

# Massive MIMO-OFDM Systems with Low Resolution ADCs: Cramér-Rao Bound, Sparse Channel Estimation, and Soft Symbol Decoding

Sai Subramanyam Thoota and Chandra R. Murthy

**Abstract**—We consider the delay-domain sparse channel estimation and data detection/decoding problems in a massive multiple-input-multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) wireless communication system with low-resolution analog-to-digital converters (ADCs). The non-linear distortion due to coarse quantization leads to severe performance degradation in conventional OFDM receivers, which necessitates novel receiver techniques. First, we derive Bayesian Cramér-Rao-lower-bounds (CRLB) on the mean squared error (MSE) in recovering jointly compressible vectors from quantized noisy underdetermined measurements. Second, we formulate the pilot-assisted channel estimation as a multiple measurement vector (MMV) sparse recovery problem, and develop a variational Bayes (VB) algorithm to infer the posterior distribution of the channel. We benchmark the MSE performance of our algorithm with that of the CRLB, and numerically show that the VB algorithm meets the CRLB. Third, we present a soft symbol decoding algorithm that infers the posterior distributions of the data symbols given the quantized observations. We utilize the posterior statistics of the detected data symbols as virtual pilots, and propose an iterative soft symbol decoding and data-aided channel estimation procedure. Finally, we present a variant of the iterative algorithm that utilizes the output bit log-likelihood ratios of the channel decoder to adapt the data prior to further improve the performance. We provide interesting insights into the impact of the various system parameters on the MSE and bit error rate of the proposed algorithms, and benchmark them against the state-of-the-art.

**Index Terms**—Channel estimation, Cramér-Rao lower bound, ADCs, massive MIMO, soft symbol decoding, variational Bayes.

## I. INTRODUCTION

Recent research in wireless communications has investigated the use of a massive number of antennas at the base station (BS) to increase the network capacity and data rates [2]. The benefits of massive multiple input multiple output (MIMO) communications are now very well understood. However, they come at the expense of high power consumption and hardware cost, which needs to be addressed to make massive MIMO commercially viable. One potential solution is to employ low-resolution analog-to-digital converters (ADCs) in the receivers [3]–[5]. The power consumption of an ADC increases exponentially with its bit-width. Hence,

in massive MIMO systems with one RF chain per antenna, employing low-resolution ADCs can result in dramatic power savings [6], [7]. Further, low resolution ADCs relax the stringent linearity range requirements on the RF circuitry, which in turn reduces the hardware cost [8]. However, they also bring new challenges in the design of receivers, as advanced signal processing techniques need to be used to counter the large quantization noise introduced by them. This paper investigates several key aspects of receiver design, and develops novel receiver architectures in the context of multiuser massive MIMO orthogonal frequency division multiplexing (OFDM) communication with low resolution ADCs.

Three main challenges arise in the use of low resolution ADCs in multi-user MIMO-OFDM systems. First, the nonlinearities introduced by coarse quantizers lead to sub-optimal performance of conventional receivers such as (regularized) zero-forcing (ZF/RZF) and minimum mean square error (MMSE) detectors [9]. In a conventional OFDM receiver, we remove the cyclic prefix (CP), decouple the subcarriers using a discrete Fourier transform (DFT), and perform frequency domain equalization on a per-subcarrier basis. However, in low resolution ADC based systems, we obtain the complex baseband time-domain samples after being coarsely quantized by the ADC, and it is not possible to decouple the subcarriers by a DFT operation, resulting in inter-carrier interference (ICI). Due to this, conventional receivers may perform poorly when low-resolution ADCs are employed.

Second, the pilot signals transmitted by the user equipment (UE) for channel estimation at the BS are also received through the low resolution ADCs. This necessitates the use of long pilot sequences for accurate channel estimation, leading to a loss in spectral efficiency [4], [10], [11].

Third, a channel encoder and decoder are integral parts of any commercial wireless communication system, and are used to correct for errors introduced by the channel. The channel decoders require the bit *log-likelihood ratios* (LLRs), rather than hard bit-decisions, to provide good performance. The bit LLRs are a function of the posterior beliefs (probabilities) of the data symbols. Therefore, the aim of the receiver is not only to detect the data symbols, but also to obtain their posterior beliefs (also known as *soft symbols*), based on the quantized observations obtained from the low resolution ADCs.

We now briefly review the existing literature on channel estimation and data detection in low resolution ADC based multi-user MIMO systems, before presenting the key contributions in this paper. Channel estimation in massive MIMO

Sai Subramanyam Thoota and Chandra R. Murthy are with the Department of ECE, Indian Institute of Science, Bangalore, India 560 012 (email: thoota@iisc.ac.in; cmurthy@iisc.ac.in).

A part of this work was published in [1].

This work was funded by research grants from Intel Inc. and the Young Faculty Research Fellowship, Govt. of India.

systems was considered in [4], [5], [12], while [13]–[20] develop data detection methods in massive MIMO single carrier (SC) and multi-carrier systems. Joint/iterative channel estimation and data detection was considered in [8], [21]–[25]. In [8], the authors develop a single iteration multiuser MIMO-OFDM channel estimator using convex optimization techniques, and a data detector using a suboptimal soft-output MMSE algorithm. A bilinear generalized approximate message passing (BiGAMP) algorithm to solve the joint channel estimation and data detection problem is developed in [21], [23]. The authors in [21] also analyze the performance of the BiGAMP algorithm using the replica method. Recently, a variational Bayesian (VB) channel estimation and data detection algorithm was developed in [25], in the context of a single-user single input single output (SU-SISO) OFDM system. While [25] restricts to a single OFDM symbol, we consider a more general multiple pilot and data symbols model in a multi-user massive MIMO-OFDM system.

An angular domain joint sparse channel estimation and data detection algorithm using the sparse Bayesian learning (SBL) framework in a hybrid millimeter wave communication system was proposed in [22]. The idea here is to utilize the decoded data symbols as virtual pilots for channel estimation. The receiver starts by forming an initial estimate of the channel using the pilot symbols, which is used to detect the data symbols. Then, in subsequent iterations, the detected data symbols are used as virtual pilots to refine the channel estimates and re-estimate the data symbols. This process is repeated until a suitable convergence condition is satisfied [26]. In [24], the authors adopt a supervised learning framework to solve the single-iteration non-sparse channel estimation and data detection problems in a massive MIMO-OFDM system with single pilot and data symbol using one-bit measurements. These approaches usually require careful parameter tuning for fast convergence and accurate data detection. Moreover, several heuristics are required to transform the detected data symbols into soft outputs which are required for the subsequent channel decoding. Furthermore, none of the above mentioned papers directly address all three challenges mentioned in the preceding paragraphs.

In this paper, we develop an iterative delay-domain *sparse channel estimation* and *soft symbol decoding* algorithm for a massive MIMO-OFDM system with low resolution ADCs. As a first step, we develop a benchmark to evaluate the performance of any sparse channel estimator with measurements acquired using low resolution ADCs. To this end, we consider a general quantized compressed sensing problem, and derive different types of Cramér-Rao lower bounds (CRLBs) on the mean squared error (MSE) performance of an estimator [27], [28]. We impose a two-stage hierarchical circularly symmetric complex Gaussian prior on the estimand (in our case, the channel) parameterized by a diagonal precision matrix. We further impose a non-informative conjugate Gamma hyper-prior on the diagonal elements of the precision matrix. This results in a Student's  $t$ -distributed marginalized prior on the estimand, which is heavy-tailed and hence promotes sparse solutions.

It is worth mentioning that CRLBs for the compressed

sensing problem with unquantized measurements have been derived in [29]. In [30], [31], the CRLB on the MSE of an estimator with 1-bit measurements is derived under a non-sparse setting. While [30] derives the CRLB in a deterministic setup, [31] obtains the Bayesian CRLB. To the best of our knowledge, different types of CRLBs for the estimation of jointly compressible vectors [32] from multi-bit quantized noisy underdetermined measurements does not exist in the literature. We develop a CRLB for this case in Sec. II. It turns out that the expectations required to obtain the Bayesian information matrix (BIM) are computationally intractable, and, consequently, the CRLB cannot be obtained in closed form. We therefore resort to numerical methods for evaluating the bound. While our CRLB for the quantized compressed sensing problem is of independent interest, we empirically illustrate its utility in the context of sparse massive MIMO-OFDM channel estimation by comparing it with the MSE performance of our algorithm.

Next, we use a statistical inference framework to compute the posterior distributions of the UEs' channels and data symbols given the quantized received pilot and data observations. We adopt a minorization-maximization based procedure called variational Bayesian (VB) inference, which is a principled approach for developing low-complexity algorithms for high-dimensional inference problems with guaranteed convergence from any initialization. The key novelty lies in how we construct the underlying probabilistic graphical models and how we identify and group the latent variables. The latent variables can also be used to compute side information such as the signal-to-noise ratio (SNR), which can, in turn, be used for link adaptation. Our main contributions are as follows:

- We derive the Bayesian CRLB for the MSE incurred by an estimator for recovering jointly compressible vectors from quantized compressed sensing measurements. Specifically, we impose a hierarchical circularly symmetric complex Gaussian prior on the estimand, parameterized by a diagonal precision matrix. The precision matrix is in turn hyper-parameterized by a Gamma distribution. Although the CRLB is not available in closed-form, it can be evaluated by numerical methods.
- We consider both deterministic and random cases for the precision matrix to obtain two different CRLBs on the MSE of jointly compressible vectors. Also, in the case of exactly-sparse signals, we derive a support-aware CRLB, which assumes the knowledge of the support set, to compute the bound. We analytically show that our derived CRLB subsumes both the unquantized and 1-bit CRLBs as special cases.
- We exploit the lag/delay domain sparsity of the channels to formulate channel estimation in a massive MIMO-OFDM system as a quantized MMV sparse signal recovery problem. We develop a VB algorithm to infer the posterior distributions of the channels. We benchmark the MSE performance of the VB channel estimator with the derived Bayesian CRLB, and empirically show that the estimator meets the CRLB.
- We then develop a quantized VB soft symbol decoding

algorithm that uses the estimated channels to obtain the posterior beliefs of the data symbols. We use these posterior statistics to generate virtual pilots, and present a data-aided channel estimation procedure to refine the initial channel estimates. Based on this, we develop an iterative algorithm that alternately runs the soft symbol decoder and data-aided channel estimator steps. Finally, we generate the bit LLRs from the posterior symbol probabilities, and input them to the channel decoder.

- We also present a variant of the iterative channel estimation and data decoding algorithm, which utilizes the a posteriori bit LLRs output from the channel decoder to adapt the prior used by the data detector. The resulting combined channel estimator, data detector and channel decoder further improves the system performance.

We evaluate the normalized MSE (NMSE) and coded bit-error-rate (BER) performance of the VB algorithms, and benchmark it against the state-of-the-art BiGAMP based joint channel estimator and data detector [21] and the conventional soft MMSE detector. Further, we study the impact of the system parameters on the performance of our algorithm, and provide several interesting insights.

One of the main takeaways from our work is that VB is a powerful and flexible technique for designing receivers in massive MIMO-OFDM systems, particularly when the BS employs low resolution ADCs. This is because the subcarriers are no longer orthogonal after the quantization step. Due to this, conventional subcarrier-by-subcarrier data detection performs poorly (See Fig. 11). Also, our choice of latent variables and approximate posterior distributions is crucial for obtaining analytically and computationally tractable solutions. Another key takeaway is that the assumption of perfect CSI at the receiver significantly overestimates the system performance, which we illustrate through empirical studies in Sec. VII. Therefore, it is important to account for channel estimation errors while designing receivers, especially when both received pilots and data are coarsely quantized.

*Notation:* We denote matrices, vectors and scalars by boldface upper case, boldface lower case, and lowercase letters, respectively.  $\mathbf{A}^T$ ,  $\mathbf{A}^H$  and  $|\mathbf{A}|$  denote the transpose, conjugate transpose, and determinant of  $\mathbf{A}$ , respectively.  $\mathbf{A} \otimes \mathbf{B}$  denotes the Kronecker product of  $\mathbf{A}$  and  $\mathbf{B}$ .  $\text{diag}(\mathbf{x})$  returns a diagonal matrix with the entries of  $\mathbf{x}$  on the diagonal.  $\mathbb{E}$  and  $\langle \cdot \rangle$  both denote the expectation operation.  $f(x) \triangleq \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$  and  $F(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{t^2}{2}) dt$  denote the probability density and cumulative distribution functions of a standard normal random variable evaluated at  $x$ .  $\Gamma(a) \triangleq \int_0^{\infty} t^{a-1} \exp(-t) dt$  denotes the Gamma function evaluated at  $a > 0$ .  $\mathbf{I}_M$ ,  $\mathbf{0}_M$  and  $\mathbf{1}_M$  denote an  $M \times M$  identity matrix,  $M \times M$  zero matrix and all-ones vector of size  $M \times 1$ , respectively.  $\Re$  and  $\Im$  are the real and imaginary part operators, respectively.

## II. QUANTIZED COMPRESSED SENSING AND BAYESIAN CRAMÉR-RAO LOWER BOUND

We consider the estimation of high-dimensional jointly compressible vectors  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T] \in \mathbb{C}^{N \times T}$  from quantized low-dimensional measurements  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_T] \in$

$\mathbb{C}^{M \times T}$ , where  $M < N$ .<sup>1</sup> The measurements are obtained as

$$\mathbf{Y} = \mathcal{Q}_b(\Phi \mathbf{X} + \mathbf{W}), \quad (1)$$

where  $\Phi \in \mathbb{C}^{M \times N}$  is a known measurement matrix, and  $\mathbf{W} \in \mathbb{C}^{M \times T}$  is the additive noise matrix whose entries are independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables with mean 0 and variance  $\sigma_w^2$ .  $\mathcal{Q}_b(\cdot)$  denotes an element-wise scalar  $b$ -bit quantizer of both real and imaginary components of its argument. We assume a common support structure on the columns of  $\mathbf{X}$ . In many applications, the signals are not exactly sparse, i.e., many entries may not be exactly equal to zero. An example is the effective wireless channel with the non-ideal transmit and receive filters. Therefore, we consider compressible signals [32] instead, where there are only a few entries with high magnitude and the remaining entries have very low magnitude. Here, by common support structure, we mean that the indices of the large magnitude entries are the same in each column of  $\mathbf{X}$ .

A  $b$ -bit quantizer on a real valued input  $z$  is defined as  $\mathcal{Q}_b(z) = L_i$ ,  $z \in [\delta_i, \delta_{i+1})$ ,  $i = 0, 1, \dots, B-1$ , where  $B = 2^b$  is the number of quantization levels,  $-\infty = \delta_0 < \delta_1 < \dots < \delta_B = \infty$  are the quantization thresholds, and  $L_0, L_1, \dots, L_{B-1}$  are the quantizer outputs. We now derive the Bayesian CRLB on the MSE of any estimator of  $\mathbf{X}$ .

To develop the CRLB, we impose a two-stage hierarchical prior on  $\mathbf{X}$  [33]. That is,  $\mathbf{x}_\ell \sim \mathcal{CN}(\mathbf{x}_\ell; \mathbf{0}, \mathbf{P}^{-1}) \forall \ell$ , where  $\mathbf{P}$  is a diagonal precision matrix containing the hyperparameters  $\alpha = [\alpha_1, \dots, \alpha_N]^T$ . As mentioned earlier, we assume a non-informative conjugate Gamma hyperprior on  $\alpha_n$ ,  $\forall n$  with shape and rate parameters  $a$  and  $r$ , respectively:

$$p(\mathbf{X} | \mathbf{P}) = \prod_{\ell=1}^T \frac{|\mathbf{P}|}{\pi^N} \exp(-\mathbf{x}_\ell^H \mathbf{P} \mathbf{x}_\ell), \quad (2)$$

$$p(\alpha; a, r) = \prod_{n=1}^N \frac{r^a}{\Gamma(a)} \alpha_n^{a-1} \exp(-r \alpha_n), \quad (3)$$

where  $|\mathbf{P}|$  denotes the determinant of  $\mathbf{P}$  and  $\Gamma(a)$  denotes the Gamma function.

