Variational Bayes' Joint Channel Estimation and Soft Symbol Decoding for Uplink Massive MIMO Systems with Low Resolution ADCs

Sai Subramanyam Thoota and Chandra R. Murthy

***Abstract*—We consider the problem of joint channel estimation and data decoding in uplink massive multiple input multiple output systems with low resolution analog-to-digital converters (ADCs) at the base station. The nonlinearities introduced by the ADCs make the problem challenging: in particular, the existing linear detectors perform poorly. Also, the channel coding used in commercial wireless systems necessitates soft symbol detection to obtain satisfactory performance. In this paper, we present a low-complexity variational Bayesian (VB) inference procedure to jointly solve the (possibly correlated) channel estimation and soft symbol decoding problem. We present the approach in progressively more complex scenarios, including the case where even the channel statistics are not available at the receiver. Finally, we combine our proposed VB procedure with a belief propagation (BP) based channel decoder, which further enhances the performance without any additional complexity. We numerically evaluate the bit error rate (BER) and the normalized mean squared error (NMSE) in the channel estimates obtained by our algorithm as a function of various system parameters, and benchmark the performance against genie-aided and state-of-the-art receivers. The results show that VB procedure is a promising technique for the design of low-complexity advanced receivers in low resolution ADC based systems.**

***Index Terms*—ADC, channel estimation, massive MIMO, soft symbol decoding, variational Bayes.**

I. Introduction

Massive multiple input multiple output (MIMO) wireless communication systems, where the base station (BS) or access point (AP) is equipped with hundreds or thousands of antennas, and simultaneously serves tens or hundreds of users, is one of the key enabling technologies to meet the increasing demand for the data rate and energy efficiency [3], [4]. However, the advantages of massive MIMO come at the cost of increased power consumption and hardware complexity due to the large number of RF chains, high precision analog-to-digital converters (ADCs), etc. In particular, the power consumption of ADCs grow exponentially with the number of quantization bits per sample [5]–[7]. For example, a commercial 1 Gsamp/s 12-bit ADC from Texas Instruments consumes over 1 W of power [8]. Also, full precision ADCs require correspondingly high rate data processing at the receiver (for example, with 100 antennas, 500 Msamp/s, the data rate at the remote radio head is more than 1 Tb/s.) This motivates the need for employing low resolution ADCs in the BS of a massive MIMO system [9]–[13]. On the other hand, the capacity analysis of coarsely quantized MIMO systems shows that, under the assumption of perfect channel state information at the receiver (CSIR), at low signal to noise ratio (SNR) and up to a first order approximation in SNR, the mutual information of a MIMO system employing 1-bit ADCs is $2/\pi$ times that of a MIMO system without quantization. Also, quardrature phase shift keying (QPSK) is the optimal modulation scheme in terms of the achievable rate at low SNRs, and under i.i.d. Rayleigh fading, the ergodic capacity of a 1-bit quantized system increases linearly with the number of receive (rx) antennas [14]–[17]. Due to this promise of close to optimal performance despite the coarse quantization introduced by the ADCs, the receiver design in multiuser massive MIMO systems with low resolution ADCs is of great practical interest.

The nonlinearities introduced by low resolution quantization leads to three challenges in receiver design. First, *linear* receive processing techniques like zero forcing (ZF), regularized ZF (RZF), and minimum mean square error (MMSE) become highly suboptimal [18]. Second, the received *training* symbols are also subject to low resolution quantization, leading to poor channel estimates. In practice, a large training overhead is needed to obtain reliable channel estimates [6], [19]. We note that although the primary task of the receiver is to correctly decode the data symbols, obtaining accurate CSIR is important for other tasks such as link adaptation. Third, practical channel decoders exhibit significantly lower codeword error rates when they are provided *soft symbol estimates* (i.e., log likelihood ratios of the coded data bits) as input, compared to the case where hard decisions are performed on the data symbols prior to channel decoding. Obtaining high-quality soft symbol estimates from coarsely quantized samples is challenging. These considerations motivate the need to devise novel techniques for joint channel estimation and soft symbol decoding with low resolution ADCs, which is the focus of this paper.

The problem of data detection in massive MIMO systems with low resolution ADCs has been studied in the literature [20]–[32]. Most of these studies assume the availability of perfect CSIR, and perform data decoding in multiuser massive MIMO systems with low resolution ADCs [20]–[26]. Techniques for joint channel estimation and data detection with low resolution ADCs have been developed based on

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

This work was financially supported by research grants from Intel Inc. and the Ministry of Electronics & Information Technology, Govt. of India.

A part of this work has appeared in [1], [2].

convex optimization [27] and approximate message passing (AMP) [28], [29]. Under the assumption of perfect CSIR, [30] and [31] explore weighted Hamming distance based soft detection and minimum mean square error (MMSE) detection, respectively, along with successive interference cancellation (SIC) under 1-bit quantization, while [32] proposes approximate belief propagation (BP) based MIMO detection in coarsely quantized systems. Variational Bayesian inference (VBI) has been used to develop high-performing algorithms for channel estimation and data decoding, without considering the nonlinearities introduced by the low resolution ADCs [33], [34]. It is worth noting that algorithms based on AMP or expectation propagation (EP) are not numerically stable, requiring heuristic modifications to address such issues [33]. In contrast, in this paper, we employ a more principled VBI based approach that is globally convergent to a local optimum.

To the best of our knowledge, most of the existing studies on massive MIMO systems do not consider joint channel estimation and soft symbol decoding, spatial correlation, and coarse quantization together. Our solution allows us to intrinsically learn the channel statistics from the quantized observations, which is potentially useful for power control and/or link adaptation also. Moreover, using VBI to compute the posterior distributions of the data leads to low complexity, high performing receivers. Preliminary versions of this work have appeared in [1], [2]. Both these works were restricted to independent and identically distributed (i.i.d.) channels. While [1] considered soft symbol decoding with perfect CSIR, [2] extended the approach to joint channel estimation and soft symbol decoding. We present several advances in this paper, including the extension to correlated Rayleigh fading channels, unknown channel statistics, etc. We also merge the VB receiver with a BP based LDPC channel decoder and utilize its extrinsic information to adapt the data prior, which further improves the performance.

We formulate the channel estimation and data decoding as a statistical inference problem in a directed probabilistic graphical model and solve it using a VBI approach. The quantized received signal is the observed variable, while the wireless channel and the $M-$QAM modulated data symbols are the latent variables. The goal is to jointly infer the posterior distributions of both the channel state and the data symbols given the quantized received signal consisting of pilot and data symbols. Since direct computation of the posterior distribution is intractable, we approximate the posterior with a factorized variational approximation over the latent variables, and compute the factors by maximizing the evidence lower bound (ELBO). The ELBO is maximized by minimizing the Kullback-Leibler (KL) divergence between the exact and the factorized distributions. The resulting iterative algorithm converges to a stationary point of the ELBO. This method of using factorized distributions originates from the mean-field approximation in statistical physics; we refer the reader to [35] for an excellent introduction to VBI. A key novelty in our solution lies in the introduction of appropriately chosen latent variables. This leads to closed-form, computationally simple updates, and the resulting iterative algorithm has low complexity and is also fast.

The main contributions of our paper are:

1) We model the channel estimation and data decoding problems in the uplink of a massive MIMO system with low resolution ADCs as a statistical inference problem. We derive the posterior distributions of the channel and the data symbols obtained from the pilot and data observations using the VB framework. We consider the following two cases:
   a) The *correlated Rayleigh fading* case, where UEs' channels are independent of each other, but the entries of the channel vector between a UE and the BS are correlated. In this case, our algorithm exploits the channel correlation to improve the channel estimation and data detection performance. In the absence of correlation, the algorithm becomes computationally very simple, with no matrix inversions.
   b) The *no CSIR* case, in which the BS is unaware of the large scale fading coefficients (LSFCs) between the UEs and the BS. We infer the LSFCs using a non-informative inverse Gamma prior. We empirically show that the resulting receiver is only marginally inferior to a receiver that has perfect knowledge of the LSFCs. Learning the LSFCs is potentially useful for downlink precoding and link adaptation.
2) We provide theoretical insights on the structure of the solution by relating it to the MMSE channel estimation and successive interference cancellation based receiver. We show that, when the interference is perfectly cancelled, the resulting channel estimate is the Bayes' optimal MMSE estimator.
3) We merge the VB receiver with a (BP) channel decoder and utilize its extrinsic information to adapt the data prior, which further improves the performance.
4) We study the bit error rate (BER) and channel normalized mean squared error (NMSE) performance of the VB algorithms with respect to different system parameters such as the ADC resolution, data power, pilot power boosting etc. and provide key insights into system design. We also evaluate the performance for both the correlated and uncorrelated fading cases as a function of channel parameters such as the coherence interval, and benchmark it against state-of-the-art bilinear generalized approximate message passing (BiGAMP) based joint channel and data estimator [28] and MMSE soft decoder [36]. Our numerical results show that the VB soft symbol decoder offers excellent performance and fast convergence, and even outperforms an unquantized BiGAMP joint channel estimator and data detector, making it an attractive choice for high data rate applications.

One of the main takeaways from our work is that VBI is a powerful and flexible technique for designing receivers for massive MIMO systems, particularly when the BS employs low resolution ADCs. Using VBI to infer the required posterior distributions helps to reduce the pilot overhead required to achieve a given BER. Further, since the algorithm yields soft symbol estimates, it is well suited for coded communications. Another crucial takeaway is that the assumption of perfect

CSI at the receiver significantly overestimates the system performance, which we illustrate through empirical studies in Sec. IV. Therefore, in a low resolution ADC setup, since both the pilots and data are coarsely quantized, it is crucial to account for channel estimation errors while designing receivers and evaluating performance.

Notation: Matrices and vectors are denoted by boldface uppercase and boldface lowercase letters. $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. $\mathbb{E}$ denotes the expectation operator. $\langle f(\mathbf{X}) \rangle$ denotes the expectation of $f(\mathbf{X})$ with respect to an approximate distribution $q(\mathbf{X})$. $\phi(x)$ and $\Phi(x)$ denote the probability density and cumulative distribution functions of a standard normal random variable evaluated at x. $\mathbf{I}_M, \mathbf{0}_M$ denote an identity and a zero matrix of size $M \times M$, respectively. $\mathcal{Q}(\cdot)$ denotes an elementwise quantization operation of the real and imaginary components of the argument.

II. System Model and Problem Statement

We consider the uplink (UL) of a single cell massive MIMO system with N_r rx antennas at the BS and K single transmit antenna user equipments (UEs), where $N_r \geq K$. The UEs encode their raw data bits using a channel code, interleave the coded bits, map the interleaved data bits to the signal constellation, insert the pilot symbols, then up-convert the signal to the carrier frequency and transmit it to the BS. The signal received at the BS is down-converted to the baseband, sampled, and quantized using a b-bit ADC before passing it to the digital front-end of the receiver, as shown in Fig. 1.

The UEs transmit data simultaneously over a frame consisting of τ_p known pilot symbols ($\tau_p \geq K$) followed by τ_d unknown data symbols. The complex baseband pilot and data symbols received at the BS, denoted by $\mathbf{Z}_p \in \mathbb{C}^{N_r \times \tau_p}$ and $\mathbf{Z}_d \in \mathbb{C}^{N_r \times \tau_d}$, respectively, are given by

$$\begin{aligned} \mathbf{Z}_p &= [\mathbf{z}_{p,1}, \ldots, \mathbf{z}_{p,\tau_p}] = \mathbf{H}\mathbf{X}_p + \mathbf{W}_p, \\ \mathbf{Z}_d &= [\mathbf{z}_{d,1}, \ldots, \mathbf{z}_{d,\tau_d}] = \mathbf{H}\mathbf{X}_d + \mathbf{W}_d, \end{aligned} \tag{1}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N_r \times K}$, $\mathbf{X}_p = \left[\mathbf{x}_{p,1}, \ldots, \mathbf{x}_{p,\tau_p}\right] \in \mathbb{C}^{K \times \tau_p}$, and $\mathbf{X}_d = [\mathbf{x}_{d,1}, \ldots, \mathbf{x}_{d,\tau_d}] \in \mathbb{M}^{K \times \tau_d}$ denote the channel matrix, pilots, and the $M-$QAM modulated transmit data symbols of the K users, respectively. Here, $\mathbf{h}_k \in \mathbb{C}^{N_r}$ is the channel between the k^{th} UE and the BS, and $\mathbf{x}_{p,t} \in \mathbb{C}^K$ and $\mathbf{x}_{d,t} \in \mathbb{M}^K$ the pilot and data symbols transmitted by the K UEs at the t^{th} symbol, respectively, where $\mathbb{M}$ denotes the M-QAM constellation. Also, $\mathbf{W}_p \in \mathbb{C}^{N_r \times \tau_p}, \mathbf{W}_d \in \mathbb{C}^{N_r \times \tau_d}$ are the additive white Gaussian noise at the receiver during the pilot and data phases, with i.i.d. $\mathcal{CN}(0, \sigma_w^2)$ entries, where $\mathcal{CN}(\mu, \sigma^2)$ denotes the circularly symmetric complex normal distribution with mean μ and variance σ^2.

