

Variational Bayes for Joint Channel Estimation and Data Detection in Few-Bit Massive MIMO Systems

Ly V. Nguyen, *Member, IEEE*, A. Lee Swindlehurst, *Fellow, IEEE*, and Duy H. N. Nguyen, *Senior Member, IEEE*

Abstract—Massive multiple-input multiple-output (MIMO) communications using low-resolution analog-to-digital converters (ADCs) is a promising technology for providing high spectral and energy efficiency with affordable hardware cost and power consumption. However, the use of low-resolution ADCs requires special signal processing methods for channel estimation and data detection since the resulting system is severely non-linear. This paper proposes joint channel estimation and data detection methods for massive MIMO systems with low-resolution ADCs based on the variational Bayes (VB) inference framework. We first derive matched-filter quantized VB (MF-QVB) and linear minimum mean-squared error quantized VB (LMMSE-QVB) detection methods assuming the channel state information (CSI) is available. Then we extend these methods to the joint channel estimation and data detection (JED) problem and propose two methods we refer to as MF-QVB-JED and LMMSE-QVB-JED. Unlike conventional VB-based detection methods that assume knowledge of the second-order statistics of the additive noise, we propose to float the elements of the noise covariance matrix as unknown random variables that are used to account for both the noise and the residual inter-user interference. We also present practical aspects of the QVB framework to improve its implementation stability. Finally, we show via numerical results that the proposed VB-based methods provide robust performance and also significantly outperform existing methods.

Index Terms—Approximate message passing, Bayesian inference, detection, estimation, massive MIMO, soft interference cancellation, variational Bayesian.

I. INTRODUCTION

Beyond-5G wireless systems will require exploitation of the large bandwidths available at THz frequencies (0.3–3 THz) [1]–[3]. An inherent challenge in operating in these bands is the strong radio frequency (RF) path loss, and while this can be effectively addressed by exploiting the beamforming gain available from large antenna arrays, scaling up existing RF technologies to very large arrays becomes complex, expensive, and demands high power consumption. Therefore, implementing massive antenna arrays for THz communications will require radical simplifications in the RF architecture. Hybrid analog-digital arrays reduce the number

of RF chains with respect to (w.r.t.) the number of antenna elements [4], but this approach yields poor spatial multiplexing and does not scale well at higher frequencies and wider bandwidths due to the need for complex analog circuitry and resource-consuming beam management schemes [5].

An alternative solution is to retain the RF chains for each antenna but reduce complexity and energy consumption through the use of low-resolution analog-to-digital converters (ADCs). It has been shown that fully digital arrays with lower-resolution data converters (even down to 1 bit) can significantly outperform hybrid analog-digital architectures in terms of beamforming flexibility and spectral/energy efficiency [6]. This is because the use of low-resolution ADCs maintains the high spatial multiplexing gains of massive arrays, and they more easily scale to higher frequencies and bandwidths with significantly reduced hardware cost and power consumption. However, the use of low-resolution quantization requires special signal processing methods for channel estimation and data detection since the resulting system is severely non-linear, and the received signals are significantly distorted.

There has been a plethora of channel estimation and data detection studies for massive MIMO systems with low-resolution ADCs. For example, one-bit maximum-likelihood (ML) and near-ML methods were proposed in [7]. The Bussgang decomposition was used to derive different linear channel estimators in [8], [9] and linear data detectors in [9]–[11]. While the ML and near-ML methods are either too complicated for practical implementation or non-robust at high signal-to-noise ratios (SNRs), the linear Bussgang-based receivers have lower complexity and are more robust, but they have limited performance. Several other detection approaches have been proposed in [12]–[15] but they require the use of either a cyclic redundancy check (CRC) or an error correcting code (ECC). The authors in [16] developed a bilinear generalized approximate message passing (BiGAMP) algorithm [17] to solve the joint channel estimation and data detection (JED) problem for few-bit MIMO systems.

Recently, machine learning for low-resolution MIMO channel estimation and data detection has gained interest and there have also been numerous results reported in the literature. In particular, the work in [18] shows how support vector machine (SVM) models can be applied to one-bit massive MIMO channel estimation and data detection. The authors of [11] exploit a deep neural network (DNN) framework to develop a special model-driven detection approach that outperforms the SVM-based methods in [18]. Deep learning-based joint pilot signal and channel estimator designs were proposed in [19] and [20]. While a conventional DNN structure was used in [19], the work in [20] employed a model-driven network similar

This work was supported by the U.S. National Science Foundation under grants CCF-2225575, CCF-2322191, ECCS-2146436, CCF-2225576, and CCF-2322190. Part of this work has been presented at the Asilomar Conf. Signals, Systems and Computers in Pacific Grove, CA, USA, in 2022.

Ly V. Nguyen was with the Computational Science Research Center, San Diego State University, San Diego, CA, USA 92182, and is now with the Center for Pervasive Communications and Computing, University of California, Irvine, CA, USA 92697 (e-mail: vanln1@uci.edu).

A. Lee Swindlehurst is with the Center for Pervasive Communications and Computing, University of California, Irvine, CA, USA 92697 (e-mail: swindle@uci.edu).

Duy H. N. Nguyen is with the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA, USA 92182 (e-mail: duy.nguyen@sdsu.edu).

to [11]. The work in [21] proposed another DNN-based detector but its computational complexity is high since the detection network must be retrained for each new channel realization. Several learning-based blind detection methods were proposed in [22]–[24] but they are restricted to small-scale systems. In [25], Bayesian inference was used to develop a JED method for quantized single-antenna systems with orthogonal frequency division multiplexing (OFDM) and time-frequency doubly selective (DS) channels where the sparsity of the DS channels was exploited. Another JED method was proposed in [26] based on the variational Bayesian (VB) inference framework, and it was shown to outperform the BiGAMP-based method in [16] for soft symbol decoding. The work in [26] was then extended to OFDM systems in [27]. The authors in [28] applied the VB method to the sparse mmWave channel estimation problem via spectral line estimation. In a recent work [29], VB inference was shown to be very efficient in MIMO data detection with infinite-resolution (perfect) ADCs. VB inference for unquantized systems was also studied for OFDM [30], [31] and orthogonal time frequency space (OTFS) [32] systems, all assuming that the noise variance or covariance matrix is fixed and known in advance.

In this paper, we develop a VB framework for channel estimation and data detection for massive MIMO systems with low-resolution ADCs. While conventional machine learning models such as SVM and DNN only provide a point estimate of the signal of interest, e.g., the channel or the data symbols, the VB approach can provide the posterior distribution of the estimate, which is important in subsequent signal processing steps such as channel decoding. Another advantage of VB is that it does not require a training process like DNNs which often suffer from performance degradation due to mismatch between the actual model and that used during training. Unlike our previous work in [33] which only considers the data detection problem and assumes perfect channel state information (CSI), we study both channel estimation and data detection in this paper. The works most closely related to ours are the Bayesian inference-based methods in [26] and [34]. However, they assume perfect knowledge of the noise variance, which is challenging to obtain and, to the best of our knowledge, remains an open problem in quantized systems. In addition, the methods in [26] and [34] do not efficiently address the problem of inter-user interference estimation and compensation. To address these challenges, the key and novel contribution of our work is developing a VB framework that can efficiently estimate and compensate for the joint effects of noise and inter-user interference through the use of a new latent variable. More specific contributions of our work are summarized as follows:

- First we consider the simpler case with known channel state information (CSI). We devise a matched-filter quantized VB (MF-QVB) detection method for few-bit MIMO systems that, unlike the VB-based detection method in [26], does not assume an *a priori* known noise variance. The proposed MF-QVB method floats the noise variance as a latent variable and uses it to also account for residual inter-user interference. This latent variable is jointly estimated with the transmitted data symbol vector. We also introduce the use of a logistic rather than Gaussian distribution for describing the tails of the distribution of the

latent variables, which dramatically improves robustness, particularly at higher SNRs. Simulation results show that with these modifications, our MF-QVB approach performs very well for both QPSK and 16QAM modulation, while the approach in [26] fails to achieve reasonable detection performance.

- Generalizing the first approach, we next consider cases where the additive noise is both unknown and spatially correlated, and we treat the entire noise covariance matrix as a latent variable rather than simply assuming it is a scaled identity matrix. We develop a linear minimum mean-squared error quantized VB (LMMSE-QVB) detector based on this approach that offers performance similar to MF-QVB for independent and identically distributed (i.i.d.) channels, but significantly outperforms MF-QVB for spatially correlated channels. Similarly, the approach in [26] is unable to achieve good results in this scenario because it ignores the correlated noise.
- Next we study the same low-resolution MIMO cases as above, but we assume that the CSI is unknown and must be estimated prior to detection. We develop two VB-based JED algorithms for this problem, referred to as MF-QVB-JED and LMMSE-QVB-JED. These algorithms jointly estimate the channel matrix, the symbol data vectors, and either the unknown noise variance (MF-QVB-JED) or the unknown noise covariance matrix (LMMSE-QVB-JED). Again, this goes well beyond the prior work in [26] that simply assumes an *a priori* known scaled identity noise covariance.
- Finally, we describe various practical aspects of our proposed VB-based algorithms to improve their numerical stability and reduce their computational complexity. We show via simulations that the proposed VB detection algorithms provide much lower symbol error rates (SERs) compared to the conventional VB-based methods in [26]. The proposed QVB-JED algorithms also outperform our previously proposed FBM-DetNet approach in [20], particularly for spatially correlated channels.

The rest of the paper is organized as follows. We present the system model and the problem of interest in Section II. Next, a brief introduction to the VB inference framework is given in Section III. Then, in Section IV, we derive the VB-based data detection method when the CSI is known. Section V proposes the VB-based JED methods. We present practical implementation aspects of the VB framework as well as numerical results in Section VI. Finally, Section VII concludes the paper.

Notation: Scalar $[\mathbf{X}]_{ij}$ denotes the element at the i th row and j th column of matrix $\mathbf{X}$; vectors $[\mathbf{X}]_{i,:}$ and $[\mathbf{X}]_{:,j}$ denote the i th row and j th column of matrix $\mathbf{X}$, respectively; the operators $\text{Tr}\{\mathbf{X}\}$ and $|\mathbf{X}|$ represent the trace and determinant of a square matrix $\mathbf{X}$, respectively; the Frobenius norm of a matrix $\mathbf{X}$ is represented by $\|\mathbf{X}\|_F$; the distribution of a K -element complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\pi^K |\boldsymbol{\Sigma}|} \exp(-(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))$, and is also written as $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; the functions $\phi(x)$ and $\Phi(x)$ denote the PDF and cumulative distribution function (CDF) of a standard Gaussian random variable $\mathcal{N}(0, 1)$; the

operators $\mathbb{E}_{p(x)}[x]$ and $\text{Var}_{p(x)}[x]$ denote the mean and variance of x w.r.t. its distribution $p(x)$; in addition, we use $\langle x \rangle$, τ_x , and $\langle |x|^2 \rangle = \langle |x| \rangle^2 + \tau_x$ to denote the mean, variance, and second moment of x w.r.t. a variational distribution $q(x)$. The symbols $\sim$ and $\propto$ indicate “distributed according to” and “proportional to”, respectively. Finally, $\mathbb{1}(\cdot)$ denotes the indicator function which equals one if the argument holds true, or zero otherwise.

II. SYSTEM MODEL AND PROBLEM STATEMENT

We consider an uplink MIMO system with K single-antenna users and an M -antenna base station (BS). The users’ transmitted symbols are collected in the vector $\mathbf{x} = [x_1, \dots, x_K]^T$, where x_i corresponds to user- i and is drawn from a discrete constellation $\mathcal{S}$. The prior distribution of x_i is $p(x_i) = \sum_{a \in \mathcal{S}} p_a \delta(x_i - a)$, where p_a corresponds to a known prior probability of the constellation point $a \in \mathcal{S}$. It is assumed that the symbols in $\mathbf{x}$ are independent of each other, i.e., $p(\mathbf{x}) = \prod_{i=1}^K p(x_i)$. For ease of presentation, we assume the same constellation $\mathcal{S}$ and the same prior distribution $p(x_i)$ for all the users. However, the methods proposed in the paper are not restricted to this assumption.

Given the uplink channel $\mathbf{H} = [\mathbf{h}_1 \dots \mathbf{h}_K] \in \mathbb{C}^{M \times K}$, the uplink MIMO transmission at a given time slot can be modeled as $\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where $\mathbf{r}$ is the unquantized received signal vector and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_M)$ models the independent and identically distributed (i.i.d.) additive white Gaussian noise at the receiver. The channel vector $\mathbf{h}_i$ from user- i to the BS is assumed to be distributed as $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_i)$, where $\mathbf{C}_i \triangleq \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^H]$ describes the spatial correlation of the channel, and is assumed to be known. The spatial statistics of the channel are relatively straightforward to obtain since they typically change much more slowly than a channel coherence interval. Thus, sample averages of the channel estimates obtained over these coherence intervals can be used to estimate $\mathbf{C}_i$. Finally, we assume that the users’ channels are independent, and hence $\mathbb{E}[\mathbf{h}_i \mathbf{h}_j^H] = \mathbf{0}$ for $i \neq j$.

We consider a block fading channel in which a pilot transmission phase over T_p time slots is followed by a data transmission phase over T_d time slots. We denote $\mathbf{X}_p \in \mathbb{C}^{K \times T_p}$ as the pilot matrix and $\mathbf{X}_d \in \mathbb{C}^{K \times T_d}$ as the data matrix. For $s \in \{p, d\}$, the unquantized received signals $\mathbf{R}_s \in \mathbb{C}^{M \times T_s}$ are given by

$$\mathbf{R}_s = \mathbf{H}\mathbf{X}_s + \mathbf{N}_s, \quad (1)$$

where $\mathbf{X}_s = [\mathbf{x}_{s,1}, \dots, \mathbf{x}_{s,T_s}]^T$ with $\mathbf{x}_{s,t} = [x_{s,t1}, \dots, x_{s,tK}]^T$, and $\mathbf{N}_s = [\mathbf{n}_{s,1}, \dots, \mathbf{n}_{s,T_s}]^T$. We assume that user- i transmits with power ρ_i during the pilot transmission phase. Each received analog signal is then quantized by a pair of b -bit ADCs to produce the quantized received signal:

$$\mathbb{R}\{\mathbf{Y}_s\} = Q_b(\mathbb{R}\{\mathbf{R}_s\}), \quad \Im\{\mathbf{Y}_s\} = Q_b(\Im\{\mathbf{R}_s\}), \quad (2)$$

where $Q_b(\cdot)$ denotes the b -bit ADC operation which is applied separately to every element of its matrix or vector argument. We assume $Q_b(\cdot)$ performs b -bit uniform scalar quantization, which is characterized by a set of $2^b - 1$ thresholds denoted as $\{d_1, \dots, d_{2^b-1}\}$. Without loss of generality, we assume $-\infty = d_0 < d_1 < \dots < d_{2^b-1} < d_{2^b} = \infty$. For a quantization

step size of Δ , the quantization thresholds are given by

$$d_k = (-2^{b-1} + k)\Delta, \text{ for } k \in \mathcal{K} = \{1, \dots, 2^b - 1\}. \quad (3)$$

The quantized output y is then defined as

$$y = Q_b(r) = \begin{cases} d_k - \frac{\Delta}{2}, & \text{if } r \in (d_{k-1}, d_k] \text{ with } k \in \mathcal{K} \\ (2^b - 1)\frac{\Delta}{2}, & \text{if } r \in (d_{2^b-1}, d_{2^b}]. \end{cases} \quad (4)$$

We also define y^{low} and y^{up} as lower and upper thresholds of the quantization bin to which y belongs.

In this paper, we first study the data detection problem with known CSI, i.e., where the problem of interest is to detect the data matrix $\mathbf{X}_d$ using the received signal matrix $\mathbf{Y}_d$ and the channel matrix $\mathbf{H}$. Then, in the following section we study the problem of joint channel estimation and data detection where $\mathbf{H}$ is estimated and $\mathbf{X}_d$ detected using knowledge of the pilot matrix $\mathbf{X}_p$ and the received signal matrices $\mathbf{Y}_p$ and $\mathbf{Y}_d$.