Now, we compute the BIM for the above model. For this, we need the joint probability distribution  $p(\mathbf{Y}, \mathbf{X}, \mathbf{P}; \Phi, \sigma_w^2)$ . Denoting the unquantized measurements by  $\mathbf{Z}$ , we write (1) as  $\mathbf{Y} = \mathcal{Q}_b(\mathbf{Z})$ , where  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_T] \in \mathbb{C}^{M \times T}$ . It is convenient to transform the system from the complex field to the real field as follows:

$$\tilde{\Phi} = \begin{bmatrix} \Re(\Phi) & -\Im(\Phi) \\ \Im(\Phi) & \Re(\Phi) \end{bmatrix}, \quad \tilde{\mathbf{x}}_\ell = \begin{bmatrix} \Re(\mathbf{x}_\ell) \\ \Im(\mathbf{x}_\ell) \end{bmatrix}, \quad \tilde{\mathbf{y}}_\ell = \begin{bmatrix} \Re(\mathbf{y}_\ell) \\ \Im(\mathbf{y}_\ell) \end{bmatrix},$$

$$\tilde{\mathbf{z}}_\ell = \begin{bmatrix} \Re(\mathbf{z}_\ell) \\ \Im(\mathbf{z}_\ell) \end{bmatrix}, \quad \tilde{\mathbf{w}}_\ell = \begin{bmatrix} \Re(\mathbf{w}_\ell) \\ \Im(\mathbf{w}_\ell) \end{bmatrix}, \quad \ell = 1, \dots, T, \quad (4)$$

where  $\Re(\cdot)$  and  $\Im(\cdot)$  denote the real and imaginary part operators, respectively. Let us denote  $\tilde{\sigma}_w^2 = \frac{\sigma_w^2}{2}$ ,  $\tilde{N} = 2N$ ,  $\tilde{M} = 2M$ . In (4),  $\mathbf{w}_\ell$  is the  $\ell^{\text{th}}$  column of  $\tilde{\mathbf{W}}$ . Now, the system model becomes  $\tilde{\mathbf{Y}} = \mathcal{Q}_b(\tilde{\mathbf{Z}}) = \mathcal{Q}_b(\tilde{\Phi} \tilde{\mathbf{X}} + \tilde{\mathbf{W}})$ .

<sup>1</sup>In the context of massive MIMO-OFDM communication systems as described in Section III,  $\mathbf{Y}$  and  $\mathbf{X}$  will denote the quantized received signal at the BS and the channel matrix, respectively.

Let us denote the precision matrix for the real field by  $\tilde{\mathbf{P}}$  ( $\triangleq \text{diag}([2\alpha^T, 2\alpha^T]^T)$ ), where  $\text{diag}(\cdot)$  returns a diagonal matrix. Now, the prior becomes

$$p(\tilde{\mathbf{X}} | \tilde{\mathbf{P}}) = \prod_{\ell=1}^T \frac{|\tilde{\mathbf{P}}|^{1/2}}{(2\pi)^{\frac{N}{2}}} \exp\left(-\frac{\tilde{\mathbf{x}}_\ell^T \tilde{\mathbf{P}} \tilde{\mathbf{x}}_\ell}{2}\right). \quad (5)$$

Since the columns of  $\tilde{\mathbf{X}}$  are independent of each other, the BIM has a block diagonal structure with the off diagonal blocks as all-zero matrices. With this prior, we present the expression for the BIM in the following theorem.

*Theorem 1:* The  $\ell^{\text{th}}$  diagonal block of the BIM required to compute the CRLB for the MSE of a Bayesian sparse signal estimator using quantized compressive measurements is given by (6) on the next page, where the expectation  $\mathbb{E}[\cdot]$  is w.r.t. the joint probability distribution  $p(\tilde{\mathbf{Y}}, \tilde{\mathbf{X}}, \tilde{\mathbf{P}}; \tilde{\Phi}, \tilde{\sigma}_w^2, a, r)$ ,

$$\tilde{\eta}_{m\ell}^{(\text{hi})} \triangleq \frac{\tilde{z}_{m\ell}^{(\text{hi})} - \sum_{n=1}^{\tilde{N}} \tilde{\Phi}_{mn} \tilde{x}_{n\ell}}{\tilde{\sigma}_w}, \quad (7)$$

$$\tilde{\eta}_{m\ell}^{(\text{lo})} \triangleq \frac{\tilde{z}_{m\ell}^{(\text{lo})} - \sum_{n=1}^{\tilde{N}} \tilde{\Phi}_{mn} \tilde{x}_{n\ell}}{\tilde{\sigma}_w}, \quad (8)$$

where  $\ell \in \{1, \dots, T\}$ ,  $\tilde{z}_{m\ell}^{(\text{lo})}$  and  $\tilde{z}_{m\ell}^{(\text{hi})}$  are the lower and upper quantization thresholds corresponding to the  $(m, \ell)^{\text{th}}$  entry of  $\tilde{\mathbf{Y}}$ , respectively.  $\tilde{\Phi}_{mn}$  and  $\tilde{x}_{n\ell}$  denote the  $(m, n)^{\text{th}}$  and  $(n, \ell)^{\text{th}}$  entries of  $\tilde{\Phi}$  and  $\tilde{\mathbf{X}}$ , respectively.  $f(\cdot)$  and  $F(\cdot)$  denote the probability density function (PDF) and cumulative distribution functions (CDF) of a standard normal random variable, respectively.

*Proof:* The result follows from direct computation of the BIM, and is detailed in the supplementary material. ■

Note that the BIM depends on the probability distribution of the hyperparameters only through their expected values.

We provide a step by step procedure to compute the Bayesian CRLB using (6). Given a realization of  $\mathbf{X}$  generated according to a prior distribution  $p(\mathbf{X}|\mathbf{P})$ , we denote the instance specific BIM at iteration  $\tau$  as  $\tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2)$  which is given by (9). Here, we use the law of iterated expectation, and the range of summation over  $\tilde{y}_{m\ell}$  is the number of discrete quantization levels. We provide a recipe to compute the instance specific BIM for the multi-bit quantized compressed sensing case in Algorithm 1.

Next, we use the chain rule to convert the real valued BIM to a complex valued BIM as

$$\begin{aligned} \mathbf{M}_\ell^{(\tau)}(\Phi, a, r, \sigma_w^2) &= \\ &= \frac{1}{4} \left( \left[ \tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2) \right]_{\Re\Re} + \left[ \tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2) \right]_{\Im\Im} \right) \\ &\quad + \frac{j}{4} \left( \left[ \tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2) \right]_{\Re\Im} - \left[ \tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2) \right]_{\Im\Re} \right), \end{aligned}$$

where  $\mathbf{M}_\ell^{(\tau)}(\Phi, a, r, \sigma_w^2)$  is the  $\ell^{\text{th}}$  diagonal block of the complex BIM. Finally, the instance specific CRLB is

$$\begin{aligned} \text{CRLB}^{(\tau)}(\Phi, a, r, \sigma_w^2) &= \\ &= \text{blkdiag} \left[ \left( \left[ \mathbf{M}_\ell^{(\tau)}(\Phi, a, r, \sigma_w^2) \right]^{-1} \right)^T \right]_{\ell=1}^T, \quad (10) \end{aligned}$$

where  $\text{blkdiag}(\cdot)$  returns a block diagonal matrix. We use

---

### Algorithm 1 Computation of $\tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2)$

---

**Input:**  $\tilde{\Phi}, \tilde{\mathbf{X}}, \{\tilde{z}_{m\ell}^{(\text{lo})}, \tilde{z}_{m\ell}^{(\text{hi})}\}_{m=1}^M, a, r, \tilde{\sigma}_w$ .

**Output:**  $\tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2)$ .

- 1: Initialize  $\mathbf{J} = \mathbf{0}_{\tilde{M}}$ .
- 2: **for**  $m = 1$  to  $\tilde{M}$  **do**
- 3:   **for**  $\tilde{y}_{m\ell} = \{L_0, \dots, L_{B-1}\}$  **do**
- 4:     Compute  $\tilde{\eta}_{m\ell}^{(\text{hi})}$  and  $\tilde{\eta}_{m\ell}^{(\text{lo})}$  using (7) and (8).
- 5:     Compute

$$\begin{aligned} \mathbf{J}(m, m) &= \mathbf{J}(m, m) + \tilde{\eta}_{m\ell}^{(\text{hi})} f(\tilde{\eta}_{m\ell}^{(\text{hi})}) - \tilde{\eta}_{m\ell}^{(\text{lo})} f(\tilde{\eta}_{m\ell}^{(\text{lo})}) \\ &\quad + \frac{\left( f(\tilde{\eta}_{m\ell}^{(\text{hi})}) - f(\tilde{\eta}_{m\ell}^{(\text{lo})}) \right)^2}{F(\tilde{\eta}_{m\ell}^{(\text{hi})}) - F(\tilde{\eta}_{m\ell}^{(\text{lo})})} \end{aligned}$$

- 6:   **end for**
  - 7:    $\mathbf{J}(m, m) = \tilde{\sigma}_w^{-2} \mathbf{J}(m, m)$ .
  - 8: **end for**
  - 9: Compute  $\tilde{\mathbf{M}}_\ell^{(\tau)}(\tilde{\Phi}, a, r, \tilde{\sigma}_w^2) = \tilde{\Phi}^T \mathbf{J} \tilde{\Phi} + \frac{a}{r} \mathbf{I}_{\tilde{N}}$ .
- 

the inverse property of block diagonal matrices to obtain (10), which reduces the complexity in computing the CRLB. We vary  $\tau$  from 1 to  $\tau_{max}$ , compute the instance specific CRLB using independent random realizations of  $\mathbf{X}$ , and then compute the average Bayesian CRLB as

$$\text{CRLB}(\Phi, a, r, \sigma_w^2) = \frac{1}{\tau_{max}} \sum_{\tau=1}^{\tau_{max}} \text{CRLB}^{(\tau)}(\Phi, a, r, \sigma_w^2). \quad (11)$$

The MSE of an estimator is lower bounded by the trace of the CRLB in (11). In the derivation above, we consider that the precision matrix  $\mathbf{P}$  is random, which leads to a Bayesian bound. We can also consider the case where the precision matrix is deterministic. In this context, we contrast three types of bounds: (a) Support-aware Bayesian CRLB: Precision matrix is random, but the support set is known. (b) Hybrid CRLB: Random  $\mathbf{X}$  parameterized by a deterministic  $\mathbf{P}$ . (c) Bayesian CRLB: Random  $\mathbf{X}$  parameterized by a random  $\mathbf{P}$  with a conjugate hyperprior.

For the support-aware Bayesian CRLB, the BIM is computed as follows: For the columns in the support set, the diagonal entries of  $\mathbb{E}[\mathbf{P}]$  in the complex BIM are equal to  $\frac{a}{r}$ , and the remaining columns are removed from the measurement matrix, to compute the CRLB. We note that the support-aware Bayesian CRLB provides a lower bound on the MSE in the estimation of exact sparse vectors.

In the case of hybrid CRLB, the expectation term  $\mathbb{E}[\tilde{\mathbf{P}}]$  in (6) is replaced by the deterministic and known  $\tilde{\mathbf{P}}$ . We generate a compressible signal using a generative model with a circularly symmetric complex normal prior parameterized by the known precision matrix, and average the CRLB over multiple realizations as in (11).

For the Bayesian CRLB, we use the mean of the Gamma hyperprior for  $\mathbb{E}[\mathbf{P}] = \frac{a}{r} \mathbf{I}_N$  in the complex case. Note that this does not require the realization of the precision parameters; it only depends on the shape and rate parameters.

Next, we consider the two special cases, namely, 1-bit and infinite-bit quantization of the noisy compressive measure-

$$\begin{aligned} \widetilde{\mathbf{M}}_\ell(\widetilde{\Phi}, a, r, \widetilde{\sigma}_w^2) &= \mathbb{E} \left[ -\frac{\partial^2}{\partial \widetilde{\mathbf{x}}_\ell \partial \widetilde{\mathbf{x}}_\ell^T} \log p(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}, \widetilde{\mathbf{P}}; \widetilde{\Phi}, \widetilde{\sigma}_w^2, a, r) \right] \\ &= \widetilde{\Phi}^T \text{diag} \left( \frac{1}{\widetilde{\sigma}_w^2} \mathbb{E} \left[ \frac{\widetilde{\eta}_{m\ell}^{(\text{hi})} f(\widetilde{\eta}_{m\ell}^{(\text{hi})}) - \widetilde{\eta}_{m\ell}^{(\text{lo})} f(\widetilde{\eta}_{m\ell}^{(\text{lo})})}{F(\widetilde{\eta}_{m\ell}^{(\text{hi})}) - F(\widetilde{\eta}_{m\ell}^{(\text{lo})})} + \left( \frac{f(\widetilde{\eta}_{m\ell}^{(\text{hi})}) - f(\widetilde{\eta}_{m\ell}^{(\text{lo})})}{F(\widetilde{\eta}_{m\ell}^{(\text{hi})}) - F(\widetilde{\eta}_{m\ell}^{(\text{lo})})} \right)^2 \right] \right)_{m=1}^{\widetilde{M}} \widetilde{\Phi} + \mathbb{E}[\widetilde{\mathbf{P}}]. \end{aligned} \quad (6)$$

$$\widetilde{\mathbf{M}}_\ell^{(\tau)}(\widetilde{\Phi}, a, r, \widetilde{\sigma}_w^2) = \widetilde{\Phi}^T \text{diag} \left( \frac{1}{\widetilde{\sigma}_w^2} \sum_{\widetilde{y}_{m\ell}=L_0}^{L_B-1} \left( \widetilde{\eta}_{m\ell}^{(\text{hi})} f(\widetilde{\eta}_{m\ell}^{(\text{hi})}) - \widetilde{\eta}_{m\ell}^{(\text{lo})} f(\widetilde{\eta}_{m\ell}^{(\text{lo})}) + \frac{(f(\widetilde{\eta}_{m\ell}^{(\text{hi})}) - f(\widetilde{\eta}_{m\ell}^{(\text{lo})}))^2}{F(\widetilde{\eta}_{m\ell}^{(\text{hi})}) - F(\widetilde{\eta}_{m\ell}^{(\text{lo})})} \right) \right)_{m=1}^{\widetilde{M}} \widetilde{\Phi} + \mathbb{E}[\widetilde{\mathbf{P}}]. \quad (9)$$

ments, and specialize the derived CRLB to these two cases. It is easy to see that the second term  $\mathbb{E}[\widetilde{\mathbf{P}}]$  in the BIM given in (6) depends only on the hyperparameters and not on the quantizer. So, we only discuss the first term in the sequel. The BIMs for the unquantized and 1-bit cases are obtained by careful algebraic manipulation of the multi-bit BIM, and we provide the details in the supplementary material.

### A. Infinite-bit Quantized Compressed Sensing

The following Lemma is useful for obtaining the BIM in the infinite-bit quantized compressed sensing case.