We assume that the data symbols are i.i.d. and are drawn from a uniform probability distribution P_X defined on the signal constellation $\mathbb{M}$. The average data transmit power $\mathbb{E}(|x_{d,kt}|^2)$ of the k^{th} UE is set to $P_{T,k}$. The pilot transmission power is set to be the same as the data transmit power. Mathematically, $P_{\mathbf{X}_d}(\mathbf{X}_d) = \prod_{k=1}^{K} \prod_{t=1}^{\tau_d} P_X(x_{d,kt})$.

The channel $\mathbf{h}_k$ is modeled as a correlated Rayleigh flat fading channel, which remains constant over a frame consisting of τ_p pilot symbols and τ_d data symbols. That is, $\mathbf{h}_k$ is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$, where $\mathbf{R}_k \in \mathbb{C}^{N_r \times N_r}$ is a positive semidefinite covariance matrix. Also, without loss of generality, we include the path loss component in $\mathbf{R}_k$. We denote $\mathbf{R} \triangleq \{\mathbf{R}_1, \ldots, \mathbf{R}_K\}$, and thus, $P_{\mathbf{H}}(\mathbf{H}; \mathbf{R}) = \prod_{k=1}^{K} \frac{1}{\pi^{N_r} \det(\mathbf{R}_k)} \exp\left(-\mathbf{h}_k^H \mathbf{R}_k^{-1} \mathbf{h}_k\right)$. The LSFC of the channel from the k^{th} UE to the BS (i.e., the diagonal entries of $\mathbf{R}_k$) is denoted by β_k. We define the SNR of the system as $\frac{\sum_k P_{T,k} \beta_k}{K \sigma_w^2}$, where $P_{T,k}$ is the transmit power of the kth UE.

Under the low resolution ADC architecture, the signals $\mathbf{Z}_p$ and $\mathbf{Z}_d$ are quantized using a $b-$bit ADC to obtain the baseband observations $\mathbf{Y}_p$ and $\mathbf{Y}_d$ as follows:

$$\begin{aligned} \mathbf{Y}_p &= \mathcal{Q}(\mathbf{Z}_p) = \mathcal{Q}(\mathbf{H}\mathbf{X}_p + \mathbf{W}_p) \\ \mathbf{Y}_d &= \mathcal{Q}(\mathbf{Z}_d) = \mathcal{Q}(\mathbf{H}\mathbf{X}_d + \mathbf{W}_d). \end{aligned} \tag{2}$$

The function $\mathcal{Q}(\cdot)$ denotes a quantizer operated element-wise on both the real and imaginary parts of the complex input. A b-bit quantizer on a real valued input z is defined as $\mathcal{Q}(z) = L_i,\ z \in [z_i, z_{i+1}),\ i = 0, 1, \ldots, B-1$, where $B = 2^b$ is the number of quantization levels, $z_0 < z_1 < \cdots < z_B$ are the quantization thresholds, and $L_0, L_1, \ldots, L_{B-1}$ are the quantizer outputs. In this paper, for simplicity and concreteness, we consider a uniform quantizer, where $z_l = (-B/2 + l)\Delta, l = 0, \ldots, B$, Δ is the quantization step size, and $L_l = (z_l + z_{l+1})/2, l = 0, \ldots, B-1$. We set the dynamic range of the real and imaginary parts of the quantizer using the expected received signal power, P_R, as $z_0 = -2.5\sqrt{P_R/2}$, $z_B = 2.5\sqrt{P_R/2}$.[1] Our choice of z_0 and z_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. However, the design of quantizers for low resolution ADC based systems is an interesting problem for future research.

Our goal is to recover the posterior probabilities of the transmitted data symbols from the quantized received signal $\mathbf{Y}_p$ and $\mathbf{Y}_d$. The posterior beliefs are fed to the channel decoder to obtain the decoded data bits. In practice, the equalizer (or detector) and channel decoder can be designed either jointly or as separate blocks. We adopt both approaches to obtain a decoder that is flexible, and comment on their relative merits.

III. Variational Bayesian Joint Channel Estimation and Soft Symbol Decoding

In this section, we derive the joint channel estimation and soft symbol decoding algorithm for an uplink massive MIMO system with low resolution ADCs. We treat the data symbols and channel as the latent variables and the quantized received samples as the observations. We represent the quantized received signal model in (2) as a Bayesian network graphical

[1] In practice, we quantize any value below z_0 to L_0, and any value above z_B to L_{B-1}. Also, in practical wireless systems, an automatic gain control (AGC) and a variable gain amplifier (VGA) are used to ensure that the power in the analog baseband signal is approximately equal to a predefined value, P_R, before quantization.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2021.3060423, IEEE Transactions on Communications

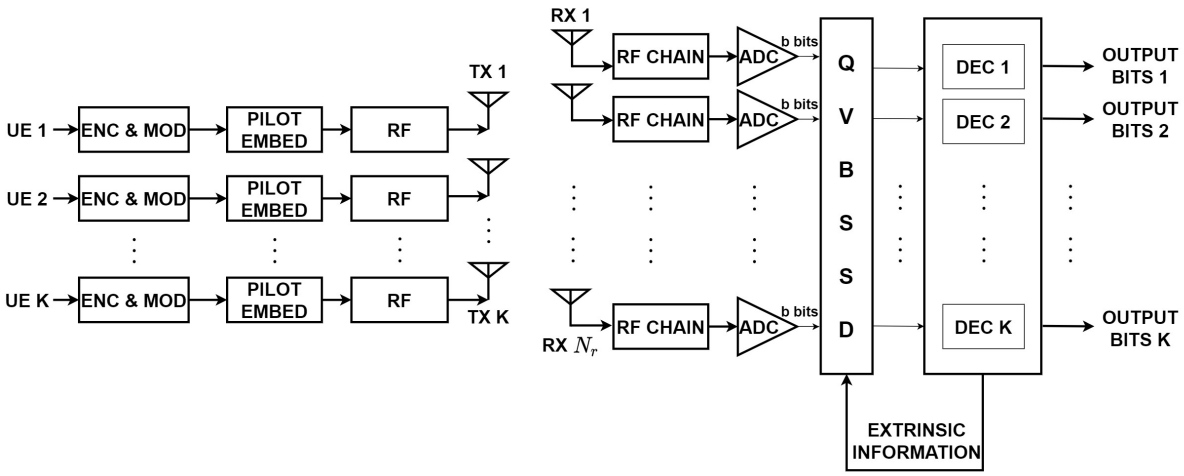


Figure 1. Massive MIMO system model with low resolution ADCs.

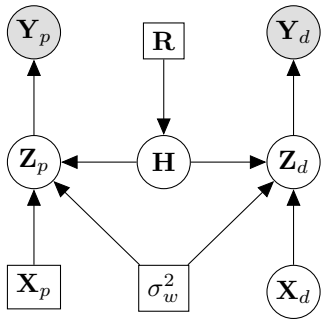


Figure 2. Bayesian network graphical model for the quantized massive MIMO wireless communication system.

model shown in Fig. 2. The pilot and data observations $\mathbf{Y}_p$ and $\mathbf{Y}_d$ are represented by shaded circles, while the latent variables $\mathbf{Z}_p$, $\mathbf{Z}_d$, $\mathbf{X}_d$, $\mathbf{H}$ are represented by circles, and deterministic variables $\mathbf{X}_p$, $\mathbf{R}$, and σ_w^2 are represented by squares. The goal is to infer the posterior distribution of the channel $\mathbf{H}$ and the data $\mathbf{X}_d$ given the observations $\mathbf{Y}_p, \mathbf{Y}_d$ and the pilots $\mathbf{X}_p$.

The posterior distribution of the channel $\mathbf{H}$ and the data $\mathbf{X}_d$ given the quantized observations $\mathbf{Y}_p, \mathbf{Y}_d$ and the pilots $\mathbf{X}_p$ is given by[2]

$$P\left(\mathbf{H}, \mathbf{X}_d | \mathbf{Y}_d, \mathbf{Y}_p; \mathbf{X}_p\right) = \frac{P(\mathbf{Y}_p|\mathbf{X}_p, \mathbf{H}) P(\mathbf{Y}_d|\mathbf{X}_d, \mathbf{H}) P(\mathbf{H}) P(\mathbf{X}_d)}{P(\mathbf{Y}_p) P(\mathbf{Y}_d)}$$

where $P(\mathbf{Y}_p) = \int P(\mathbf{Y}_p|\mathbf{X}_p, \mathbf{H}) P(\mathbf{H}) \, \mathrm{d}\mathbf{H}$ and $P(\mathbf{Y}_d) = \int P(\mathbf{Y}_d|\mathbf{X}_d, \mathbf{H}) P(\mathbf{H}) P(\mathbf{X}_d) \, \mathrm{d}\mathbf{H} \, \mathrm{d}\mathbf{X}_d$ are the marginal likelihoods of $\mathbf{Y}_p$ and $\mathbf{Y}_d$, respectively.

Exact computation of the posterior distributions using the above is computationally intractable, as it requires solving high dimensional integrals over $\mathbf{H}$ and $\mathbf{X}_d$ to obtain the partition functions $P(\mathbf{Y}_p)$ and $P(\mathbf{Y}_d)$. To circumvent this problem, we approach the joint channel estimation and data detection using VBI, in which the posterior distributions of the latent variables are obtained by solving an optimization problem. We approximate the exact posterior using a structured factorized distribution that can be computed in polynomial time in each iteration. We explain this in detail in the forthcoming subsections.

The rest of this section is organized as follows. We first present our algorithm in the simplest case of perfect CSIR with unquantized received data signal (Sec. III-A), followed by the case with statistical CSIR and unquantized pilot and data signals (Sec. III-B), then the case of statistical CSIR with quantized received signals (Sec. III-C), and finally no CSIR with quantized pilot and data signals (Sec. III-D). In Sec. III-E, we combine the VB receiver with a belief propagation (BP) based channel decoder to further improve the coded BER. As explained earlier, this gives flexibility in configuring the VB and BP either jointly or separately. Our approach of presenting our VB based algorithms by adding each imperfection one after the other not only eases the exposition, it also allows us to benchmark results, analytically interpret the algorithms, and makes it convenient to apply the algorithms in a variety of settings. In Sec. III-F, we elucidate the computational complexity of the variational Bayesian algorithm in the quantized case, and discuss ways to further reduce the complexity.

A. Perfect CSIR and Unquantized Observations

We assume that the receiver has perfect knowledge of the channel and has access to unquantized observations $\mathbf{Z}_d = \mathbf{H}\mathbf{X}_d + \mathbf{W}_d$. Here, the transmit data symbols $\mathbf{X}_d$ are the latent variables whose posterior distribution needs to be inferred. The logarithm of the joint distribution of the observations and the latent variables can be written as

$$\ln p(\mathbf{Z}_d, \mathbf{X}_d | \mathbf{H}, \sigma_w^2) = \ln p(\mathbf{Z}_d | \mathbf{X}_d, \mathbf{H}, \sigma_w^2) + \ln p(\mathbf{X}_d). \quad (3)$$

Since the additive noise is complex Gaussian distributed, $p\left(\mathbf{Z}_d | \mathbf{X}_d, \mathbf{H}; \sigma_w^2\right) = \frac{1}{(\pi\sigma_w^2)^{\tau_d N_r}} \times \exp\left(-\frac{1}{\sigma_w^2} \sum_{t=1}^{\tau_d} \|\mathbf{z}_{d,t} - \mathbf{H} x_{d,t}\|_2^2\right)$, and since $\mathbf{X}_d$ is uniformly distributed over the M-QAM constellation, $p(\mathbf{X}_d) = \frac{1}{M^{(\tau_d K)}}$. Our goal is to find the posterior distribution $p(\mathbf{X}_d | \mathbf{Z}_d, \mathbf{H}, \sigma_w^2)$. We write the logarithm of the

[2] $\mathbf{X}_d$ comes from a discrete constellation, but we use integrals here for convenience. In practice, the integrals are replaced by summations.

model evidence, $\ln p(\mathbf{Z}_d)$, as follows:

$$\ln p(\mathbf{Z}_d) = \mathcal{L}(q) + \mathrm{KL}(q \,\|\, p) \geq \mathcal{L}(q), \tag{4}$$

where $\mathcal{L}(q) \triangleq \int q(\mathbf{X}_d) \ln \left\{ \frac{p(\mathbf{X}_d, \mathbf{Z}_d)}{q(\mathbf{X}_d)} \right\} d\mathbf{X}_d$ and $\mathrm{KL}(q\|p) \triangleq -\int q(\mathbf{X}_d) \ln \left\{ \frac{p(\mathbf{X}_d|\mathbf{Z}_d)}{q(\mathbf{X}_d)} \right\} d\mathbf{X}_d \geq 0$ are the evidence lower bound (ELBO) and the Kullback-Leibler (KL) divergence terms, respectively. Here, $q(\mathbf{X}_d)$ is an approximate posterior distribution, which is arbitrary and can be optimized. We formally state the ELBO optimization problem as

$$q_{\text{opt}} = \underset{q \in \mathcal{P}}{\arg\max}\ \mathcal{L}(q) = \underset{q \in \mathcal{P}}{\arg\min}\ \mathrm{KL}(q \,\|\, p), \tag{5}$$

where $\mathcal{P}$ is the space of probability distributions.