III. BACKGROUND ON VARIATIONAL BAYES INFERENCE

Here we present a brief background on the VB method for approximate inference that will be used to solve the problems of interest in this paper. In variational inference, the posterior distribution $p(\mathbf{x}|\mathbf{y})$ over a set of latent variables $\mathbf{x}$ given some observed data $\mathbf{y}$ is approximated by a variational distribution $q(\mathbf{x})$. A set of variational parameters describing $q(\mathbf{x})$ within a family of densities $\mathcal{Q}$ that minimize the Kullback-Leibler (KL) divergence from $q(\mathbf{x})$ to $p(\mathbf{x}|\mathbf{y})$ are determined [35], [36]. Minimizing the KL divergence is equivalent to maximizing the evidence lower bound (ELBO), which is defined as

$$\text{ELBO}(q) = \mathbb{E}_{q(\mathbf{x})}[\ln p(\mathbf{x}, \mathbf{y})] - \mathbb{E}_{q(\mathbf{x})}[\ln q(\mathbf{x})]. \quad (5)$$

The VB method assumes the *mean field variational family*, for which $q(\mathbf{x}) = \prod_{i=1}^m q_i(x_i)$. The general expression for the optimal solution $q_i(x_i)$ can be obtained as [35]

$$q_i(x_i) \propto \exp \left\{ \langle \ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x}) \rangle_{-x_i} \right\}. \quad (6)$$

Here, $\langle \cdot \rangle_{-x_i}$ denotes the expectation w.r.t. all latent variables except x_i using the variational distribution $q_{-i}(\mathbf{x}_{-i}) = \prod_{j \neq i} q_j(x_j)$. In the following, if $\langle \cdot \rangle$ is used, it means the variational expectation is taken w.r.t. all the latent variables in the argument. By iterating the update of $q_i(x_i)$ sequentially over all i , the $\text{ELBO}(q)$ objective function can be monotonically improved, guaranteeing the convergence to at least a local optimum of $\text{ELBO}(q)$ [35], [36].

The following lemma on the variational posterior mean of multiple random variables will be used throughout the paper.

Lemma 1. Suppose that $\{\mathbf{A}, \mathbf{y}, \mathbf{x}\}$ are respectively of size $\{m \times n, m \times 1, n \times 1\}$ and are independent random variables w.r.t. variational distribution $q_{\mathbf{A}, \mathbf{y}, \mathbf{x}}(\mathbf{A}, \mathbf{y}, \mathbf{x}) = q(\mathbf{A})q(\mathbf{y})q(\mathbf{x})$. Assume $\mathbf{A}$ is column-wise independent and let $\langle \mathbf{a}_i \rangle$ and $\Sigma_{\mathbf{a}_i}$ be the variational mean and covariance of the i th column of $\mathbf{A}$. Let $\langle \mathbf{x} \rangle$ and $\Sigma_{\mathbf{x}}$ (and $\langle \mathbf{y} \rangle$ and $\Sigma_{\mathbf{y}}$) be the variational mean and covariance of $\mathbf{x}$ (and $\mathbf{y}$), respectively. For an arbitrary Hermitian matrix $\mathbf{B}$, let $\langle (\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{B} (\mathbf{y} - \mathbf{A}\mathbf{x}) \rangle$ be the expectation of $(\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{B} (\mathbf{y} - \mathbf{A}\mathbf{x})$ w.r.t. $q_{\mathbf{A}, \mathbf{y}, \mathbf{x}}(\mathbf{A}, \mathbf{y}, \mathbf{x})$.

Then

$$\begin{aligned} & \langle (\mathbf{y} - \mathbf{Ax})^H \mathbf{B}(\mathbf{y} - \mathbf{Ax}) \rangle \\ &= (\langle \mathbf{y} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle)^H \mathbf{B}(\langle \mathbf{y} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle) + \text{Tr}\{\mathbf{B}\Sigma_{\mathbf{y}}\} \\ & \quad + \langle \mathbf{x} \rangle^H \mathbf{D} \langle \mathbf{x} \rangle + \text{Tr}\{\Sigma_{\mathbf{x}} \mathbf{D}\} + \text{Tr}\{\Sigma_{\mathbf{x}} \langle \mathbf{A}^H \rangle \mathbf{B} \langle \mathbf{A} \rangle\}, \end{aligned} \quad (7)$$

where $\mathbf{D} = \text{diag}(\text{Tr}\{\mathbf{B}\Sigma_{\mathbf{a}_1}\}, \dots, \text{Tr}\{\mathbf{B}\Sigma_{\mathbf{a}_n}\})$.

Proof: The proof of this lemma is straightforward and similar to the proof of Lemma 1 in [29], except that $\mathbf{y}$ is now a random vector. We omit the proof for brevity. ■

We note that if any of $\mathbf{A}$, $\mathbf{y}$, and $\mathbf{x}$ is deterministic, the corresponding covariance matrices $\{\Sigma_{\mathbf{a}_i}\}$, $\Sigma_{\mathbf{y}}$, and $\Sigma_{\mathbf{x}}$ will be set to $\mathbf{0}$ and the expectation of $(\mathbf{y} - \mathbf{Ax})^H \mathbf{B}(\mathbf{y} - \mathbf{Ax})$ given in (7) can be simplified accordingly.

IV. VB FOR DATA DETECTION IN FEW-BIT MIMO SYSTEMS WITH CSIR

The VB-based methods proposed in [26] assume prior information about the noise variance N_0 . However, by fixing the noise variance to N_0 , the conventional VB methods in [26] are unable to properly account for the residual inter-user interference, and therefore the obtained variational distribution will in general deviate significantly from the true posterior [29]. Here, we float the noise variance/covariance as an unknown parameter that is postulated by the VB estimation framework. To this end, we develop two new VB-based algorithms for solving the data detection problem in few-bit MIMO systems with known channel $\mathbf{H}$. For ease of presentation, we drop the subscript d, t indicating the data transmission at time index t .

A. Proposed MF-QVB For Few-Bit MIMO Detection

In this algorithm, we consider the residual interference-plus-noise as an unknown parameter N_0^{post} that is postulated by the VB estimation framework. For ease of computation, we use $\gamma = 1/N_0^{\text{post}}$ to denote the precision to be estimated.

The joint distribution $p(\mathbf{y}, \mathbf{r}, \mathbf{x}; \gamma, \mathbf{H})$ of the observed variable $\mathbf{y}$ and the latent variables $\mathbf{r}$ and $\mathbf{x}$ can be factored as

$$\begin{aligned} p(\mathbf{y}, \mathbf{r}, \mathbf{x}; \gamma, \mathbf{H}) &= p(\mathbf{y}|\mathbf{r})p(\mathbf{r}|\mathbf{x}; \gamma, \mathbf{H})p(\mathbf{x}) \\ &= \left[\prod_{m=1}^M p(y_m|r_m) \right] p(\mathbf{r}|\mathbf{x}; \gamma, \mathbf{H}) \left[\prod_{i=1}^K p(x_i) \right], \end{aligned} \quad (8)$$

where $p(y_m|r_m) = \mathbf{1}(r_m \in [y_m^{\text{low}}, y_m^{\text{up}}])$ and $p(\mathbf{r}|\mathbf{x}; \gamma, \mathbf{H}) = \mathcal{CN}(\mathbf{r}; \mathbf{H}\mathbf{x}, \gamma^{-1}\mathbf{I}_M)$. We note that the random vector $\mathbf{r}$ is comprised of conditional independent elements due to the same noise variance being imposed on the M receive antennas.

In the **E-step**, for a currently fixed estimate $\hat{\gamma}$ of γ , we aim to derive the mean field variational distribution $q(\mathbf{r}, \mathbf{x})$ of $\mathbf{r}$ and $\mathbf{x}$ given $\mathbf{y}$ such that

$$p(\mathbf{r}, \mathbf{x}|\mathbf{y}; \hat{\gamma}, \mathbf{H}) \approx q(\mathbf{r}, \mathbf{x}) = q(\mathbf{r}) \left[\prod_{i=1}^K q(x_i) \right]. \quad (9)$$

1) *Updating r.* The variational distribution $q(\mathbf{r})$ is obtained

Algorithm 1 – MF-QVB for Few-Bit MIMO Detection

```

1: Input:  $\mathbf{y}, \mathbf{H}$ 
2: Output:  $\hat{\mathbf{x}}$ 
3: Initialize  $\hat{x}_i^1 = 0$  and  $\tau_{x_i}^1 = \text{Var}_{p(x_i)}[x_i], \forall i, \hat{r}_m^1 = y_m$  and
    $\tau_{r_m}^1 = 0, \forall m, \mathbf{e} = \hat{\mathbf{r}}^1 - \mathbf{H}\hat{\mathbf{x}}^1$ , and compute  $g_i = \|\mathbf{h}_i\|^2$ 
4: for  $\ell = 1, 2, \dots$  do
5:    $\hat{\gamma}^\ell \leftarrow M / (\|\mathbf{e}\|^2 + \sum_{m=1}^M \tau_{r_m}^\ell + \sum_{i=1}^M g_i \tau_{x_i}^\ell)$ 
6:    $\mathbf{s}^\ell \leftarrow \hat{\mathbf{r}}^\ell - \mathbf{e}$ 
7:    $\hat{\mathbf{r}}^{\ell+1} \leftarrow \mathbf{F}_r(\mathbf{s}^\ell, \hat{\gamma}^\ell, \mathbf{y}^{\text{low}}, \mathbf{y}^{\text{up}})$  ▷ Update  $\mathbf{r}$ 
8:    $\tau_{\mathbf{r}}^{\ell+1} \leftarrow \mathbf{G}_r(\mathbf{s}^\ell, \hat{\gamma}^\ell, \mathbf{y}^{\text{low}}, \mathbf{y}^{\text{up}})$ 
9:    $\mathbf{e} \leftarrow \mathbf{e} - \hat{\mathbf{r}}^\ell + \hat{\mathbf{r}}^{\ell+1}$  ▷ Update residual
10:  for  $i = 1, \dots, K$  do ▷ Update  $\mathbf{x}$ 
11:     $z_i^\ell \leftarrow \hat{x}_i^\ell + \mathbf{h}_i^H \mathbf{e} / g_i$  ▷ Matched filter
12:     $\hat{x}_i^{\ell+1} \leftarrow \mathbf{F}_x(z_i^\ell, g_i \hat{\gamma}^\ell)$ 
13:     $\tau_{x_i}^{\ell+1} \leftarrow \mathbf{G}_x(z_i^\ell, g_i \hat{\gamma}^\ell)$ 
14:     $\mathbf{e} \leftarrow \mathbf{e} + \mathbf{h}_i(\hat{x}_i^\ell - \hat{x}_i^{\ell+1})$  ▷ Update residual
15:  $\forall i : \hat{x}_i \leftarrow \arg \max_{a \in \mathcal{S}} p_a \mathcal{CN}(z_i^\ell; a, 1/(g_i \hat{\gamma}^\ell))$ 

```

by taking the expectation of the conditional in (8) w.r.t. $q(\mathbf{x})$:

$$\begin{aligned} q(\mathbf{r}) &\propto \exp\left\{ \langle \ln p(\mathbf{y}|\mathbf{r}) + \ln p(\mathbf{r}|\mathbf{x}; \hat{\gamma}, \mathbf{H}) \rangle_{-\mathbf{r}} \right\} \\ &\propto \exp\left\{ \langle \ln \mathbf{1}(\mathbf{r} \in [\mathbf{y}^{\text{low}}, \mathbf{y}^{\text{up}}]) - \hat{\gamma} \|\mathbf{r} - \mathbf{H}\langle \mathbf{x} \rangle\|^2 \rangle_{-\mathbf{r}} \right\} \\ &\propto \mathbf{1}(\mathbf{r} \in [\mathbf{y}^{\text{low}}, \mathbf{y}^{\text{up}}]) \times \exp\{-\hat{\gamma} \|\mathbf{r} - \mathbf{H}\langle \mathbf{x} \rangle\|^2\} \\ &\propto \mathbf{1}(\mathbf{r} \in [\mathbf{y}^{\text{low}}, \mathbf{y}^{\text{up}}]) \times \mathcal{CN}(\mathbf{r}; \mathbf{H}\langle \mathbf{x} \rangle, \hat{\gamma}^{-1}\mathbf{I}_M). \end{aligned} \quad (10)$$

We note that variational distribution $q(\mathbf{r})$ is inherently separable as $\prod_{m=1}^M q(r_m)$ without enforcing the mean field approximation on $q(\mathbf{r})$. Thus, the variational mean and variance can be obtained concurrently for all the elements of $\mathbf{r}$. We see in (10) that $q(r_m)$ is the truncated complex normal distribution obtained from bounding $r_m \sim \mathcal{CN}(s_m, \hat{\gamma}^{-1})$, where $s_m = \mathbf{H}_{m,:} \langle \mathbf{x} \rangle$, to the interval $(y_m^{\text{low}}, y_m^{\text{up}})$. Thus, its mean $\langle r_m \rangle$ and variance τ_{r_m} are given by $\mathbf{F}_r(s_m, \hat{\gamma}, y_m^{\text{low}}, y_m^{\text{up}})$ and $\mathbf{G}_r(s_m, \hat{\gamma}, y_m^{\text{low}}, y_m^{\text{up}})$, respectively.¹ Here, s_m can be considered as a linear estimate of r_m and $\langle r_m \rangle$ can then be considered as a *denoised* version of s_m through the variational distribution $q(r_m)$.

2) *Updating x_i .* The variational distribution $q(x_i)$ is obtained by taking the expectation of the conditional in (8) w.r.t. $q(\mathbf{r}) \prod_{j \neq i} q(x_j)$:

$$\begin{aligned} q(x_i) &\propto \exp\left\{ \langle \ln p(\mathbf{r}|\mathbf{x}; \hat{\gamma}, \mathbf{H}) + \ln p(x_i) \rangle_{-\mathbf{r}} \right\} \\ &\propto p(x_i) \exp\left\{ -\hat{\gamma} \langle \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \rangle_{-\mathbf{r}} \right\} \\ &\propto p(x_i) \exp\left\{ -\hat{\gamma} \left[\|\mathbf{h}_i\|^2 |x_i|^2 \right. \right. \\ & \quad \left. \left. - 2 \Re \left\{ \mathbf{h}_i^H \left(\langle \mathbf{r} \rangle - \sum_{j \neq i}^K \mathbf{h}_j \langle x_j \rangle \right) x_i^* \right\} \right] \right\} \\ &\propto p(x_i) \exp\left\{ -\hat{\gamma} \|\mathbf{h}_i\|^2 (|x_i|^2 - 2 \Re\{x_i^* z_i\}) \right\} \\ &\propto p(x_i) \exp\left\{ -\hat{\gamma} \|\mathbf{h}_i\|^2 |x_i - z_i|^2 \right\} \\ &\propto p(x_i) \mathcal{CN}(z_i; x_i, 1/(\hat{\gamma} \|\mathbf{h}_i\|^2)), \end{aligned} \quad (11)$$

¹The computations of the mean $\mathbf{F}_r(\mu, \gamma, a, b)$ and variance $\mathbf{G}_r(\mu, \gamma, a, b)$ of an arbitrary complex normal distribution $\mathcal{CN}(\mu, \gamma^{-1})$ truncated to an interval (a, b) are presented in Appendix A.

where we define

$$\begin{aligned} z_i &= \frac{\mathbf{h}_i^H}{\|\mathbf{h}_i\|^2} \left(\langle \mathbf{r} \rangle - \sum_{j \neq i}^K \mathbf{h}_j \langle x_j \rangle \right) \\ &= \langle x_i \rangle + \frac{\mathbf{h}_i^H}{\|\mathbf{h}_i\|^2} (\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle) \end{aligned} \quad (12)$$

with $\langle x_i \rangle$ being the *currently fixed* nonlinear estimate of $x_i, \forall i$. We can see in (11) that the mean field VB approximation decouples the few-bit MIMO system into an AWGN channel $z_i = x_i + \mathcal{CN}(0, 1/(\langle \gamma \rangle \|\mathbf{h}_i\|^2))$ for user- i . The variational distribution $q(x_i)$ can be realized by normalizing $p(x_i) \mathcal{CN}(z_i; x_i, 1/(\langle \gamma \rangle \|\mathbf{h}_i\|^2))$. The variational mean $\langle x_i \rangle$ and variance τ_{x_i} are now updated as $F_x(z_i, \hat{\gamma} \|\mathbf{h}_i\|^2)$ and $G_x(z_i, \hat{\gamma} \|\mathbf{h}_i\|^2)$, respectively.²

In the **M-step**, the estimate of γ is updated to maximize $\ln p(\mathbf{y}, \mathbf{r}, \mathbf{x}; \gamma, \mathbf{H})$ w.r.t. $q(\mathbf{r}, \mathbf{x})$, i.e.,

$$\begin{aligned} \hat{\gamma} &= \arg \max_{\gamma} \langle \ln p(\mathbf{r} | \mathbf{x}; \gamma; \mathbf{H}) \rangle \\ &= \arg \max_{\gamma} M \ln \gamma - \gamma \langle \|\mathbf{r} - \mathbf{H} \mathbf{x}\|^2 \rangle. \end{aligned} \quad (13)$$

Applying Lemma 1 to evaluate the expectation $\langle \|\mathbf{r} - \mathbf{H} \mathbf{x}\|^2 \rangle$, the new estimate of γ is given by

$$\hat{\gamma} = \frac{M}{\langle \|\mathbf{r} - \mathbf{H} \langle \mathbf{x} \rangle\|^2 \rangle + \sum_{m=1}^M \tau_{r_m} + \sum_{i=1}^K \tau_{x_i} \|\mathbf{h}_i\|^2}. \quad (14)$$

By iteratively optimizing $q(\mathbf{r})$, $\{q(x_i)\}$, and updating $\hat{\gamma}$, we obtain the variational Bayes expectation-maximization (VBEM) algorithm for estimating $\mathbf{r}$, $\mathbf{x}$, and γ . Similar to our previous work [29], we refer to this scheme as the **MF-QVB algorithm** due to the use of the matched-filter $\mathbf{h}_i^H / \|\mathbf{h}_i\|^2$ to obtain the linear estimate z_i of x_i in (12). Then, $\langle x_i \rangle$ can be considered as a *denoised* estimate of z_i through the variational distribution $q(x_i)$. If γ is fixed to N_0^{-1} , the **MF-QVB algorithm** will be referred to as the **conv-QVB algorithm**, that was referred to as the QVB-CSIR algorithm in [26]. In MF-QVB, the estimate $\hat{\gamma}$ in (14) factors in the effect on the background noise, i.e., $\langle \|\mathbf{r} - \mathbf{H} \langle \mathbf{x} \rangle\|^2 \rangle$, as well as the errors of estimating $\{r_m\}$ and $\{x_i\}$, i.e., $\{\tau_{r_m}\}$ and $\{\tau_{x_i}\}$. Accounting for these estimation error terms in MF-QVB will result in a variational distribution $q(x_i)$ that provides a closer approximation to the true posterior $p(x_i | \mathbf{y})$, when compared to conv-QVB.