*Lemma 1:* For  $\eta, \delta \in \mathbb{R}$ ,

$$\lim_{\delta \rightarrow 0} \frac{(\eta + \delta) f(\eta + \delta) - \eta f(\eta)}{F(\eta + \delta) - F(\eta)} = 1 - \eta^2, \quad (12)$$

and

$$\lim_{\delta \rightarrow 0} \left( \frac{f(\eta + \delta) - f(\eta)}{F(\eta + \delta) - F(\eta)} \right)^2 = \eta^2, \quad (13)$$

where  $f(\cdot)$  and  $F(\cdot)$  are as defined earlier. ■

As the number of bits  $b$  increases, the difference between the quantization intervals  $\widetilde{z}_{m\ell}^{(\text{lo})}$  and  $\widetilde{z}_{m\ell}^{(\text{hi})}$  decreases, and tends to zero as  $b \rightarrow \infty$ . Therefore, if we apply Lemma 1 to each term inside the expectation in (6), it becomes unity, which results in the  $\ell^{\text{th}}$  diagonal block of the BIM as

$$\widetilde{\mathbf{M}}_\ell(\widetilde{\Phi}, a, r, \widetilde{\sigma}_w^2) = \frac{\widetilde{\Phi}^T \widetilde{\Phi}}{\widetilde{\sigma}_w^2} + \mathbb{E}[\widetilde{\mathbf{P}}]. \quad (14)$$

Next, we compute the BIM for the 1-bit quantization case.

### B. 1-bit Quantized Compressed Sensing

For the 1-bit quantized compressed sensing case, we consider the output of the quantizer as the sign of its input. Thus, if we denote the  $(m, \ell)^{\text{th}}$  entry of  $\widetilde{\mathbf{Y}}$  and  $\widetilde{\mathbf{Z}}$  as  $\widetilde{y}_{m\ell}$  and  $\widetilde{z}_{m\ell}$ , respectively, then  $\widetilde{y}_{m\ell} = +1$  if  $\widetilde{z}_{m\ell} \geq 0$ , and  $\widetilde{y}_{m\ell} = -1$  otherwise. We simplify the BIM in the Lemma below.

*Lemma 2:* The  $\ell^{\text{th}}$  diagonal block of the BIM required to compute the CRLB for the MSE of a Bayesian sparse signal estimator using 1-bit compressive measurements is given by

$$\begin{aligned} \widetilde{\mathbf{M}}_\ell(\widetilde{\Phi}, a, r, \widetilde{\sigma}_w^2) &= \mathbb{E}[\widetilde{\mathbf{P}}] + \\ &\widetilde{\Phi}^T \text{diag} \left( \frac{1}{\widetilde{\sigma}_w^2} \mathbb{E} \left[ \frac{\widetilde{\xi}_{m\ell} f(\widetilde{\xi}_{m\ell})}{F(\widetilde{\xi}_{m\ell})} + \left( \frac{f(\widetilde{\xi}_{m\ell})}{F(\widetilde{\xi}_{m\ell})} \right)^2 \right] \right)_{m=1}^{\widetilde{M}} \widetilde{\Phi}, \end{aligned} \quad (15)$$

where  $\mathbb{E}[\cdot]$  is w.r.t.  $p(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}, \widetilde{\mathbf{P}}; \widetilde{\Phi}, \widetilde{\sigma}_w^2, a, r)$ . Here,  $\widetilde{\xi}_{m\ell} \triangleq \frac{\widetilde{y}_{m\ell} \sum_{n=1}^{\widetilde{N}} \widetilde{\Phi}_{mn} \widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}$ ,  $\ell \in \{1, \dots, T\}$ , and the other notation are as in Theorem 1. ■

It is worth mentioning that the BIMs for the unquantized compressed sensing in [29] and 1-bit compressed sensing in [30] (Fisher information matrix in [30]), are special cases of the BIM for multi-bit compressed sensing. The BIM for the 1-bit compressed sensing case in (15) can be further simplified (along the same lines as in [30]) to get

$$\begin{aligned} \widetilde{\mathbf{M}}_\ell(\widetilde{\Phi}, a, r, \widetilde{\sigma}_w^2) &= \widetilde{\Phi}^T \text{diag} \left( \frac{1}{\widetilde{\sigma}_w^2} \mathbb{E} \left[ \frac{(f(\widetilde{\nu}_{m\ell}))^2}{F(\widetilde{\nu}_{m\ell})(1 - F(\widetilde{\nu}_{m\ell}))} \right] \right)_{m=1}^{\widetilde{M}} \widetilde{\Phi} + \mathbb{E}[\widetilde{\mathbf{P}}] \end{aligned} \quad (16)$$

where  $\widetilde{\nu}_{m\ell} = \frac{\sum_{n=1}^{\widetilde{N}} \widetilde{\Phi}_{mn} \widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}$ , and the first expectation is w.r.t.  $p(\widetilde{\mathbf{X}}, \widetilde{\mathbf{P}})$ .

We now turn to the massive MIMO-OFDM wireless communication system, and present the system model for the channel estimation and soft symbol decoding problems.

## III. DESCRIPTION OF MASSIVE MIMO-OFDM SYSTEM AND PROBLEM STATEMENTS

We consider the uplink (UL) of a single cell massive MIMO-OFDM system with  $N_r$  antennas at the BS and  $K$  single antenna user equipments (UEs), where  $N_r \geq K$ . Fig. 1 shows the system model. Each UE encodes and interleaves its information bits, and maps them to constellation symbols. The symbols are then loaded onto the subcarriers and OFDM modulated using an inverse discrete Fourier transform (IDFT). After passing the OFDM modulated data symbols through a parallel to serial converter, a cyclic prefix (CP) is added, RF up-converted to the passband, and transmitted over a frequency-selective wireless channel to the BS. At the BS, the received RF signal is down-converted to baseband, the CP is removed, the signal is sampled, and quantized using  $b$ -bit ADCs to obtain the complex baseband signal.

We assume that the coherence interval of the channel is at least  $\tau_p + \tau_d$  OFDM symbols. Each UE transmits  $\tau_p$  pilot OFDM symbols followed by  $\tau_d$  data OFDM symbols. In conventional OFDM systems, pilots are embedded in the same OFDM symbol along with the data, at specific subcarriers. However, due to the quantization errors introduced by the low resolution ADCs, this results in severe inter-carrier interference between the pilot and data subcarriers, which cannot

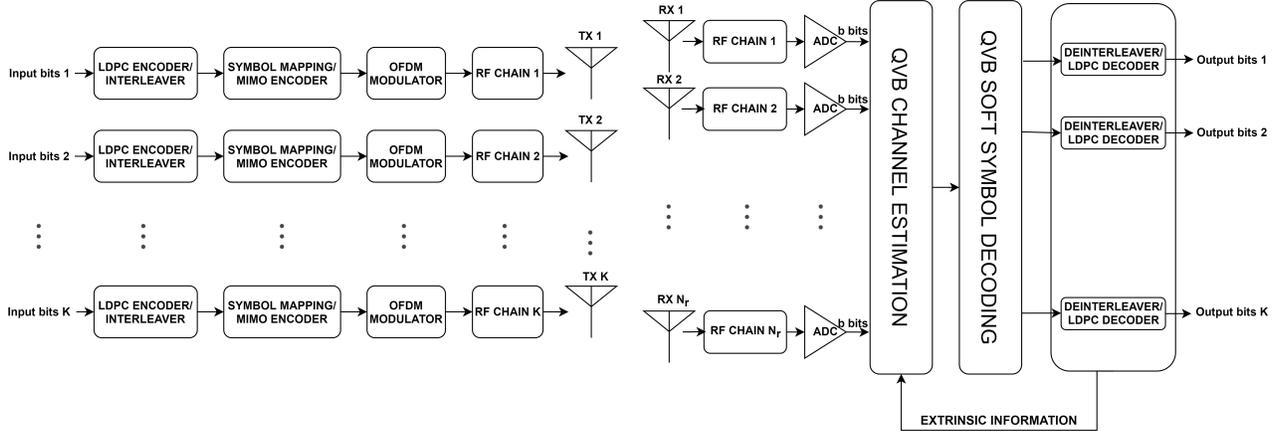


Figure 1. Massive MIMO-OFDM wireless communication system model.

be canceled to obtain channel estimates. So, in this work, we consider a model where the pilot OFDM symbols are distinct from the data OFDM symbols. We denote the number of subcarriers by  $N_c$ .<sup>2</sup> The unquantized received pilot and data signals at the  $n_r^{\text{th}}$  receive antenna in the  $n^{\text{th}}$  symbol interval ( $0 \leq n \leq N_c - 1$ ) within the  $t^{\text{th}}$  pilot and data OFDM symbol durations, respectively, are <sup>3</sup>

$$\begin{aligned} z_{n_r}^{(p)}[t][n] &= \sum_{k=1}^K \sum_{\ell=0}^{L-1} h_{n_r,k}[\ell] \bar{x}_k^{(p)}[t][n-\ell] + w_{n_r}^{(p)}[t][n], \\ z_{n_r}^{(d)}[t][n] &= \sum_{k=1}^K \sum_{\ell=0}^{L-1} h_{n_r,k}[\ell] \bar{x}_k^{(d)}[t][n-\ell] + w_{n_r}^{(d)}[t][n], \end{aligned} \quad (17)$$

where  $t \in \{1, \dots, \tau_p\}$  and  $t \in \{\tau_p + 1, \dots, \tau_p + \tau_d\}$  for the pilot and data phases, respectively,  $h_{n_r,k}[\ell]$  is the complex channel gain of the  $\ell^{\text{th}}$  delay tap of the channel between the  $k^{\text{th}}$  UE and the  $n_r^{\text{th}}$  receive antenna at the BS, distributed as  $\mathcal{CN}(h_{n_r,k}[\ell]; 0, \beta_{k\ell})$ , where  $\beta_{k\ell}$  is the large scale fading coefficient (LSFC),  $L$  is the total number of delay taps of the frequency selective channel,  $\bar{\mathbf{x}}_k^{(p)}[t] = [\bar{x}_k^{(p)}[t][0], \dots, \bar{x}_k^{(p)}[t][N_c - 1]]$  and  $\bar{\mathbf{x}}_k^{(d)}[t] = [\bar{x}_k^{(d)}[t][0], \dots, \bar{x}_k^{(d)}[t][N_c - 1]]$  are the pilot and data symbols transmitted in the time domain by the  $k^{\text{th}}$  UE in the  $t^{\text{th}}$  OFDM symbol, respectively, and  $w_{n_r}^{(p)}[t][n]$  and  $w_{n_r}^{(d)}[t][n]$  are the complex additive white Gaussian noise during the pilot and data phases, respectively, with mean 0 and variance  $\sigma_w^2$ .

#### A. Delay-Domain Sparse Channel Estimation Model

In this subsection, we reformulate the received signal model above to utilize the lag-domain sparsity for channel estimation. We denote the channel sparsity, i.e., the maximum number of nonzero delay taps in the channel, by  $L_{\text{sp}}$ , where  $L_{\text{sp}} \ll L$ . We vectorize the unquantized received pilot signal

<sup>2</sup>We do not explicitly include any guard subcarriers in our system model. However, we note that  $N_c$  can be considered the number of active subcarriers without loss of generality.

<sup>3</sup>We do not explicitly include the effects of the transmit and receive filters in our system model. However, we consider their effect in one of the simulation results, and observe only a marginal performance loss.

in the  $t^{\text{th}}$  OFDM symbol at the  $n_r^{\text{th}}$  receive antenna as

$$\begin{aligned} \mathbf{z}_{n_r}^{(p)}[t] &= [z_{n_r}^{(p)}[t][0] \quad z_{n_r}^{(p)}[t][1] \quad \dots \quad z_{n_r}^{(p)}[t][N_c - 1]]^T \\ &= \sum_{k=1}^K \bar{\mathbf{X}}_k^{(p)}[t] \bar{\mathbf{h}}_{n_r,k} + \mathbf{w}_{n_r}^{(p)}[t] \in \mathbb{C}^{N_c \times 1}, \end{aligned} \quad (18)$$

where  $\bar{\mathbf{h}}_{n_r,k} = [\mathbf{h}_{n_r,k}^T, \mathbf{0}_{N_c-L}^T]^T \in \mathbb{C}^{N_c \times 1}$ ,  $\mathbf{h}_{n_r,k} = [h_{n_r,k}[0], \dots, h_{n_r,k}[L-1]]^T \in \mathbb{C}^{L \times 1}$  is the  $k^{\text{th}}$  UE's frequency selective channel, and  $\bar{\mathbf{X}}_k^{(p)}[t] \in \mathbb{C}^{N_c \times N_c}$  is a circulant matrix whose first column is  $\bar{\mathbf{x}}_k^{(p)}[t]$ . Since any circulant matrix is diagonalized by the unitary DFT matrix with the frequency domain coefficients as the eigenvalues, we represent (18) as

$$\begin{aligned} \mathbf{z}_{n_r}^{(p)}[t] &= \sum_{k=1}^K \mathbf{F}_{N_c}^H \mathbf{X}_k^{(p)}[t] \mathbf{F}_{N_c,L} \mathbf{h}_{n_r,k} + \mathbf{w}_{n_r}^{(p)}[t] \\ &= (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \mathbf{X}^{(p)}[t] (\mathbf{I}_K \otimes \mathbf{F}_{N_c,L}) \mathbf{h}_{n_r} + \mathbf{w}_{n_r}^{(p)}[t], \end{aligned}$$

where  $\otimes$  denotes the matrix Kronecker product operator,  $\mathbf{X}_k^{(p)}[t] = \mathbf{F}_{N_c} \bar{\mathbf{X}}_k^{(p)}[t] \mathbf{F}_{N_c}^H$  is a diagonal matrix with its entries as the pilots loaded on the subcarriers,  $\mathbf{X}^{(p)}[t] = \text{diag}(\mathbf{X}_1^{(p)}[t], \dots, \mathbf{X}_K^{(p)}[t]) \in \mathbb{C}^{K N_c \times K N_c}$ ,  $\mathbf{h}_{n_r} = [\mathbf{h}_{n_r,1}^T, \dots, \mathbf{h}_{n_r,K}^T]^T \in \mathbb{C}^{K L \times 1}$  is the vectorized lag domain frequency selective channel between all the users and the  $n_r^{\text{th}}$  BS antenna.  $\mathbf{F}_{N_c} \in \mathbb{C}^{N_c \times N_c}$  and  $\mathbf{F}_{N_c,L} \in \mathbb{C}^{N_c \times L}$  are the DFT and the  $L$  column truncated DFT matrices, respectively. Note that  $\mathbf{h}_{n_r}$  is a sparse vector with sparsity  $K L_{\text{sp}}$ . We stack the received vector of all the  $N_r$  antennas and  $\tau_p$  pilot OFDM symbols to obtain the unquantized received pilot matrix as shown in (19) on the next page, where  $\mathbf{Z}^{(p)} \in \mathbb{C}^{\tau_p N_c \times N_r}$ ,  $\Phi^{(p)} \in \mathbb{C}^{\tau_p N_c \times K L}$ ,  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_{N_r}] \in \mathbb{C}^{K L \times N_r}$  is a row sparse channel matrix, and  $\mathbf{W}^{(p)}$  is the additive noise matrix.