In the above, maximizing the ELBO $\mathcal{L}(q)$ would render an approximate distribution q that is close to the original model evidence, because the KL divergence is non-negative. The maximum of $\mathcal{L}(q)$ occurs when $q(\mathbf{X}_d) = p(\mathbf{X}_d|\mathbf{Z}_d)$, but $p(\mathbf{X}_d|\mathbf{Z}_d)$ is computationally intractable. Therefore, we impose a factorized posterior structure on q, i.e., $q(\mathbf{X}_d) = \prod_{k=1}^{K} \prod_{t=1}^{\tau_d} q_{kt}(x_{d,kt})$. Substituting this into the ELBO, and considering one of the factors, say $q_{kt} \triangleq q(x_{d,kt})$, we get

$$\mathcal{L}(q) = -\mathrm{KL}\left(q_{kt} \| \tilde{p}(\mathbf{Z}_d, x_{d,kt})\right) + \text{const.} \tag{6}$$

where the `const.` terms do not depend on q_{kt}, and $\tilde{p}(\mathbf{Z}_d, x_{d,kt})$ is defined using $\ln \tilde{p}(\mathbf{Z}_d, x_{d,kt}) \triangleq \mathbb{E}_{(i,l)\neq(k,t)} [\ln p(\mathbf{Z}_d, \mathbf{X}_d)] + \text{const.}$, where the notation $\mathbb{E}_{(i,l)\neq(k,t)}[\cdot]$ denotes the expectation with respect to the distributions $q_{11}(x_{d,11})$, …, $q_{K\tau_d}(x_{d,K\tau_d})$ except $q_{kt}(x_{d,kt})$. Now, $\mathcal{L}(q)$ is maximized with respect to q_{kt} when the KL divergence term in (6) is minimized, which in turn occurs when $q_{kt}(x_{d,kt}) = \tilde{p}(\mathbf{Z}_d, x_{d,kt})$. Therefore, the optimal marginal distribution is

$$q_{kt}(x_{d,kt}) = \text{const.} \times \exp\left(\mathbb{E}_{(i,l)\neq(k,t)} [\ln p(\mathbf{Z}_d, \mathbf{X}_d)]\right), \tag{7}$$

where the constant is chosen such that q_{kt} becomes a probability distribution. From (7), the optimal q_{kt} depends on the distributions $\{q_{i\ell}\}_{(i,\ell)\neq(k,t)}$, which also need to be evaluated. The VBI algorithm proceeds by initializing latent distributions $\{q_{i\ell}\}_{(i,\ell)\neq(k,t)}$ and finding q_{kt} in an iterative manner across all k and t. This recipe falls in the category of minorization-maximization (MM), which solves a maximization problem by iteratively obtaining a lower bound to the objective function and maximizing it [37]. It is known that MM based optimization is guaranteed to converge to a stationary point of the original optimization problem from any initialization.

Thus, based on the above discussion, we impose a fully factorized structure on the approximate posterior, namely, $p\left(\mathbf{X}_d|\mathbf{Z}_d, \mathbf{H}, \sigma_w^2\right) \approx \prod_{k=1}^{K} \prod_{t}^{\tau_d} q_{x_{d,kt}}(x_{d,kt})$. Now, we compute the approximate posterior $q(x_{d,kt})$, $k = 1, \ldots, K$, $t = 1, \ldots, \tau_d$. To this end, we compute the following expectation using the joint distribution in (3):

$$\ln q_{x_{d,kt}}(x_{d,kt}) = \left\langle \ln p\left(\mathbf{Z}_d|\mathbf{X}_d, \mathbf{H}; \sigma_w^2\right) + \ln p\left(\mathbf{X}_d\right) \right\rangle, \tag{8}$$

where $\langle\cdot\rangle$ denotes the expectation with respect to all the latent variables except $x_{d,kt}$, and is computed using the approximate posterior distribution $q(\mathbf{X})$. Taking the exponential on both sides of (8) and normalizing to obtain a probability distribution, after some algebra, we get

$$q_{x_{d,kt}}(s_m) = \frac{\exp\Big(f_{k,t}(s_m)\Big)}{\sum_{s' \in \mathbb{M}} \exp\Big(f_{k,t}(s')\Big)}, \quad m = 1, \ldots, M, \tag{9}$$

with $f_{k,t}(s)$ defined in (10). From (9), we see that $q_{x_{d,kt}}$ is Boltzmann distributed (it is also known as the *softmax* function and is widely used in machine learning applications as an output activation function.) The mean and mean squared values of $x_{d,kt}$ under the distribution $q_{x_{d,kt}}$ can be computed as $\langle x_{d,kt}\rangle = \sum_{m=1}^{M} s_m\, q_{x_{d,kt}}(s_m)$, $\langle |x_{d,kt}|^2\rangle = \sum_{m=1}^{M} |s_m|^2\, q_{x_{d,kt}}(s_m)$.

Algorithm 1 VB Soft Symbol Decoding with Perfect CSIR

Input: $\mathbf{Z}_d, \{\mathbf{h}_1, \ldots, \mathbf{h}_K\}, M, \mathbb{M} = \{s_1, \ldots, s_M\}, \tau_d, \sigma_w$
Output: $q_{x_{d,kt}}, \langle x_{d,kt}\rangle\ \forall k \in [K], t \in [\tau_d]$

```
Initialize q_{x_{d,kt}}(x_{d,kt}), ⟨x_{d,kt}⟩ = 0, ∀k ∈ [K], t ∈ [τ_d]
repeat
    for k = 1 to K, t = 1 to τ_d do
        for m = 1 to M do
            Compute q_{x_{d,kt}}(s_m) using (9) and (10).
        end for
        ⟨x_{d,kt}⟩ = Σ_{s∈M} s q_{x_{d,kt}}(s).
        ⟨|x_{d,kt}|^2⟩ = Σ_{s∈M} |s|^2 q_{x_{d,kt}}(s).
    end for
until a stopping condition is met.
```

The distribution $q_{x_{d,kt}}$ depends only on the means of the other latent variables $\{x_{d,k't}\}_{k'\neq k}$ taken with respect to their approximate distributions. Therefore, we initialize the means of the distributions $q_{x_{d,kt}}$, $k = 1, \ldots, K$ and $t = 1, \ldots, \tau_d$ arbitrarily and apply (9) iteratively across k and t till the log likelihood in (8) converges for all k and t. As mentioned earlier, this procedure converges to a stationary point of the original optimization problem of choosing q to maximize the model evidence in (4). The pseudocode for the VB procedure for the soft symbol decoding is shown in Algorithm 1.

Remark 1: In an uncoded system, the receiver decodes the transmitted symbol by solving an optimization problem mentioned in (11). Given the aprori distribution of the data symbols $p(x_{d,kt} = s)$, the above equation can be interpreted as successive interference cancellation followed by maximum aposteriori probability (MAP) decoding of a single user's data symbol. This reduces the exponential complexity of a MAP receiver to linear complexity in the number of UEs. However, in low resolution quantized systems (discussed later), the analog input to the receiver is not observed, and the above approach may lead to suboptimal performance. In this case, VBI provides a systematic approach to updating the soft symbol estimates by incorporating the posterior distribution induced by the quantization.

In the next subsection, we remove the perfect CSIR assumption and infer the posterior distributions of the channel as well as the data symbols given the unquantized observations.

$$f_{k,t}(s) \triangleq -\frac{1}{\sigma_w^2}\Big(\|\mathbf{h}_k\|^2 |s|^2 - 2\Re\Big[\mathbf{h}_k^H\Big(\mathbf{z}_{d,t} - \sum_{\substack{k'=1\\k'\neq k}}^{K} \mathbf{h}_{k'}\langle x_{d,k't}\rangle\Big)s^*\Big]\Big) + \ln p(x_{d,kt}=s). \tag{10}$$

$$s_{k,t}^{\text{opt}} = \arg\min_{s\in\mathbb{M}} \frac{1}{\sigma_w^2}\Big(\|\mathbf{h}_k\|^2 |s|^2 - 2\Re\Big[\mathbf{h}_k^H\Big(\mathbf{z}_{d,t} - \sum_{\substack{k'=1\\k'\neq k}}^{K} \mathbf{h}_{k'}\langle x_{d,k't}\rangle\Big)s^*\Big]\Big) - \ln p(x_{d,kt}=s). \tag{11}$$

B. Statistical CSIR and Unquantized Observations

In this subsection, we solve the joint channel estimation and soft symbol decoding problem in an unquantized setup. Therefore, the unquantized received signals $\mathbf{Z}_p = [\mathbf{z}_{p,1},\ldots,\mathbf{z}_{p,\tau_p}] = \mathbf{H}\mathbf{X}_p + \mathbf{W}_p$ and $\mathbf{Z}_d = [\mathbf{z}_{d,1},\ldots,\mathbf{z}_{d,\tau_d}] = \mathbf{H}\mathbf{X}_d + \mathbf{W}_d$ are the observations. The channel $\mathbf{H}$ and the data symbols $\mathbf{X}_d$ are the latent variables whose posterior distributions need to be inferred. For this problem, a message passing based approach is adopted in [33], [34] in a MIMO OFDM setup where the channel has a diagonal structure, whereas we use an iterative VBI framework which can be used for arbitrary channels. Similar to the perfect CSIR case, we impose a factorized structure on the posterior as follows: $p\left(\mathbf{X}_d,\mathbf{H}|\mathbf{Z}_p,\mathbf{Z}_d,\mathbf{X}_p;\mathbf{R},\sigma_w^2\right) \approx q\left(\mathbf{X}_d\right) q\left(\mathbf{H}\right)$,[3] where $q\left(\mathbf{H}\right) = \prod_{k=1}^{K} q\left(\mathbf{h}_k\right), q\left(\mathbf{X}_d\right) = \prod_{k=1}^{K}\prod_{t=1}^{\tau_d} q_{x_{d,kt}}\left(x_{d,kt}\right)$.

Using the structure of the model in Fig. 2, we factorize the joint distribution of the observations and latent variables as

$$\begin{aligned} &P\left(\mathbf{Z}_p,\mathbf{Z}_d,\mathbf{H},\mathbf{X}_d|\mathbf{X}_p;\mathbf{R},\sigma_w^2\right) \\ &= P\left(\mathbf{Z}_p|\mathbf{X}_p,\mathbf{H};\sigma_w^2\right) P\left(\mathbf{Z}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right) P\left(\mathbf{H}|\mathbf{R}\right) P\left(\mathbf{X}_d\right). \end{aligned} \tag{12}$$

The likelihood functions of the pilot and data observations $\mathbf{Z}_p$ and $\mathbf{Z}_d$ given the channel $\mathbf{H}$, the pilots $\mathbf{X}_p$, and the data $\mathbf{X}_d$ are

$$\begin{aligned} &P\left(\mathbf{Z}_p|\mathbf{X}_p,\mathbf{H};\sigma_w^2\right) \\ &= \frac{1}{(\pi\sigma_w^2)^{\tau_p N_r}} \exp\left(-\frac{1}{\sigma_w^2}\sum_{t=1}^{\tau_p}\|\mathbf{z}_{p,t} - \mathbf{H}\mathbf{x}_{p,t}\|^2\right), \\ &P\left(\mathbf{Z}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right) \\ &= \frac{1}{(\pi\sigma_w^2)^{\tau_d N_r}} \exp\left(-\frac{1}{\sigma_w^2}\sum_{t=1}^{\tau_d}\|\mathbf{z}_{d,t} - \mathbf{H}\mathbf{x}_{d,t}\|^2\right), \end{aligned}$$

respectively. Also, the prior distribution on the channel $\mathbf{H}$ is

$$p\left(\mathbf{H}|\mathbf{R}\right) = \prod_{k=1}^{K} \frac{1}{\pi^{N_r}\det(\mathbf{R}_k)} \exp\left(-\sum_{k=1}^{K}\mathbf{h}_k^H\mathbf{R}_k^{-1}\mathbf{h}_k\right).$$

As before, the data symbols $\mathbf{X}_d$ are drawn i.i.d. uniformly from the M-QAM constellation.

Our goal is to compute the approximate posterior distributions $q(\mathbf{X}_d)$ and $q(\mathbf{H})$. As mentioned in the previous section, we optimize the distributions by minimizing the KL divergence between the original and approximate posterior. We present the steps associated in obtaining the approximate posterior distributions $q(\mathbf{X}_d)$ and $q(\mathbf{H})$ below.

[3]We drop the subscripts on q for notational simplicity.

1) Computation of $q(\mathbf{h}_k)$: We compute the approximate posterior distribution $q(\mathbf{h}_k)$ of the channel between the k^{th} user and the BS using the joint distribution in (12), as follows:

$$\begin{aligned} \ln q\left(\mathbf{h}_k\right) \propto \big\langle \ln p\left(\mathbf{Z}_p|\mathbf{X}_p,\mathbf{H};\sigma_w^2\right) + \ln p\left(\mathbf{Z}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right) \\ + \ln p\left(\mathbf{H}|\mathbf{R}\right)\big\rangle, \end{aligned} \tag{13}$$

where $\langle\cdot\rangle$ denotes the expectation of the joint distribution over all the latent variables excluding $\mathbf{h}_k$. In (13), we only include the terms involving $\mathbf{h}_k$. Upon simplification, we get (14). Taking the exponential on both sides of (14), we see that the structure is that of a complex normal distribution with covariance and mean given by (15) and (16), respectively.