The MF-QVB approach is summarized in Algorithm 1. Here, we use $\hat{\mathbf{r}}^\ell$, $\hat{x}_i^\ell$, and $\hat{\gamma}^\ell$ to replace $\langle \mathbf{r} \rangle$, $\langle x_i \rangle$, and $\hat{\gamma}$ at iteration ℓ and each iteration consists of one round of updating the estimation of $\mathbf{r}$, $\mathbf{x}$, and γ . To reduce the complexity of the algorithm, we include the residual term $\mathbf{e}$, which is initialized as $\hat{\mathbf{r}}^1 - \mathbf{H} \hat{\mathbf{x}}^1$, and we pre-compute $g_i = \|\mathbf{h}_i\|^2$. We also define $\mathbf{s}^\ell = \hat{\mathbf{r}}^\ell - \mathbf{e}$ as an efficient way to compute $\mathbf{H} \hat{\mathbf{x}}^\ell$ at iteration- ℓ , which requires a complexity of only $\mathcal{O}(M)$ instead of $\mathcal{O}(MK)$. The computation of $\mathbf{r}^{\ell+1}$ (and $\tau_{\mathbf{r}}^{\ell+1}$) can be carried out element-wise in parallel. The complexity of updating $\mathbf{e}$ in step 9 of the algorithm with the latest value $\mathbf{r}^{\ell+1}$ is also $\mathcal{O}(M)$. In addition, computing z_i^ℓ in step 11 and updating $\mathbf{e}$ in step 14 of the algorithm only require a per-user-complexity of $\mathcal{O}(M)$.

²The computations of the mean $F_x(z, \gamma)$ and variance $G_x(z, \gamma)$ of a discrete random variable x given a prior distribution $p(x)$ and the observation $z = x + \mathcal{CN}(0, \gamma^{-1})$ are presented in Appendix B.

Thus, by properly using and updating $\mathbf{e}$, we can reduce the complexity from $\mathcal{O}(MK)$ to $\mathcal{O}(M)$ of computing $\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle$ from the latest update of $\mathbf{r}$ and $\mathbf{x}$.

B. Proposed LMMSE-QVB For Few-Bit MIMO Detection

While MF-QVB relies on a latent random variable to capture the second-order statistics of the noise and residual inter-user interference, the LMMSE-QVB developed in this section utilizes a latent matrix variable for this purpose. To account for the spatial correlation in the residual inter-user interference $\sum_{i=1}^K \mathbf{h}_i(x_i - \hat{x}_i)$, especially in scenarios with correlated channels or few users, we propose using a postulated noise plus inter-user interference covariance matrix $\mathbf{C}^{\text{post}}$. By adding more degrees of freedom through $\mathbf{C}^{\text{post}}$, we can get a variational distribution $q(x_i)$ that better matches the real posterior distribution $p(x_i | \mathbf{y})$. This helps to reduce symbol estimation errors compared to the MF-QVB method. For ease of computation, we use $\mathbf{\Gamma} = (\mathbf{C}^{\text{post}})^{-1}$ as the precision matrix for the noise and residual inter-user interference to be estimated.

The joint distribution $p(\mathbf{y}, \mathbf{r}, \mathbf{x}; \mathbf{\Gamma}, \mathbf{H})$ of the observed variable $\mathbf{y}$ and the latent variables $\mathbf{r}$ and $\mathbf{x}$ at time slot t can be factored as

$$\begin{aligned} p(\mathbf{y}, \mathbf{r}, \mathbf{x}; \mathbf{\Gamma}, \mathbf{H}) &= p(\mathbf{y} | \mathbf{r}) p(\mathbf{r} | \mathbf{x}; \mathbf{\Gamma}, \mathbf{H}) p(\mathbf{x}) \\ &= \left[\prod_{m=1}^M p(y_m | r_m) \right] p(\mathbf{r} | \mathbf{x}; \mathbf{\Gamma}, \mathbf{H}) \left[\prod_{i=1}^K p(x_i) \right], \end{aligned} \quad (15)$$

where $p(\mathbf{r} | \mathbf{x}; \mathbf{\Gamma}, \mathbf{H}) = \mathcal{CN}(\mathbf{r}; \mathbf{H} \mathbf{x}, \mathbf{\Gamma}^{-1})$. We note that the random vector $\mathbf{r}$ is no longer comprised of conditional independent elements since the noise covariance matrix $\mathbf{\Gamma}^{-1}$ is in general non-diagonal.

In the **E-step**, for a currently fixed estimate $\hat{\mathbf{\Gamma}}$ of $\mathbf{\Gamma}$, we aim to derive the mean field variational distribution $q(\mathbf{r}, \mathbf{x})$ of $\mathbf{r}$ and $\mathbf{x}$ given $\mathbf{y}$ such that

$$p(\mathbf{r}, \mathbf{x} | \mathbf{y}; \hat{\mathbf{\Gamma}}, \mathbf{H}) \approx q(\mathbf{r}, \mathbf{x}) = \left[\prod_{m=1}^M q(r_m) \right] \left[\prod_{i=1}^K q(x_i) \right]. \quad (16)$$

1) *Updating r_m* . The variational distribution $q(r_m)$ is obtained by taking the expectation of the conditional in (15) w.r.t. $q(\mathbf{x}) \prod_{n \neq m} q(r_n)$:

$$\begin{aligned} q(r_m) &\propto \exp \left\{ \langle \ln p(y_m | r_m) + \ln p(\mathbf{r} | \mathbf{x}; \hat{\mathbf{\Gamma}}, \mathbf{H}) \rangle_{-r_m} \right\} \\ &\propto \exp \left\{ \ln \mathbb{1}(r_m \in [y_m^{\text{low}}, y_m^{\text{up}}]) \right. \\ &\quad \left. - \langle (\mathbf{r} - \mathbf{H} \mathbf{x})^H \hat{\mathbf{\Gamma}} (\mathbf{r} - \mathbf{H} \mathbf{x}) \rangle_{-r_m} \right\} \\ &\propto \mathbb{1}(r_m \in [y_m^{\text{low}}, y_m^{\text{up}}]) \\ &\quad \times \exp \left\{ - [\hat{\mathbf{\Gamma}}]_{mm} (|r_m|^2 - 2 \Re\{r_m^* \mathbf{H}_{m,:} \langle \mathbf{x} \rangle\})^2 \right. \\ &\quad \left. - 2 \sum_{n \neq m}^M \Re\{r_m^* [\hat{\mathbf{\Gamma}}]_{mn} (\langle r_n \rangle - \mathbf{H}_{n,:} \langle \mathbf{x} \rangle)\} \right\} \\ &\propto \mathbb{1}(r_m \in [y_m^{\text{low}}, y_m^{\text{up}}]) \times \exp \left\{ - [\hat{\mathbf{\Gamma}}]_{mm} |r_m - s_m|^2 \right\} \\ &\propto \mathbb{1}(r_m \in [y_m^{\text{low}}, y_m^{\text{up}}]) \times \mathcal{CN}(r_m; s_m, [\hat{\mathbf{\Gamma}}]_{mm}^{-1}), \end{aligned} \quad (17)$$

where s_m is now defined as

$$\begin{aligned} s_m &= \mathbf{H}_{m,:} \langle \mathbf{x} \rangle - [\hat{\Gamma}]_{mm}^{-1} \sum_{n \neq m}^M [\hat{\Gamma}]_{mn} (\langle r_n \rangle - \mathbf{H}_{n,:} \langle \mathbf{x} \rangle) \\ &= \langle r_m \rangle - \frac{[\hat{\Gamma}]_{m,:}}{[\hat{\Gamma}]_{mm}} (\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle), \end{aligned} \quad (18)$$

$\langle r_m \rangle$ is the *currently fixed* nonlinear estimate of r_m . We can see in (17) that the variational distribution $q(r_m)$ is the truncated complex normal distribution obtained from bounding $r_m \sim \mathcal{CN}(s_m, [\hat{\Gamma}]_{mm}^{-1})$ to the interval $(y_m^{\text{low}}, y_m^{\text{up}})$. Thus, its mean $\langle r_m \rangle$ and variance τ_{r_m} are updated as $F_r(s_m, [\hat{\Gamma}]_{mm}, y_m^{\text{low}}, y_m^{\text{up}})$ and $G_r(s_m, [\hat{\Gamma}]_{mm}, y_m^{\text{low}}, y_m^{\text{up}})$, respectively.

2) *Updating x_i* . The variational distribution $q(x_i)$ is obtained by taking the expectation of the conditional in (8) w.r.t. $q(\mathbf{r}) \prod_{j \neq i} q(x_j)$:

$$\begin{aligned} q(x_i) &\propto \exp \left\{ \langle \ln p(\mathbf{r} | \mathbf{x}; \hat{\Gamma}, \mathbf{H}) + \ln p(x_i) \rangle_{-x_i} \right\} \\ &\propto p(x_i) \exp \left\{ - \langle (\mathbf{r} - \mathbf{H}\mathbf{x})^H \hat{\Gamma} (\mathbf{r} - \mathbf{H}\mathbf{x}) \rangle_{-x_i} \right\} \\ &\propto p(x_i) \exp \left\{ - \mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i |x_i - z_i|^2 \right\} \\ &\propto p(x_i) \mathcal{CN}(z_i; x_i, 1/(\mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i)), \end{aligned} \quad (19)$$

where z_i is a linear estimate of x_i that is now defined as

$$\begin{aligned} z_i &= \frac{\mathbf{h}_i^H \hat{\Gamma}}{\mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i} \left(\langle \mathbf{r} \rangle - \sum_{j \neq i}^K \mathbf{h}_j \langle x_j \rangle \right) \\ &= \langle x_i \rangle + \frac{\mathbf{h}_i^H \hat{\Gamma}}{\mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i} (\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle), \end{aligned} \quad (20)$$

and $\langle x_i \rangle$ is the *current* nonlinear estimate of x_i . Here, z_i is the LMMSE estimate of x_i using the LMMSE filter $\mathbf{h}_i^H \hat{\Gamma} / (\mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i)$. The variational distribution $q(x_i)$ can be realized by normalizing $p(x_i) \mathcal{CN}(z_i; x_i, 1/(\mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i))$. The variational mean $\langle x_i \rangle$ and variance τ_{x_i} are updated as $F_x(z_i, \mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i)$ and $G_x(z_i, \mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i)$, respectively.

In the **M-step**, the estimate of Γ is updated to maximize $\ln p(\mathbf{y}, \mathbf{r}, \mathbf{x}; \Gamma, \mathbf{H})$ w.r.t. $q(\mathbf{r}, \mathbf{x})$, i.e.,

$$\begin{aligned} \hat{\Gamma} &= \arg \max_{\Gamma} \langle \ln p(\mathbf{r} | \mathbf{x}; \Gamma, \mathbf{H}) \rangle \\ &= \arg \max_{\Gamma} \ln |\Gamma| - \langle (\mathbf{r} - \mathbf{H}\mathbf{x})^H \Gamma (\mathbf{r} - \mathbf{H}\mathbf{x}) \rangle. \end{aligned} \quad (21)$$

By applying Lemma 1, we have

$$\begin{aligned} &\langle (\mathbf{r} - \mathbf{H}\mathbf{x})^H \Gamma (\mathbf{r} - \mathbf{H}\mathbf{x}) \rangle \\ &= \text{Tr} \left\{ \left[(\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle) (\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle)^H + \Sigma_{\mathbf{r}} + \mathbf{H} \Sigma_{\mathbf{x}} \mathbf{H}^H \right] \Gamma \right\}, \end{aligned} \quad (22)$$

where $\Sigma_{\mathbf{r}} = \text{diag}(\tau_{r_1}, \dots, \tau_{r_M})$ and $\Sigma_{\mathbf{x}} = \text{diag}(\tau_{x_1}, \dots, \tau_{x_K})$. Thus, a new estimate of Γ is given by

$$\hat{\Gamma} = \left((\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle) (\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle)^H + \Sigma_{\mathbf{r}} + \mathbf{H} \Sigma_{\mathbf{x}} \mathbf{H}^H \right)^{-1}. \quad (23)$$

We note that the matrix inversion in (23) often results in numerical errors due to the rank deficiency of $(\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle) (\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle)^H + \Sigma_{\mathbf{r}} + \mathbf{H} \Sigma_{\mathbf{x}} \mathbf{H}^H$. Similar to the approach in [29],

Algorithm 2 – LMMSE-QVB for Few-Bit MIMO Detection

```

1: Input:  $\mathbf{y}, \mathbf{H}$ 
2: Output:  $\hat{\mathbf{x}}$ 
3: Initialize  $\hat{x}_i^1 = 0$  and  $\tau_{x_i}^1 = \text{Var}_{p(x_i)}[x_i], \forall i$ ,  $\hat{r}_m^1 = y_m$  and  $\tau_{r_m}^1 = 0, \forall m$ , and  $\mathbf{e} = \hat{\mathbf{r}}^1 - \mathbf{H} \hat{\mathbf{x}}^1$ 
4: for  $\ell = 1, 2, \dots$  do
5:    $\Sigma_{\mathbf{r}} \leftarrow \text{diag}(\tau_{r_1}^\ell, \dots, \tau_{r_M}^\ell)$ 
6:    $\Sigma_{\mathbf{x}} \leftarrow \text{diag}(\tau_{x_1}^\ell, \dots, \tau_{x_K}^\ell)$ 
7:    $\hat{\Gamma}^\ell \leftarrow ((\|\mathbf{e}\|^2/M) \mathbf{I}_M + \Sigma_{\mathbf{r}} + \mathbf{H} \Sigma_{\mathbf{x}} \mathbf{H}^H)^{-1}$ 
8:   for  $m = 1, \dots, M$  do ▷ Update  $\mathbf{r}$ 
9:      $s_m^\ell \leftarrow \hat{r}_m^\ell - [\hat{\Gamma}^\ell]_{m,:} \mathbf{e} / [\hat{\Gamma}^\ell]_{mm}$ 
10:     $\hat{r}_m^{\ell+1} \leftarrow F_r(s_m^\ell, [\hat{\Gamma}^\ell]_{mm}, y_m^{\text{low}}, y_m^{\text{up}})$ 
11:     $\tau_{r_m}^{\ell+1} \leftarrow G_r(s_m^\ell, [\hat{\Gamma}^\ell]_{mm}, y_m^{\text{low}}, y_m^{\text{up}})$ 
12:     $e_m \leftarrow e_m - \hat{r}_m^\ell + \hat{r}_m^{\ell+1}$  ▷ Update residual
13:   for  $i = 1, \dots, K$  do ▷ Update  $\mathbf{x}$ 
14:      $z_i^\ell \leftarrow \hat{x}_i^\ell + \mathbf{h}_i^H \hat{\Gamma}^\ell \mathbf{e} / (\mathbf{h}_i^H \hat{\Gamma}^\ell \mathbf{h}_i)$  ▷ LMMSE
15:      $\hat{x}_i^{\ell+1} \leftarrow F_x(z_i^\ell, \mathbf{h}_i^H \hat{\Gamma}^\ell \mathbf{h}_i)$ 
16:      $\tau_{x_i}^{\ell+1} \leftarrow G_x(z_i^\ell, \mathbf{h}_i^H \hat{\Gamma}^\ell \mathbf{h}_i)$ 
17:      $\mathbf{e} \leftarrow \mathbf{e} + \mathbf{h}_i (\hat{x}_i^\ell - \hat{x}_i^{\ell+1})$  ▷ Update residual
18:  $\forall i: \hat{x}_i \leftarrow \arg \max_{a \in \mathcal{S}} p_a \mathcal{CN}(z_i^\ell; a, 1/(\mathbf{h}_i^H \hat{\Gamma}^\ell \mathbf{h}_i)).$ 

```

we propose to use the following estimator

$$\hat{\Gamma} = \left(\frac{\|\langle \mathbf{r} \rangle - \mathbf{H} \langle \mathbf{x} \rangle\|^2}{M} \mathbf{I}_M + \Sigma_{\mathbf{r}} + \mathbf{H} \Sigma_{\mathbf{x}} \mathbf{H}^H \right)^{-1}. \quad (24)$$

for the precision matrix Γ .