Now, we quantize the received signal using low-resolution ADCs. A  $b$ -bit quantizer on a real valued input  $z$  is defined as  $\mathcal{Q}_b(z) = L_i$ ,  $z \in [\delta_i, \delta_{i+1})$ ,  $i = 0, 1, \dots, B-1$ , where  $B = 2^b$  is the number of quantization levels,  $\delta_0 < \delta_1 < \dots < \delta_B$  are the quantization thresholds, and  $L_0, L_1, \dots, L_{B-1}$  are the quantizer outputs. In this paper, for simplicity and concreteness, we consider a uniform quantizer, where  $\delta_l = (-B/2 + l)\Delta$ ,  $l = 1, \dots, B-1$ ,  $\Delta$  is the quantization step

$$\mathbf{Z}^{(p)} = \begin{bmatrix} \mathbf{z}_1^{(p)}[1] & \dots & \mathbf{z}_{N_r}^{(p)}[1] \\ \vdots & \ddots & \vdots \\ \mathbf{z}_1^{(p)}[\tau_p] & \dots & \mathbf{z}_{N_r}^{(p)}[\tau_p] \end{bmatrix} = \begin{bmatrix} (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \mathbf{X}^{(p)}[1] (\mathbf{I}_K \otimes \mathbf{F}_{N_c, L}) \\ \vdots \\ (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \mathbf{X}^{(p)}[\tau_p] (\mathbf{I}_K \otimes \mathbf{F}_{N_c, L}) \end{bmatrix} \mathbf{H} + \mathbf{W}^{(p)} \triangleq \mathbf{\Phi}^{(p)} \mathbf{H} + \mathbf{W}^{(p)}. \quad (19)$$

size, and  $L_l = (\delta_l + \delta_{l+1})/2, l = 0, \dots, B-1$ . However, we note that the foregoing development is applicable to any other quantizer also. We set the dynamic range of the real and imaginary parts of the quantizer using the expected received signal power,  $P_R$ , as  $\delta_0 = -2.5\sqrt{P_R/2}$ ,  $\delta_B = 2.5\sqrt{P_R/2}$ . In practice, we quantize any value below  $\delta_0$  to  $L_0$ , and any value above  $\delta_B$  to  $L_{B-1}$ . Also, in practical systems, an automatic gain control unit is used to ensure that the power in the analog baseband signal is approximately equal to a predefined value,  $P_R$ , before quantization. Our choice of  $\delta_0$  and  $\delta_B$  is motivated by the fact that the absolute value of a Gaussian distributed zero mean real-valued random variable with variance  $P_R/2$  exceeds  $2.5\sqrt{P_R/2}$  with probability less than 0.01, i.e., the quantizer gets overloaded with low probability. We quantize the received pilots in (19) using the  $b$ -bit ADCs to obtain the quantized received pilots as

$$\mathbf{Y}^{(p)} = \mathcal{Q}_b(\mathbf{Z}^{(p)}) = \mathcal{Q}_b(\mathbf{\Phi}^{(p)} \mathbf{H} + \mathbf{W}^{(p)}) \in \mathbb{C}^{\tau_p N_c \times N_r}. \quad (20)$$

Our first goal is to obtain an estimate of  $\mathbf{H}$  given  $\mathbf{Y}^{(p)}$  and  $\mathbf{\Phi}^{(p)}$  in (20). Note that, if  $\tau_p N_c < KL$ , (20) represents an underdetermined system of equations. In order to exploit the lag-domain sparsity in the channel, as in [34], we use a two stage hierarchical prior on  $\mathbf{H}$  i.e.,  $\forall i, \mathbf{h}_i \sim \mathcal{CN}(\mathbf{h}_i; \mathbf{0}, \mathbf{P}^{-1})$ , where the precision matrix  $\mathbf{P}$  is diagonal and contains the hyperparameters  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{KL}]^T$  as its diagonal elements. Further, we impose a Gamma hyperprior on  $\boldsymbol{\alpha}$ . This results in a Student's  $t$ -distributed prior on  $\mathbf{h}_i$ , which is known to promote sparse channel estimates [33]. After estimating  $\mathbf{H}$ , our goal is to decode the data symbols. Next, we describe the signal model in the data transmission phase.

### B. MIMO-OFDM Data Detection Model

We vectorize the unquantized received data at the  $n_r^{\text{th}}$  receive antenna during the  $t^{\text{th}}$  OFDM symbol in (17) as

$$\begin{aligned} \mathbf{z}_{n_r}^{(d)}[t] &= \begin{bmatrix} z_{n_r}^{(d)}[t][0] & z_{n_r}^{(d)}[t][1] & \dots & z_{n_r}^{(d)}[t][N_c - 1] \end{bmatrix}^T \\ &= \sum_{k=1}^K \mathbf{H}_{n_r, k}^{\text{time}} \mathbf{F}_{N_c}^H \mathbf{x}_k^{(d)}[t] + \mathbf{w}_{n_r}^{(d)}[t], \end{aligned} \quad (21)$$

where  $t \in \{\tau_p + 1, \dots, \tau_p + \tau_d\}$  and  $\mathbf{x}_k^{(d)}[t] = \begin{bmatrix} x_k^{(d)}[t][0] & \dots & x_k^{(d)}[t][N_c - 1] \end{bmatrix}^T = \mathbf{F}_{N_c} \bar{\mathbf{x}}_k^{(d)}[t] \in \mathbb{C}^{N_c \times 1}$  is the  $M$ -QAM modulated data symbols loaded on the subcarriers, where  $\bar{\mathbf{x}}_k^{(d)}[t] = \begin{bmatrix} \bar{x}_k^{(d)}[t][0] & \dots & \bar{x}_k^{(d)}[t][N_c - 1] \end{bmatrix}^T \in \mathbb{C}^{N_c \times 1}$  is the time domain transmitted signal of the  $k^{\text{th}}$  user. Also,  $\mathbf{H}_{n_r, k}^{\text{time}} \in \mathbb{C}^{N_c \times N_c}$  is a circulant matrix with the first column as  $\bar{\mathbf{h}}_{n_r, k}$  (from (18)). Using the diagonalizability property of a circulant matrix, we represent (21) as shown in (22) on the next page, where  $\mathbf{H}_{n_r, k}^{\text{req}} = \mathbf{F}_{N_c} \mathbf{H}_{n_r, k}^{\text{time}} \mathbf{F}_{N_c}^H \in \mathbb{C}^{N_c \times N_c}$  is diagonal, containing the frequency domain representation of  $\bar{\mathbf{h}}_{n_r, k}$ ,  $\mathbf{x}^{(d)}[t] \triangleq \begin{bmatrix} \mathbf{x}_1^{(d)T}[t] & \dots & \mathbf{x}_K^{(d)T}[t] \end{bmatrix}^T$ ,  $\mathbf{1}_K$  is the  $K \times 1$

all-ones vector, and  $\mathbf{0}$  is an  $N_c \times N_c$  all-zero matrix. Now, we vectorize and stack the signal received over the  $N_r$  receive antennas and  $\tau_d$  OFDM data symbols to obtain  $\mathbf{Z}^{(d)}$  as shown in (23) on the next page, where  $\mathbf{D} \in \mathbb{C}^{N_r N_c \times K N_c}$  is the measurement matrix for data detection,  $\mathbf{X}^{(d)} \in \mathbb{C}^{K N_c \times \tau_d}$  is the transmit data matrix, and  $\mathbf{W}^{(d)}$  is the additive white Gaussian noise matrix during the data phase. Now, we quantize the received signal (23) using the  $b$ -bit ADCs to obtain

$$\mathbf{Y}^{(d)} = \mathcal{Q}_b(\mathbf{Z}^{(d)}) = \mathcal{Q}_b(\mathbf{D} \mathbf{X}^{(d)} + \mathbf{W}^{(d)}). \quad (24)$$

Our goal in this part is to decode the data symbols  $\mathbf{X}^{(d)}$  given  $\mathbf{Y}^{(d)}$  and  $\mathbf{D}$ . With the posterior distribution of  $\mathbf{X}^{(d)}$  in hand, our next task is to perform data-aided channel estimation to refine the channel estimates. We explain the model for this problem in the next subsection.

### C. Virtual Pilots-Aided MIMO-OFDM Channel Estimation

From section III-A, we write the unquantized pilot received signal as  $\mathbf{Z}^{(p)} = \mathbf{\Phi}^{(p)} \mathbf{H} + \mathbf{W}^{(p)}$ . Similar to the pilot reception phase, if we consider the decoded data as known virtual pilot symbols, then we can write the received data signal as  $\mathbf{Z}^{(d)}[t] = (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \langle \mathbf{X}^{(d)}[t] \rangle (\mathbf{I}_K \otimes \mathbf{F}_{N_c, L}) \mathbf{H} + \mathbf{W}^{(d)}[t]$ , where  $t = \{\tau_p + 1, \dots, \tau_p + \tau_d\}$ ,  $\langle \mathbf{X}^{(d)}[t] \rangle = \text{diag}(\langle \mathbf{X}_1^{(d)}[t] \rangle, \dots, \langle \mathbf{X}_K^{(d)}[t] \rangle) \in \mathbb{C}^{K N_c \times K N_c}$ , and  $\langle \mathbf{X}_k^{(d)}[t] \rangle = \text{diag}(\langle \mathbf{x}_k^{(d)}[t] \rangle) \in \mathbb{C}^{N_c \times N_c}$ . Here,  $\langle \mathbf{x}_k^{(d)}[t] \rangle$  are the posterior means of the decoded data symbols of the  $k^{\text{th}}$  user during the  $t^{\text{th}}$  OFDM symbol. We stack  $\mathbf{Z}^{(p)}$  and  $\mathbf{Z}^{(d)}[t]$  to obtain an expression for the unquantized received signal over one coherence interval as shown in (25) on the next page, where  $\mathbf{\Phi} \in \mathbb{C}^{(\tau_p + \tau_d) N_c \times KL}$  is the augmented measurement matrix and  $\mathbf{W} \in \mathbb{C}^{(\tau_p + \tau_d) N_c \times N_r}$  is the additive white Gaussian noise matrix. The  $b$ -bit quantized received signal after the ADCs then reads

$$\mathbf{Y} = \mathcal{Q}_b(\mathbf{Z}) = \mathcal{Q}_b(\mathbf{\Phi} \mathbf{H} + \mathbf{W}) \in \mathbb{C}^{(\tau_p + \tau_d) N_c \times N_r}. \quad (26)$$

Our goal is to estimate of  $\mathbf{H}$  given  $\mathbf{Y}$  and  $\mathbf{\Phi}$ . Once we estimate  $\mathbf{H}$ , we use it to obtain  $\mathbf{D}$  as in (24), which in turn is used to refine the posterior beliefs of the  $M$ -QAM modulated data symbols in the next data decoding iteration.

In the next section, we present our solutions to the above channel estimation and data detection problems.

## IV. QUANTIZED VB CHANNEL ESTIMATION

Our goal is to infer the posterior distributions of the channels and the LLRs of the data symbols, given the quantized pilot and data observations. To this end, we adopt a statistical inference approach, where we represent the received pilot and data signals in a probabilistic graphical model. Exact computation of the posterior distributions is computationally intractable, as it requires solving high dimensional integrals

$$\begin{aligned} \mathbf{z}_{n_r}^{(d)}[t] &= \sum_{k=1}^K \mathbf{F}_{N_c}^H \mathbf{H}_{n_r,k}^{\text{freq}} \mathbf{x}_k^{(d)}[t] + \mathbf{w}_{n_r}^{(d)}[t] = (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \begin{bmatrix} \mathbf{H}_{n_r,1}^{\text{freq}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{n_r,2}^{\text{freq}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_{n_r,K}^{\text{freq}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^{(d)}[t] \\ \mathbf{x}_2^{(d)}[t] \\ \vdots \\ \mathbf{x}_K^{(d)}[t] \end{bmatrix} + \mathbf{w}_{n_r}^{(d)}[t] \\ &= (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \mathbf{H}_{n_r}^{\text{freq}} \mathbf{x}^{(d)}[t] + \mathbf{w}_{n_r}^{(d)}[t]. \end{aligned} \quad (22)$$

$$\begin{aligned} \mathbf{Z}^{(d)} &= \begin{bmatrix} \mathbf{z}_1^{(d)}[\tau_p + 1] & \cdots & \mathbf{z}_1^{(d)}[\tau_p + \tau_d] \\ \vdots & \ddots & \vdots \\ \mathbf{z}_{N_r}^{(d)}[\tau_p + 1] & \cdots & \mathbf{z}_{N_r}^{(d)}[\tau_p + \tau_d] \end{bmatrix} = \begin{bmatrix} (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \mathbf{H}_1^{\text{freq}} \\ (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \mathbf{H}_2^{\text{freq}} \\ \vdots \\ (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \mathbf{H}_{N_r}^{\text{freq}} \end{bmatrix} [\mathbf{x}^{(d)}[\tau_p + 1] \quad \cdots \quad \mathbf{x}^{(d)}[\tau_p + \tau_d]] + \mathbf{W}^{(d)} \\ &= \mathbf{D} \mathbf{X}^{(d)} + \mathbf{W}^{(d)}, \end{aligned} \quad (23)$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}^{(p)} \\ \mathbf{Z}^{(d)}[\tau_p + 1] \\ \vdots \\ \mathbf{Z}^{(d)}[\tau_p + \tau_d] \end{bmatrix} = \begin{bmatrix} \Phi^{(p)} \\ (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \langle \mathbf{X}^{(d)}[\tau_p + 1] \rangle (\mathbf{I}_K \otimes \mathbf{F}_{N_c,L}) \\ \vdots \\ (\mathbf{1}_K^T \otimes \mathbf{F}_{N_c}^H) \langle \mathbf{X}^{(d)}[\tau_p + \tau_d] \rangle (\mathbf{I}_K \otimes \mathbf{F}_{N_c,L}) \end{bmatrix} \mathbf{H} + \mathbf{W} = \begin{bmatrix} \Phi^{(p)} \\ \Phi^{(d)} \end{bmatrix} \mathbf{H} + \mathbf{W} \triangleq \Phi \mathbf{H} + \mathbf{W}. \quad (25)$$

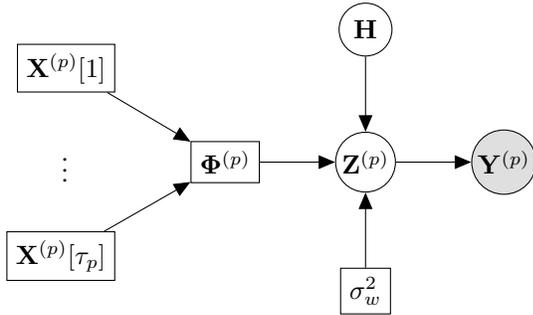


Figure 2. Bayesian network model for the channel estimation problem (20).

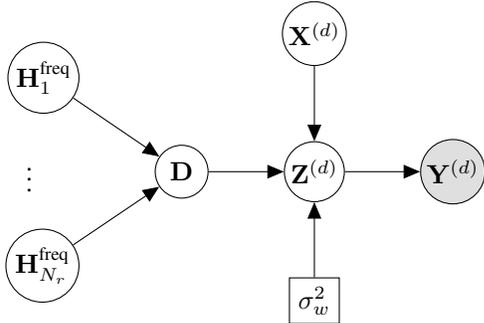


Figure 3. Bayesian network model for the data detection problem (24).

over  $\mathbf{H}$ ,  $\mathbf{x}_1^{(d)}[\tau_p + 1], \dots, \mathbf{x}_K^{(d)}[\tau_p + \tau_d]$  to obtain the partition functions  $P(\mathbf{Y}^{(p)})$  and  $P(\mathbf{Y}^{(d)})$ . This motivates the need for approximate inference techniques, where we replace the exact posterior distribution with a distribution that is close to the original in a particular distance measure, and is also easy to compute. We will show that this leads to computationally tractable algorithms for the problem at hand. An excellent introduction to approximate inference can be found in [35].

We present Bayesian network graphical models for the channel estimation, data detection and data-aided channel estimation problems in Figures 2, 3, and 4, respectively. We use shaded circles, transparent circles, and squares to represent the observations, latent variables, and deterministic variables. In our channel estimation and data detection problems, the

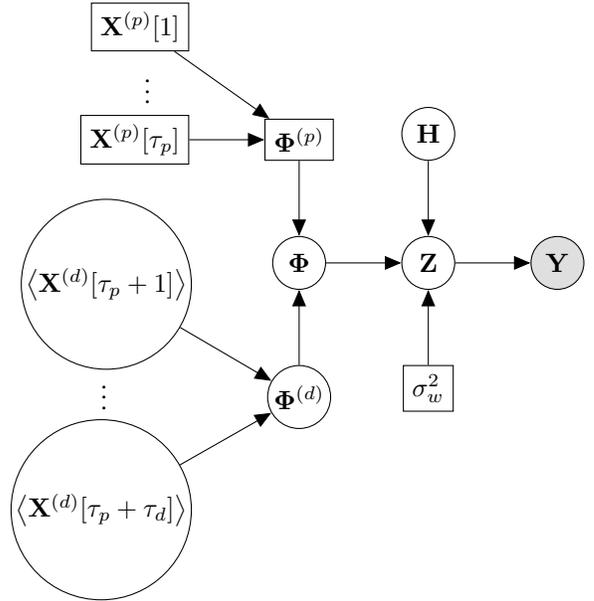


Figure 4. Bayesian network graphical model for the data-aided channel estimation problem in (26).

quantized received pilot and data signals are the observations, and the channel and data symbols are the latent variables. Also, the pilot symbols and noise variance are deterministic and known. In the following paragraphs, we briefly describe variational Bayesian (VB) inference, which is the approximate inference technique adopted in this paper to solve the pilot-aided channel estimation problem given in (20).