Remark 2: We provide an interesting interpretation of the channel estimate in (16). Consider a single user MIMO channel estimation problem with pilots $\mathbf{x}_p \in \mathbb{C}^{\tau_p\times 1}$ transmitted over a duration of τ_p symbols. The received pilot symbols are given by $\mathbf{Y}_p = \mathbf{h}\mathbf{x}_p^H + \mathbf{W}_p \in \mathbb{C}^{N_r\times\tau_p}$, where $\mathbf{h}\in\mathbb{C}^{N_r\times 1}$ is the channel distributed as $\mathcal{CN}(\mathbf{0},\mathbf{R})$ and $\mathbf{W}_p \in \mathbb{C}^{N_r\times\tau_p}$ is the additive noise whose entries are i.i.d. $\mathcal{CN}(0,\sigma_w^2)$. Post-multiplying $\mathbf{Y}_p$ with $\mathbf{x}_p$, we get $\mathbf{Y}_p\mathbf{x}_p = \mathbf{h}\|\mathbf{x}_p\|_2^2 + \mathbf{W}_p\mathbf{x}_p$. The MMSE estimate of $\mathbf{h}$ can be computed to be

$$\widehat{\mathbf{h}}^{\mathbf{MMSE}} = \left(\|\mathbf{x}_p\|_2^2\mathbf{I}_{N_r} + \sigma_w^2\mathbf{R}^{-1}\right)^{-1}\mathbf{Y}_p\mathbf{x}_p. \tag{17}$$

Now, comparing the solution obtained in (16) and (15) with (17), we see that the channel estimate $\langle\mathbf{h}_k\rangle$ of the k^{th} user in (16) is an MMSE estimate assuming the interference caused by the remaining $K-1$ users is cancelled successfully. Therefore, by using a factorized structure on the posterior distribution, and assuming successful interference cancellation, the channel estimate obtained in (16) matches with the *Bayes' optimal MMSE estimator*.

2) Computation of $q_{x_{d,kt}}(x_{d,kt})$: In this subsection, we present the steps involved in the computation of the approximate posterior distribution $q_{x_{d,kt}}(x_{d,kt})$ of the k^{th} user's data symbol transmitted during the t^{th} symbol interval. Similar to the computation of $q(\mathbf{h}_k)$, we use the joint distribution (12) to compute the expectation in (18) with respect to all the latent variables except $x_{d,kt}$. We substitute the values of $x_{d,kt}$ from the $M-$QAM constellation in the above equation to get $q_{x_{d,kt}}(x_{d,kt}=s)$ in the same form as (9), with $f_{k,t}(s)$ defined in (19), where s belongs to a symbol from the $M-$QAM constellation. Similar to the perfect CSIR case, $q_{x_{d,kt}}(x_{d,kt})$ is Boltzmann distributed. We compute the mean and mean squared values of $x_{d,kt}$ in the same manner as in the previous subsection. We present the pseudocode for the above procedure in Algorithm 2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2021.3060423, IEEE Transactions on Communications

$$\ln q\left(\mathbf{h}_k\right) \propto -\Bigg[\mathbf{h}_k^H \left\{\frac{\sum_{t=1}^{\tau_p}|x_{p,kt}|^2+\sum_{t=1}^{\tau_d}\langle|x_{d,kt}|^2\rangle}{\sigma_w^2}\mathbf{I}_{N_r}+\mathbf{R}_k^{-1}\right\}\mathbf{h}_k - 2\Re\Bigg\{\mathbf{h}_k^H\left(\frac{1}{\sigma_w^2}\left(\sum_{t=1}^{\tau_p}\left(\mathbf{z}_{p,t}-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle x_{p,k't}\right)x_{p,kt}^*+\sum_{t=1}^{\tau_d}\left(\mathbf{z}_{d,t}-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle\langle x_{d,k't}\rangle\right)\langle x_{d,kt}^*\rangle\right)\right)\Bigg\}\Bigg]. \quad (14)$$

$$\mathbf{\Sigma}_{\mathbf{h}_k}=\left(\frac{\sum_{t=1}^{\tau_p}|x_{p,kt}|^2+\sum_{t=1}^{\tau_d}\langle|x_{d,kt}|^2\rangle}{\sigma_w^2}\mathbf{I}_{N_r}+\mathbf{R}_k^{-1}\right)^{-1}, \quad (15)$$

$$\langle\mathbf{h}_k\rangle=\frac{1}{\sigma_w^2}\mathbf{\Sigma}_{\mathbf{h}_k}\left(\sum_{t=1}^{\tau_p}\left(\mathbf{z}_{p,t}-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle x_{p,k't}\right)x_{p,kt}^*+\sum_{t=1}^{\tau_d}\left(\mathbf{z}_{d,t}-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle\langle x_{d,k't}\rangle\right)\langle x_{d,kt}^*\rangle\right). \quad (16)$$

$$\ln q_{x_{d,kt}}\left(x_{d,kt}\right)\propto-\frac{1}{\sigma_w^2}\left(\left\langle\left\|\mathbf{h}_k\right\|^2\right\rangle|x_{d,kt}|^2-2\Re\left[\left\langle\mathbf{h}_k\right\rangle^H\left(\mathbf{z}_{d,t}-\sum_{\substack{k'=1\\k'\neq k}}^{K}\left\langle\mathbf{h}_{k'}\right\rangle\left\langle x_{d,k't}\right\rangle\right)x_{d,kt}^*\right]\right)+\ln p(x_{d,kt}). \quad (18)$$

$$f_{k,t}(s)=-\frac{1}{\sigma_w^2}\left(\left\langle\left\|\mathbf{h}_k\right\|^2\right\rangle|s|^2-2\Re\left[\left\langle\mathbf{h}_k\right\rangle^H\left(\mathbf{z}_{d,t}-\sum_{\substack{k'=1\\k'\neq k}}^{K}\left\langle\mathbf{h}_{k'}\right\rangle\left\langle x_{d,k't}\right\rangle\right)s^*\right]\right)+\ln p(x_{d,kt}=s). \quad (19)$$

Remark 3: Both the channel estimate in (16) and the soft symbol estimator in (19) depend on all the data symbols. Since the channel remains constant over the entire frame, iterative channel estimation and data detection entails using all the data symbols. Symbol-by-symbol detection is not optimal here, unlike the case where either perfect CSIR is assumed to be available or pilot-only based channel estimates are used for data detection.

Note that the approximate marginal posterior distributions of the latent variables are dependent on each other. The algorithm runs by randomly initializing the statistics of the factorized distributions of the latent variables, and cycling through to iteratively update the distributions.

Algorithm 2 VB Joint Channel Estimation and Soft Symbol Decoding with Statistical CSIR

Input: $\mathbf{Z}_p, \mathbf{Z}_d, \{\mathbf{R}_1,\ldots,\mathbf{R}_K\}, \mathbf{X}_p, M, \mathbb{M}=\{s_1,\ldots,s_M\}$, τ_p, τ_d, σ_w

Output: $\{\langle\mathbf{h}_1\rangle,\ldots,\langle\mathbf{h}_K\rangle\}, q_{x_{d,kt}}, \langle x_{d,kt}\rangle\ \forall k\in[K], t\in[\tau_d]$

Initialize $q_{x_{d,kt}}(x_{d,kt}), \langle x_{d,kt}\rangle=0, \forall k\in[K], t\in[\tau_d]$
repeat
  for $k=1$ to K, $t=1$ to τ_d **do**
    for $m=1$ to M **do**
      Compute $q_{x_{d,kt}}(s_m)$ using (9) and (19).
    end for
    $\langle x_{d,kt}\rangle=\sum_{s\in\mathbb{M}} s\, q_{x_{d,kt}}(s)$.
    $\langle|x_{d,kt}|^2\rangle=\sum_{s\in\mathbb{M}}|s|^2\, q_{x_{d,kt}}(s)$.
  end for
  for $k=1$ to K **do**
    Compute $\mathbf{\Sigma}_{\mathbf{h}_k}$ and $\langle\mathbf{h}_k\rangle$ using (15) and (16), respectively.
  end for
until a stopping condition is met.

C. Statistical CSIR and Quantized Observations

In this subsection, we infer the marginal posterior distributions of the data symbols and the channel given the *quantized* observations $\mathbf{Y}_p=\mathcal{Q}(\mathbf{Z}_p)=\mathcal{Q}\left(\mathbf{H}\mathbf{X}_p+\mathbf{W}_p\right)$, $\mathbf{Y}_d=\mathcal{Q}(\mathbf{Z}_d)=\mathcal{Q}\left(\mathbf{H}\mathbf{X}_d+\mathbf{W}_d\right)$ and the pilot symbols $\mathbf{X}_p$. The joint distribution of the observations and latent variables is factorized as

$$p\left(\mathbf{Y}_p,\mathbf{Y}_d,\mathbf{H},\mathbf{X}_d|\mathbf{X}_p;\sigma_w^2,\mathbf{R}\right) = p\left(\mathbf{Y}_p|\mathbf{X}_p,\mathbf{H};\sigma_w^2\right)p\left(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right)p\left(\mathbf{H}|\mathbf{R}\right)p\left(\mathbf{X}_d\right).$$

The conditional distribution of the quantized observations $\mathbf{Y}_d$ given $\mathbf{H},\mathbf{X}_d$ is given by

$$p\left(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right)=\int_{\mathbf{Z}_d}p\left(\mathbf{Y}_d\,|\,\mathbf{Z}_d\right)p\left(\mathbf{Z}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right)d\,\mathbf{Z}_d = \int_{\mathbf{Z}_d^{(\mathrm{lo})}}^{\mathbf{Z}_d^{(\mathrm{hi})}}\frac{1}{(\pi\sigma_w^2)^{N_r\tau_d}}\exp\left(-\frac{1}{\sigma_w^2}\left\|\mathbf{Z}_d-\mathbf{H}\mathbf{X}_d\right\|_F^2\right)d\,\mathbf{Z}_d, \quad (20)$$

where $\mathbf{Z}_d^{(\mathrm{lo})}$ and $\mathbf{Z}_d^{(\mathrm{hi})}$ are the lower and upper thresholds of the quantizer corresponding to the observation $\mathbf{Y}_d$. The conditional distribution of the quantized pilot observations also has a similar structure as (20). Now, to obtain a closed form expression for the approximate posterior distributions of the latent variables using VBI, it is necessary to compute the expectation of the logarithm of the difference of two cumulative distribution functions of a complex normal random vector, which is not straightforward. We circumvent this problem by introducing the unquantized pilot and data observations as latent variables. We will see that this leads to a convenient analytical expression for the posterior distribution.

The conditional distributions of the unquantized observations and the channel are as derived in Sec. III-B. The conditional distributions of the quantized observations, given the unquantized received signals, are given by

$$\begin{aligned} p\left(\mathbf{Y}_d|\mathbf{Z}_d\right) &= \mathbb{1}\big(\mathbf{Z}_d \in [\mathbf{Z}_d^{(\text{lo})}, \mathbf{Z}_d^{(\text{hi})}]\big), \\ p\left(\mathbf{Y}_p|\mathbf{Z}_p\right) &= \mathbb{1}\big(\mathbf{Z}_p \in [\mathbf{Z}_p^{(\text{lo})}, \mathbf{Z}_p^{(\text{hi})}]\big), \end{aligned} \tag{21}$$

where $\mathbb{1}(\cdot)$ is the indicator function and the arguments of the indicator function $\mathbf{Z}_d^{(\text{lo})}, \mathbf{Z}_d^{(\text{hi})}$ and $\mathbf{Z}_p^{(\text{lo})}, \mathbf{Z}_p^{(\text{hi})}$ are the lower and upper limits of the quantizer corresponding to the observations $\mathbf{Y}_d$ and $\mathbf{Y}_p$, respectively. To elaborate, let us take an example of quantization of a scalar z ($z^{(\text{lo})} \leq z \leq z^{(\text{hi})}$) to a value y that takes values from a discrete set $\mathbb{Q}$ with cardinality 2^b, where b is the number of bits of the quantizer. Note that this discrete set depends on the choice of the quantizer. For example, a uniform quantizer contains uniformly spaced values. Now, the value to be quantized fully specifies the output of the quantizer, i.e., the probability mass function of the quantizer output is a Kronecker delta function.

