log(2)$ and replacing LB from equation (15) on page 7 and UB from equation (16) on page 7 into equation (48), after some simplifications, we obtain equation (18) on page 7. ■

REFERENCES

- [1] A. Gallyas-Sanhueza and C. Studer, "Blind SNR estimation and non-parametric channel denoising in multi-antenna mmWave systems," in *IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–7.
- [2] H. Song, X. You, C. Zhang, O. Tirkkonen, and C. Studer, "Minimizing pilot overhead in cell-free massive MIMO systems via joint estimation and detection," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, May 2020, pp. 1–5.
- [3] T. Schenck, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer Science & Business Media, 2008.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[5] F. Rusek, D. Persson, B. Kiong, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[6] 3GPP, "5G; NR; user equipment (UE) radio transmission and reception; part 1: Range 1 standalone," Nov. 2020, TS 38.101-1 version 16.5.0 Rel. 16.

[7] R. Ghods, A. Gallyas-Sanhueza, S. H. Mirfarshbafan, and C. Studer, "BEACHES: Beamspace channel estimation for multi-antenna mmWave systems and beyond," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1–5.

[8] S. H. Mirfarshbafan, A. Gallyas-Sanhueza, R. Ghods, and C. Studer, "Beamspace channel estimation for massive MIMO mmWave systems: Algorithm and VLSI design," *IEEE Trans. Circuits Syst. I (TCAS-I)*, vol. 67, no. 12, pp. 5482–5495, Sep. 2020.

[9] A. Gallyas-Sanhueza, S. H. Mirfarshbafan, R. Ghods, and C. Studer, "Sparsity-adaptive beamspace channel estimation for 1-bit mmWave massive MIMO systems," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, May 2020, pp. 1–5.

[10] K. Upadhyay, C. S. Seelamantula, and K. Hari, "A risk minimization framework for channel estimation in OFDM systems," *Elsevier Signal Process.*, vol. 128, pp. 78–87, Nov. 2016.

[11] R. Gholamipourfard, "Cell-free massive MIMO receiver design and channel estimation," Ph.D. dissertation, Sorbonne Université, Paris, France, Dec. 2021.

[12] F. Socheleau, D. Pastor, A. Aissa-El-Bey, and S. Houcke, "Blind noise variance estimation for OFDMA signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2009, pp. 2581–2584.

[13] K. Wang and X. Zhang, "Blind noise variance and SNR estimation for OFDM systems based on information theoretic criteria," *Elsevier J. Signal Process.*, vol. 90, no. 9, pp. 2766 – 2772, Mar. 2010.

[14] M. Zivkovic and R. Mathar, "Preamble-based SNR estimation in frequency selective channels for wireless OFDM systems," in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, Apr. 2009, pp. 1–5.

[15] A. Das and B. D. Rao, "SNR and noise variance estimation for MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3929–3941, Apr. 2012.

[16] S. Huang and T. D. Tran, "1-bit compressive sensing via approximate message passing with built-in parameter estimation," *arXiv preprint: 2007.07679*, Jul. 2020.

[17] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Jul. 2004.

[18] S. Chrétienn and S. Darses, "Sparse recovery with unknown variance: A LASSO-type approach," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3970–3988, Mar. 2014.

[19] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," in *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, Nov. 2009, pp. 18914–18919.

[20] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2011, pp. 2168–2172.

[21] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.

[22] R. Tibshirani, "Stein's unbiased risk estimate," Course Notes for Statistical Machine Learning, Carnegie Mellon University, Spring 2015.

[23] A. Mousavi, A. Maleki, and R. G. Baraniuk, "Consistent parameter estimation for LASSO and approximate message passing," *Ann. Statist.*, vol. 45, no. 6, pp. 2427–2454, Dec. 2017.

[24] R. Ghods, C. Jeon, G. Mirza, A. Maleki, and C. Studer, "Optimally-tuned nonparametric linear equalization for massive MU-MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2118–2122.

[25] P. Maechler, C. Studer, D. E. Bellasi, A. Maleki, A. Burg, N. Felber, H. Kaeslin, and R. G. Baraniuk, "VLSI design of approximate message passing for signal restoration and compressive sensing," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 2, no. 3, pp. 579–590, Sep. 2012.

[26] L. Bai, P. Maechler, M. Muehlerhuber, and H. Kaeslin, "High-speed compressed sensing reconstruction on FPGA using OMP and AMP," in *Proc. IEEE Int. Conf. Electron., Circuits, Syst. (ICECS)*, Dec. 2012, pp. 53–56.

[27] H. Yan, S. Ramesh, T. Gallagher, C. Ling, and D. Cabric, "Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures," *IEEE Circuits Syst. Mag.*, vol. 19, no. 2, pp. 33–58, May 2019.

[28] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek, "A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 484–498, Jun. 2018.