By iteratively optimizing $\{q(r_m)\}$, $\{q(x_i)\}$, and $\hat{\Gamma}$, we obtain the VBEM algorithm for estimating $\mathbf{r}$, $\mathbf{x}$, and Γ . We refer to this scheme as the **LMMSE-QVB algorithm** due to the use of the LMMSE filter $\mathbf{h}_i^H \hat{\Gamma} / (\mathbf{h}_i^H \hat{\Gamma} \mathbf{h}_i)$ to obtain the linear estimate z_i of x_i in (20). The LMMSE-QVB approach is summarized in Algorithm 2. As before, we use $\hat{r}_m^\ell$, $\hat{x}_i^\ell$, and $\hat{\Gamma}^\ell$ to replace $\langle r_m \rangle$, $\langle x_i \rangle$, and $\hat{\Gamma}$ at iteration ℓ , and each iteration consists of one update of the estimates for $\mathbf{r}$, $\mathbf{x}$, and Γ . Unlike MF-QVB, the LMMSE-QVB algorithm requires sequential updates over $\{\hat{r}_m^\ell\}$. Note that LMMSE-QVB is equivalent to MF-QVB in Algorithm 1 if $\hat{\Gamma}^\ell = \hat{\gamma}^\ell \mathbf{I}_M$.

C. Practical Aspects of Implementing MF/LMMSE-QVB

1) *Computing $\langle r_m \rangle$ and τ_{r_m}* : In MF-QVB, $\langle r_m \rangle$ and τ_{r_m} are updated as $F_r(s_m, \hat{\gamma}, y_m^{\text{low}}, y_m^{\text{up}})$ and $G_r(s_m, \hat{\gamma}, y_m^{\text{low}}, y_m^{\text{up}})$, respectively. The computation of these two terms, presented in Appendix A, can result in catastrophic cancellation when $\Phi(\beta) = \Phi(\alpha)$, even if $\beta = \sqrt{2\hat{\gamma}}(y_m^{\text{low}} - s_m)$ and $\alpha = \sqrt{2\hat{\gamma}}(y_m^{\text{up}} - s_m)$ are different. This numerical error often occurs when s_m is outside the interval $(y_m^{\text{low}}, y_m^{\text{up}})$ and $\langle \gamma \rangle$ is large, i.e., for high SNR. To improve the robustness of the computation, we use the logistic CDF $F(x; c) = \frac{1}{1+e^{-cx}}$ and logistic PDF $p(x; c) = cF(x; c)(1 - F(x; c))$ in place of the normal CDF $\Phi(x)$ and normal PDF $\phi(x)$. We choose $c = 3/\sqrt{\pi}$ to impose a unit variance on the logistic distribution. Note that $F(x)$ is much easier to compute than $\Phi(x)$. We have observed through numerous simulations that this modification eliminates numerical errors since the tail of the logistic distribution is heavier than that of the normal distribution. Interestingly, the

detection accuracy is also slightly better when using $F(x; c)$ and $p(x; c)$ rather than $\Phi(x)$ and $\phi(x)$, even when no numerical errors occur using the latter approach. We also use $F(x; c)$ and $p(x; c)$ in LMMSE-QVB and observe the same effect.

2) *Using e*: The residual term $\mathbf{e}$ is included in MF/LMMSE-QVB to reduce the computational complexity of these algorithms. Due to the sequential nature of VB, s_m and z_i are computed using the latest updated values of $\langle \mathbf{r} \rangle$ and $\langle \mathbf{x} \rangle$. Instead of computing $\langle \mathbf{r} \rangle - \mathbf{H}\langle \mathbf{x} \rangle$ each time before computing s_m or z_i , which induces a complexity of $\mathcal{O}(MK)$, we use the current value of the residual term $\mathbf{e}$. The residual term $\mathbf{e}$ is then recomputed to reflect any update on the estimation of r_m or x_i only induces a complexity of $\mathcal{O}(M)$.

V. VB FOR JOINT CHANNEL ESTIMATION AND DATA DETECTION IN FEW-BIT MIMO SYSTEMS

The algorithms in the previous section assumed that the CSI was already obtained prior to data detection. In this section, we generalize the MF-QVB and LMMSE-QVB approaches to perform joint channel estimation and data detection.

A. Proposed MF-QVB-JED Algorithm

We denote γ_p and $\gamma_d = [\gamma_{d,1}, \dots, \gamma_{d,T_d}]^T$ as the precision of the noise in the pilot transmission phase and the T_d data transmission time slots, respectively. Denote $\{\mathbf{C}_i\}$ as the set of covariance matrices $\mathbf{C}_1, \dots, \mathbf{C}_K$. The factorization of the joint distribution of all the observed and latent variables in the system model (1)–(2) is given in (25), where $p(\mathbf{r}_{p,t}|\mathbf{H}; \gamma_p, \mathbf{x}_{p,t}) = \mathcal{CN}(\mathbf{r}_{p,t}; \mathbf{H}\mathbf{x}_{p,t}, \gamma_p^{-1}\mathbf{I}_M)$ and $p(\mathbf{r}_{d,t}|\mathbf{H}, \mathbf{x}_{d,t}; \gamma_d, t) = \mathcal{CN}(\mathbf{r}_{d,t}; \mathbf{H}\mathbf{x}_{d,t}, \gamma_{d,t}^{-1}\mathbf{I}_M)$.

In the **E-step**, for currently fixed estimates $\hat{\gamma}_p$ and $\hat{\gamma}_d$ of γ_p and γ_d , respectively, we obtain the mean field variational distribution $q(\mathbf{R}_p, \mathbf{R}_d, \mathbf{H}, \mathbf{X}_d)$ given $\mathbf{Y}_p$ and $\mathbf{Y}_d$ such that

$$\begin{aligned} p(\mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H}|\mathbf{Y}_p, \mathbf{Y}_d; \hat{\gamma}_p, \hat{\gamma}_d, \mathbf{X}_p, \{\mathbf{C}_i\}) \\ \approx q(\mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H}) \\ = \left[\prod_{t=1}^{T_p} q(\mathbf{r}_{p,t}) \right] \left[\prod_{t=1}^{T_d} q(\mathbf{r}_{d,t}) \right] \left[\prod_{i=1}^K q(\mathbf{h}_i) \right] \left[\prod_{i=1}^K \prod_{t=1}^{T_d} q(x_{d,ti}) \right]. \end{aligned} \quad (26)$$

1) *Updating $\mathbf{r}_{p,t}$* . Taking the expectation of the conditional (25) w.r.t. all latent variables except $\mathbf{r}_{p,t}$, the variational distribution $q(\mathbf{r}_{p,t})$ is given by

$$\begin{aligned} q(\mathbf{r}_{p,t}) \\ \propto \exp \left\{ \left\langle \ln p(\mathbf{y}_{p,t}|\mathbf{r}_{p,t}) + \ln p(\mathbf{r}_{p,t}|\mathbf{H}; \hat{\gamma}_p, \mathbf{x}_{p,t}) \right\rangle_{-\mathbf{r}_{p,t}} \right\} \\ \propto \exp \left\{ \left\langle \ln \mathbb{1}(\mathbf{r}_{p,t} \in [\mathbf{y}_{p,t}^{\text{low}}, \mathbf{y}_{p,t}^{\text{up}}]) - \right. \right. \\ \left. \left. \hat{\gamma}_p \|\mathbf{r}_{p,t} - \mathbf{H}\mathbf{x}_{p,t}\|^2 \right\rangle_{-\mathbf{r}_{p,t}} \right\} \\ \propto \mathbb{1}(\mathbf{r}_{p,t} \in [\mathbf{y}_{p,t}^{\text{low}}, \mathbf{y}_{p,t}^{\text{up}}]) \times \exp \left\{ -\hat{\gamma}_p \|\mathbf{r}_{p,t} - \langle \mathbf{H} \rangle \mathbf{x}_{p,t}\|^2 \right\}. \end{aligned} \quad (27)$$

We note that the variational distribution $q(\mathbf{r}_{p,t})$ is inherently separable as $\prod_{m=1}^M q(r_{p,tm})$ and the variational distribution $q(\mathbf{r}_{p,tm})$ is the complex normal distribution obtained from bounding $r_{p,tm} \sim \mathcal{CN}(\langle \mathbf{H}_{m,:} \rangle \mathbf{x}_{p,t}, \hat{\gamma}_p^{-1})$ to the interval $(y_{p,tm}^{\text{low}}, y_{p,tm}^{\text{up}})$. The variational mean $\langle r_{p,tm} \rangle$ and variance $\tau_{r_{p,tm}}$ are given by $F_r(\langle \mathbf{H}_{m,:} \rangle \mathbf{x}_{p,t}, \hat{\gamma}_p^{-1}, y_{p,tm}^{\text{low}}, y_{p,tm}^{\text{up}})$ and $G_r(\langle \mathbf{H}_{m,:} \rangle \mathbf{x}_{p,t}, \hat{\gamma}_p^{-1}, y_{p,tm}^{\text{low}}, y_{p,tm}^{\text{up}})$, respectively.

2) *Updating $\mathbf{r}_{d,t}$* . Taking the expectation of the conditional (25) w.r.t. all latent variables except $\mathbf{r}_{d,t}$, the variational distribution $q(\mathbf{r}_{p,t})$ is given by

$$\begin{aligned} q(\mathbf{r}_{d,t}) \\ \propto \exp \left\{ \left\langle \ln p(\mathbf{y}_{d,t}|\mathbf{r}_{d,t}) + \ln p(\mathbf{r}_{d,t}|\mathbf{H}, \mathbf{x}_{d,t}; \hat{\gamma}_{d,t}) \right\rangle_{-\mathbf{r}_{d,t}} \right\} \\ \propto \exp \left\{ \left\langle \ln \mathbb{1}(\mathbf{r}_{d,t} \in [\mathbf{y}_{d,t}^{\text{low}}, \mathbf{y}_{d,t}^{\text{up}}]) - \right. \right. \\ \left. \left. \hat{\gamma}_{d,t} \|\mathbf{r}_{d,t} - \mathbf{H}\mathbf{x}_{d,t}\|^2 \right\rangle_{-\mathbf{r}_{d,t}} \right\} \\ \propto \mathbb{1}(\mathbf{r}_{d,t} \in [\mathbf{y}_{d,t}^{\text{low}}, \mathbf{y}_{d,t}^{\text{up}}]) \times \exp \left\{ -\hat{\gamma}_{d,t} \|\mathbf{r}_{p,t} - \langle \mathbf{H} \rangle \mathbf{x}_{d,t}\|^2 \right\}. \end{aligned} \quad (28)$$

The update of $\mathbf{r}_{d,t}$ is similar to that of $\mathbf{r}_{p,t}$. Due to the inherent decoupling of $q(\mathbf{r}_{d,t})$, the variational mean $\langle \mathbf{r}_{d,tm} \rangle$ and variance $\tau_{r_{d,tm}}$ are given by $F_r(\langle \mathbf{H}_{m,:} \rangle \mathbf{x}_{d,t}, \hat{\gamma}_{d,t}, y_{d,tm}^{\text{low}}, y_{d,tm}^{\text{up}})$ and $G_r(\langle \mathbf{H}_{m,:} \rangle \mathbf{x}_{d,t}, \hat{\gamma}_{d,t}, y_{d,tm}^{\text{low}}, y_{d,tm}^{\text{up}})$, respectively.

3) *Updating $\mathbf{h}_i$* . Taking the expectation of the conditional (25) w.r.t. all latent variables except $\mathbf{h}_i$, the variational distribution $q(\mathbf{h}_i)$ is given by

$$\begin{aligned} q(\mathbf{h}_i) \propto \exp \left\{ \left\langle \ln p(\mathbf{R}_p|\mathbf{H}; \hat{\gamma}_p, \mathbf{X}_p) + \ln p(\mathbf{R}_d|\mathbf{X}_d, \mathbf{H}; \hat{\gamma}_d) \right. \right. \\ \left. \left. + \ln p(\mathbf{h}_i; \mathbf{C}_i) \right\rangle_{-\mathbf{h}_i} \right\}, \end{aligned} \quad (29)$$

which is expanded into (30). Thus, the variational distribution $q_{\mathbf{h}_i}(\mathbf{h}_i)$ is the pdf of a Gaussian random vector with covariance matrix $\Sigma_{\mathbf{h}_i}$ given in (31) and mean $\langle \mathbf{h}_i \rangle$ given in (32).

4) *Updating $x_{d,ti}$* . Taking the expectation of the conditional (25) w.r.t. all latent variables except $x_{d,ti}$, the variational distribution $q(x_{d,ti})$ is given by

$$\begin{aligned} q(x_{d,ti}) \propto \exp \left\{ \left\langle \ln p(\mathbf{r}_{d,t}|\mathbf{H}, \mathbf{x}_{d,t}; \hat{\gamma}_{d,t}) + \ln p(x_{d,ti}) \right\rangle_{-x_{d,ti}} \right\} \\ \propto p(x_{d,ti}) \exp \left\{ -\hat{\gamma}_{d,t} \langle \|\mathbf{r}_{d,t} - \mathbf{H}\mathbf{x}_{d,t}\|^2 \rangle_{-x_{d,ti}} \right\}. \end{aligned} \quad (33)$$

Note that (33) can be expanded into (34), in which we define

$$\begin{aligned} z_{d,ti} &= \frac{\langle \mathbf{h}_i^H \rangle}{\langle \|\mathbf{h}_i\|^2 \rangle} \left(\langle \mathbf{r}_{d,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle \langle x_{d,jt} \rangle \right) \\ &= \frac{\|\langle \mathbf{h}_i \rangle\|^2 \langle x_{d,ti} \rangle + \langle \mathbf{h}_i^H \rangle (\langle \mathbf{r}_{d,t} \rangle - \langle \mathbf{H} \rangle \langle \mathbf{x}_{d,t} \rangle)}{\langle \|\mathbf{h}_i\|^2 \rangle} \end{aligned} \quad (35)$$

as a linear estimate of $x_{d,ti}$. We note that $\langle \|\mathbf{h}_i\|^2 \rangle = \|\langle \mathbf{h}_i \rangle\|^2 + \text{Tr}\{\Sigma_{\mathbf{h}_i}\}$. The variational mean and variance of $x_{d,ti}$ are given by $F_x(z_{d,ti}, \hat{\gamma}_{d,t} \langle \|\mathbf{h}_i\|^2 \rangle)$ and $G_x(z_{d,ti}, \hat{\gamma}_{d,t} \langle \|\mathbf{h}_i\|^2 \rangle)$, respectively.