We impose a factorized structure on the posterior distribution of the latent variables:

$$\begin{aligned} &p\left(\mathbf{Z}_p, \mathbf{Z}_d, \mathbf{X}_d, \mathbf{H}|\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{X}_p, \mathbf{R}, \sigma_w^2\right) \\ &\qquad \approx q\left(\mathbf{Z}_p\right) q\left(\mathbf{Z}_d\right) q\left(\mathbf{X}_d\right) q\left(\mathbf{H}\right), \end{aligned} \tag{22}$$

where $q\left(\mathbf{H}\right) = \prod_{k=1}^{K} q\left(\mathbf{h}_k\right)$, $q\left(\mathbf{X}_d\right) = \prod_{k=1}^{K}\prod_{t=1}^{\tau_d} q_{x_{d,kt}}\left(x_{d,kt}\right)$, $q\left(\mathbf{Z}_d\right) = \prod_{t=1}^{\tau_d} q\left(\mathbf{z}_{d,t}\right)$, and $q\left(\mathbf{Z}_p\right) = \prod_{t=1}^{\tau_p} q\left(\mathbf{z}_{p,t}\right)$. We compute the approximate marginal posterior distributions of the latent variables in a similar manner as in Sections III-A and III-B, and outline the differences below.

1) Computation of $q(\mathbf{h}_k)$: We compute the approximate posterior distribution $q(\mathbf{h}_k)$ of the channel between the k^{th} user and the BS by taking expectation of the natural logarithm of the joint probability distribution with respect to the approximate distributions of all the latent variables excluding $\mathbf{h}_k$. This computation is similar to that in Sec. III-B, except that the unquantized observations $\mathbf{Z}_p$ and $\mathbf{Z}_d$ are latent variables. The final expression is as shown in (23).

Taking the exponential on both sides of (23), we see that the structure is that of a complex normal distribution with covariance and mean given by (24) and (25), respectively.

2) Computation of $q_{x_{d,kt}}(x_{d,kt})$: Following a similar approach as in Sections III-A and III-B, $q_{x_{d,kt}}(x_{d,kt})$, $k = 1, \ldots, K, t = 1, \ldots, \tau_d$ is given by the Boltzmann distribution in (9), where s is an $M-$QAM symbol, with $f_{k,t}(s)$ defined in (26). We compute the mean and mean square values of $q_{x_{d,kt}}$ in a similar manner as in Sec. III-A.

3) Computation of $q(\mathbf{z}_{d,t})$ and $q(\mathbf{z}_{p,t})$: We obtain the distribution $q(\mathbf{z}_{d,t})$ as follows:

$$\begin{aligned} &\ln q\left(\mathbf{z}_{d,t}\right) \\ &= \big\langle \ln p\left(\mathbf{Y}_p|\mathbf{Z}_p\right) + \ln p\left(\mathbf{Y}_d|\mathbf{Z}_d\right) + \ln p\left(\mathbf{Z}_p|\mathbf{X}_p, \mathbf{H}; \sigma_w^2\right) \\ &\qquad + \ln p\left(\mathbf{Z}_d|\mathbf{X}_d, \mathbf{H}; \sigma_w^2\right) + \ln p\left(\mathbf{H}|\mathbf{R}\right) + \ln p\left(\mathbf{X}_d\right)\big\rangle, \\ &\propto \left\langle \ln \mathbb{1}\big(\mathbf{z}_{d,t} \in [\mathbf{z}_{d,t}^{(\text{lo})}, \mathbf{z}_{d,t}^{(\text{hi})}]\big) - \frac{1}{\sigma_w^2}\left\|\mathbf{z}_{d,t} - \mathbf{H}\mathbf{x}_{d,t}\right\|^2 \right\rangle. \end{aligned}$$

Thus, we see that $q\left(\mathbf{z}_{d,t}\right)$ is a truncated complex normal distribution with mean

$$\left\langle \mathbf{z}_{d,t} \right\rangle = \boldsymbol{\mu}_{\mathbf{z}_{d,t}} + \frac{\phi\left(\frac{\mathbf{z}_{d,t}^{(\text{lo})} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right) - \phi\left(\frac{\mathbf{z}_{d,t}^{(\text{hi})} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right)}{\Phi\left(\frac{\mathbf{z}_{d,t}^{(\text{hi})} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right) - \Phi\left(\frac{\mathbf{z}_{d,t}^{(\text{lo})} - \boldsymbol{\mu}_{\mathbf{z}_{d,t}}}{\sigma_w/\sqrt{2}}\right)} \frac{\sigma_w}{\sqrt{2}}, \tag{27}$$

where $\boldsymbol{\mu}_{\mathbf{z}_{d,t}} = \langle \mathbf{H} \rangle \langle \mathbf{x}_{d,t} \rangle$. The division operation is a scalar element-wise division operated individually on real and imaginary components. We compute the approximate posterior distribution $q(\mathbf{z}_{p,t})$ of the unquantized received pilot symbols in the same manner as $q(\mathbf{z}_{d,t})$, and its mean is given by

$$\left\langle \mathbf{z}_{p,t} \right\rangle = \boldsymbol{\mu}_{\mathbf{z}_{p,t}} + \frac{\phi\left(\frac{\mathbf{z}_{p,t}^{(\text{lo})} - \boldsymbol{\mu}_{\mathbf{z}_{p,t}}}{\sigma_w/\sqrt{2}}\right) - \phi\left(\frac{\mathbf{z}_{p,t}^{(\text{hi})} - \boldsymbol{\mu}_{\mathbf{z}_{p,t}}}{\sigma_w/\sqrt{2}}\right)}{\Phi\left(\frac{\mathbf{z}_{p,t}^{(\text{hi})} - \boldsymbol{\mu}_{\mathbf{z}_{p,t}}}{\sigma_w/\sqrt{2}}\right) - \Phi\left(\frac{\mathbf{z}_{p,t}^{(\text{lo})} - \boldsymbol{\mu}_{\mathbf{z}_{p,t}}}{\sigma_w/\sqrt{2}}\right)} \frac{\sigma_w}{\sqrt{2}}, \tag{28}$$

where $\boldsymbol{\mu}_{\mathbf{z}_{p,t}} = \langle \mathbf{H} \rangle \mathbf{x}_{p,t}$. Note that computing $q(\mathbf{z}_{p,t})$ and $q(\mathbf{z}_{d,t})$ does not involve an expectation of the logarithm of integrals, as mentioned earlier.

The pseudocode for the VB procedure for the joint channel estimation and soft symbol decoding is shown in Algorithm 3. The VBI algorithm starts by randomly initializing the latent variables and maximizing the ELBO by fixing all but one hidden variable. Once the probabilities $q_{x_{d,kt}}$ are obtained, we choose the symbol with the highest probability as the decoded symbol for each $k \in \{1, \ldots, K\}$ and $t \in \{1, \ldots, \tau_d\}$ in the case of uncoded communication. For coded communication applications, it is straightforward to compute the LLRs of the data bits from $q_{x_{d,kt}}$ and pass them as inputs to the channel decoder [36]. Also, the mean of the marginal posterior of the channel can be used as a channel estimate for SNR computation, CSI feedback etc.

D. No CSIR and Quantized Observations

In this subsection, we extend the VB approach to the case where there is no information about the channel statistics also. This situation may arise in massive machine type communication applications where a large number of low power UEs sporadically wake up and transmit data to the BS or AP. In such scenarios, the assumption of knowledge of channel statistics at the BS or AP may not be appropriate. We assume that the channels between a UE and the different antennas at the BS are uncorrelated and that the LSFCs are the same across all BS antennas. We impose a non-informative conjugate Gamma prior for the inverse of the LSFCs. Let α_k denote the inverse of the LSFC between the k^{th} UE and the BS and let $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^T$. The prior on α_k is given by

$$p(\alpha_k; a, b) = \frac{b^a}{\Gamma(a)} \alpha_k^{a-1} \exp(-b\alpha_k), \tag{29}$$

where a and b are the parameters of the Gamma distribution. Conditioned on α_k, the channel $\mathbf{h}_k$ of the k^{th} UE is distributed as $\mathcal{CN}(\mathbf{0}, \alpha_k^{-1}\mathbf{I}_{N_r})$. The joint distribution of the observed and

$$\ln q\left(\mathbf{h}_k\right) \propto -\Bigg[\mathbf{h}_k^H\left\{\frac{\sum_{t=1}^{\tau_p}|x_{p,kt}|^2+\sum_{t=1}^{\tau_d}\langle|x_{d,kt}|^2\rangle}{\sigma_w^2}\mathbf{I}_{N_r}+\mathbf{R}_k^{-1}\right\}\mathbf{h}_k - 2\Re\left\{\mathbf{h}_k^H\left(\frac{1}{\sigma_w^2}\Big(\sum_{t=1}^{\tau_p}\Big(\langle\mathbf{z}_{p,t}\rangle-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle x_{p,k't}\Big)x_{p,kt}^*+\sum_{t=1}^{\tau_d}\Big(\langle\mathbf{z}_{d,t}\rangle-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle\langle x_{d,k't}\rangle\Big)\langle x_{d,kt}^*\rangle\Big)\right)\right\}\Bigg]. \tag{23}$$

$$\boldsymbol{\Sigma}_{\mathbf{h}_k}=\left(\frac{\sum_{t=1}^{\tau_p}|x_{p,kt}|^2+\sum_{t=1}^{\tau_d}\langle|x_{d,kt}|^2\rangle}{\sigma_w^2}\mathbf{I}_{N_r}+\mathbf{R}_k^{-1}\right)^{-1}, \tag{24}$$

$$\langle\mathbf{h}_k\rangle=\frac{1}{\sigma_w^2}\boldsymbol{\Sigma}_{\mathbf{h}_k}\left(\sum_{t=1}^{\tau_p}\Big(\langle\mathbf{z}_{p,t}\rangle-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle x_{p,k't}\Big)x_{p,kt}^*+\sum_{t=1}^{\tau_d}\Big(\langle\mathbf{z}_{d,t}\rangle-\sum_{\substack{k'=1\\k'\neq k}}^{K}\langle\mathbf{h}_{k'}\rangle\langle x_{d,k't}\rangle\Big)\langle x_{d,kt}^*\rangle\right). \tag{25}$$

$$f_{k,t}(s)=-\frac{1}{\sigma_w^2}\Big(\big\langle\,\|\mathbf{h}_k\|^2\,\big\rangle\,|s|^2-2\Re\Big[\big\langle\mathbf{h}_k\big\rangle^H\Big(\big\langle\mathbf{z}_{d,t}\big\rangle-\sum_{\substack{k'=1\\k'\neq k}}^{K}\big\langle\mathbf{h}_{k'}\big\rangle\big\langle x_{d,k't}\big\rangle\Big)s^*\Big]\Big)+\ln p(x_{d,kt}=s). \tag{26}$$

Algorithm 3 QVB Joint Channel Estimation and Soft Symbol Decoding with Statistical CSIR

Input: $\mathbf{Y}_p, \mathbf{Y}_d, \{\mathbf{R}_1,\ldots,\mathbf{R}_K\}, \mathbf{X}_p, M, \mathbb{M}=\{s_1,\ldots,s_M\}$, τ_p, τ_d, σ_w

Output: $\{\langle\mathbf{h}_1\rangle,\ldots,\langle\mathbf{h}_K\rangle\}, q_{x_{d,kt}}, \langle x_{d,kt}\rangle\ \forall k\in[K], t\in[\tau_d]$

Initialize $\langle\mathbf{Z}_p\rangle, \langle\mathbf{Z}_d\rangle, q_{x_{d,kt}}(x_{d,kt}), \langle x_{d,kt}\rangle = 0, \forall k \in [K], t\in[\tau_d]$
repeat
  for $k=1$ to K, $t=1$ to τ_d **do**
    for $m=1$ to M **do**
      Compute $q_{x_{d,kt}}(s_m)$ using (9) and (26).
    end for
    $\langle x_{d,kt}\rangle = \sum_{s\in\mathbb{M}} s\, q_{x_{d,kt}}(s)$.
    $\langle|x_{d,kt}|^2\rangle = \sum_{s\in\mathbb{M}}|s|^2\, q_{x_{d,kt}}(s)$.
  end for
  for $k=1$ to K **do**
    Compute $\boldsymbol{\Sigma}_{\mathbf{h}_k}$ and $\langle\mathbf{h}_k\rangle$ using (24) and (25), respectively.
  end for
  for $t=1$ to τ_p **do**
    Compute $\langle\mathbf{z}_{p,t}\rangle$ using (28).
  end for
  for $t=1$ to τ_d **do**
    Compute $\langle\mathbf{z}_{d,t}\rangle$ using (27).
  end for
until a stopping condition is met.

the latent variables is given by

$$\begin{aligned}&p\left(\mathbf{Y}_p,\mathbf{Y}_d,\mathbf{Z}_p,\mathbf{Z}_d,\mathbf{H},\mathbf{X}_d,\boldsymbol{\alpha};\mathbf{X}_p,\sigma_w^2,a,b\right)\\&=p\left(\mathbf{Y}_p|\mathbf{Z}_p\right)p\left(\mathbf{Y}_d|\mathbf{Z}_d\right)p\left(\mathbf{Z}_p|\mathbf{X}_p,\mathbf{H};\sigma_w^2\right)\\&\quad\times p\left(\mathbf{Z}_d|\mathbf{X}_d,\mathbf{H};\sigma_w^2\right)p\left(\mathbf{H}|\boldsymbol{\alpha}\right)p\left(\mathbf{X}_d\right)p\left(\boldsymbol{\alpha};a,b\right).\end{aligned} \tag{30}$$

We impose a factorized structure on the posterior distribution:

$$\begin{aligned}&p\left(\mathbf{Z}_p,\mathbf{Z}_d,\mathbf{X}_d,\mathbf{H},\boldsymbol{\alpha}|\mathbf{Y}_p,\mathbf{Y}_d;\mathbf{X}_p,\sigma_w^2,a,b\right)\\&\approx q\left(\mathbf{Z}_p\right)q\left(\mathbf{Z}_d\right)q\left(\mathbf{X}_d\right)q\left(\mathbf{H}\right)q\left(\boldsymbol{\alpha}\right),\end{aligned} \tag{31}$$

where $q\left(\mathbf{H}\right)=\prod_{k=1}^{K}q\left(\mathbf{h}_k\right)$, $q\left(\mathbf{X}_d\right)=\prod_{k=1}^{K}\prod_{t=1}^{\tau_d}q\left(x_{d,kt}\right)$, $q\left(\mathbf{Z}_d\right)=\prod_{t=1}^{\tau_d}q\left(\mathbf{z}_{d,t}\right)$, $q\left(\mathbf{Z}_p\right)=\prod_{t=1}^{\tau_p}q\left(\mathbf{z}_{p,t}\right)$, and $q\left(\boldsymbol{\alpha}\right)=\prod_{k=1}^{K}q\left(\alpha_k\right)$.