VB is an iterative procedure to compute approximate posterior distributions of the latent variables given the observations. In this, we first write the logarithm of the observations  $\mathbf{Y}^{(p)}$  as the sum of two terms and lower-bound it as

$$\ln p(\mathbf{Y}^{(p)}) = \mathcal{L}(q) + \text{KL}(q \| p) \geq \mathcal{L}(q), \quad (27)$$

where  $\mathcal{L}(q) \triangleq \int q(\mathbf{H}) \ln \left\{ \frac{p(\mathbf{Y}^{(p)}, \mathbf{H}; \Phi^{(p)}, \sigma_w^2)}{q(\mathbf{H})} \right\} d\mathbf{H}$  and  $\text{KL}(q \| p) \triangleq - \int q(\mathbf{H}) \ln \left\{ \frac{p(\mathbf{H} | \mathbf{Y}^{(p)}, \Phi^{(p)}, \sigma_w^2)}{q(\mathbf{H})} \right\} d\mathbf{H} \geq 0$  are the

evidence lower bound (ELBO) and non-negative Kullback Leibler (KL) divergence terms, respectively. Here,  $q(\mathbf{H})$  is a posterior distribution which is arbitrary, and can be approximated and optimized. Note that  $q(\mathbf{H})$  depends on  $\mathbf{Y}^{(p)}$ , but we do not explicitly include it in the notation for brevity. In the above, maximizing the ELBO  $\mathcal{L}(q)$  would render a distribution  $q$  that is close to the original model evidence. We formally state the ELBO maximization problem as

$$q_{\text{opt}} = \arg \max_{q \in \mathcal{P}} \mathcal{L}(q) = \arg \min_{q \in \mathcal{P}} \text{KL}(q \| p), \quad (28)$$

where  $\mathcal{P}$  is the space of probability distributions. The maximum of  $\mathcal{L}(q)$  occurs when  $q(\mathbf{H}) = p(\mathbf{H} | \mathbf{Y}^{(p)}, \Phi^{(p)}, \sigma_w^2)$ , but computing it is intractable. Therefore, we impose a factorized structure on each column of  $\mathbf{H}$  i.e.,  $q(\mathbf{H}) = \prod_{i=1}^{N_r} q_i(\mathbf{h}_i)$ . Substituting this in the ELBO, and simplifying it by fixing one of the factors, say  $q_j(\mathbf{h}_j)$ , we get

$$\mathcal{L}(q) = -\text{KL}\left(q_j \| \tilde{p}(\mathbf{Y}^{(p)}, \mathbf{h}_j; \Phi^{(p)}, \sigma_w^2)\right) + \text{constant}, \quad (29)$$

where the constant terms do not depend on  $q_j(\mathbf{h}_j)$ , and  $\tilde{p}(\mathbf{Y}^{(p)}, \mathbf{h}_j; \Phi^{(p)}, \sigma_w^2)$  is defined using  $\ln \tilde{p}(\mathbf{Y}^{(p)}, \mathbf{h}_j; \Phi^{(p)}, \sigma_w^2) \triangleq \mathbb{E}_{i \neq j} [\ln p(\mathbf{Y}^{(p)}, \mathbf{H}; \Phi^{(p)}, \sigma_w^2)] + \text{constant}$ , where the notation  $\mathbb{E}_{i \neq j}[\cdot]$  denotes the expectation with respect to the distributions  $q_1(\mathbf{h}_1), \dots, q_{N_r}(\mathbf{h}_{N_r})$  except  $q_j(\mathbf{h}_j)$ . Now,  $\mathcal{L}(q)$  is maximized when the KL divergence term in (29) is minimized, which happens when  $q_j(\mathbf{h}_j) = \tilde{p}(\mathbf{Y}^{(p)}, \mathbf{h}_j; \Phi^{(p)}, \sigma_w^2)$ . Therefore, the optimal marginal distribution is

$$q_j(\mathbf{h}_j) = \text{const} \times \exp\left(\mathbb{E}_{i \neq j} \left[\ln p(\mathbf{Y}^{(p)}, \mathbf{H}; \Phi^{(p)}, \sigma_w^2)\right]\right), \quad (30)$$

where the constant is chosen such that  $q_j$  becomes a probability distribution. Thus, VB is an iterative algorithm that falls in the category of minorization-maximization (MM), which solves a maximization problem by iteratively obtaining a lower bound on the objective function as in (27), and maximizing it. It is known that MM based optimization converges to a stationary point of the original optimization problem from any initialization [36].

We note that the marginal distribution in (30) is still hard to compute, as  $p(\mathbf{Y}^{(p)}, \mathbf{H}; \Phi^{(p)}, \sigma_w^2)$  contains terms involving the difference of the CDF of complex Gaussian random vectors. Hence, we add  $\mathbf{Z}^{(p)}$  also as a latent variable. This leads to a closed form solution as described below.

We use the Bayesian network in Fig. 2 to express the logarithm of the joint probability distribution of the observations and latent variables as

$$\begin{aligned} \ln p(\mathbf{Y}^{(p)}, \mathbf{Z}^{(p)}, \mathbf{H}, \alpha; \Phi^{(p)}, \sigma_w^2, a, r) \\ = \ln p(\mathbf{Y}^{(p)} | \mathbf{Z}^{(p)}) + \ln p(\mathbf{Z}^{(p)} | \mathbf{H}; \Phi^{(p)}, \sigma_w^2) \\ + \ln p(\mathbf{H} | \mathbf{P}) + \ln p(\alpha; a, r), \end{aligned} \quad (31)$$

where the prior distributions of  $\mathbf{H}$  and  $\alpha$  are

$$p(\mathbf{H} | \mathbf{P}) = \prod_{n=1}^{N_r} \frac{|\mathbf{P}|}{\pi^{KL}} \exp(-\mathbf{h}_n^H \mathbf{P} \mathbf{h}_n), \quad (32)$$

$$p(\alpha; a, r) = \prod_{k=1}^{KL} \frac{r^a}{\Gamma(a)} \alpha_k^{a-1} \exp(-r\alpha_k), \quad (33)$$

respectively. We set  $a$  and  $r$  to small values (say,  $10^{-4}$ ) such that the hyperprior  $p(\alpha; a, r)$  is non-informative. We approximate the posterior  $p(\mathbf{Z}^{(p)}, \mathbf{H}, \alpha | \mathbf{Y}^{(p)}; \Phi^{(p)}, \sigma_w^2, a, r)$  of the latent variables as the factorized distribution:

$$\begin{aligned} p(\mathbf{Z}^{(p)}, \mathbf{H}, \alpha | \mathbf{Y}^{(p)}; \Phi^{(p)}, \sigma_w^2, a, r) \\ \approx q_{\mathbf{H}}(\mathbf{H}) q_{\mathbf{Z}}(\mathbf{Z}^{(p)}) q_{\alpha}(\alpha) \\ = \prod_{n=1}^{N_r} q_{\mathbf{h}_n}(\mathbf{h}_n) \prod_{n=1}^{N_r} q_{\mathbf{z}_n}(\mathbf{z}_n^{(p)}) \prod_{k=1}^{KL} q_{\alpha_k}(\alpha_k), \end{aligned} \quad (34)$$

where we define  $\mathbf{Z}^{(p)} \triangleq [\mathbf{z}_1^{(p)}, \dots, \mathbf{z}_{N_r}^{(p)}]$  and  $\mathbf{H} \triangleq [\mathbf{h}_1, \dots, \mathbf{h}_{N_r}]$ . Next, we express the conditional probability distributions of the observations and latent variables that are needed to compute the posterior distributions under the factorized structure as

$$\begin{aligned} p(\mathbf{Y}^{(p)} | \mathbf{Z}^{(p)}) &= \prod_{t=1}^{\tau_p N_c} \prod_{n=1}^{N_r} \mathbb{1}\left(\Re(z_{tn}^{(p)}) \in \left(\Re(z_{tn}^{(\text{lo})}), \Re(z_{tn}^{(\text{hi})})\right)\right) \\ &\quad \times \mathbb{1}\left(\Im(z_{tn}^{(p)}) \in \left(\Im(z_{tn}^{(\text{lo})}), \Im(z_{tn}^{(\text{hi})})\right)\right) \\ &\triangleq \prod_{n=1}^{N_r} \mathbb{1}\left(\mathbf{z}_n^{(p)} \in \left(\mathbf{z}_n^{(\text{lo})}, \mathbf{z}_n^{(\text{hi})}\right)\right), \end{aligned} \quad (35)$$

$$\begin{aligned} p(\mathbf{Z}^{(p)} | \mathbf{H}; \Phi^{(p)}, \sigma_w^2) \\ = \prod_{n=1}^{N_r} \frac{1}{(\pi \sigma_w^2)^{\tau_p N_c}} \exp\left(-\frac{1}{\sigma_w^2} \|\mathbf{z}_n^{(p)} - \Phi^{(p)} \mathbf{h}_n\|^2\right), \end{aligned} \quad (36)$$

where  $z_{tn}^{(p)}$  is the  $(t, n)^{\text{th}}$  entry of  $\mathbf{Z}^{(p)}$ ,  $\mathbb{1}(\cdot)$  is the indicator function,  $z_{tn}^{(\text{lo})}$  and  $z_{tn}^{(\text{hi})}$  are the lower and upper quantization thresholds corresponding to the  $(t, n)^{\text{th}}$  entry of  $\mathbf{Y}^{(p)}$ , respectively. The posterior distributions of the latent variables are computed by finding the expectations of the logarithm of the joint distribution (31) with respect to the latent variables, and are provided in closed form in the following three Lemmas. The proofs for the Lemmas follow from (30), and are provided in the supplementary material.

*Lemma 3 (Computation of  $q_{\mathbf{H}}(\mathbf{H})$ ):* The posterior distribution  $q_{\mathbf{H}}(\mathbf{H})$  is complex normal with the covariance matrix of each of its columns and mean given by

$$\Sigma_{\mathbf{H}} = \left(\frac{1}{\sigma_w^2} \Phi^{(p)H} \Phi^{(p)} + \langle \mathbf{P} \rangle\right)^{-1}, \quad (37)$$

$$\langle \mathbf{H} \rangle = \frac{1}{\sigma_w^2} \Sigma_{\mathbf{H}} \Phi^{(p)H} \langle \mathbf{Z}^{(p)} \rangle, \quad (38)$$

respectively. Here,  $\langle \mathbf{P} \rangle = \text{diag}(\langle \alpha \rangle)$ , and  $\langle \mathbf{Z}^{(p)} \rangle$  and  $\langle \alpha \rangle$  are the posterior means of  $q_{\mathbf{Z}}(\mathbf{Z}^{(p)})$  and  $q_{\alpha}(\alpha)$ , respectively. ■

*Lemma 4 (Computation of  $q_{\mathbf{Z}}(\mathbf{Z}^{(p)})$ ):* The posterior distribution  $q_{\mathbf{Z}}(\mathbf{Z}^{(p)})$  is truncated complex normal with mean

$$\begin{aligned} \langle \mathbf{Z}^{(p)} \rangle &= \Phi^{(p)} \langle \mathbf{H} \rangle \\ &+ \frac{\sigma_w}{\sqrt{2}} \frac{f\left(\frac{\mathbf{Z}^{(\text{lo})} - \Phi^{(p)} \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right) - f\left(\frac{\mathbf{Z}^{(\text{hi})} - \Phi^{(p)} \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right)}{F\left(\frac{\mathbf{Z}^{(\text{hi})} - \Phi^{(p)} \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right) - F\left(\frac{\mathbf{Z}^{(\text{lo})} - \Phi^{(p)} \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right)}, \end{aligned} \quad (39)$$

---

**Algorithm 2** Quantized VB Channel Estimation
 

---

**Input:**  $\mathbf{Y}^{(p)}$ ,  $\Phi^{(p)}$ ,  $\tau_p$ ,  $\sigma_w$ ,  $N_r$ ,  $K$ ,  $L$ ,  $N_c$ 
**Output:**  $\langle \mathbf{H} \rangle$ 

- 1: Initialize  $\langle \mathbf{Z}^{(p)} \rangle$ ,  $\langle \alpha \rangle$ ,  $a$ ,  $r$ .
  - 2: **repeat**
  - 3:    $\langle \mathbf{P} \rangle = \text{diag}(\langle \alpha \rangle)$
  - 4:   Compute  $\Sigma_{\mathbf{H}}$  using (37).
  - 5:   Compute  $\langle \mathbf{H} \rangle$  using (38).
  - 6:   Compute  $\langle \mathbf{Z}^{(p)} \rangle$  using (39).
  - 7:   Compute  $\langle \alpha_k \rangle$  using (41),  $k = 1, \dots, KL$ .
  - 8: **until** stopping condition is met
- 

where  $\mathbf{Z}^{(\text{lo})}$  and  $\mathbf{Z}^{(\text{hi})}$  are the lower and upper quantization levels corresponding to the observation  $\mathbf{Y}^{(p)}$ , respectively, and  $\langle \mathbf{H} \rangle$  is the posterior mean of  $q_{\mathbf{H}}(\mathbf{H})$ . Also,  $f(\cdot)$  and  $F(\cdot)$  are the PDF and CDF of a standard normal random variable, respectively, computed element-wise on the real and imaginary parts of the argument. The division operation in (39) is also performed element-wise. ■

*Lemma 5 (Computation of  $q_{\alpha_k}(\alpha_k)$ ,  $k = 1, \dots, KL$ ):*

The posterior distribution  $q_{\alpha_k}(\alpha_k)$  follows a Gamma distribution with shape and rate parameters given by

$$\tilde{a}_k = a + N_r \quad \text{and} \quad \tilde{r}_k = r + \sum_{n=1}^{N_r} \langle |h_{kn}|^2 \rangle, \quad (40)$$

respectively. Its mean is given by

$$\langle \alpha_k \rangle = \frac{a + N_r}{r + \sum_{n=1}^{N_r} \langle |h_{kn}|^2 \rangle}, \quad (41)$$

where  $h_{kn}$  is the  $(k, n)$ <sup>th</sup> element of  $\mathbf{H}$ , and  $\langle |h_{kn}|^2 \rangle = \langle |h_{kn}|^2 \rangle + \Sigma_{\mathbf{H}}[k, k]$ . ■

Note that we have included the subscript  $k$  in  $\tilde{a}_k$  for consistency of notation, even though it is independent of  $k$ . From (37), (38), (39), and (41), we see that the statistics of the posterior distributions  $q_{\mathbf{H}}(\mathbf{H})$ ,  $q_{\mathbf{Z}}(\mathbf{Z}^{(p)})$ , and  $q_{\alpha}(\alpha)$  depend on each other. The VB algorithm proceeds iteratively by randomly initializing the posteriors and alternately computing each of the posterior distributions until a suitable convergence condition is satisfied. Once the algorithm converges, we use the posterior mean from (38) as the final channel estimate. Then, we compute the DFT of the lag domain channel estimates, and use them for data decoding. We present VB channel estimation procedure in Algorithm 2.