In the **M-step**, the estimates of γ_p and $\gamma_{d,t}$ are updated to maximize $\ln p(\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H}; \gamma_p, \gamma_d, \mathbf{X}_p, \{\mathbf{C}_i\})$ w.r.t. the variational distribution $q(\mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H})$, i.e.,

$$\begin{aligned} \hat{\gamma}_p &= \arg \max_{\gamma_p} \langle \ln p(\mathbf{R}_p|\mathbf{H}; \gamma_p, \mathbf{X}_p) \rangle \\ &= \arg \max_{\gamma_p} MT_p - \gamma_p \langle \|\mathbf{R}_p - \mathbf{H}\mathbf{X}_p\|^2 \rangle \\ &= \frac{MT_p}{\sum_{t=1}^{T_p} \langle \|\mathbf{r}_{p,t} - \mathbf{H}\mathbf{x}_{p,t}\|^2 \rangle}, \end{aligned} \quad (36)$$

$$\begin{aligned}
& p(\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H}; \gamma_p, \gamma_d, \mathbf{X}_p, \{\mathbf{C}_i\}) \\
&= p(\mathbf{Y}_p | \mathbf{R}_p) p(\mathbf{R}_p | \mathbf{H}; \gamma_p, \mathbf{X}_p) p(\mathbf{Y}_d | \mathbf{R}_d) p(\mathbf{R}_d | \mathbf{X}_d, \mathbf{H}; \gamma_d) p(\mathbf{X}_d) p(\mathbf{H} | \{\mathbf{C}_i\}) \\
&= \left[\prod_{t=1}^{T_p} p(\mathbf{y}_{p,t} | \mathbf{r}_{p,t}) p(\mathbf{r}_{p,t} | \mathbf{H}; \gamma_p, \mathbf{x}_{p,t}) \right] \left[\prod_{t=1}^{T_d} p(\mathbf{y}_{d,t} | \mathbf{r}_{d,t}) p(\mathbf{r}_{d,t} | \mathbf{H}, \mathbf{x}_{d,t}; \gamma_d, t) p(\mathbf{x}_{d,t}) \right] \left[\prod_{i=1}^K p(\mathbf{h}_i | \mathbf{C}_i) \right]. \quad (25)
\end{aligned}$$

$$\begin{aligned}
q(\mathbf{h}_i) &\propto \exp \left\{ - \left\langle \hat{\gamma}_p \sum_{t=1}^{T_p} \|\mathbf{r}_{p,t} - \mathbf{H} \mathbf{x}_{p,t}\|^2 + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \|\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t}\|^2 + \mathbf{h}_i^H \mathbf{C}_i^{-1} \mathbf{h}_i \right\rangle_{-\mathbf{h}_i} \right\} \\
&\propto \exp \left\{ - \mathbf{h}_i^H \left[\left(\hat{\gamma}_p \sum_{t=1}^{T_p} |x_{p,ti}|^2 + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \langle |x_{d,ti}|^2 \rangle \right) \mathbf{I}_M + \mathbf{C}_i^{-1} \right] \mathbf{h}_i \right. \\
&\quad \left. + 2 \Re \left\{ \mathbf{h}_i^H \left[\hat{\gamma}_p \sum_{t=1}^{T_p} \left(\langle \mathbf{r}_{p,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle x_{p,tj} \right) x_{p,ti}^* + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \left(\langle \mathbf{r}_{d,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle \langle x_{d,tj} \rangle \right) x_{d,ti}^* \right] \right\} \right\}. \quad (30)
\end{aligned}$$

$$\boldsymbol{\Sigma}_{\mathbf{h}_i} = \left[\left(\hat{\gamma}_p \sum_{t=1}^{T_p} |x_{p,ti}|^2 + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \langle |x_{d,ti}|^2 \rangle \right) \mathbf{I}_M + \mathbf{C}_i^{-1} \right]^{-1} \quad (31)$$

$$\langle \mathbf{h}_i \rangle = \boldsymbol{\Sigma}_{\mathbf{h}_i} \left[\hat{\gamma}_p \sum_{t=1}^{T_p} \left(\langle \mathbf{r}_{p,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle x_{p,tj} \right) x_{p,ti}^* + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \left(\langle \mathbf{r}_{d,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle \langle x_{d,tj} \rangle \right) x_{d,ti}^* \right]. \quad (32)$$

$$\begin{aligned}
q(x_{d,ti}) &\propto p(x_{d,ti}) \exp \left\{ - \hat{\gamma}_{d,t} \left[\langle \|\mathbf{h}_i\|^2 \rangle |x_{d,ti}|^2 - 2 \Re \left\{ \langle \mathbf{h}_i^H \rangle \left(\langle \mathbf{r}_{d,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle \langle x_{d,tj} \rangle \right) x_{d,ti}^* \right\} \right] \right\} \\
&\propto p(x_{d,ti}) \exp \left\{ - \hat{\gamma}_{d,t} \langle \|\mathbf{h}_i\|^2 \rangle |x_{d,ti} - z_{d,ti}|^2 \right\} \\
&\propto p(x_{d,ti}) \mathcal{CN}(z_{d,ti}; x_{d,ti}, 1 / (\hat{\gamma}_{d,t} \langle \|\mathbf{h}_i\|^2 \rangle)). \quad (34)
\end{aligned}$$

and for $t = 1, \dots, T_d$

$$\begin{aligned}
\hat{\gamma}_{d,t} &= \arg \max_{\gamma_{d,t}} \langle \ln p(\mathbf{r}_{d,t} | \mathbf{H}, \mathbf{x}_{d,t}; \gamma_{d,t}) \rangle \\
&= \arg \max_{\gamma_{d,t}} T_p - \gamma_{d,t} \langle \|\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t}\|^2 \rangle \\
&= \frac{T_p}{\langle \|\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t}\|^2 \rangle}. \quad (37)
\end{aligned}$$

Applying Lemma 1, we have

$$\begin{aligned}
\langle \|\mathbf{r}_{p,t} - \mathbf{H} \mathbf{x}_{p,t}\|^2 \rangle &= \|\langle \mathbf{r}_{p,t} \rangle - \langle \mathbf{H} \rangle \mathbf{x}_{p,t}\|^2 + \sum_{m=1}^M \tau_{r_{p,tm}} \\
&\quad + \sum_{i=1}^K |x_{p,ti}|^2 \text{Tr}\{\boldsymbol{\Sigma}_{\mathbf{h}_i}\} \\
\langle \|\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t}\|^2 \rangle &= \|\langle \mathbf{r}_{d,t} \rangle - \langle \mathbf{H} \rangle \langle \mathbf{x}_{d,t} \rangle\|^2 + \sum_{m=1}^M \tau_{r_{d,tm}} \\
&\quad + \sum_{i=1}^K [\langle |x_{d,ti}|^2 \rangle \text{Tr}\{\boldsymbol{\Sigma}_{\mathbf{h}_i}\} + \tau_{x_{d,ti}} \langle \|\mathbf{h}_i\|^2 \rangle]. \quad (39)
\end{aligned}$$

By iteratively optimizing $\{q(\mathbf{r}_{p,t})\}$, $\{q(\mathbf{r}_{d,t})\}$, $\{q(\mathbf{h}_i)\}$, $\{q(x_{d,ti})\}$, $\hat{\gamma}_p$, and $\{\hat{\gamma}_{d,t}\}$, we obtain the VBEM algorithm for estimating $\mathbf{R}_p$, $\mathbf{R}_d$, $\mathbf{H}$, $\mathbf{X}_d$, γ_p , and $\{\gamma_{d,t}\}$. We refer to this scheme as the **MF-QVB-JED algorithm** for joint channel estimation and data detection.

Remark 1: If γ_p and $\{\gamma_{d,t}\}$ are all set to N_0^{-1} , the MF-QVB-JED algorithm is equivalent to the VB-based joint channel estimation and data detection approach in [26]. We will refer to the algorithm in [26] as **conv-QVB-JED**, whose variational covariance matrix for $\mathbf{h}_i$, given by

$$\boldsymbol{\Sigma}_{\mathbf{h}_i} = \left[N_0^{-1} \left(\sum_{t=1}^{T_p} |x_{p,ti}|^2 + \sum_{t=1}^{T_d} \langle |x_{d,ti}|^2 \rangle \right) \mathbf{I}_M + \mathbf{C}_i^{-1} \right]^{-1}, \quad (40)$$

becomes smaller with increasing T_d or $\langle |x_{d,ti}|^2 \rangle, \forall t$. This result, however, implies that the estimation of $\mathbf{h}_i$ becomes more accurate with a longer transmission phase or even with an unreliable estimate of $x_{d,ti}$ reflected through large $\tau_{x_{d,ti}}$ (and $\langle |x_{d,ti}|^2 \rangle$). In MF-QVB-JED, an unreliable estimation of $x_{d,ti}$ will decrease $\hat{\gamma}_{d,t}$ in (37). Evidently, the effect of $\langle |x_{d,ti}|^2 \rangle$ on the variational covariance matrix $\boldsymbol{\Sigma}_{\mathbf{h}_i}$ of $\mathbf{h}_i$ in (31) is less pronounced than its effect on $\boldsymbol{\Sigma}_{\mathbf{h}_i}$ in (40). Therefore, in the MF-QVB-JED algorithm, an unreliable estimate of $x_{d,ti}$ will not increase the accuracy of estimating $\mathbf{h}_i$. This is one of the explanations for the superior performance of MF-QVB-JED compared to conv-QVB-JED.

Remark 2: Denoting

$$\gamma_i = \hat{\gamma}_p \sum_{t=1}^{T_p} |x_{p,ti}|^2 + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \langle |x_{d,ti}|^2 \rangle \quad (41)$$

$$\mathbf{k}_i = \gamma_i^{-1} \left[\hat{\gamma}_p \sum_{t=1}^{T_p} \left(\langle \mathbf{r}_{p,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle x_{p,tj} \right) x_{p,ti}^* \right. \\ \left. + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \left(\langle \mathbf{r}_{d,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle \langle x_{d,tj} \rangle \right) \langle x_{d,ti}^* \rangle \right], \quad (42)$$

the variational distribution $q(\mathbf{h}_i)$ in (30) can be expressed as

$$q(\mathbf{h}_i) \propto \mathcal{CN}(\mathbf{h}_i; \mathbf{k}_i, \gamma_i^{-1} \mathbf{I}_M) \mathcal{CN}(\mathbf{h}_i; \mathbf{0}, \mathbf{C}_i) \\ = \mathcal{CN}(\mathbf{h}_i; \gamma_i (\gamma_i \mathbf{I}_M + \mathbf{C}_i^{-1})^{-1} \mathbf{k}_i, (\gamma_i \mathbf{I}_M + \mathbf{C}_i^{-1})^{-1}),$$

which then explains the results $\Sigma_{\mathbf{h}_i} = (\gamma_i \mathbf{I}_M + \mathbf{C}_i^{-1})^{-1}$ and $\langle \mathbf{h}_i \rangle = \gamma_i \Sigma_{\mathbf{h}_i} \mathbf{k}_i$ in (31) and (32), respectively. We also note that $\mathbf{k}_i$ can be also written as

$$\mathbf{k}_i = \left(1 - \frac{\sum_{t=1}^{T_d} \hat{\gamma}_{d,t} \tau_{x_{d,ti}}}{\gamma_i} \right) \langle \mathbf{h}_i \rangle \\ + \gamma_i^{-1} \left[\hat{\gamma}_p \mathbf{E}_p [\mathbf{X}_p]_{i,:}^H + \mathbf{E}_d ([\mathbf{X}_p]_{i,:}^H \odot \hat{\gamma}_d) \right], \quad (43)$$

enabling its efficient computation using the residual matrices $\mathbf{E}_p = \langle \mathbf{R}_p \rangle - \langle \mathbf{H} \rangle \mathbf{X}_p$ and $\mathbf{E}_d = \langle \mathbf{R}_d \rangle - \langle \mathbf{H} \rangle \mathbf{X}_d$. Here, $\mathbf{k}_i$ can be considered as a linear estimate of $\mathbf{h}_i$ and $\langle \mathbf{h}_i \rangle = \gamma_i \Sigma_{\mathbf{h}_i} \mathbf{k}_i$ can be considered as the denoised estimate of $\mathbf{h}_i$ given the variational distribution $q(\mathbf{h}_i)$.

The proposed MF-QVB-JED algorithm is summarized in Algorithm 3. Here, we use $\hat{\mathbf{r}}_{p,t}^\ell$, $\hat{\mathbf{r}}_{d,t}^\ell$, $\hat{\mathbf{h}}_i^\ell$, $\hat{x}_{d,ti}^\ell$, $\hat{\gamma}_p^\ell$, and $\hat{\gamma}_{d,t}^\ell$ to replace $\langle \mathbf{r}_{p,t} \rangle$, $\langle \mathbf{r}_{d,t} \rangle$, $\langle \mathbf{h}_i \rangle$, $\langle x_{d,ti} \rangle$, $\hat{\gamma}_p$, and $\hat{\gamma}_{d,t}$ at iteration ℓ . We also include in the algorithm the residual terms $\mathbf{E}_p$ and $\mathbf{E}_d$, which are adjusted to reflect any update to the estimates of $\mathbf{R}_p$, $\mathbf{R}_d$, $\mathbf{H}$, and $\mathbf{X}_d$.

B. Proposed LMMSE-QVB-JED Algorithm

This section extends the LMMSE-QVB algorithm to the case of joint channel estimation and data detection. We denote γ_p and $\{\Gamma_t\} = \{\Gamma_1, \dots, \Gamma_{T_d}\}$ as the precision of the noise during the pilot transmission phase and the T_d data transmission slots, respectively. The joint distribution of all the observations and latent variables in (25) are now factored as given in (44), where $p(\mathbf{r}_{p,t} | \mathbf{H}; \gamma_p, \mathbf{x}_{p,t}) = \mathcal{CN}(\mathbf{r}_{p,t}; \mathbf{H} \mathbf{x}_{p,t}, \gamma_p^{-1} \mathbf{I}_M)$ and $p(\mathbf{r}_{d,t} | \mathbf{H}, \mathbf{x}_{d,t}; \Gamma_t) = \mathcal{CN}(\mathbf{r}_{d,t}; \mathbf{H} \mathbf{x}_{d,t}, \Gamma_t^{-1})$.

In the **E-step**, for currently fixed estimates $\hat{\gamma}_p$ and $\{\hat{\Gamma}_t\}$ of γ_p and $\{\Gamma_t\}$, respectively, we aim to obtain the mean field variational distribution $q(\mathbf{R}_p, \mathbf{R}_d, \mathbf{H}, \mathbf{X}_d)$ of $\mathbf{R}_p$, $\mathbf{R}_d$, $\mathbf{H}$, and $\mathbf{X}_d$ given $\mathbf{Y}_p$ and $\mathbf{Y}_d$ such that

$$p(\mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H} | \mathbf{Y}_p, \mathbf{Y}_d; \hat{\gamma}_p, \{\hat{\Gamma}_t\}, \mathbf{X}_p, \{\mathbf{C}_i\}) \\ \approx q(\mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H}) \\ = \left[\prod_{t=1}^{T_p} q(\mathbf{r}_{p,t}) \right] \left[\prod_{t=1}^{T_d} q(\mathbf{r}_{d,t}) \right] \left[\prod_{i=1}^K q(\mathbf{h}_i) \right] \left[\prod_{i=1}^K \prod_{t=1}^{T_d} q(x_{d,ti}) \right]. \quad (45)$$

1) *Updating $\mathbf{r}_{p,t}$.* Similar to the MF-QVB-JED algorithm.