The computation of $q(\mathbf{Z}_p)$, $q(\mathbf{Z}_d)$, $q(\mathbf{H})$ and $q(\mathbf{X}_d)$ is the same as in Sec. III-C. We obtain $q(\boldsymbol{\alpha})$ by taking the expectation of the logarithm of the joint distribution in (30) with respect to all the latent variables except $\boldsymbol{\alpha}$ to get

$$\ln q(\alpha_k)\propto(a+N_r-1)\ln\alpha_k-\alpha_k(b+\langle\|\mathbf{h}_k\|^2\rangle). \tag{32}$$

Taking the exponential on both sides, we observe that $q(\alpha_k)$ is Gamma distributed with shape and rate parameters $(a+N_r)$ and $(b+\langle\|\mathbf{h}_k\|^2\rangle)$, respectively. The mean of $q(\alpha_k)$ is

$$\langle\alpha_k\rangle=\frac{a+N_r}{b+\langle\|\mathbf{h}_k\|^2\rangle}. \tag{33}$$

If we denote the LSFC of the k^{th} UE by β_k, then $\langle\beta_k\rangle = 1/\langle\alpha_k\rangle = \big(b+\langle\|\mathbf{h}_k\|^2\rangle\big)/\big](a+N_r)$. We set $a=0$ and $b=10^{-4}$ in our simulations. The pseudocode for joint channel estimation and soft symbol decoding with no CSIR is similar to Algorithm 3; we highlight the changes in Algorithm 4. Note that the QVB with no CSIR algorithm does not have any matrix inverse operations, which makes it computationally attractive. Also, the estimation of $\langle\alpha_k\rangle$ involves only K scalar divisions and does not add much complexity to the procedure.

E. VB-BP Receiver

In order to improve the BER of a communication system, a typical engineering approach is to employ an iterative receiver architecture where the posterior beliefs of the output bits

Algorithm 4 QVB Joint Channel Estimation and Soft Symbol Decoding with No CSIR

Input: $\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{X}_p, M, \mathbb{M} = \{s_1, \ldots, s_M\}, \tau_p, \tau_d, \sigma_w, a, b, N_r$
Output: $\{\langle \mathbf{h}_1 \rangle, \ldots, \langle \mathbf{h}_K \rangle\}, q_{x_{d,kt}}, \langle x_{d,kt} \rangle \; \forall k \in [K], t \in [\tau_d]$

Initialize $\langle \mathbf{Z}_p \rangle, \langle \mathbf{Z}_d \rangle, q_{x_{d,kt}}(x_{d,kt}), \langle x_{d,kt} \rangle = 0, \forall k \in [K], t \in [\tau_d], \{\langle \alpha_1 \rangle, \ldots, \langle \alpha_K \rangle\}$
repeat
  Computation of $q_{x_{d,kt}}, \langle x_{d,kt} \rangle, \langle |x_{d,kt}|^2 \rangle$: Follow steps 3 to 8 of Algorithm 3.
  for $k = 1$ to K **do**
    Compute $\mathbf{\Sigma}_{\mathbf{h}_k} = \left(\frac{\sum_{t=1}^{\tau_p} |x_{p,kt}|^2 + \sum_{t=1}^{\tau_d} \langle |x_{d,kt}|^2 \rangle}{\sigma_w^2} + \langle \alpha_k \rangle \right)^{-1} \mathbf{I}_{N_r}$.
    Compute $\langle \mathbf{h}_k \rangle$ using (25).
    Compute $\langle \|\mathbf{h}_k\|^2 \rangle = \text{Trace}(\mathbf{\Sigma}_{\mathbf{h}_k}) + \|\langle \mathbf{h}_k \rangle\|^2$ and $\langle \alpha_k \rangle = \frac{a+N_r}{b+\langle \|\mathbf{h}_k\|^2 \rangle}$.
  end for
  Computation of $\langle \mathbf{z}_{p,t} \rangle, \langle \mathbf{z}_{d,t} \rangle \; \forall t$: Follow steps 13 to 18 of Algorithm 3.
until a stopping condition is met.

of the channel decoder (belief propagation (BP) decoder in our case) are fed back to the detector block. However, in commercial wireless systems, the stringent time requirements to complete the detection and channel decoding task within the transmission interval needs fast detection and decoding. Our VB approach makes it convenient to implement such a receiver without additional complexity. Instead of running the VB algorithm till convergence, we execute only one iteration of Algorithms 2, 3 or 4 and feed the LLRs obtained to the channel decoder. The BP decoder also runs only one iteration and outputs the aposteriori LLRs that are used to compute the extrinsic information.[4] We use this to adapt the data prior probability distribution in the next VB iteration. This approach is also guaranteed to converge as the VB converges to a stationary point from any initialization, and BP also converges if the associated factor graph is cycle-free. We illustrate this using an LDPC code with a cycle-free sparse parity check matrix. We find that the approach converges fast and also leads to a performance improvement of around 0.5 dB compared to performing soft symbol estimation and channel decoding one after the other. Hence, we use this receiver architecture in our simulations.

F. Computational Complexity

In this subsection, we analyze the per-iteration computational complexity of the variational Bayesian algorithms. We provide the complexity of the VB algorithm with statistical CSIR and quantized observations, as the algorithms for the perfect CSIR and unquantized observations are special cases with lower complexity. Table I shows the order ($\mathcal{O}$) of the per-iteration computational complexity of the steps involved in one iteration of the VB algorithm. The per-iteration complexity of the overall algorithm scales cubically with the number of rx antennas, quadratically with the number of users, and linearly with the constellation size and number of pilot and data symbols. In particular, the complexity is linear in the number of data symbols τ_d, unlike maximum likelihood approaches where the complexity grows exponentially with τ_d. The total time taken by VB algorithm is low due to its fast and guaranteed convergence to a local optimum. In our experiments, we find that the VB algorithm typically converges within 16 iterations, and the improvement from running further iterations is negligible.

[4]The extrinsic information is obtained by subtracting the LLRs output by the VB receiver from the LLRs output by the channel decoder.

TABLE I.
PER-ITERATION COMPLEXITY OF THE VB ALGORITHM

Matrix	Order Complexity
$\langle \mathbf{X}_d \rangle$	$MK^2 N_r \tau_d$
$\{\mathbf{\Sigma}_{\mathbf{h}_k}\}_{k=1}^K$	$K(N_r^3 + \tau_d)$
$\langle \mathbf{H} \rangle$	$KN_r^2 + K^2 N_r(\tau_p + \tau_d)$
$\langle \mathbf{Z}_p \rangle$	$KN_r \tau_p$
$\langle \mathbf{Z}_d \rangle$	$KN_r \tau_d$

IV. SIMULATION RESULTS

In this section, we evaluate the normalized mean square error (NMSE) in channel estimation and the data bit error rate (BER) of the VBI algorithms in an uplink massive MIMO wireless communication system with low resolution ADCs at the BS. We use an LDPC channel code from 3GPP 5G NR specifications [38]. 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. We set $N_r = \{100, 200\}, K = \{25, 50\}$. We vary τ_d from 100 to 450 and set the ADC resolution to 3 bits. We also evaluate the performance with pilot power boosting in which pilots are transmitted at a slightly higher power level to improve the channel estimation (see Fig. 9). Throughout this section, the data symbols are drawn i.i.d. from a 4-QAM constellation with unit energy. With higher order constellations, the performance is similar, with an expected shift in the SNR required to achieve a given BER. Except in Fig. 7b, we fix the maximum number of iterations to 16, as the BER improvement beyond 16 iterations is marginal.

The channel coefficient between the k^{th} UE and the n^{th} rx antenna at the BS is denoted by h_{nk}, and $\mathbf{h}_k \triangleq [h_{1k}, h_{2k}, \ldots, h_{N_r k}]^T$. We assume that the LSFCs between the k^{th} UE and each antenna at the BS are the same due to the close spacing between the BS antennas. We also assume that the transmit antennas at each UE are spatially uncorrelated. We model the spatial correlation between the rx antennas at the BS using a Kronecker spatial fading correlation model [39]–[41]. The channel vector $\mathbf{h}_k$ is expressed as $\mathbf{h}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{h}_k^{\text{iid}}$, where $\mathbf{h}_k^{\text{iid}}$ is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$, and $\mathbf{R}_k \in \mathbb{C}^{N_r \times N_r}$ is the spatial correlation matrix of the k^{th} UE's channel. The $(m,n)^{\text{th}}$ element of $\mathbf{R}_k$ is given in (34), where β_k is the LSFC of the k^{th} UE, $a = \frac{2\pi d}{\lambda} \sqrt{\nu_k^{\mathsf{el}}} (n-m) \cos\left(\theta_k^{\mathsf{el}}\right)$, $b = 1 + \nu_k^{\mathsf{az}} a^2 \sin^2\left(\theta_k^{\mathsf{az}}\right)$ and $c = \frac{2\pi d}{\lambda} (n-m) \sin\left(\theta_k^{\mathsf{el}}\right)$. Here, λ is the carrier wavelength, d is the antenna spacing; θ_k^{az} and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2021.3060423, IEEE Transactions on Communications

$$[\boldsymbol{R}_k]_{mn} = \frac{\beta_k}{\sqrt{b}} \exp\left(-\frac{1}{2b}\left[a^2 \cos^2\left(\theta_k^{\text{az}}\right) - 2jc\cos\left(\theta_k^{\text{az}}\right) + \nu_k^{\text{az}} c^2 \sin^2\left(\theta_k^{\text{az}}\right) \right] \right). \tag{34}$$

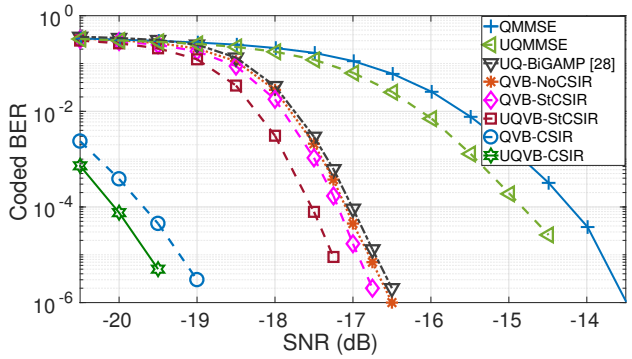


Figure 3. Coded BER vs. SNR (dB), with $N_r = 200$, $K = 50$, $\tau_d = 450$, $\tau_p = 50$, and 3 bits quantization.

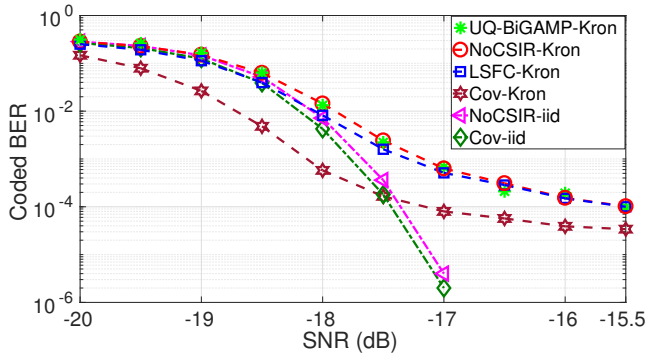


Figure 4. Coded BER vs. SNR (dB), with $N_r = 200$, $K = 50$, $\tau_d = 450$, $\tau_p = 50$, and 3 bits quantization.

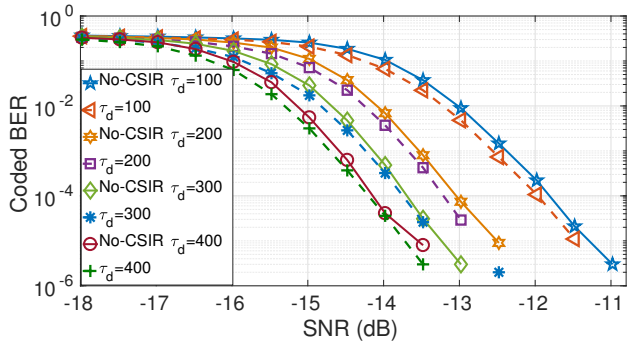


Figure 5. Coded BER vs. SNR (dB), with $N_r = 100$, $K = 25$, $\tau_p = 25$, 3 bits quantization; comparing QVB with statistical CSIR and no CSIR when the channel correlation matrix is set to $\mathbf{I}_{N_r}$.