## V. QUANTIZED VB SOFT SYMBOL DECODING

In this section, we develop a VB algorithm for soft symbol decoding in MIMO-OFDM systems using the system model in (24), reproduced here for convenience:

$$\mathbf{Y}^{(d)} = \mathcal{Q}_b(\mathbf{Z}^{(d)}) = \mathcal{Q}_b(\mathbf{D}\mathbf{X}^{(d)} + \mathbf{W}^{(d)}), \quad (42)$$

where  $\mathbf{Y}^{(d)} = [\mathbf{y}^{(d)}[\tau_p + 1], \dots, \mathbf{y}^{(d)}[\tau_p + \tau_d]] \in \mathbb{C}^{N_r N_c \times \tau_d}$ ,  $\mathbf{Z}^{(d)} = [\mathbf{z}^{(d)}[\tau_p + 1], \dots, \mathbf{z}^{(d)}[\tau_p + \tau_d]] \in \mathbb{C}^{N_r N_c \times \tau_d}$ ,  $\mathbf{X}^{(d)} = [\mathbf{x}^{(d)}[\tau_p + 1], \dots, \mathbf{x}^{(d)}[\tau_p + \tau_d]] \in \mathbb{C}^{K N_c \times \tau_d}$ ,  $\mathbf{x}^{(d)}[t] = [x_{1t}^{(d)}, \dots, x_{K N_c, t}^{(d)}]^T$ ,  $\mathbf{D} \in \mathbb{C}^{N_r N_c \times K N_c}$ ,  $\mathbf{W}^{(d)} \in \mathbb{C}^{N_r N_c \times \tau_d}$ . We represent the corresponding Bayesian network in Fig. 3.

Similar to Sec. IV, we consider the unquantized received data signal as a latent variable, and express the logarithm of the joint probability distribution of the observations and the latent variables as

$$\begin{aligned} \ln p(\mathbf{Y}^{(d)}, \mathbf{Z}^{(d)}, \mathbf{X}^{(d)} | \mathbf{D}, \sigma_w^2) &= \ln p(\mathbf{Y}^{(d)} | \mathbf{Z}^{(d)}) \\ &+ \ln p(\mathbf{Z}^{(d)} | \mathbf{X}^{(d)}, \mathbf{D}, \sigma_w^2) + \ln p(\mathbf{X}^{(d)}). \end{aligned} \quad (43)$$

We factorize the posterior distribution of  $\mathbf{Z}^{(d)}$  and  $\mathbf{X}^{(d)}$  as

$$\begin{aligned} p(\mathbf{Z}^{(d)}, \mathbf{X}^{(d)} | \mathbf{Y}^{(d)}, \mathbf{D}, \sigma_w^2) \\ \approx q_{\mathbf{Z}}(\mathbf{Z}^{(d)}) \prod_{t=\tau_p+1}^{\tau_p+\tau_d} \prod_{k=1}^{K N_c} q_{x_{kt}}(x_{kt}^{(d)}), \end{aligned} \quad (44)$$

where  $\mathbf{Z}^{(d)} = [\mathbf{z}_{\tau_p+1}^{(d)}, \dots, \mathbf{z}_{\tau_p+\tau_d}^{(d)}]$ , and  $x_{kt}^{(d)}$  is the  $k$ <sup>th</sup> component of  $\mathbf{x}^{(d)}[t]$ . We write the conditional probability distributions in (43) as follows:

$$p(\mathbf{Y}^{(d)} | \mathbf{Z}^{(d)}) = \mathbb{1}(\mathbf{Z}^{(d)} \in (\mathbf{Z}^{(\text{lo})}, \mathbf{Z}^{(\text{hi})})), \quad (45)$$

$$\begin{aligned} p(\mathbf{Z}^{(d)} | \mathbf{X}^{(d)}; \mathbf{D}, \sigma_w^2) \\ = \prod_{t=\tau_p+1}^{\tau_p+\tau_d} \frac{1}{(\pi \sigma_w^2)^{N_r N_c}} \exp\left(-\frac{1}{\sigma_w^2} \left\| \mathbf{z}^{(d)}[t] - \mathbf{D}\mathbf{x}^{(d)}[t] \right\|_2^2\right), \end{aligned}$$

where  $\mathbb{1}(\cdot)$  is the indicator function,  $\mathbf{Z}^{(\text{lo})}$ ,  $\mathbf{Z}^{(\text{hi})}$  are the entry-wise lower and upper quantization intervals of the real and imaginary components of  $\mathbf{Y}^{(d)}$ . We present the posterior distributions that maximize the ELBO in the following two Lemmas. The computation of the posterior distribution  $q_{\mathbf{Z}}(\mathbf{Z}^{(d)})$  is similar to the computation of the posterior distribution of the unquantized pilot received signal in Lemma 4. Therefore, we omit the proof of Lemma 7 to avoid repetition. The proof of Lemma 6 is available in the supplementary material.

*Lemma 6 (Computation of  $q_{x_{kt}}(x_{kt}^{(d)})$ ):* The posterior  $q_{x_{kt}}(x_{kt}^{(d)})$  follows a Boltzmann distribution with the probability mass function

$$q_{x_{kt}}(x_{kt}^{(d)} = s_m) = \frac{\exp(f_{kt}(s_m))}{\sum_{s' \in \mathbb{M}} \exp(f_{kt}(s'))} \quad (46)$$

for  $m = 1, \dots, M$ , where  $k \in \{1, \dots, K N_c\}$ ,  $t \in \{\tau_p + 1, \dots, \tau_p + \tau_d\}$ ,  $\mathbb{M} = \{s_1, \dots, s_M\}$  is the signal constellation set of cardinality  $M$ , and  $f_{kt}(s)$  is shown in (47) on the next page, where  $\Re$  and  $*$  denote the real part and complex conjugate operators, respectively,  $\mathbf{D}_{:,k}$  is the  $k$ <sup>th</sup> column of  $\mathbf{D}$ ,  $\langle \mathbf{z}^{(d)}[t] \rangle$  and  $\langle x_{k't}^{(d)} \rangle$  are the posterior means of  $q_{\mathbf{z}_t}(\mathbf{z}^{(d)}[t])$  and  $q_{x_{k't}}(x_{k't}^{(d)})$ , respectively. ■

We compute the mean and mean square value of  $q_{x_{kt}}(x_{d,kt})$  as follows:

$$\langle x_{kt}^{(d)} \rangle = \sum_{s \in \mathbb{M}} s q_{x_{kt}}(s), \quad \langle |x_{kt}^{(d)}|^2 \rangle = \sum_{s \in \mathbb{M}} |s|^2 q_{x_{kt}}(s).$$

*Lemma 7 (Computation of  $q_{\mathbf{Z}}(\mathbf{Z}^{(d)})$ ):* The posterior distribution  $q_{\mathbf{Z}}(\mathbf{Z}^{(d)})$  is truncated complex normal, with mean

$$f_{kt}(s) = -\frac{1}{\sigma_w^2} \left( \|\mathbf{D}_{:,k}\|^2 |s|^2 - 2\Re \left[ \mathbf{D}_{:,k}^H \left( \langle \mathbf{z}^{(d)}[t] \rangle - \sum_{\substack{k'=1 \\ k' \neq k}}^{KN_c} \mathbf{D}_{:,k'} \langle x_{k't}^{(d)} \rangle \right) s^* \right] \right) + \ln p(x_{kt}^{(d)} = s), \quad (47)$$

---

**Algorithm 3** Quantized VB Soft Symbol Decoding
 

---

**Input:**  $\mathbf{Y}^{(d)}$ ,  $\mathbf{D}$ ,  $\mathbb{M} = \{s_1, \dots, s_M\}$ ,  $\tau_p$ ,  $\tau_d$ ,  $\sigma_w$ ,  $K$ ,  $N_c$ .

**Output:**  $q_{\mathbf{X}}(\mathbf{X}^{(d)})$ ,  $\langle \mathbf{X}^{(d)} \rangle$

- 1: Initialize  $\langle \mathbf{Z}^{(d)} \rangle$ ,  $\langle \mathbf{X}^{(d)} \rangle = \mathbf{0}_{KN_c \times \tau_d}$
  - 2: **repeat**
  - 3:   **for**  $k = 1$  to  $KN_c$  **do**
  - 4:     **for**  $t = \tau_p + 1$  to  $\tau_p + \tau_d$  **do**
  - 5:       Compute  $q_{x_{kt}}(x_{kt}^{(d)} = s)$  using (46)  $\forall s \in \mathbb{M}$ .
  - 6:       Compute  $\langle x_{kt}^{(d)} \rangle = \sum_{s \in \mathbb{M}} s q_{x_{kt}}(x_{kt}^{(d)} = s)$ .
  - 7:     **end for**
  - 8:   **end for**
  - 9:   Compute  $\langle \mathbf{Z}^{(d)} \rangle$  using (48).
  - 10: **until** stopping condition is met
- 

$$\begin{aligned} \langle \mathbf{Z}^{(d)} \rangle &= \mathbf{D} \langle \mathbf{X}^{(d)} \rangle \\ &+ \frac{\sigma_w}{\sqrt{2}} \frac{f\left(\frac{\mathbf{z}^{(lo)} - \mathbf{D} \langle \mathbf{X}^{(d)} \rangle}{\sigma_w / \sqrt{2}}\right) - f\left(\frac{\mathbf{z}^{(hi)} - \mathbf{D} \langle \mathbf{X}^{(d)} \rangle}{\sigma_w / \sqrt{2}}\right)}{F\left(\frac{\mathbf{z}^{(hi)} - \mathbf{D} \langle \mathbf{X}^{(d)} \rangle}{\sigma_w / \sqrt{2}}\right) - F\left(\frac{\mathbf{z}^{(lo)} - \mathbf{D} \langle \mathbf{X}^{(d)} \rangle}{\sigma_w / \sqrt{2}}\right)}, \quad (48) \end{aligned}$$

where  $\mathbf{Z}^{(lo)}$  and  $\mathbf{Z}^{(hi)}$  are defined in (45),  $\langle \mathbf{X}^{(d)} \rangle$  contains the posterior means of  $q_{x_{kt}}(x_{kt}^{(d)}) \forall k, t$  as its entries,  $f(\cdot)$ ,  $F(\cdot)$ , and the division operation are as defined in Lemma 4. ■

As mentioned in Section IV, the VB algorithm starts by randomly initializing the latent variables, and iteratively computes the posterior distributions of data symbols. We use the posterior distributions of the data symbols in (46) to calculate the bit LLRs. We present the quantized VB soft symbol decoding procedure in Algorithm 3. Next, we describe the data-aided channel estimation procedure.

## VI. ITERATIVE QUANTIZED VB CHANNEL ESTIMATION AND SOFT SYMBOL DECODING

In this section, we merge the channel estimation and soft symbol decoding into an iterative algorithm that improves on the system performance obtained by only executing Algorithms 2 and 3. We utilize the data-aided channel estimation system model to refine the channel estimates in an iterative fashion. Recall our system model (26) from section III-C:  $\mathbf{Y} = \mathcal{Q}_b(\mathbf{Z}) = \mathcal{Q}_b(\Phi \mathbf{H} + \mathbf{W})$ . We start with the pilot based channel estimation Algorithm 2 followed by the soft symbol decoding Algorithm 3. Now, we utilize the posterior means of the decoded data symbols to form a new measurement matrix  $\Phi$  that is input to the channel estimation block. In a VB procedure, we obtain the posterior distribution of a given latent variable by computing the expectation of the joint probability distribution w.r.t. the posterior distributions of all the other latent variables. This in turn means that its posterior distribution depends only on the posterior statistics of the other latent variables. Moreover, in the context of soft symbol decoding, the posterior statistics of the data symbol enter the equivalent measurement equation through their posterior

TABLE I.  
PER-ITERATION COMPLEXITY OF THE QVB ALGORITHM 4

Matrix	Order Complexity
$\langle \mathbf{X}^{(d)} \rangle$	$MK^2N_c^3N_r\tau_d$
$\Sigma_{\mathbf{H}}$	$K^2L^2N_c(\tau_p + \tau_d)$
$\langle \mathbf{H} \rangle$	$KLN_c(KL + N_r)(\tau_p + \tau_d)$
$\langle \mathbf{Z}^{(d)} \rangle$	$KN_c^2N_r\tau_d$
$\langle \mathbf{Z} \rangle$	$KLN_cN_r(\tau_p + \tau_d)$

means (see (47) and (48)). Therefore, we use the posterior means of the data symbols to construct a new measurement matrix for iterative channel estimation.

The data-aided channel estimation procedure follows the same steps as in Algorithm 2 except that its inputs  $\mathbf{Y}^{(p)}$  and  $\Phi^{(p)}$  are replaced by  $\mathbf{Y}$  and  $\Phi$ , respectively. The derivation is similar to Sec. IV; we provide the final expressions of the posterior statistics of the latent variables below.

$$\Sigma_{\mathbf{H}} = \left( \frac{1}{\sigma_w^2} \Phi^H \Phi + \langle \mathbf{P} \rangle \right)^{-1}, \quad \langle \mathbf{H} \rangle = \frac{1}{\sigma_w^2} \Sigma_{\mathbf{H}} \Phi^H \langle \mathbf{Z} \rangle, \quad (49)$$

$$\langle \alpha_k \rangle = \frac{a + N_r}{r + \sum_{n=1}^{N_r} \langle |h_{kn}|^2 \rangle}, \quad k = 1, \dots, KL, \quad (50)$$

$$\langle \mathbf{Z} \rangle = \Phi \langle \mathbf{H} \rangle + \frac{\sigma_w}{\sqrt{2}} \frac{f\left(\frac{\mathbf{z}^{(lo)} - \Phi \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right) - f\left(\frac{\mathbf{z}^{(hi)} - \Phi \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right)}{F\left(\frac{\mathbf{z}^{(hi)} - \Phi \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right) - F\left(\frac{\mathbf{z}^{(lo)} - \Phi \langle \mathbf{H} \rangle}{\sigma_w / \sqrt{2}}\right)}, \quad (51)$$

where  $\mathbf{Z}^{(lo)}$  and  $\mathbf{Z}^{(hi)}$  are the lower and upper quantization thresholds corresponding to  $\mathbf{Y}$ . The other notations are as defined in Lemma 4. We repeat this process of channel estimation and data decoding for a fixed number of iterations. Finally, we use the posterior distribution of the transmit symbols to obtain the bit LLRs, which are deinterleaved and input to the channel decoder. We present the iterative VB channel estimation and soft symbol decoding in Algorithm 4.

Now, we present a variant of the iterative channel estimation and soft symbol decoding Algorithm 4 that marginally improves the performance. We use the a posteriori bit LLRs from the channel decoder to adapt the data prior that is input to the soft symbol detector in the next iteration. We interleave the posterior bit LLRs output by the channel decoder, and generate the extrinsic information to compute the symbol LLRs. Then, instead of using a uniform prior on the data symbols, we bias the data detector by a non-uniform data prior. In every outer iteration of the iterative channel estimator and soft symbol decoder, we increase the probability mass on the data symbol output by the extrinsic information progressively by a judiciously chosen step size. At lower SNRs, such prior adaptation may lead to error propagation effects, but at SNRs of interest, this leads to performance improvement. We show a block diagram for one outer iteration of the iterative channel estimation and soft symbol decoding algorithm with data prior

---

**Algorithm 4** Iterative Quantized VB Channel Estimation and Soft Symbol Decoding
 

---

**Input:**  $\mathbf{Y}^{(p)}$ ,  $\mathbf{Y}^{(d)}$ ,  $\mathbf{X}^{(p)}$ ,  $\mathbb{M} = \{s_1, \dots, s_M\}$ ,  $M$ ,  $\tau_p$ ,  $\tau_d$ ,  $\sigma_w$ ,  $N_r$ ,  $K$ ,  $L$ ,  $N_c$ , MAX\_ITER.