2) *Updating $\mathbf{r}_{d,tm}$.* Similar to the LMMSE-QVB algorithm, the variational mean $\langle \mathbf{r}_{d,tm} \rangle$ and variance $\tau_{\mathbf{r}_{d,tm}}$ are determined by $F_r(s_{d,tm}, [\hat{\Gamma}_t]_{mm}, y_{d,tm}^{\text{low}}, y_{d,tm}^{\text{up}})$ and

Algorithm 3 – MF-QVB-JED for Few-Bit MIMO Joint Channel Estimation and Data Detection

```

1: Input:  $\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{X}_p, \mathbf{C}_i, \forall i$ 
2: Output:  $\hat{\mathbf{H}}, \hat{\mathbf{X}}_d$ 
3: Initialize  $\hat{\mathbf{H}}^1 = \mathbf{0}, \hat{\mathbf{X}}_d^1 = \mathbf{0}, \tau_{x_{d,ti}}^1 = \text{Var}_{p(x_i)}[x_i], \forall i, \forall t,$ 
    $\hat{\mathbf{R}}_p^1 = \mathbf{Y}_p, \hat{\mathbf{R}}_d^1 = \mathbf{Y}_d, \tau_{\mathbf{r}_{p,tm}}^1 = 0, \tau_{\mathbf{r}_{d,tm}}^1 = 0, \forall m, \forall t,$ 
    $\mathbf{E}_p = \hat{\mathbf{R}}_p^1 - \hat{\mathbf{H}}^1 \mathbf{X}_p$ , and  $\mathbf{E}_d = \hat{\mathbf{R}}_d^1 - \hat{\mathbf{H}}^1 \hat{\mathbf{X}}_d^1$ 
4: for  $\ell = 1, 2, \dots$  do
5:   Update  $\hat{\gamma}_p^\ell$  using (36) and (38)
6:   for  $t = 1, \dots, T_p$  do ▷ Update  $\mathbf{R}_p$ 
7:      $\mathbf{s}_{p,t}^\ell \leftarrow \hat{\mathbf{r}}_{p,t}^\ell - [\mathbf{E}_p]_{:,t}$ 
8:      $\hat{\mathbf{r}}_{p,t}^{\ell+1} \leftarrow F_r(\mathbf{s}_{p,t}^\ell, \hat{\gamma}_p^\ell, \mathbf{y}_{p,t}^{\text{low}}, \mathbf{y}_{p,t}^{\text{up}})$ 
9:      $\tau_{\mathbf{r}_{p,t}}^{\ell+1} \leftarrow G_r(\mathbf{s}_{p,t}^\ell, \hat{\gamma}_p^\ell, \mathbf{y}_{p,t}^{\text{low}}, \mathbf{y}_{p,t}^{\text{up}})$ 
10:    $\mathbf{E}_p \leftarrow \mathbf{E}_p - \hat{\mathbf{R}}_p^\ell + \hat{\mathbf{R}}_p^{\ell+1}$ 
11:   for  $t = 1, \dots, T_d$  do ▷ Update  $\mathbf{R}_d$ 
12:     Update  $\hat{\gamma}_{d,t}^\ell$  using (37) and (39)
13:      $\mathbf{s}_{d,t}^\ell \leftarrow \hat{\mathbf{r}}_{d,t}^\ell - [\mathbf{E}_d]_{:,t}$ 
14:      $\hat{\mathbf{r}}_{d,t}^{\ell+1} \leftarrow F_r(\mathbf{s}_{d,t}^\ell, \hat{\gamma}_{d,t}^\ell, \mathbf{y}_{d,t}^{\text{low}}, \mathbf{y}_{d,t}^{\text{up}})$ 
15:      $\tau_{\mathbf{r}_{d,t}}^{\ell+1} \leftarrow G_r(\mathbf{s}_{d,t}^\ell, \hat{\gamma}_{d,t}^\ell, \mathbf{y}_{d,t}^{\text{low}}, \mathbf{y}_{d,t}^{\text{up}})$ 
16:    $\mathbf{E}_d \leftarrow \mathbf{E}_d - \hat{\mathbf{R}}_d^\ell + \hat{\mathbf{R}}_d^{\ell+1}$ 
17:   for  $i = 1, \dots, K$  do ▷ Update  $\mathbf{H}$ 
18:      $\gamma_i^\ell \leftarrow \hat{\gamma}_p \sum_{t=1}^{T_p} |x_{p,ti}|^2 + \sum_{t=1}^{T_d} \hat{\gamma}_{d,t}^\ell (|\hat{x}_{d,ti}^\ell|^2 + \tau_{x_{d,ti}}^\ell)$ 
19:      $\mathbf{k}_i^\ell \leftarrow \left( 1 - \frac{\sum_{t=1}^{T_d} \hat{\gamma}_{d,t}^\ell \tau_{x_{d,ti}}^\ell}{\gamma_i^\ell} \right) \hat{\mathbf{h}}_i^\ell$ 
        $+ (\gamma_i^\ell)^{-1} [\hat{\gamma}_p \mathbf{E}_p [\mathbf{X}_p]_{i,:}^H + \mathbf{E}_d ([\mathbf{X}_d]_{i,:}^H \odot \hat{\gamma}_d^\ell)]$ 
20:      $\Sigma_{\mathbf{h}_i}^{\ell+1} \leftarrow (\gamma_i^\ell \mathbf{I}_M + \mathbf{C}_i^{-1})^{-1}$ 
21:      $\hat{\mathbf{h}}_i^{\ell+1} \leftarrow \gamma_i^\ell \Sigma_{\mathbf{h}_i}^{\ell+1} \mathbf{k}_i^\ell$ 
22:      $\mathbf{E}_p \leftarrow \mathbf{E}_p + (\hat{\mathbf{h}}_i^\ell - \hat{\mathbf{h}}_i^{\ell+1}) [\mathbf{X}_p]_{i,:}$ 
23:      $\mathbf{E}_d \leftarrow \mathbf{E}_d + (\hat{\mathbf{h}}_i^\ell - \hat{\mathbf{h}}_i^{\ell+1}) [\hat{\mathbf{X}}_d]_{i,:}$ 
24:      $g_i^\ell \leftarrow \|\hat{\mathbf{h}}_i^{\ell+1}\|^2 + \text{Tr}\{\Sigma_{\mathbf{h}_i}^{\ell+1}\}$ 
25:   for  $t = 1, \dots, T_d$  do ▷ Update  $\mathbf{X}_d$ 
26:     for  $i = 1, \dots, K$  do
27:        $z_{d,ti}^\ell \leftarrow (\|\hat{\mathbf{h}}_i^{\ell+1}\|^2 \hat{x}_{d,ti}^\ell + (\hat{\mathbf{h}}_i^{\ell+1})^H [\mathbf{E}_d]_{:,t}) / g_i^\ell$ 
28:        $\hat{x}_{d,ti}^{\ell+1} \leftarrow F_x(z_{d,ti}^\ell, g_i^\ell \hat{\gamma}_{d,t}^\ell)$ 
29:        $\tau_{x_{d,ti}}^{\ell+1} \leftarrow G_x(z_{d,ti}^\ell, g_i^\ell \hat{\gamma}_{d,t}^\ell)$ 
30:        $[\mathbf{E}_d]_{:,t} \leftarrow [\mathbf{E}_d]_{:,t} + \hat{\mathbf{h}}_i^\ell (\hat{x}_{d,ti}^\ell - \hat{x}_{d,ti}^{\ell+1})$ 
31:    $\forall t, \forall i : \hat{x}_{d,ti} \leftarrow \arg \max_{a \in \mathcal{S}} p_a \mathcal{CN}(z_{d,ti}^\ell; a, 1/(g_i^\ell \hat{\gamma}_{d,t}^\ell)).$ 

```

$G_r(s_{d,tm}, \hat{\gamma}_{t,mm}, y_{d,tm}^{\text{low}}, y_{d,tm}^{\text{up}})$, where

$$s_{d,tm} = \langle \mathbf{r}_{d,tm} \rangle - \frac{[\hat{\Gamma}_t]_{m,:}}{[\hat{\Gamma}_t]_{mm}} (\langle \mathbf{r}_{d,t} \rangle - \langle \mathbf{H} \rangle \langle \mathbf{x}_{d,t} \rangle). \quad (46)$$

3) *Updating $\mathbf{h}_i$.* Taking the expectation of the conditional (44) w.r.t. all latent variables except $\mathbf{h}_i$, the variational distribution $q(\mathbf{h}_i)$ is given by

$$q(\mathbf{h}_i) \propto \exp \left\{ \langle \ln p(\mathbf{R}_p | \mathbf{H}; \hat{\gamma}_p, \mathbf{X}_p) + \ln p(\mathbf{R}_d | \mathbf{X}_d, \mathbf{H}; \{\hat{\Gamma}_t\}) \right. \\ \left. + \ln p(\mathbf{h}_i; \mathbf{C}_i) \right\rangle_{-\mathbf{h}_i}. \quad (47)$$

Following the same procedure to obtain $q(\mathbf{h}_i)$ as in the MF-

$$\begin{aligned}
& p(\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H}; \gamma_p, \{\Gamma_t\}, \mathbf{X}_p, \{\mathbf{C}_i\}) \\
&= p(\mathbf{Y}_p | \mathbf{R}_p) p(\mathbf{R}_p | \mathbf{H}; \gamma_p, \mathbf{X}_p) p(\mathbf{Y}_d | \mathbf{R}_d) p(\mathbf{R}_d | \mathbf{X}_d, \mathbf{H}; \{\Gamma_t\}) p(\mathbf{X}_d) p(\mathbf{H} | \{\mathbf{C}_i\}) \\
&= \left[\prod_{t=1}^{T_p} p(\mathbf{y}_{p,t} | \mathbf{r}_{p,t}) p(\mathbf{r}_{p,t} | \mathbf{H}; \gamma_p, \mathbf{x}_{p,t}) \right] \left[\prod_{t=1}^{T_d} p(\mathbf{y}_{d,t} | \mathbf{r}_{d,t}) p(\mathbf{r}_{d,t} | \mathbf{H}, \mathbf{x}_{d,t}; \Gamma_t) p(\mathbf{x}_{d,t}) \right] \prod_{i=1}^K p(\mathbf{h}_i | \mathbf{C}_i) \quad (44)
\end{aligned}$$

QVB-JED algorithm, we have

$$\begin{aligned}
q(\mathbf{h}_i) &\propto \mathcal{CN}(\mathbf{h}_i; \mathbf{k}_i, \Gamma_i^{-1}) \mathcal{CN}(\mathbf{h}_i; \mathbf{0}, \mathbf{C}_i) \\
&= \mathcal{CN}(\mathbf{h}_i; (\Gamma_i + \mathbf{C}_i^{-1})^{-1} \Gamma_i \mathbf{k}_i, (\Gamma_i + \mathbf{C}_i^{-1})^{-1}), \quad (48)
\end{aligned}$$

where Γ_i and $\mathbf{k}_i$ are defined as

$$\begin{aligned}
\Gamma_i &= \hat{\gamma}_p \sum_{t=1}^{T_p} |x_{p,ti}|^2 \mathbf{I}_M + \sum_{t=1}^{T_d} \langle |x_{d,it}|^2 \rangle \hat{\Gamma}_t, \quad (49) \\
\mathbf{k}_i &= \left(\mathbf{I}_M - \Gamma_i^{-1} \sum_{t=1}^{T_d} \hat{\Gamma}_t \tau_{x_{d,ti}} \right) \langle \mathbf{h}_i \rangle \\
&\quad + \Gamma_i^{-1} \left[\hat{\gamma}_p \mathbf{E}_p [\mathbf{X}_p]_{i,:}^H + \sum_{t=1}^{T_d} \hat{\Gamma}_t [\mathbf{E}_d]_{:,t} x_{d,ti}^* \right], \quad (50)
\end{aligned}$$

and where $\mathbf{E}_p = \langle \mathbf{R}_p \rangle - \langle \mathbf{H} \rangle \mathbf{X}_p$ and $\mathbf{E}_d = \langle \mathbf{R}_d \rangle - \langle \mathbf{H} \rangle \langle \mathbf{X}_d \rangle$ are the residual terms. The variational covariance matrix and mean of $\mathbf{h}_i$ are now given by $\Sigma_{\mathbf{h}_i} = (\Gamma_i + \mathbf{C}_i^{-1})^{-1}$ and $\langle \mathbf{h}_i \rangle = \Sigma_{\mathbf{h}_i} \Gamma_i \mathbf{k}_i$.

4) *Updating $x_{d,it}$.* Taking the expectation of the conditional (25) w.r.t. all latent variables except $x_{d,ti}$, the variational distribution $q(x_{d,ti})$ is given by

$$q(x_{d,ti}) \propto \exp \left\{ \langle \ln p(\mathbf{r}_{d,t} | \mathbf{H}, \mathbf{x}_{d,t}; \hat{\Gamma}_t) + \ln p(x_{d,ti}) \rangle_{-x_{d,ti}} \right\}. \quad (51)$$

Similar to the procedure in the LMMSE-QVB and MF-QVB-JED algorithms, we obtain

$$q(x_{d,ti}) \propto p(x_{d,ti}) \mathcal{CN}(z_{d,ti}; x_{d,ti}, 1 / \langle \mathbf{h}_i^H \hat{\Gamma}_t \mathbf{h}_i \rangle), \quad (52)$$

where

$$\begin{aligned}
z_{d,ti} &= \frac{\langle \mathbf{h}_i^H \rangle \hat{\Gamma}_t}{\langle \mathbf{h}_i^H \hat{\Gamma}_t \mathbf{h}_i \rangle} \left(\langle \mathbf{r}_{d,t} \rangle - \sum_{j \neq i}^K \langle \mathbf{h}_j \rangle \langle x_{d,tj} \rangle \right) \\
&= \frac{\langle \mathbf{h}_i^H \rangle \hat{\Gamma}_t \langle \mathbf{h}_i \rangle \langle x_{d,ti} \rangle + \langle \mathbf{h}_i^H \rangle \hat{\Gamma}_t [\mathbf{E}_d]_{:,t}}{\langle \mathbf{h}_i^H \hat{\Gamma}_t \mathbf{h}_i \rangle} \quad (53)
\end{aligned}$$

is a linear estimate of $x_{d,ti}$. We note that $\langle \mathbf{h}_i^H \hat{\Gamma}_t \mathbf{h}_i \rangle = \langle \mathbf{h}_i \rangle \hat{\Gamma}_t \langle \mathbf{h}_i \rangle + \text{Tr}\{\hat{\Gamma}_t \Sigma_{\mathbf{h}_i}\}$.

In the **M-step**, γ_p and $\{\Gamma_t\}$ are estimated to maximize $\ln p(\mathbf{Y}_p, \mathbf{Y}_d, \mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H}; \gamma_p, \{\Gamma_t\}, \mathbf{X}_p, \{\mathbf{C}_i\})$ w.r.t. the variational distribution $q(\mathbf{R}_p, \mathbf{R}_d, \mathbf{X}_d, \mathbf{H})$. The update of $\hat{\gamma}_p$ is similar to the procedure in the MF-QVB-JED algorithm and is given in (36) and (38). The update of $\hat{\Gamma}_t$ is given by

$$\begin{aligned}
\hat{\Gamma}_t &= \arg \max_{\Gamma_t} \langle \ln p(\mathbf{r}_{d,t} | \mathbf{H}, \mathbf{x}_{d,t}; \Gamma_t) \rangle \\
&= \arg \max_{\Gamma_t} \left[\ln |\Gamma_t| \right. \\
&\quad \left. - \langle (\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t})^H \Gamma_t (\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t}) \rangle \right].
\end{aligned}$$

Applying Lemma 1, we have

$$\begin{aligned}
& \langle (\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t})^H \Gamma_t (\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t}) \rangle \\
&= \text{Tr} \left\{ \left[(\langle \mathbf{r}_{d,t} \rangle - \langle \mathbf{H} \rangle \langle \mathbf{x}_{d,t} \rangle) (\langle \mathbf{r}_{d,t} \rangle - \langle \mathbf{H} \rangle \langle \mathbf{x}_{d,t} \rangle)^H + \Sigma_{\mathbf{r}_{d,t}} \right. \right. \\
&\quad \left. \left. + \sum_{i=1}^K \langle |x_{d,ti}|^2 \rangle \Sigma_{\mathbf{h}_i} + \langle \mathbf{H} \rangle \Sigma_{\mathbf{x}_{d,t}} \langle \mathbf{H} \rangle^H \right] \Gamma_t \right\}. \quad (54)
\end{aligned}$$

Due to the rank deficiency of $\langle (\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t})^H \Gamma_t (\mathbf{r}_{d,t} - \mathbf{H} \mathbf{x}_{d,t}) \rangle_{-\Gamma_t}$, we propose the following estimator for Γ_t :

$$\begin{aligned}
\hat{\Gamma}_t &= \left(\frac{\| \langle \mathbf{r}_{d,t} \rangle - \langle \mathbf{H} \rangle \langle \mathbf{x}_{d,t} \rangle \|^2}{M} \mathbf{I}_M + \Sigma_{\mathbf{r}_{d,t}} \right. \\
&\quad \left. + \sum_{i=1}^K \langle |x_{d,ti}|^2 \rangle \Sigma_{\mathbf{h}_i} + \langle \mathbf{H} \rangle \Sigma_{\mathbf{x}_{d,t}} \langle \mathbf{H} \rangle^H \right)^{-1}. \quad (55)
\end{aligned}$$

By iteratively optimizing $\{q(\mathbf{r}_{p,t})\}$, $\{q(\mathbf{r}_{d,t})\}$, $\{q(\mathbf{h}_i)\}$, $\{q(x_{d,ti})\}$, $\hat{\gamma}_p$, and $\{\hat{\Gamma}_t\}$, we obtain the VBEM algorithm for estimating $\mathbf{R}_p$, $\mathbf{R}_d$, $\mathbf{H}$, $\mathbf{X}_d$, γ_p , and $\{\Gamma_t\}$. We refer to this scheme as the **LMMSE-QVB-JED algorithm** for joint channel estimation and data detection. The implementation of the LMMSE-QVB-JED algorithm is similar to that of MF-QVB-JED presented in Algorithm 3, and thus we skip a summary of its implementation.