θ_k^{el} are the means of horizontal angle of arrival (AoA) and vertical AoA, respectively; ν_k^{az} and ν_k^{el} are the variances of horizontal AoA and vertical AoA, respectively. We consider a uniform linear antenna array with the spacing between its elements set to λ. We set the mean and standard deviation of the horizontal and vertical AoA to be uniformly distributed between $-\pi/3$ and $\pi/3$ radians, and 0 and $\pi/6$ radians, respectively. The UEs adopt path loss inversion based transmit

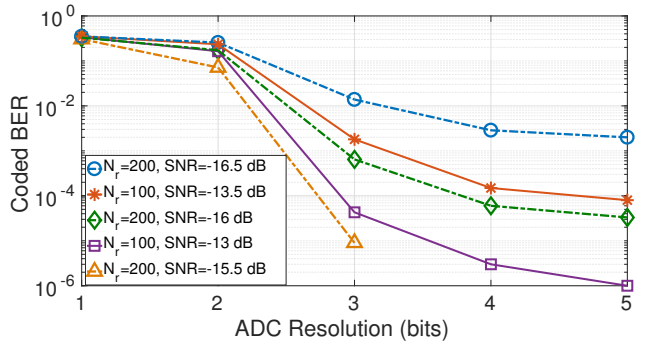


Figure 6. Coded BER vs. ADC resolution (bits), with $K = 25$, $\tau_d = 100$, and $\mathbf{R}_k = \mathbf{I}_{N_r}$ $\forall k$.

power control that compensates for the LSFCs,[5] i.e., the diagonal entries of the channel covariance matrix are close to 1 [42], [43]. Without power control, if the LSFCs of the UEs' channels are very different, the UEs that have small LSFCs will suffer high quantization noise. This leads to low signal to quantization noise ratio and therefore poor BER. Taking into account possibly imperfect power control, we assume that the diagonal entries of the channel covariance matrix are uniformly distributed between 1 and 1.2. For the uncorrelated case, we set the channel covariance to be $\mathbf{I}_{N_r}$.

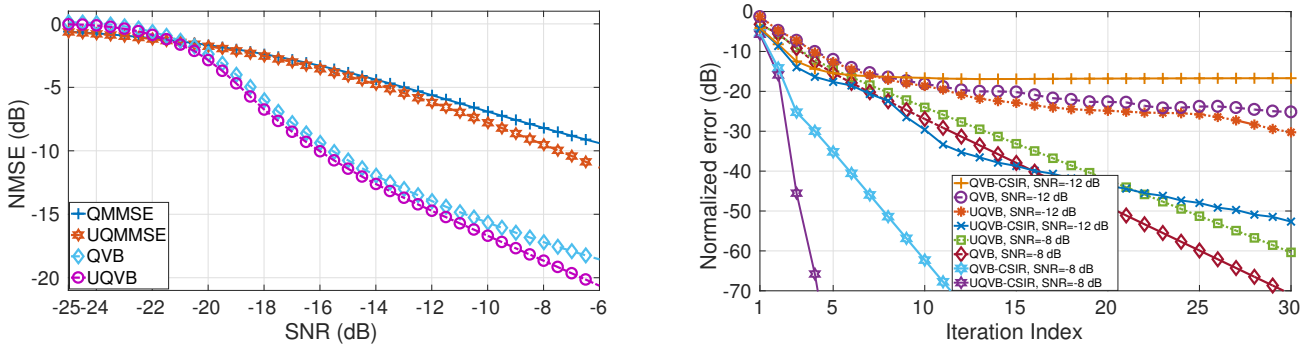
We benchmark the coded BER and NMSE performance of the quantized VB algorithm with that of a genie channel aided unquantized VB algorithm, an unquantized BiGAMP based joint channel and data estimator [28] and a unquantized linear MMSE soft-decoder [36]. We set the maximum number of iterations for the BiGAMP receiver to 500. The unquantized VB algorithm with perfect CSIR serves as a lower bound for the BER of the quantized algorithm. We note that there are variants of MMSE decoder such as MMSE-SIC, but their computational complexities are much higher due to the matrix inversions involved in every iteration of SIC [31]. Hence, we do not compare against these methods here.

Figure 3 shows the *coded* BER when SNR (dB) is varied for $N_r = 200$, $K = 50$, $\tau_d = 450$ and 3 bits quantization when the channel covariance matrix is set to $\mathbf{I}_{N_r}$. We compare the performance of the VB algorithm with perfect CSIR (labeled UQVB-CSIR, QVB-CSIR for the unquantized and quantized cases, respectively) with the performance with statistical CSIR (labeled UQVB-StCSIR, QVB-StCSIR in the unquantized and quantized cases, respectively) and no CSIR case (labeled QVB-NoCSIR).

There are three groups of curves. The best performance is achieved by the genie-aided receivers that have perfect CSIR (QVB-CSIR and UQVB-CSIR). With 3-bit quantization in the ADCs, the gap between QVB-CSIR and UQVB-CSIR is less than 0.4 dB. The next set of curves correspond to the VB algorithm with statistical CSIR and no CSIR. There

[5]The LFSCs can be estimated at the UEs, for example, using the primary synchronization signals that are periodically transmitted by the BS.

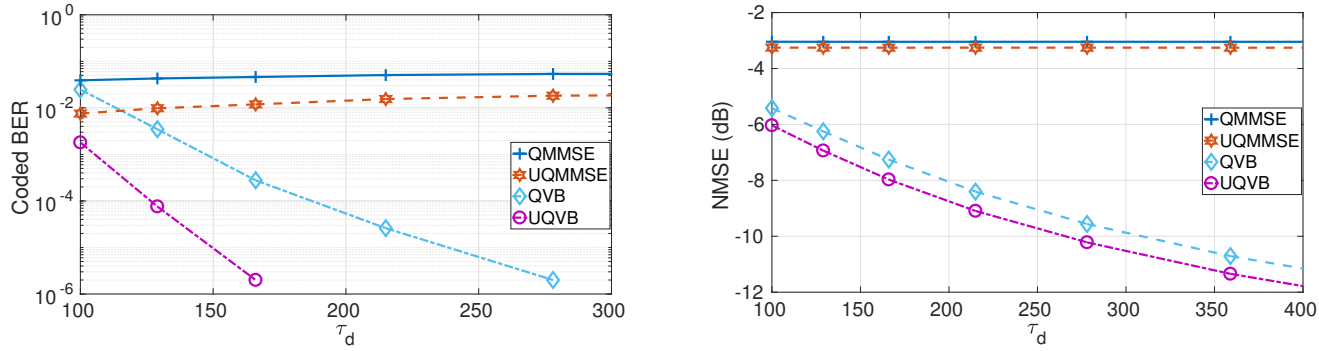
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2021.3060423, IEEE Transactions on Communications



(a) NMSE vs. SNR (dB) for $N_r = 200$, $K = 50$, $\tau_d = 450$, $\tau_p = 50$.

(b) Convergence of VB algorithm. $N_r = 100$, $K = 25$, $\tau_d = 200$.

Figure 7. NMSE performance and convergence behavior of VB algorithm, with 3 bits quantization.



(a) Coded BER vs. τ_d.

(b) NMSE vs. τ_d.

Figure 8. Coded BER as a function of the data duration, with $N_r = 100$, $K = 25$, SNR=-13.5 dB, $\tau_p = 25$, and 3 bits quantization.

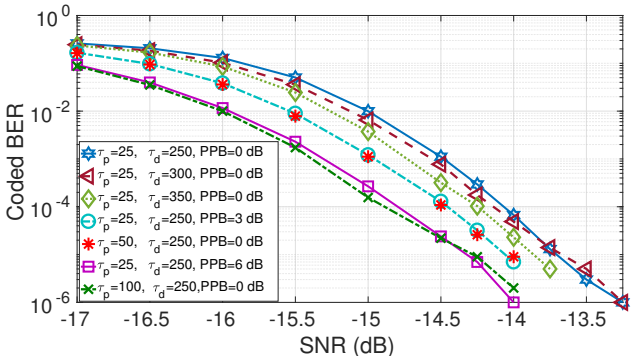
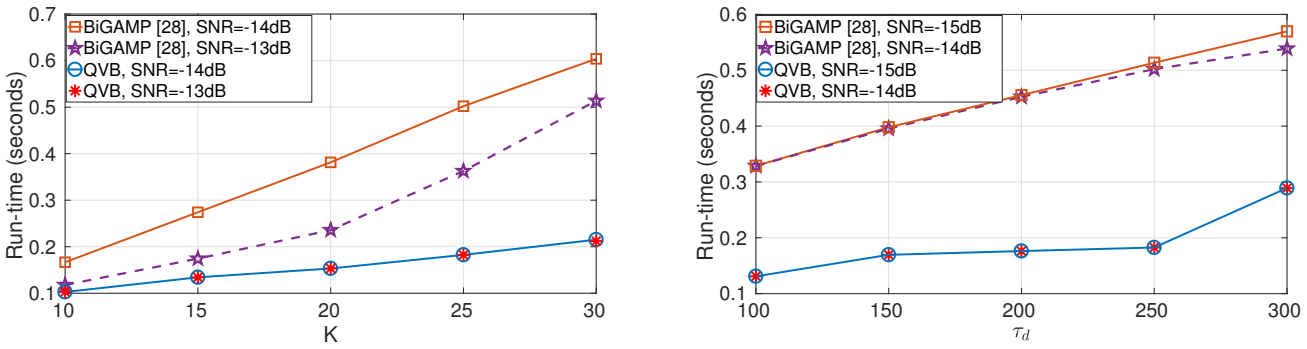


Figure 9. Coded BER vs. SNR (dB), with $N_r = 100$, $K = 25$, 3 bits quantization, pilot power boosting and increased τ_p.

is almost no performance gap between the quantized VB algorithm with and without statistical CSIR. Further, as in the perfect CSIR case, the loss due to 3-bit quantization compared to the unquantized VB algorithm is less than 0.4 dB. We also note that the performance gap between the first and second set of curves is around 2.5 dB, in both the quantized and unquantized cases. This illustrates that it is important to account for the effect of channel estimation errors, in order to realistically estimate the performance. The performance gap can be reduced by employing pilot power boosting or longer pilot sequences. UQVB-StCSIR outperforms unquantized BiGAMP (labeled UQ-BiGAMP) by more than 0.5 dB. Moreover, with only 3 bits quantization, both QVB-StCSIR and QVB-NoCSIR marginally outperform UQ-BiGAMP. Finally, the worst performance is achieved by the MMSE based receivers [36], with the gap between the quantized MMSE receiver and the quantized VB algorithm with no CSIR being nearly 3 dB at a BER of 10^{-4}. The significantly better performance achieved by the VB algorithms is clear from the plot. In addition, the gap between the quantized MMSE and the unquantized MMSE receivers shows that linear receivers can result in suboptimal performance, even if the channel state is made available to the receiver. For the quantized and unquantized MMSE detectors, the channels are estimated using the quantized and unquantized pilot received signals, respectively.

Figure 4 shows the BER vs. SNR performance of the VB algorithm with correlated channels with and without CSIR (curves labeled Kron). We also show the performance under i.i.d. channels (curves labeled iid). We observe that, in a spatially correlated scenario, the algorithm that has the knowledge of the channel covariance matrices (Cov-Kron) performs around 1 dB better at a BER of 10^{-3} than the algorithms that do not have the knowledge of the channel covariance matrices (NoCSIR-Kron and LSFC-Kron). Note that the VB algorithm with NoCSIR assumes i.i.d. channels, which results in a degraded performance under spatially correlated scenarios. This shows the importance of utilizing the correlation information when designing receiver schemes. We observe that the performance of the VB algorithm when the BS has the knowledge of only the LSFCs of the UEs' channels (LSFC-Kron) is poor compared to Cov-Kronecker. Thus, utilizing only the knowledge of LSFCs is not sufficient, and the full correlation information is necessary to obtain better performance. Also, under an i.i.d. channel scenario, VB

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2021.3060423, IEEE Transactions on Communications



(a) Run-time (s) vs. K for $N_r = 100$, $\tau_d = 250$, 3 bits quantization.

(b) Run-time (s) vs. τ_d for $N_r = 100$, $K = 25$, 3 bits quantization.

Figure 10. Run-time (seconds) comparison of QVB and unquantized BiGAMP algorithms.

with NoCSIR (NoCSIR-iid) performs very close to VB with complete knowledge of channel covariance matrices (Cov-iid). This shows that VB with NoCSIR is able to estimate the LSFCs accurately with very little additional computational complexity. We see an error floor in the spatially correlated case at high SNRs. This is because, the channel AoAs are spread within a narrow range (standard deviation of $\pi/6$), which results in non-negligible multiuser interference at high SNRs. We also observe that, in the spatially correlated scenario, the unquantized BiGAMP based receiver (labeled UQ-BiGAMP-Kron) performs about 1 dB worse than Cov-Kron with 3 bit ADCs, and marginally worse than LSFC-Kron with 3 bit ADCs. The BiGAMP receiver assumes i.i.d. channels but has the knowledge of the LSFCs. Finally, the NoCSIR-Kron curve, which is obtained without knowledge of the channel covariance matrices or the LSFCs, but 3 bits quantization, performs the same as UQ-BiGAMP-Kron.