**Output:**  $\langle \mathbf{H} \rangle$ ,  $q_{\mathbf{X}}(\mathbf{X}^{(d)})$ ,  $\langle \mathbf{X}^{(d)} \rangle$

- 1: Initialize  $\langle \mathbf{H} \rangle$ ,  $\langle \boldsymbol{\alpha} \rangle$ ,  $\langle \mathbf{X}^{(d)} \rangle = \mathbf{0}_{KN_c \times \tau_d}$ ,  $a$ ,  $r$
- 2: **Initial Channel Estimation:** Run Algorithm 2.
- 3: **repeat**
- 4:   **Soft Symbol Decoding:**
- 5:   **repeat**
- 6:     Compute  $\mathbf{D}$  using (23).
- 7:     **for**  $k = 1$  to  $KN_c$  **do**
- 8:       **for**  $t = \tau_p + 1$  to  $\tau_p + \tau_d$  **do**
- 9:          Compute  $q_{x_{kt}}(x_{kt}^{(d)} = s)$  using (46)  $\forall s \in \mathbb{M}$ .
- 10:          Compute  $\langle x_{kt}^{(d)} \rangle = \sum_{s \in \mathbb{M}} s q_{x_{kt}}(x_{kt}^{(d)} = s)$ .
- 11:       **end for**
- 12:     **end for**
- 13:     Compute  $\langle \mathbf{Z}^{(d)} \rangle$  using (48).
- 14:   **until** stopping condition is met
- 15:   **Data-Aided Channel Estimation:**
- 16:   Compute  $\Phi$  using (25),  $\tilde{a}_k = a + N_r$ ,  $k = 1, \dots, KL$ .
- 17:   Initialize  $\langle \mathbf{Z} \rangle$ ,  $\langle \boldsymbol{\alpha} \rangle$ .
- 18:   **repeat**
- 19:      $\langle \mathbf{P} \rangle = \text{diag}(\langle \boldsymbol{\alpha} \rangle)$
- 20:     Compute  $\Sigma_{\mathbf{H}}$  and  $\langle \mathbf{H} \rangle$  using (49).
- 21:     Compute  $\langle \alpha_k \rangle$  using (50),  $k = 1, \dots, KL$ .
- 22:     Compute  $\langle \mathbf{Z} \rangle$  using (51).
- 23:   **until** stopping condition is met
- 24: **until** MAX\_ITER times

---

adaptation in Fig. 5. We use this algorithm in our simulations, unless specified otherwise.

#### A. Computational Complexity

In this subsection, we analyze the per-iteration number of floating point operations (flops) in the VB algorithm as a function of the system dimensions. Table I shows the order ( $\mathcal{O}$ ) of the per-iteration computational complexity of the steps involved in one iteration of Algorithm 4. The complexity scales cubically with the number of subcarriers, as the square of the number of users and the channel length, and linearly with the number of receive antennas, constellation size, and number of pilot and data symbols. In particular, the complexity is linear in the number of data symbols  $\tau_d$ , unlike maximum likelihood approaches where the complexity grows exponentially with  $\tau_d$ . The complexity of the conventional MIMO-OFDM MMSE channel estimation algorithm scales cubically with the number of users and the channel length, whereas the complexity of BiGAMP scales linearly with the pilot and data symbols, subcarriers, number of users and channel length, and returns hard decisions of the data symbols. Note that, in order to speed up computations, we can precompute the PDF and CDF of a standard normal variable and store them in a lookup table.

## VII. SIMULATION RESULTS

In the following subsections, we first evaluate the Bayesian and hybrid CRLBs developed in Sec. II, and then study the NMSE and coded BER performance of the iterative VB channel estimation and soft symbol decoding algorithm.

#### A. Cramér-Rao Lower Bounds

In this subsection, we evaluate the Bayesian and hybrid CRLBs on the NMSE of any estimator of a compressible signal using quantized measurements. We compute the CRLBs for the recovery of joint sparse vectors from compressive measurements acquired using a random underdetermined measurement matrix  $\Phi \in \mathbb{C}^{M \times N}$ , whose entries are i.i.d. and complex normal distributed with mean 0 and variance 1. We generate each column of a compressible signal  $\mathbf{X} \in \mathbb{C}^{N \times L}$  by sampling from a complex normal distribution with mean  $\mathbf{0}$  and precision matrix  $\mathbf{P} = \text{diag}(\boldsymbol{\alpha})$ . The precision matrix is generated using a Gamma distribution with shape and rate parameters  $a$  and  $r$ , respectively. We show the decay profile of the sorted magnitudes of compressible signals of length  $N = 512$ , generated using different shape and rate parameters, in Fig. 6. The rapid decay of the coefficients shows that the signals are compressible. We fix  $N = 512$ ,  $T = 20$ , and set the rate parameter to  $10^{-8}$  for all the simulations in this subsection. We define the SNR as  $1/\sigma_w^2$ . In this case, since the support set of the jointly compressible signals contains all the indices, the support-aware CRLB and Bayesian CRLB coincide. Hence, we evaluate only the Bayesian and hybrid CRLBs.

Figure 7 shows the NMSE of the VB algorithm for the unquantized (labeled “UQ”) and 3 bits quantization (labeled “3 bits”) cases, and the Bayesian (labeled “BCRLB”) and hybrid CRLBs (labeled “HCRLB”), as a function of the shape parameter of the Gamma hyperprior of the precision matrix. We set the number of measurements to 250 and the SNR to 40 dB. In the case of hybrid CRLB, we know the generative model of the compressible signal, which provides extra information. Hence, the hybrid CRLB is a tight lower bound on the performance of the VB algorithm. The Bayesian CRLB only uses the parameters for the Gamma hyperprior and is therefore looser. Also, the Bayesian CRLB for the unquantized and 3 bits almost overlap on each other, whereas there is a small gap between the two in the case of hybrid CRLB. As the shape parameter increases, the peak value of the compressible signal decreases, and hence the compressibility of the signal decreases. Due to this, the NMSE and the CRLBs both increase with the shape parameter.

Figure 8 shows the NMSE and hybrid CRLB as a function of the number of measurements for the unquantized and  $\{1, 2, 3\}$  bits quantized cases. We do not include the Bayesian CRLB in this figure to avoid clutter. The shape parameter is set to 0.55 and the SNR to 30 dB. We observe that the gap between the NMSE and CRLB for the 1-bit quantization is higher compared to the  $\{2, 3\}$  bits and the unquantized cases. This shows that having only sign measurements leads to severe performance loss due to the large quantization noise, which results in higher NMSE. We also see that the gap

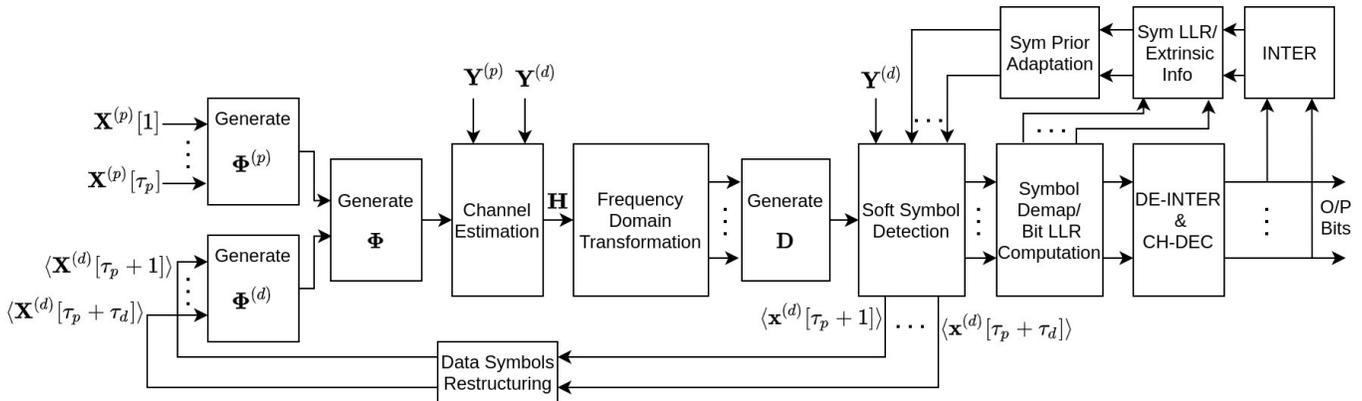


Figure 5. Quantized VB iterative channel estimation and soft symbol decoding with data prior adaptation.

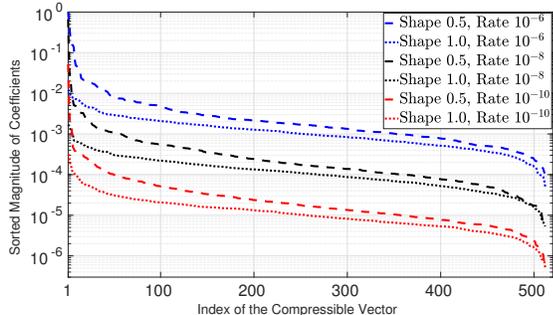


Figure 6. Magnitude decay profile of the sorted magnitudes of i.i.d. samples drawn from a complex normal distribution parameterized by a Gamma distributed precision matrix.

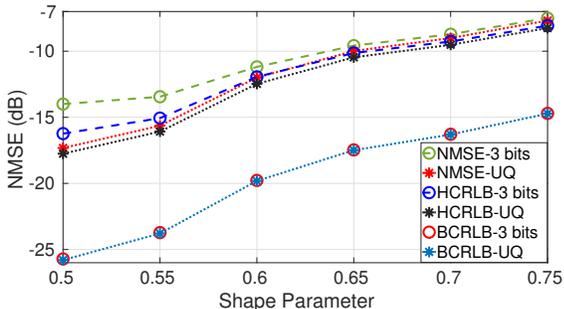


Figure 7. NMSE (dB) vs. the shape parameter of the Gamma hyperprior of the precision matrix, with  $M = 250$ ,  $N = 512$ ,  $T = 20$ ,  $\text{SNR} = 40$  dB.

between the unquantized case and the 3-bits quantization is very small, which empirically shows that a 3-bit quantizer provides a good trade-off between the performance and system complexity. We thus set the ADC resolution to 3 bits in all the further simulations, unless specified otherwise.

### B. QVB Channel Estimation and Soft Symbol Decoding

In this subsection, we evaluate the NMSE and coded BER performance of the iterative VB channel estimation and soft symbol decoding algorithm. We also study the impact of the various system parameters on the support-aware Bayesian CRLB derived in Section II. The data bits are generated i.i.d. from a uniform distribution. Each UE's data bits are encoded with an LDPC channel code from 3GPP 5G NR specifications [37]. We use the parity check matrix from LDPC base graph 0 with a lifting size  $Z_c$  set to 8 and set index 0, which results in 176 message bits and 544 coded bits per block. The coded bits are interleaved by a

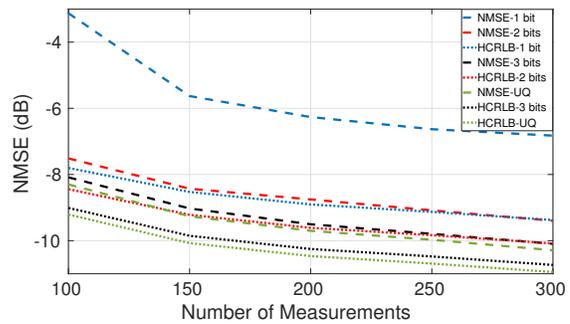


Figure 8. NMSE (dB) as a function of the number of quantized measurements.  $N = 512$ ,  $T = 20$ ,  $\text{shape} = 0.55$ ,  $\text{SNR} = 30$  dB.

random interleaver which is known to both the UE and the BS, mapped to 4-QAM constellation of unit energy, OFDM modulated, and transmitted over frequency-selective wireless channels.<sup>4</sup> We assume that the LSFCs between the  $k^{\text{th}}$  UE and the BS antennas are the same due to the close spacing between the antennas compared to the BS-UE distance. The UEs adopt path loss inversion based transmit power control that compensates for the LSFCs, and therefore we set them to 1 in all our simulations.<sup>5</sup> Each tap of the frequency-selective channels is i.i.d. circularly symmetric complex normal distributed with mean 0 and variance 1. We include the details of the system parameters used for simulations in the captions of each simulation plot. We define the SNR as  $1/\sigma_w^2$ . We use the Frobenius norm of the difference between the channel estimates (and estimates of the unquantized received data symbols) in consecutive iterations as the stopping condition for the VB channel estimation (and soft symbol decoding) procedures. We set the maximum number of iterations for VB channel estimation and data detection algorithms to 25, and the total number of iterations in Algorithm 4 to four.

Fig. 9 compares the CRLB derived for a quantized system with the analytical CRLB for an unquantized system (labeled UQ, from [29].) We see that, as the ADC resolution increases, the gap between the quantized and unquantized Bayesian CRLBs decreases, and the bounds meet beyond an ADC resolution of 4 bits. In fact, if the ADC resolution is infinite,

<sup>4</sup>With higher order constellations, the performance is similar, with an expected shift in the SNR required to achieve a given coded BER.

<sup>5</sup>The LFSCs can be estimated at the UEs, for example, using the synchronization signals that are periodically transmitted by the BS.

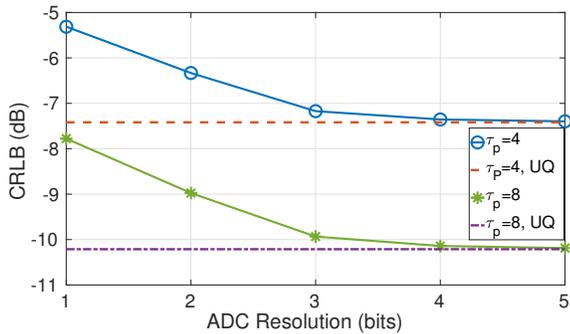


Figure 9. CRLB (dB) as a function of ADC resolution (bits) for  $N_r = 64$ ,  $K = 16$ ,  $L = 64$ ,  $L_{sp} = 8$ ,  $\text{SNR} = 1$  dB.

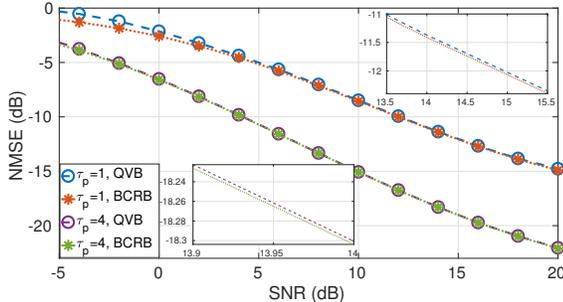


Figure 10. NMSE (dB) as a function of SNR (dB) for  $N_r = 40$ ,  $K = 10$ ,  $N_c = 256$ ,  $L = 64$ ,  $L_{sp} = 8$ , 3-bits quantization.

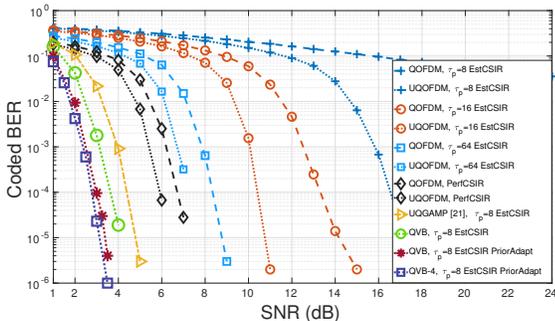


Figure 11. Coded BER as a function of SNR (dB) for  $N_r = 32$ ,  $K = 8$ ,  $N_c = 256$ ,  $L = 32$ ,  $L_{sp} = 8$ ,  $\tau_d = 10$ , 3-bits quantization. The conventional OFDM receiver (curves labelled ‘‘UQOFDM’’ and ‘‘QOFDM’’) uses the soft-detection procedure from [38].