C. Practical Aspects of Implementing MF/LMMSE-QVB-JED

1) *Computing $\langle |x_{d,ti}|^2 \rangle$:* For PSK signaling, the variational second moment $\langle |x_{d,ti}|^2 \rangle$ is constant and need not be updated in each iteration of the algorithms. We present the proof for this observation in Appendix B.

2) *Computing $\Sigma_{\mathbf{h}_i}$ with uncorrelated channels:* When $\mathbf{C}_i$ is a diagonal matrix, the variational covariance matrix $\Sigma_{\mathbf{h}_i}$ in (40) is also a diagonal matrix and its computation does not require matrix inversion. Thus, the MF-QVB-JED algorithm can be implemented without any matrix inversion. This property does not hold for the LMMSE-QVB-JED algorithm, since Γ_i in (49) is not, in general, a diagonal matrix.

3) *Estimating $\mathbf{h}_i$ without its prior distribution:* While the prior distribution of $\mathbf{h}_i$ is assumed to be known as $\mathcal{CN}(\mathbf{0}, \mathbf{C}_i)$, the proposed MF/LMMSE-QVB-JED algorithms can be adapted to cases without this knowledge. For example, we can simply assume an uninformative prior for $\mathbf{h}_i$ and thus use $\mathbf{k}_i$ in (43) or $\mathbf{k}_i$ in (50) as a point estimate of $\mathbf{h}_i$ in the **M-step** of these two algorithms.

4) *Lite implementation of MF-QVB-JED:* Instead of using the latent variable $\gamma_{d,t}$ as the precision at time slot t , we can impose a single latent variable γ_d as the precision for all time slots during the data transmission phase. A lite version of MF-QVB-JED can be devised using the same procedure as in Section V-A where $\{\gamma_{d,t}\}$ is replaced by γ_d . In the **M-step**,

the estimate of γ_d can be found as

$$\hat{\gamma}_d = \frac{MT_d}{\sum_{t=1}^{T_d} \langle \|\mathbf{r}_{d,t} - \mathbf{H}\mathbf{x}_{d,t}\|^2 \rangle}, \quad (56)$$

where $\langle \|\mathbf{r}_{d,t} - \mathbf{H}\mathbf{x}_{d,t}\|^2 \rangle$ is given in (39).

5) *Lite implementation of LMMSE-QVB-JED*: Instead of using the latent variable $\mathbf{\Gamma}_t$ as the precision matrix at time slot t , we can use the same precision matrix $\mathbf{\Gamma}$ for all time slots. A lite version of LMMSE-QVB-JED can be devised using the same procedure as in Section V-B where $\{\mathbf{\Gamma}_t\}$ is replaced by $\mathbf{\Gamma}$. In the **M-step**, we propose the following estimator for $\mathbf{\Gamma}$:

$$\hat{\mathbf{\Gamma}} = T_d \left(\frac{\langle \|\mathbf{R}_d\rangle - \langle \mathbf{H} \rangle \langle \mathbf{X}_d \rangle^2_F}{M} \mathbf{I}_M + \mathbf{\Sigma}_{\mathbf{r}_d} + \sum_{i=1}^K \langle \|\mathbf{x}_{d,i}\|^2 \rangle \mathbf{\Sigma}_{\mathbf{h}_i} + \langle \mathbf{H} \rangle \mathbf{\Sigma}_{\mathbf{x}_d} \langle \mathbf{H} \rangle^H \right)^{-1}, \quad (57)$$

where we denote $\mathbf{\Sigma}_{\mathbf{r}_d} = \sum_{t=1}^{T_d} \mathbf{\Sigma}_{\mathbf{r}_{d,t}}$, $\mathbf{\Sigma}_{\mathbf{x}_d} = \sum_{t=1}^{T_d} \mathbf{\Sigma}_{\mathbf{x}_{d,t}}$ and $\langle \|\mathbf{x}_{d,i}\|^2 \rangle = \sum_{t=1}^{T_d} \langle \|x_{d,ti}\|^2 \rangle$.

We observe in our simulations that the lite version of MF/LMMSE-QVB-JED slightly increases the detection error compared to the original version. However, the lite version can significantly reduce the computational complexity, especially for the LMMSE-QVB-JED algorithm. LMMSE-QVB-JED requires one matrix inversion in (57) for computing $\hat{\mathbf{\Gamma}}$ in the lite version, while requiring T_d matrix inversions to compute $\{\hat{\mathbf{\Gamma}}_t\}$ in the original version. In the numerical results, we will use the lite version of these algorithms.

VI. NUMERICAL RESULTS

This section presents numerical results comparing the performance of the proposed VB-based methods with the conventional quantized VB-based approach with perfect knowledge of the noise variance, denoted as conv-QVB in [26], the GAMP-based approach in [34], and FBM-DetNet in [20], which are the most recent and related methods to our work. The maximum number of iterations is set to 50 for all iterative algorithms. The covariance matrices $\mathbf{C}_i$ are normalized so that their diagonal elements are 1, which implies $\mathbb{E}[\|\mathbf{h}_i\|^2] = M$, $\forall i$. The noise variance N_0 is set based on the operating SNR, defined as

$$\text{SNR} = \frac{\mathbb{E}[\|\mathbf{H}\mathbf{x}\|^2]}{\mathbb{E}[\|\mathbf{n}\|^2]} = \frac{\sum_{i=1}^K \text{Tr}\{\mathbf{C}_i\}}{MN_0} = \frac{K}{N_0}. \quad (58)$$

For i.i.d. channels, we set $\mathbf{C}_i = \mathbf{I}$, $\forall i$. For spatially correlated channels, we use the typical urban channel model in [8] where the power angle spectrum of the channel model follows a Laplacian distribution with an angle spread of 10° . The covariance matrix $\mathbf{C}_i$ is obtained according to [37, Eq. (2)]. Unless otherwise stated, we set the training length $T_p = 2K$ and the data transmission length $T_d = 100$. The pilot sequences from the K users are chosen to be orthogonal to each other.

First, we investigate data detection for the case of perfect CSI. For the conv-QVB approach, we also examine a scenario in which the noise variance is unknown and estimated by the median absolute deviation (MAD) method [38], [39]. Since the unquantized received signal $\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n}$ is unobserved, we consider two MAD-based estimators for the noise standard

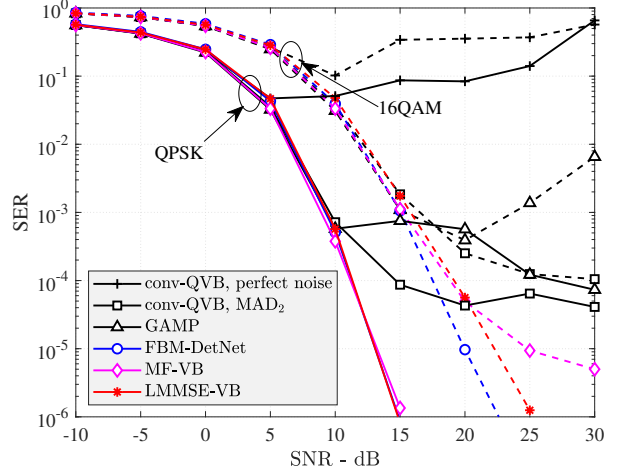


Fig. 1: Data detection performance comparison for i.i.d. channels with perfect CSI, $b = 3$ bits, $K = 16$, and with $M = 32$ for QPSK signaling (solid lines) and $M = 64$ for 16QAM signaling (dashed lines), respectively.

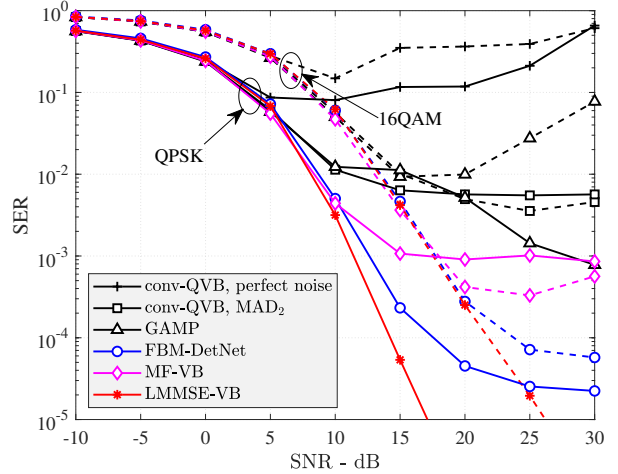


Fig. 2: Data detection performance comparison for spatially correlated channels with perfect CSI, $b = 3$ bits, $K = 16$, and with $M = 32$ for QPSK signaling (solid lines) and $M = 64$ for 16QAM signaling (dashed lines), respectively.

deviation σ :

$$\text{MAD}_1 : \hat{\sigma} = 1.4826 \times \text{med} \{ |\hat{\mathbf{r}} - \text{med}\{\hat{\mathbf{r}}\}| \}, \quad (59)$$

which is based on the current estimator $\hat{\mathbf{r}}$ of $\mathbf{r}$, and

$$\text{MAD}_2 : \hat{\sigma} = 1.4826 \times \text{med} \{ |\mathbf{e} - \text{med}\{\mathbf{e}\}| \}, \quad (60)$$

which is based on the current residual vector $\mathbf{e} = \hat{\mathbf{r}} - \mathbf{H}\hat{\mathbf{x}}$. We then apply these MAD estimates to the MF-QVB algorithm with $\hat{\gamma} = 1/\hat{\sigma}_{\{1,2\}}^2$ in place of the EM-based estimate in (14). In numerous simulations, we have consistently observed that MAD_2 outperforms MAD_1 . Therefore, we will only present the SER performance results for the MAD_2 scheme.

SER performances in i.i.d. and spatially correlated channels are shown in Fig. 1 and Fig. 2, respectively. It can be seen that, for both i.i.d. and correlated channels, the conv-QVB and GAMP-based methods are outperformed by all other methods, and their performance is severely degraded at high SNRs. This is because they do not take into account the residual inter-user

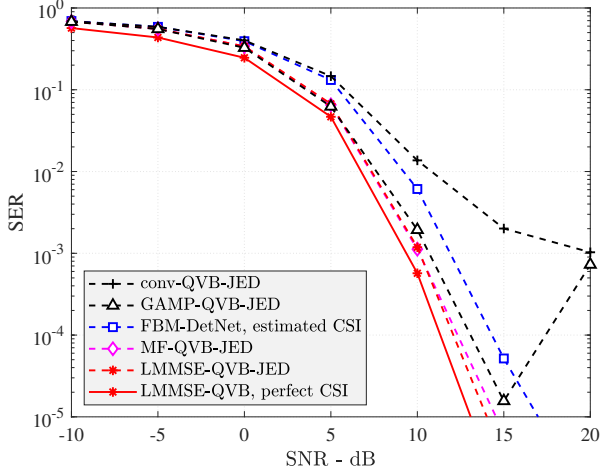


Fig. 3: Data detection comparison between the proposed MF-QVB-JED, LMMSE-QVB-JED, and other existing methods for i.i.d. channels with $K = 16$, $M = 32$, $b = 3$ bits, and QPSK signaling.

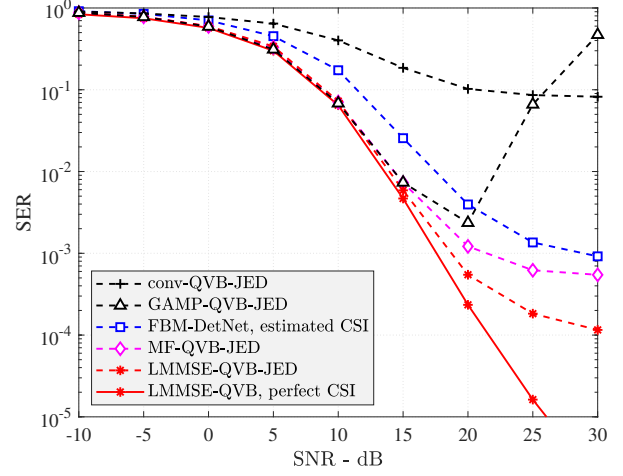


Fig. 4: Data detection performance comparison between the proposed MF-QVB-JED, LMMSE-QVB-JED, and other existing methods for spatially correlated channels with $K = 16$, $M = 64$, $b = 3$ bits, and 16QAM signaling.

interference and often encounter catastrophic cancellation issues at high SNR. The conv-QVB scheme with the MAD-based noise variance estimator significantly outperforms the conv-QVB scheme that uses the known noise variance, especially at high SNR. However, the performance of the MAD-based scheme is much worse than that of the proposed QVB schemes which use the EM-based noise variance/covariance estimators. For i.i.d. channels, FBM-DetNet, MF-QVB, and LMMSE-QVB all yield the same performance for QPSK signals, while for 16QAM FBM-DetNet and LMMSE-QVB are similar and both outperform MF-QVB. For spatially correlated channels, LMMSE-QVB provides a significantly lower SER than the other methods due to its estimation of the precision matrix Γ , which can better compensate for the effect of the noise and the strong residual inter-user interference in spatially correlated channels. It is also worth noting that in the case of 16QAM, the inter-user interference effect is more severe than QPSK, and therefore the advantage of LMMSE-QVB compared to MF-QVB is better seen.

Fig. 3 presents results for data detection with estimated CSI and i.i.d. channels. Both MF-QVB-JED and LMMSE-QVB-JED outperform the conv-QVB-JED, GAMP-based, and FBM-DetNet methods. Note that FBM-DetNet uses estimated CSI provided by FBM-CENet, a channel estimation network also proposed in [20] and designed to estimate the CSI using only the pilot sequence. MF-QVB-JED and LMMSE-QVB-JED both yield the same SER, which is about 2–3 dB better than FBM-DetNet at an SER of 10^{-3} and 10^{-5} , respectively. The performance of MF-QVB-JED and LMMSE-QVB-JED is also quite close to that of LMMSE-QVB with perfect CSI.

Results for data detection with estimated CSI and spatially correlated channels are given in Fig. 4, where we see that the proposed MF-QVB-JED and LMMSE-QVB-JED methods outperform conv-QVB-JED, GAMP-QVB-JED, and FBM-DetNet since the effects of both inter-user interference and spatial channel correlation are taken into account. However, unlike the case of i.i.d. channels where MF-QVB-JED and LMMSE-QVB-JED give the same performance, the LMMSE-

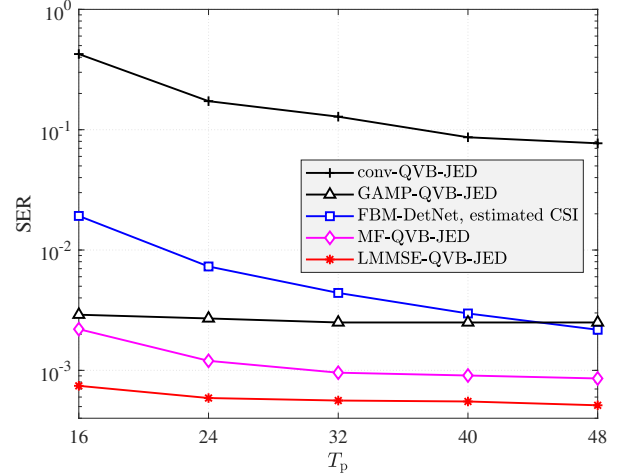


Fig. 5: Data detection performance versus the pilot length T_p for spatially correlated channels with $K = 16$, $M = 64$, $b = 3$ bits, and 16QAM signaling at SNR = 20 dB.

QVB-JED method provides a significantly lower SER than MF-QVB-JED at high SNRs for spatially correlated channels. For example, at 30dB, the SER of LMMSE-QVB-JED is about 10 times lower than that of MF-QVB-JED, which is already better than FBM-DetNet.

In Fig. 5 we present the data detection performance versus the pilot length T_p . It can be seen that the SER improves remarkably as the pilot length increases beyond the minimum value of K , consistent with prior work involving coarsely quantized observations. The SER improvement with T_p for FBM-DetNet is more noticeable than the other methods because FBM-DetNet uses channel estimates that are solely obtained from the pilot signals, while the other methods perform joint channel estimation and data detection, i.e., both pilot and data signals are used to improve the channel estimation.