Figure 5 compares the BER of the QVB algorithms with and without the knowledge of statistical CSIR for various values of τ_d, for uncorrelated channels ($\mathbf{R}_k = \mathbf{I}_{N_r}$). The performance of QVB with no CSIR is only marginally worse than that of QVB with statistical CSIR. Also, we see a roughly 10-fold improvement BER when the number of data symbols is doubled. This shows that the VB algorithm is able to effectively use the data symbols to improve the channel estimates. We recall that the computational complexity of the receiver for uncorrelated channels is lower than the correlated channel case, as the channel covariance matrix is diagonal and all the matrix operations can be computed using scalar computations.

Figure 6 shows the BER vs. ADC resolution (in bits) for various rx antenna and SNR configurations. We set $\tau_d = 100$, $K = 25$ and the spatial correlation matrices to $\mathbf{I}_{N_r}$. The BER improves as the ADC resolution increases, as expected, but the slope of the BER curve decreases and becomes almost 0 beyond $4-$bits resolution. This illustrates that low resolution ADCs are relevant in wireless communication systems, especially in massive MIMO systems where the number of antennas are large compared to the number of users. For example, in order to achieve a BER of 10^{-3}, UEs have to expend twice the transmit power if the BS is equipped with 100 antennas and 3 bit ADCs compared to a system with 200 antennas at the BS and 5 bit ADCs. Thus, a 3 dB higher transmit power at the UEs can lead to significant power savings at the BS. On the other hand, if a UE does not have the power budget to increase its transmit power, using additional antennas at the BS can provide the required rx antenna gain. Such tradeoff analyses can be used by a system designer to configure the system parameters based on the bit budget and power constraints at the BS and UEs.

We now turn to the channel estimation performance of the VB algorithms. Figure 7a shows the NMSE of channel estimation as a function of the SNR (dB), for $N_r = 200$, $K = 50$, $\tau_d = 450$ and 3 bits quantization. The VBI algorithms *with quantized observations* provide around 8 dB improvement at an NMSE of -10 dB compared to the MMSE estimation based on *unquantized observations*. We also see that the NMSE of the unquantized VB (UQVB) is almost the same as the 3-bit quantized VB (QVB) algorithm. This is because the VBI algorithms refine the channel estimates based on the posterior beliefs of the data symbols. This feature can be directly translated to a reduction in the training overhead required in massive MIMO systems with low resolution ADCs, and thereby improve the achievable spectral efficiency.

Figure 7b shows the convergence behavior of the VB algorithms, with $N_r = 100$, $K = 25$, $\tau_d = 200$ for both unquantized and 3 bits quantization cases. The convergence behaviors for both the unknown CSIR and genie aided case (with legend suffixed with `CSIR`) are shown. We use the means of the estimated data symbols and channel to compute the normalized error for the perfect CSIR and the unknown CSIR cases, respectively. We see that the VBI algorithms converge to a normalized error below -20 dB within about 20 iterations. This illustrates that the proposed algorithms are of polynomial complexity with fast convergence.

Figures 8a and 8b show the coded BER and NMSE performance of the VB algorithms, respectively, with $N_r = 100$, $K = 25$, SNR $= -13.5$ dB and 3 bits quantization, as a function of the data duration τ_d. As τ_d increases, the BER and NMSE of the VBI algorithms improves, unlike the MMSE receivers. Again, this is because the VBI algorithm uses the posterior beliefs of the data symbols to refine its channel estimates, which in turn improves the quality of the posterior beliefs of the data symbols. Therefore, the performance can be dramatically improved by jointly decoding a larger number of data symbols (up to the coherence time of the channel), leading to a reduced training overhead even in the presence of low resolution ADCs.

Figure 9 compares the BER across different algorithms when the pilot transmit power is boosted (PPB) by a certain amount above that of the data transmit power, and with τ_p set to K, $2K$ and $4K$. We see that, as the pilot duration increases, the BS is able to estimate the channels better, resulting in improved performance. A similar performance improvement occurs with PPB. For example, PPB of 3 dB results in nearly the same performance as the case when the pilot duration is doubled. We also show the BER when $\tau_d = \{250, 300, 350\}$. If the coherence interval is large, the VBI approach provides a system designer the option to avoid PPB or increasing τ_p while still meeting the QoS requirements.

Figures 10a and 10b compare the average run times of the QVB algorithm based on quantized observations with that of unquantized BiGAMP procedure for various values of the number of users K and the data duration τ_d, respectively. The simulations were executed using MATLAB R2020b in an Intel core i7, 3 GHz $\times$ 8 CPU with 64 GB RAM running an Ubuntu 18.04 LTS 64 bits operating system. We use the normalized mean squared difference in the channel estimate between two successive iterations as the convergence metric, and set it to 10^{-5}. We see that the total run time taken by quantized VB algorithm is far less than that of unquantized BiGAMP, even though the per iteration complexity of BiGAMP scales linearly with system dimensions. This shows that our proposed quantized VB-BP based joint channel estimation and soft symbol decoding not only performs better than unquantized BiGAMP, but is also faster.

V. Conclusions

We considered joint channel estimation and soft symbol decoding in a single carrier uplink massive multiple input multiple output (MIMO) receiver with low resolution ADCs. We proposed a novel, low-complexity VB procedure that directly outputs the posterior beliefs of the data symbols. The channel estimates obtained can potentially be used for signal to interference noise (SINR) computation and link adaptation. We provided a flexible approach to integrate the VB receiver with a BP channel decoder via extrinsic information feedback. We evaluated the coded data BER and the NMSE in the channel estimates obtained by our algorithm using Monte Carlo simulations and benchmarked it against the state-of-the-art receivers. Our future work would involve combining various approximate inference techniques for the joint channel estimation and data decoding, and perform a comparative study of their performance. We would also like to unfold the VB receiver in a deep neural network to reduce the computation time further, and study its performance under various system configurations.

References

[1] S. S. Thoota and C. R. Murthy, "Variational Bayesian inference based soft-symbol decoding for uplink massive MIMO systems with low resolution ADCs," in *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, 2019.

[2] S. S. Thoota, C. R. Murthy, and R. Annavajjala, "Quantized variational Bayesian joint channel estimation and data detection for uplink massive MIMO systems with low resolution ADCs," in *Proc. MLSP*, 2019, pp. 1–6.

[3] 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.

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

[5] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[6] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *ArXiv*, p. arXiv:1404.7736, Apr. 2014.

[7] B. Murmann, "Energy limits in A/D converters," in *2013 IEEE Faible Tension Faible Consommation*, 2013, pp. 1–4.

[8] *ADS54J20 Dual-Channel, 12-Bit, 1.0-GSPS, Analog-to-Digital Converter (Rev. B)*, Texas Instruments, 2011, rev. B. [Online]. Available: https://www.ti.com/document-viewer/ADS54J20/datasheet

[9] 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.

[10] 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.

[11] Y. Li, C. Tao, A. L. Swindlehurst, A. Mezghani, and L. Liu, "Downlink achievable rate analysis in massive MIMO systems with one-bit DACs," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1669–1672, Jul. 2017.

[12] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, Jun. 2018.

[13] 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.

[14] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2007, pp. 1286–1289.

[15] J. A. Nossek and M. T. Ivrlač, "Capacity and coding for quantized MIMO systems," in *Proc. Int. Conf. Wireless. Commun., and Mobile Comput. Conf. (IWCMC)*, 2006, pp. 1387–1392.

[16] J. Singh, O. Dabeer, and U. Madhow, "Capacity of the discrete-time AWGN channel under output quantization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2008, pp. 1218–1222.

[17] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[18] 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.

[19] 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.

[20] S.-N. Hong, S. Kim, and N. Lee, "A weighted minimum distance decoding for uplink multiuser MIMO systems with low-resolution ADCs," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 1912–1924, May 2018.

[21] S.-N. Hong and N. Lee, "Soft-output detector for uplink MU-MIMO systems with one-bit ADCs," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 930–933, May 2018.

[22] 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.

[23] S. Wang, Y. Li, and J. Wang, "Multiuser detection for uplink large-scale MIMO under one-bit quantization," in *Proc. ICC*, 2014, pp. 4460–4465.

[24] 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, Mar. 2018.

[25] H. Wang, W.-T. Shih, C.-K. Wen, and S. Jin, "Reliable OFDM receiver with ultra-low resolution ADC," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3566–3579, Jan. 2019.

[26] 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.

[27] 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.

[28] 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.
[29] P. Sun, Z. Wang, R. W. Heather, 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, Dec. 2018.
[30] Y. Cho and S.-N. Hong, "One-bit successive-cancellation soft-output (OSS) detector for uplink MU-MIMO systems with one-bit ADCs," *IEEE Access.*, vol. 7, pp. 27 172–27 182, Feb. 2019.
[31] Z. Shao, R. C. de Lamare, and L. T. N. Landau, "Iterative detection and decoding for large-scale multiple-antenna systems with 1-bit ADCs," *IEEE Wireless Comm. Letters*, vol. 7, no. 3, pp. 476–479, Jun. 2018.
[32] 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.
[33] M.-A. Badiu, G. E. Kirkelund, C. N. Manchón, E. Riegler, and B. H. Fleury, "Message-passing algorithms for channel estimation and decoding using approximate inference," in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2376–2380.
[34] G. E. Kirkelund, C. N. Manchón, L. P. Christensen, E. Riegler, and B. H. Fleury, "Variational message-passing for joint channel estimation and decoding in MIMO-OFDM," in *Proc. Globecom*, 2010, pp. 1–6.
[35] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2003.
[36] 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.
[37] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, 2004.
[38] "3GPP TS 38.212 V15.7.0 (2019-09); NR Multiplexing and channel coding," 2019.
[39] L. Schumacher, K. I. Pedersen, and P. E. Mogensen, "From antenna spacings to theoretical capacities-guidelines for simulating MIMO systems," in *Proc. PIMRC*, vol. 2, 2002, pp. 587–592.
[40] S. Wu, L. Kuang, Z. Ni, D. Huang, Q. Guo, and J. Lu, "Message-passing receiver for joint channel estimation and decoding in 3d massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8122–8138, Dec. 2016.
[41] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. ICC*, Jun. 2014, pp. 5865–5870.
[42] S. Weber, J. G. Andrews, and N. Jindal, "The effect of fading, channel inversion, and threshold scheduling on ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4127–4149, Feb. 2007.
[43] P. Herath, C. Tellambura, and W. A. Krzymień, "Coverage probability analysis of three uplink power control schemes: Stochastic geometry approach," *EURASIP J. on Wireless Commun. and Netw.*, vol. 2018, no. 1, p. 141, 2018.

Sai Subramanyam Thoota (S'16) received the B. E. degree in ECE from the College of Engg., Guindy, Anna University, Chennai, India, in 2003, and the M. Tech. degree in EE from the Indian Institute of Technology Madras, Chennai, India, in 2006. From 2008 to 2015, he worked as a Chief engineer for Samsung Research India Bangalore, India, where he worked on GSM/GPRS/EDGE baseband transceiver design. From 2015 to 2016, he worked as a Project Associate at the Dept. of ECE, Indian Institute of Science, Bangalore, India. Since 2016, he is pursuing Ph. D at the Department of ECE, Indian Institute of Science, Bangalore, India. His research interests include massive MIMO communication systems, sparse signal recovery and statistical signal processing.

Chandra R. Murthy (S'03–M'06–SM'11) received the B. Tech. degree in Electrical Engineering from the Indian Institute of Technology, Madras in 1998, the M. S. and Ph. D. degrees in Electrical and Computer Engineering from Purdue University and the University of California, San Diego, in 2000 and 2006, respectively. From 2000 to 2002, he worked as an engineer for Qualcomm Inc., where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From Aug. 2006 to Aug. 2007, he worked as a staff engineer at Beceem Communications Inc. on advanced receiver architectures for the 802.16e Mobile WiMAX standard. In Sept. 2007, he joined the Department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India, where he is currently working as a Professor.

His research interests are in the areas of energy harvesting communications, 5G/6G technologies and compressed sensing. He has over 69 journal and 98 conference papers to his credit. He is a recipient of the MeitY Young Faculty Fellowship from the Govt. of India and the Prof. Satish Dhawan state award for engineering from the Karnataka State Government. He was an associate editor for the IEEE SIGNAL PROCESSING LETTERS during 2012-16 and the SADHANA ACADEMY PROCEEDINGS IN ENGINEERING SCIENCES during 2017-18. He is a past Chair of the IEEE Signal Processing Society, Bangalore Chapter. He was an elected member of the IEEE SPCOM Technical Committee during 2014-19. He is currently serving as a senior area editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and as an associate editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON INFORMATION THEORY.