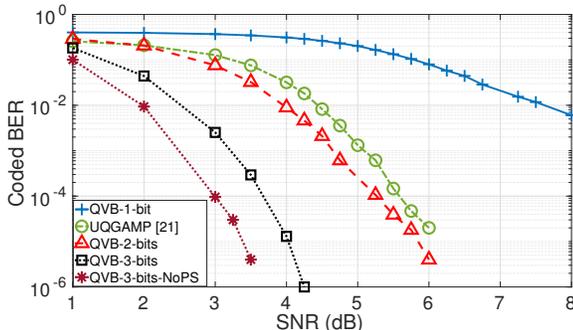


Figure 12. Coded BER as a function of SNR (dB) for  $N_r = 32$ ,  $K = 8$ ,  $N_c = 256$ ,  $L = 32$ ,  $L_{sp} = 8$ ,  $\tau_d = 10$  with a square root raised cosine pulse shaping transmit and receive filters.

our CRLB matches with the analytical expression derived for an unquantized system. Therefore, our derived Bayesian CRLB captures the effect of low resolution ADCs well, and can serve as a benchmark to evaluate the NMSE performance

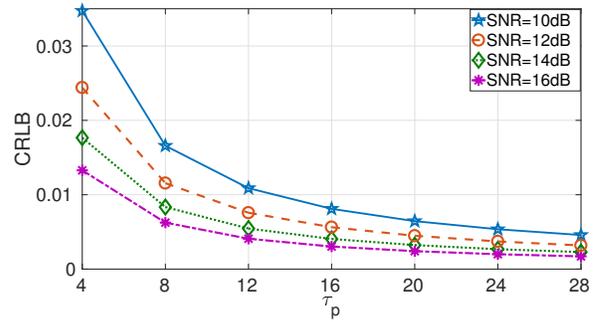


Figure 13. CRLB (dB) as a function of  $\tau_p$  for  $N_r = 32$ ,  $K = 8$ ,  $L = 64$ ,  $L_{sp} = 16$ ,  $N_c = 256$ , 3-bits quantization.

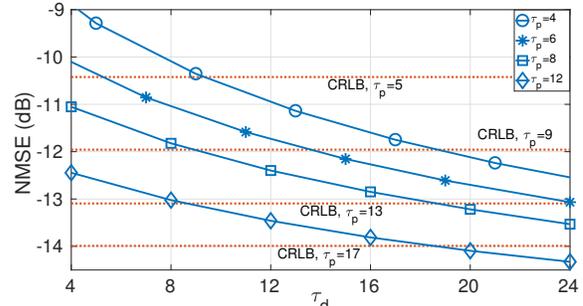


Figure 14. NMSE (dB) as a function of  $\tau_d$  for  $N_r = 64$ ,  $K = 16$ ,  $L = 64$ ,  $L_{sp} = 8$ ,  $N_c = 256$ ,  $\text{SNR} = 1$  dB, 3-bits quantization.

of any estimator in a quantized setup.

Fig. 10 shows the NMSE performance of the quantized VB algorithm and the Bayesian CRLB when  $N_r = 40$ ,  $K = 10$ ,  $L_{sp} = 8$ , and 3 bits quantization. We observe that when  $\tau_p = 1$  and  $\tau_p = 4$  OFDM symbols, the NMSE of VB overlaps with the Bayesian CRLB beyond 4 dB and  $-4$  dB SNR, respectively. At low SNRs and  $\tau_p = 1$ , the gap between the Bayesian CRLB and the NMSE of VB is slightly more than that at high SNRs, which can be attributed to the fact that there can be support recovery errors in VB due to high noise. In our wireless communication application, we typically operate at medium to high SNRs, where the Bayesian CRLB and VB almost overlap. Therefore, the BCRB serves as a good benchmark to characterize the NMSE performance of a Bayesian-inspired channel estimator.

Fig. 11 compares the coded BER performance of the VB algorithm with that of an unquantized joint channel estimation and data detection algorithm based on BiGAMP [21], and MMSE channel estimator and soft-detector [38]. For the quantized MMSE receiver, we compute the DFT after the quantization, and perform the equalization. We set the number of outer iterations of the iterative channel estimator and soft symbol decoder to 8. An advantage of the VB algorithm is that it can recover the channel with only one pilot OFDM symbol. However, for fair comparison, we set  $\tau_p = 8$  because the conventional OFDM receiver cannot estimate the channel in an underdetermined setting. We see that, at a BER of  $10^{-4}$ , the quantized VB algorithm (labelled ‘‘QVB  $\tau_p = 8$ , EstCSIR’’) outperforms conventional OFDM receiver with unquantized observations and channels estimated using  $\tau_p = 8$  pilot OFDM symbols by around 13 dB. In fact, it *even outperforms* the conventional OFDM receiver with *unquan-*

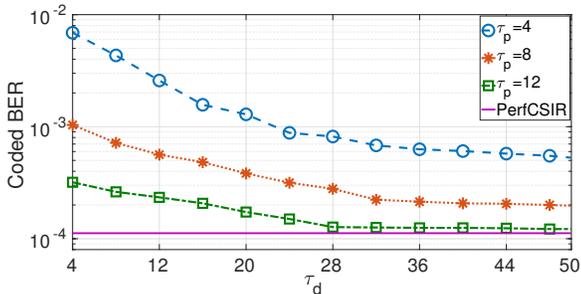


Figure 15. Coded BER as a function of  $\tau_d$  for  $N_r = 64$ ,  $K = 16$ ,  $L = 64$ ,  $L_{sp} = 8$ ,  $\text{SNR} = 0$  dB, 3-bits quantization.

tized observations and perfect CSIR by 2.5 dB. Moreover, the VB algorithm with only 3 bits quantization performs better than an unquantized GAMP by around 1 dB. This shows the importance of directly inferring the posterior distributions of the data symbols. We also see that the VB algorithm with data prior adaptation performs better than the VB algorithm without any adaptation in this scenario by more than 0.5 dB. Finally, the VB algorithm with 4-bits quantization (labelled “QVB-4  $\tau_p = 4$  EstCSIR PriorAdapt”) is only marginally better than the VB with 3 bits quantization. Therefore, a 3 bits quantizer is sufficient to achieve good performance in a low-resolution ADC based MIMO-OFDM system.

Fig. 12 compares the coded BER of the VB algorithm for the ADC resolution set to  $\{1, 2, 3\}$  bits with square root raised cosine transmit and receive pulse shaping filters. The roll-off factors for the transmit and receive filters are set to 0.3. The system bandwidth is set to 2 GHz, so the sampling period  $T_s$  is 0.5 ns. We set the cyclic prefix length to the maximum delay spread of  $L = 32$  symbols. The number of nonzero taps  $L_{sp}$  is set to 8, with the corresponding delays generated uniformly at random between 0 and  $(L - 2)T_s$ . The channel gains of the nonzero taps are i.i.d. complex normal with zero mean and unit variance. For comparison, we include the quantized VB algorithm with ideal pulse shaping filters (labeled “QVB-3-bits-NoPS”) and the BiGAMP algorithm [21] with unquantized pilot and data received signals (labeled “UQGAMP”). Although the pulse shaping filters introduce inter-symbol interference and noise correlation, it only results in a marginal performance loss of around 0.5 – 0.7 dB in the VB algorithm with 3 bits quantization. This shows that the VB algorithm is robust to non-idealities. Moreover, VB algorithm with only 2 bits quantization outperforms UQGAMP.

Fig. 13 shows the CRLB as a function of the pilot length  $\tau_p$  for various values of SNR. As  $\tau_p$  increases, the CRLB decreases, which is due to the increase in the number of observations. The slope of the CRLB curves decreases as  $\tau_p$  increases, and asymptotically becomes zero, which follows the law of diminishing returns. That is, if we vary only  $\tau_p$  by fixing all the other parameters, we do not see any significant performance improvement beyond a point. This is because, irrespective of the number of measurements, the quantization noise floor limits the improvement obtainable by increasing  $\tau_p$ . Also, the value of the threshold decreases as SNR increases, which shows that when the noise power is low, we can potentially achieve better spectral efficiency with fewer number of pilots.

Figures 14 and 15 show the NMSE (dB) and coded BER, respectively, as a function of the data duration,  $\tau_d$ , for the iterative VB algorithm. We also plot the CRLB in Fig. 14. The NMSE decreases with  $\tau_d$ , as expected, due to the increase in the number of virtual pilot symbols. In fact, the NMSE goes below the CRLB beyond a particular  $\tau_d$  in all the  $\tau_p$  configurations. For e.g., when  $\tau_p$  is 12, the NMSE is around  $-13$  dB for  $\tau_d = 8$ , whereas the CRLB of  $-13$  dB is achieved at  $\tau_p$  set to 13. That is, the VB channel estimator can attain an NMSE even lower than the CRLB computed using a larger training overhead, since it uses the data symbols as virtual pilots. On the other hand, when  $\tau_p$  is 12 and  $\tau_d$  is 5, the VB algorithm achieves an NMSE of around  $-12.5$  dB, which is higher than the CRLB when  $\tau_p$  is 17 by around 1.5 dB. Thus, the NMSE of the channel estimator is higher than the CRLB computed using  $\tau_p + \tau_d$  as the pilot duration. This is because the data symbols are also estimated using the received symbols. Nonetheless, the iterative data-aided channel estimation assists in reducing the training overhead and increasing the spectral efficiency. Further, the slope of the NMSE curves decreases with  $\tau_d$ , which reiterates our observation in Fig. 13 about the error floor due to the quantization noise.

In Fig. 15, we see that, as  $\tau_p$  increases, the coded BER decreases due to better channel estimation performance. We also include the coded BER performance for a quantized VB soft symbol decoding algorithm with perfect CSIR, which serves as a lower bound for the iterative quantized VB channel estimator and soft symbol decoder. We see that, when  $\tau_p$  is 12, and  $\tau_d$  is greater than 28, our iterative VB algorithm almost meets the performance of the perfect CSIR case. Given the coherence interval, such studies can guide system designers to configure  $\tau_p$  (and thus  $\tau_d$ ) and obtain the same BER performance as with perfect CSIR, or to choose  $\tau_p$  to attain the right trade off between training overhead and data duration and thereby achieve maximal spectral efficiency.

## VIII. CONCLUSIONS

We derived the Bayesian, hybrid, and support-aware CRLBs for an estimator of a compressible signal using quantized lower dimensional measurements. Next, we developed a pilot-based channel estimator and a soft symbol decoder using a VB framework, which directly infers the posterior distributions of the channel and data given the quantized received signals. We utilized the posterior statistics of the decoded data symbols to develop an iterative VB data-aided channel estimator and soft symbol decoder. We marginally improved the performance by proposing a variant of the iterative algorithm that used the posterior bit LLRs from the channel decoder for data prior adaptation. We benchmarked the NMSE performance of the VB estimator with that of the derived Bayesian CRLB, and numerically showed that it is efficient. We also evaluated the NMSE and coded BER performances of the iterative VB channel estimator and soft symbol decoder, and compared with the state-of-the-art. Finally, we provided interesting insights into the impact of various parameters on the system performance. Future work could consider

extending these ideas to millimeter-wave channels exploiting spatial sparsity, or account for carrier frequency and timing offsets across users by modeling them using latent variables that are estimated using the VB framework, and so on.

#### REFERENCES

- [1] S. S. Thoota and C. R. Murthy, "Cramér-Rao lower bound for Bayesian estimation of quantized MMV sparse signals," in *Proc. SPAWC*, 2021, pp. 171–175.
- [2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [3] Z. Jiayi, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Comm. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.
- [4] 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.
- [5] H. Kim and J. Choi, "Channel estimation for spatially/temporally correlated massive MIMO systems with one-bit ADCs," *EURASIP J. on Wireless Commun. and Netw.*, vol. 2019, no. 1, p. 267, Dec. 2019.
- [6] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.
- [7] B. Murmann, "Energy limits in A/D converters," in *2013 IEEE Faible Tension Faible Consommation*. IEEE, 2013, pp. 1–4.
- [8] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.
- [9] J. Liu, Z. Luo, and X. Xiong, "Low-resolution ADCs for wireless communication: A comprehensive survey," *IEEE Access*, vol. 7, pp. 91 291–91 324, Jul. 2019.
- [10] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *ArXiv*, p. arXiv:1404.7736, Apr. 2014.
- [11] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3111–3124, Apr. 2014.
- [12] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.
- [13] A. Mezghani and J. A. Nossek, "Belief propagation based MIMO detection operating on quantized channel output," in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 2113–2117.
- [14] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.
- [15] H. Wang, C.-K. Wen, and S. Jin, "Bayesian optimal data detector for mmWave OFDM system with low-resolution ADC," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1962–1979, Sep. 2017.
- [16] Z. Zhang, X. Cai, C. Li, C. Zhong, and H. Dai, "One-bit quantized massive MIMO detection based on variational approximate message passing," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2358–2373, May 2018.
- [17] H. He, C.-K. Wen, and S. Jin, "Bayesian optimal data detector for hybrid mmwave MIMO-OFDM systems with low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 469–483, Jun. 2018.
- [18] Y.-S. Jeon, N. Lee, S.-N. Hong, and R. W. Heath, "One-bit sphere decoding for uplink massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, Jul. 2018.
- [19] Y.-S. Jeon, H. Do, S.-N. Hong, and N. Lee, "Soft-output detection methods for sparse millimeter-wave MIMO systems with low-precision ADCs," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2822–2836, Apr. 2019.
- [20] H. Wang, W. Shih, C. Wen, and S. Jin, "Reliable OFDM receiver with ultra-low resolution ADC," *IEEE Trans. Wireless Commun.*, vol. 67, no. 5, pp. 3566–3579, May 2019.
- [21] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, pp. 2541–2556, May 2016.
- [22] Y. Ding, S.-E. Chiu, and B. D. Rao, "Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog-digital processing and low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 499–513, Jun. 2018.
- [23] P. Sun, Z. Wang, R. W. Heath, and P. Schniter, "Joint channel-estimation/decoding with frequency-selective channels and few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 899–914, Feb. 2019.
- [24] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "Svm-based channel estimation and data detection for one-bit massive mimo systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2086–2099, 2021.
- [25] X. Cheng, B. Xia, K. Xu, and S. Li, "Bayesian channel estimation and data detection in oversampled OFDM receiver with low-resolution ADC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5558–5571, Sep. 2021.
- [26] S. S. Thoota and C. R. Murthy, "Variational Bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low resolution ADCs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3467–3481, May 2021.
- [27] H. L. V. Trees, K. L. Bell, and Z. Tian, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2013.
- [28] M. P. Shützenberger, "A generalization of the Fréchet-Cramér inequality in the case of Bayes estimation," *Bull. Am. Math. Soc.*, vol. 63 (142), 1957.
- [29] R. Prasad and C. R. Murthy, "Cramér-Rao-Type bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 622–632, Feb. 2013.
- [30] J. García, J. Munir, K. Roth, and J. A. Nossek, "Channel estimation and data equalization in frequency-selective MIMO systems with one-bit quantization," *ArXiv*, p. arXiv:1609.04536, Mar. 2021.
- [31] Z. Shao, L. T. N. Landau, and R. C. De Lamare, "Channel estimation for large-scale multiple-antenna systems using 1-bit ADCs and oversampling," *IEEE Access*, vol. 8, pp. 85 243–85 256, May 2020.
- [32] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5016–5034, Aug. 2012.
- [33] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
- [34] Y. Ding, S.-E. Chiu, and B. D. Rao, "Sparse recovery with quantized multiple measurement vectors," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 845–849.
- [35] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer New York, 2006.
- [36] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, 2004.
- [37] "3GPP TS 38.212 V15.7.0 (2019-09); NR; Multiplexing and channel coding," 2019.
- [38] N. K. Chavali and A. R. Yalla, "A soft-demapper for coded MIMO-OFDM system," in *Proc. Int. Conf. on Contemp. Comput. and Inform. (IC3I)*, 2014, pp. 451–457.