[29] P. Skrimponis, S. Dutta, M. Mezzavilla, S. Rangan, S. H. Mirfarshbafan, C. Studer, J. Buckwalter, and M. Rodwell, "Power consumption analysis for mobile mmWave and sub-THz receivers," in *6G Wireless Summit (6G SUMMIT)*, Mar. 2020, pp. 1–5.

[30] S. Haene, A. Burg, P. Luethi, N. Felber, and W. Fichtner, "FFT processor for OFDM channel estimation," in *Proc. IEEE Int. Symp. Circuits and Syst. (ISCAS)*, May 2007, pp. 1417–1420.

[31] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Sep. 1994.

[32] H. Nahrstaedt and B. Strunz. PyYAWT. Accessed Sep. 2, 2021. [Online]. Available: <https://pyyawt.readthedocs.io>

[33] M. Alexander, R. Baumgartner, C. Windischberger, E. Moser, and R. Somorjai, "Wavelet domain de-noising of time-courses in MR image sequences," *Magn. Reson. Imag.*, vol. 18, no. 9, pp. 1129–1134, Aug. 2000.

[34] S. Jacobsson, U. Gustavsson, G. Durisi, and C. Studer, "Massive MU-MIMO-OFDM uplink with hardware impairments: Modeling and analysis," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1829–1835.

[35] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[36] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Apr. 2017.

[37] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *J. American Stat. Association*, vol. 88, no. 424, pp. 1273–1283, Sep. 1993.

[38] P. J. Huber, *Robust statistics*. Wiley, 2004.

[39] R. J. Tibshirani, "Fast computation of the median by successive binning," *arXiv preprint: 0806.3301*, May 2009.

[40] A. Alexandrescu, "Fast deterministic selection," in *Int. Symp. Exp. Algor. (SEA)*, vol. 75, Aug. 2017, pp. 24:1–24:19.

[41] T. Pham-Gia and T. L. Hung, "The mean and median absolute deviations," *Math. Comput. Model.*, vol. 34, no. 7, pp. 921–936, Oct. 2001.

[42] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Series in Statistics, 2001.

[43] A. Merberg and S. J. Miller, "The sample distribution of the median," Course Notes for Math 162: Mathematical Statistics, Williams College, Feb. 2008.

[44] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2011, pp. 799–803.

[45] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 2013.

[46] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "QuaDRiGa - quasi deterministic radio channel generator, user manual and documentation," Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2.2.0, Jun. 2019.

[47] L. V. Nguyen and D. H. Nguyen, "Linear receivers for massive MIMO systems with one-bit ADCs," *arXiv preprint: 1907.06664*, Jul. 2019.

[48] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[49] 3GPP, "5G; NR; base station (BS) radio transmission and reception," May 2019, TS 38.104 version 15.5.0 Rel. 15.

[50] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Stat.*, vol. 35, no. 1, pp. 73 – 101, Mar. 1964.



Alexandra Gallyas-Sanhueza (Graduate Student Member, IEEE) received the B.Sc. degree in Electrical Engineering from Pontificia Universidad Católica de Chile, Santiago, Chile, in 2015. She is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. From 2019 to 2020, she was at Cornell Tech, New York City, NY, USA. Her research interests include signal processing, algorithm design, statistics, and wireless communications. In 2019, she received the Qualcomm Innovation Fellowship.



Christoph Studer (Senior Member, IEEE) is an Associate Professor at the Department of Information Technology and Electrical Engineering at ETH Zurich in Switzerland. He received his M.S. and Ph.D. degrees in Electrical Engineering from ETH Zurich in 2006 and 2009, respectively. From 2009 to 2012, he was a Postdoctoral Researcher at ETH Zurich and Rice University in Houston, TX. In 2013, he was a Research Scientist at Rice University. From 2014 to 2019, he was an Assistant Professor at Cornell University, Ithaca, NY. From 2019 to 2020, he was an Associate Professor at Cornell University and at Cornell Tech in New York City. In 2020, Dr. Studer joined ETH Zurich.

Dr. Studer's research interests are at the intersection of wireless communications, digital signal processing, machine learning, and digital VLSI design.

Dr. Studer received ETH Medals for his M.S. and Ph.D. theses in 2006 and 2009, respectively. He received a Swiss National Science Foundation fellowship for Advanced Researchers in 2011 and a US National Science Foundation CAREER Award in 2017. He shared the Swisscom/ICTnet Innovations Award in both 2010 and 2013. Dr. Studer was the winner of the Student Paper Contest of the 2007 Asilomar Conf. on Signals, Systems, and Computers and received a Best Student Paper Award of the 2008 IEEE Int. Symp. on Circuits and Systems (ISCAS). He shared the best Live Demonstration Award at the IEEE ISCAS in 2013 and the European Solid-state Devices and Circuits Conference Best Paper Award in 2021. Dr. Studer is an Associate Editor for the IEEE Open Journal of Circuits and Systems, the IEEE Communications Letters, and the IEEE Transactions on Circuits and Systems II: Express Briefs. In 2019 and 2022, he was the Technical Program Chair and the General Chair of the Asilomar Conference on Signals, Systems, and Computers, respectively.