We provide a channel estimation comparison in Fig. 6 where i.i.d. channels are considered in Fig. 6a and spatially correlated channels are considered in Fig. 6b. The normalized

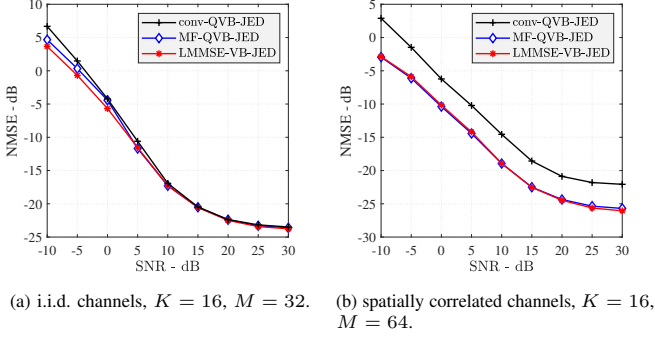


Fig. 6: Channel estimation performance comparison.

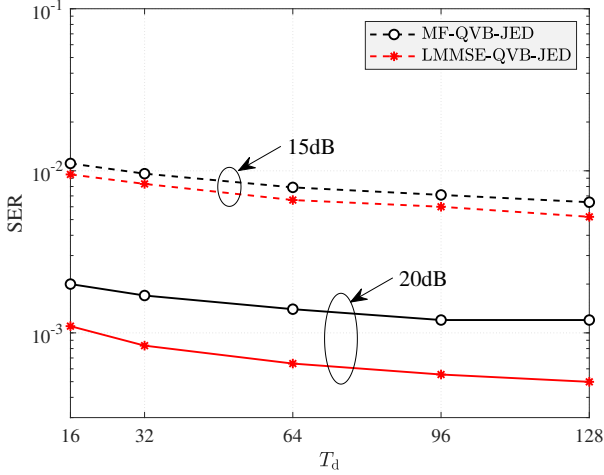


Fig. 7: Detection performance of MF-QVB-JED and LMMSE-QVB-JED versus T_d with $K = 16$, $M = 64$, $b = 3$ bits, and 16QAM signaling.

mean squared error (NMSE) in these figures is defined as $\text{NMSE} = \mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2 / \|\mathbf{H}\|_F^2]$. For i.i.d. channels, all three VB-based methods conv-QVB-JED, MF-QVB-JED, and LMMSE-QVB-JED give similar performance but for spatially correlated channels, the proposed MF-QVB-JED and LMMSE-QVB-JED methods are seen to provide lower NMSEs compared to the conv-QVB-JED method.

Fig. 7 presents the SER performance of the proposed MF-QVB-JED and LMMSE-QVB-JED methods w.r.t. T_d . We observe that the SER performance improves with increasing T_d since more received signals are combined to achieve a more accurate channel estimate. Consequently, the data detection phase can result in a lower detection error.

In Fig. 8, we evaluate the data detection performance of the proposed MF-QVB-JED and LMMSE-QVB-JED methods for different ADC bit resolutions. As expected, increasing the resolution b significantly helps improve the detection performance. It is observed that lower SNRs require a lower bit resolution for the best performance, e.g., 4-bit ADCs are sufficient to obtain the lowest SER at 10dB. Increasing the ADC bit resolution to values higher than 4 does not result in a lower SER. It is also interesting to note that at high SNRs, LMMSE-QVB-JED can provide much lower SERs compared to MF-QVB-JED as the bit resolution increases.

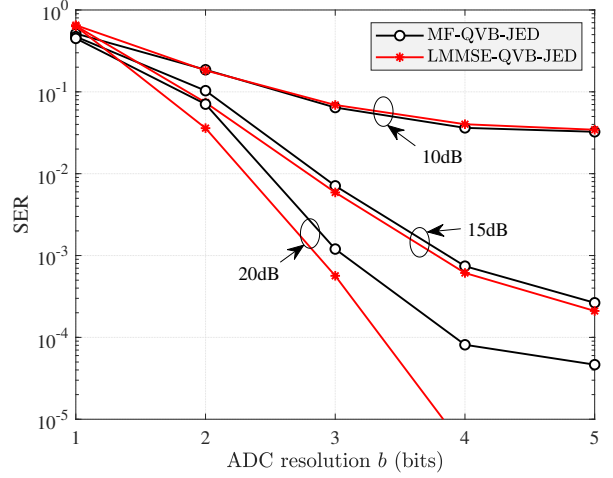


Fig. 8: Detection performance of MF-QVB-JED and LMMSE-QVB-JED versus the ADC bit resolution b with $K = 16$, $M = 64$, and 16QAM signaling.

VII. CONCLUSION

In this paper, we exploited the VB inference framework to propose different channel estimation and data detection methods for massive MIMO systems with low-resolution ADCs. In particular, we proposed new VB-based algorithms referred to as MF-QVB and LMMSE-QVB for data detection with known CSI, and MF-QVB-JED and LMMSE-QVB-JED for joint channel estimation and data detection. In the proposed QVB framework, we proposed to float the noise covariance matrix as unknown random variables which also allows the algorithms to take into account the residual inter-user interference. Numerous practical aspects of the QVB framework were studied to improve the implementation stability. It was also shown via a number of simulation studies that the proposed methods provide robust performance and significantly outperform existing methods, particularly when the channels are spatially correlated.

APPENDIX A

COMPUTATION OF $F_r(\mu, \gamma, a, b)$ AND $G_r(\mu, \gamma, a, b)$

For ease of presentation, we denote $\alpha = \sqrt{2\gamma}(a - \mu)$ and $\beta = \sqrt{2\gamma}(b - \mu)$. For an arbitrary complex random variable $\mathcal{CN}(\mu, \gamma^{-1})$ whose real and imaginary parts are both truncated on the interval (a, b) , the mean $F_r(\mu, \gamma, a, b)$ and variance $G_r(\mu, \gamma, a, b)$ are computed as

$$F_r(\mu, \gamma, a, b) = \mu - \frac{1}{\sqrt{2\gamma}} \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \quad (61)$$

$$G_r(\mu, \gamma, a, b) = \frac{1}{2\langle\gamma\rangle} \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right], \quad (62)$$

where the PDF and CDF operators $\phi(\cdot)$ and $\Phi(\cdot)$, as well as the multiplication, division, and square operations are applied individually on the real and imaginary components. The variance $G_r(\mu, \gamma, a, b)$ is computed by adding the variances of

the two components.

APPENDIX B

COMPUTATION OF $F_x(z, \gamma)$ AND $G_x(z, \gamma)$

Given $z = x + \mathcal{CN}(0, \gamma^{-1})$, the posterior distribution of x given z is $p(x|z; \gamma) \propto p(x) \mathcal{CN}(z; x, \gamma^{-1})$. For $a \in \mathcal{S}$, we have

$$p(x = a|z; \gamma) = (1/Z) p_a \exp(-\gamma|z - a|^2),$$

where $Z = \sum_{b \in \mathcal{S}} p_b \exp(-\gamma|z - b|^2)$ is a normalization factor. The corresponding posterior mean $F_x(z, \gamma)$ and variance $G_x(z, \gamma)$ are computed as

$$F_x(z, \gamma) = \sum_{a \in \mathcal{S}} a \times p(x = a|z, \gamma)$$

$$G_x(z, \gamma) = \sum_{a \in \mathcal{S}} |a|^2 \times p(x = a|z, \gamma) - |F_x(z, \gamma)|^2.$$

We note that $\mathbb{E}[|x|^2|z; \gamma]$ is equal to $|a|^2$ for PSK signaling with transmit energy $|a|^2$, as shown below:

$$\mathbb{E}[|x|^2|z; \gamma] = \sum_{a \in \mathcal{S}} |a|^2 \frac{1}{Z} p_a \exp(-\gamma|z - a|^2)$$

$$= |a|^2 \frac{\sum_{a \in \mathcal{S}} p_a \exp(-\gamma|z - a|^2)}{Z} = |a|^2.$$

REFERENCES

- [1] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Commun.*, vol. 12, pp. 16–32, 2014.
- [2] —, "TeraNets: Ultra-broadband communication networks in the terahertz band," *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 130–135, Aug. 2014.
- [3] N. Rajatheva, I. Atzeni, E. Bjornson, A. Bourdoux, S. Buzzi, J.-B. Dore, S. Erkkucuk, M. Fuentes, K. Guan, Y. Hu *et al.*, "White paper on broadband connectivity in 6G," *arXiv preprint arXiv:2004.14247*, 2020.
- [4] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Select. Topics in Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [5] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4624–4634, Sept. 2017.
- [6] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek, "A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI," *IEEE J. Select. Topics in Signal Process.*, vol. 12, no. 3, pp. 484–498, June 2018.
- [7] 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.
- [8] 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.
- [9] N. Kolomvakis, T. Eriksson, M. Coldrey, and M. Viberg, "Quantized uplink massive MIMO systems with linear receivers," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, June 2020.
- [10] A. S. Lan, M. Chiang, and C. Studer, "Linearized binary regression," in *Proc. Annual Conf. on Inform. Sciences and Systems*, Princeton, NJ, USA, Mar. 2018.
- [11] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "Linear and deep neural network-based receivers for massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7333–7345, Nov. 2021.
- [12] Y.-S. Jeon, N. Lee, and H. V. Poor, "Robust data detection for MIMO systems with one-bit ADCs: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1663–1676, Mar. 2020.
- [13] S. H. Song, S. Lim, G. Kwon, and H. Park, "CRC-aided soft-output detection for uplink multi-user MIMO systems with one-bit ADCs," in *Proc. IEEE Wireless Commun. and Networking Conf.*, Marrakesh, Morocco, Apr. 2019.
- [14] Y. Cho and S. 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.
- [15] 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 Commun. Letters*, vol. 7, no. 3, pp. 476–479, June 2018.
- [16] 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, no. 10, pp. 2541–2556, May 2016.
- [17] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—Part I: Derivation," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5839–5853, 2014.
- [18] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "SVM-based channel estimation and data detection for one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2086–2099, 2021.
- [19] D. H. N. Nguyen, "Neural network-optimized channel estimator and training signal design for MIMO systems with few-bit ADCs," *IEEE Signal Process. Letters*, vol. 27, pp. 1370–1374, 2020.
- [20] L. V. Nguyen, D. H. N. Nguyen, and A. L. Swindlehurst, "Deep learning for estimation and pilot signal design in few-bit massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 379–392, Jan. 2023.
- [21] S. Khobahi, N. Shlezinger, M. Soltanalian, and Y. C. Eldar, "LoRD-Net: Unfolded deep detection network with low-resolution receivers," *IEEE Trans. Signal Process.*, vol. 69, pp. 5651–5664, 2021.
- [22] Y. Jeon, S. Hong, and N. Lee, "Supervised-learning-aided communication framework for MIMO systems with low-resolution ADCs," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7299–7313, Aug. 2018.
- [23] L. V. Nguyen, D. T. Ngo, N. H. Tran, A. L. Swindlehurst, and D. H. N. Nguyen, "Supervised and semi-supervised learning for MIMO blind detection with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2427–2442, Apr. 2020.
- [24] S. Kim, J. Chae, and S.-N. Hong, "Machine learning detectors for MU-MIMO systems with one-bit ADCs," *IEEE Access*, vol. 8, pp. 86 608–86 616, Apr. 2020.
- [25] Y. Xiang, K. Xu, B. Xia, and X. Cheng, "Bayesian joint channel-and-data estimation for quantized OFDM over doubly selective channels," *IEEE Trans. Wireless Commun. (Early Access)*, 2022.
- [26] S. S. Thootta and C. R. Murthy, "Variational Bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low resolution ADCs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3467–3481, May 2021.
- [27] —, "Massive MIMO-OFDM systems with low resolution ADCs: Cramér-Rao bound, sparse channel estimation, and soft symbol decoding," *IEEE Trans. Signal Process.*, vol. 70, pp. 4835–4850, 2022.
- [28] J. Zhu, C.-K. Wen, J. Tong, C. Xu, and S. Jin, "Grid-less variational bayesian channel estimation for antenna array systems with low resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1549–1562, 2020.
- [29] D. H. Nguyen, I. Atzeni, A. Tölle, and A. L. Swindlehurst, "A variational Bayesian perspective on massive MIMO detection," *arXiv preprint arXiv:2205.11649*, 2022.
- [30] F. Li, Z. Xu, and S. Zhu, "Variational-inference-based data detection for OFDM systems with imperfect channel estimation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1394–1399, 2013.
- [31] X.-Y. Zhang, D.-G. Wang, and J.-B. Wei, "Joint symbol detection and channel estimation for MIMO-OFDM systems via the variational Bayesian EM algorithm," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, 2008, pp. 13–17.
- [32] W. Yuan, Z. Wei, J. Yuan, and D. W. K. Ng, "A simple variational Bayes detector for orthogonal time frequency space (OTFS) modulation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 7976–7980, 2020.
- [33] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "A variational bayesian perspective on MIMO detection with low-resolution ADCs," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, USA, 2022.
- [34] C. Wen, C. Wang, S. Jin, 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, no. 10, pp. 2541–2556, May 2016.
- [35] C. M. Bishop, *Pattern recognition and machine learning*. New York, NY, USA: Springer, 2006.
- [36] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. and Trends® Mach. Learn.*, vol. 1, no. 1–2, pp. 1–305, Jan. 2008.

- [37] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, June 2015.
- [38] P. J. Huber, *Robust statistics*. New York: John Wiley, 1981.
- [39] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.



Ly V. Nguyen (Member, IEEE) received the B.Eng. degree in Electronics and Telecommunications from the University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam, in 2014, the M.Sc. degree in Advanced Wireless Communications Systems from CentraleSupélec, Paris-Saclay University, France, in 2016, and the Ph.D. degree in a joint doctoral program in Computational Science with San Diego State University and University of California, Irvine, CA, USA, in 2022. He received a Best Paper Award at the 2020 IEEE International Conference on Communications (ICC). His research interests include wireless communications, signal processing, and machine learning.



Duy H. N. Nguyen (Senior Member, IEEE) received the B.Eng. degree (Hons.) from the Swinburne University of Technology, Hawthorn, VIC, Australia, in 2005, the M.Sc. degree from the University of Saskatchewan, Saskatoon, SK, Canada, in 2009, and the Ph.D. degree from McGill University, Montréal, QC, Canada, in 2013, all in electrical engineering. From 2013 to 2015, he held a joint appointment as a Research Associate with McGill University and a Post-doctoral Research Fellow with the Institut National de la Recherche Scientifique, Université du Québec, Montréal, QC, Canada. He was a Research Assistant with the University of Houston, Houston, TX, USA, in 2015, and a Post-doctoral Research Fellow with the University of Texas at Austin, Austin, TX, USA, in 2016. He is an Associate Professor at the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA, USA. His current research interests include resource allocation in wireless networks, signal processing for communications, convex optimization, game theory, and machine learning. Dr. Nguyen has been a TPC member for several flagship IEEE conferences, including ICC, GLOBECOM, and INFOCOM. He received the Australian Development Scholarship, the FRQNT Doctoral and Post-doctoral Fellowship, and the NSERC Post-doctoral Fellowship. He received a Best Paper Award at the 2020 IEEE International Conference on Communications (ICC). He is currently an Associate Editor for the *IEEE Open Journal of the Communications Society*.



Lee Swindlehurst (Fellow, IEEE) received the B.S. (1985) and M.S. (1986) degrees in Electrical Engineering from Brigham Young University (BYU), and the PhD (1991) degree in Electrical Engineering from Stanford University. He was with the Department of Electrical and Computer Engineering at BYU from 1990–2007, where he served as Department Chair from 2003–06. During 1996–97, he held a joint appointment as a visiting scholar at Uppsala University and the Royal Institute of Technology in Sweden. From 2006–07, he was on leave working as

Vice President of Research for ArrayComm LLC in San Jose, California. Since 2007 he has been with the Electrical Engineering and Computer Science (EECS) Department at the University of California Irvine, where he is a Distinguished Professor and currently serving as Department Chair. Dr. Swindlehurst is a Fellow of the IEEE, during 2014–17 he was also a Hans Fischer Senior Fellow in the Institute for Advanced Studies at the Technical University of Munich, and in 2016, he was elected as a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA). He received the 2000 IEEE W. R. G. Baker Prize Paper Award, the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory, the 2006, 2010 and 2021 IEEE Signal Processing Society's Best Paper Awards, the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award, a Best Paper award at the 2020 and 2024 IEEE International Conferences on Communications, the 2022 Claude Shannon-Harry Nyquist Technical Achievement Award from the IEEE Signal Processing Society, and the 2024 Fred W. Ellersick Prize from the IEEE Communications Society. His research focuses on array signal processing for radar, wireless communications, and biomedical applications.