Novel Receiver Designs for Low-resolution ADC-Based Uplink MU-MIMO Systems with Unknown Channel Covariance Information

Rubin Jose Peter, *Member, IEEE* and Chandra R. Murthy, *Fellow, IEEE*

***Abstract*—It is known that joint channel estimation and soft symbol decoding enhance the performance of multiple-input multiple-output (MIMO) communication systems with low-resolution analog-to-digital converters (ADCs). However, existing techniques require accurate knowledge of channel statistics to be effective. In this paper, we consider an uplink multi-user MIMO system and develop novel algorithms for joint estimation of channel coefficients and statistics along with soft symbol decoding. Specifically, we develop a sequential processing technique, based on variational Bayesian inference, for jointly estimating the covariance matrices of all users' channels, the corresponding channel matrices, and soft symbols. We also develop an alternative block processing scheme for the same problem. Corresponding algorithms for sequential and block processing of unquantized data can be obtained as simplifications of our algorithms. On the analytical side, we derive a marginalized Bayesian Cramér-Rao Lower Bound (MB-CRLB) for covariance matrix estimation and elucidate the effect of the number of snapshots and SNR on the estimation error. Finally, we empirically evaluate the new algorithms and demonstrate their superior performance in scenarios where the channel statistics are unknown. The results also show that the mean squared error in channel covariance matrix estimation via the proposed algorithm closely follows the behavior of the MB-CRLB as a function of the SNR and number of frames.**

***Index Terms*—variational Bayesian inference, sequential processing, low-resolution ADCs, multiple-input multiple-output communications, Cramér-Rao lower bound.**

I. Introduction

The use of large multiple-input multiple-output (MIMO) arrays is now a well-accepted means to provide high data rate through spatial multiplexing and to compensate for the propagation loss in millimeter (mmWave) and terahertz (THz) bands communication [1]. The realization of such systems demands a large number of radio-frequency chains, increased power consumption, and high-precision analog-to-digital converters (ADCs). The power consumption of ADCs scales with the sampling rate and the number of quantization bits [2]. Hence, low-resolution ADCs are the practical solution for the energy-efficient communication in multi-user MIMO (MU-MIMO) systems [3].

In general, MIMO wireless channels exhibit spatial correlation, and knowledge of the spatial covariance matrix can help significantly boost the channel estimation and data detection performance. However, in low-resolution ADC based systems, all three tasks: channel covariance matrix estimation, channel estimation, and data detection, have to be performed using the received coarsely quantized samples. The goal of this paper is to address this challenge.

In a MU-MIMO uplink system with single-bit ADCs, convex relaxation of the exhaustive search based maximum likelihood detector can help to realize a computationally efficient receiver [4]. Joint channel estimation and symbol decoding using both quantized pilot and data symbols can help in improving the channel estimation accuracy. In this context, the generalized approximate message passing (GAMP) algorithm has been extended to joint channel estimation and soft symbol decoding [5]. Further performance improvement is obtained by exploiting the sparsity of the channel impulse response [6]. The bilinear GAMP (Bi-GAMP) algorithm has been extended to joint channel estimation and symbol decoding in reconfigurable intelligent surface (RIS)-aided massive MIMO systems with low-resolution ADCs [7]. An analytical framework was developed to evaluate the theoretical performance of the estimator in the large-system limit. The supervised learning technique, support vector machine (SVM), has been exploited for joint channel estimation and data detection using single-bit ADC measurements [8], with extensions to orthogonal frequency-division multiplexing (OFDM) systems and frequency-selective channels. The AdaBoost-based channel estimation and data detection can overcome the computational complexity of the SVM based method in single-bit MIMO OFDM systems [9]. Other approaches include Bussgang's decomposition-based channel estimation and symbol decoding to mitigate the effects of quantization noise [10], [11].

Joint channel estimation and soft symbol decoding using variational Bayesian framework was first introduced in [12]. A two-phase Bayesian inference-based approach for channel estimation followed by data detection in OFDM receivers with time-domain oversampling using low-resolution ADCs was presented in [13]. In variational Bayesian inference, the unquantized measurements, channel vectors and data symbols are considered as latent variables and an approximate posterior probability distribution is computed for parameter estimation and soft symbol decoding [14], [15]. If the variance of the additive noise in the received measurements is unknown, the variance can be treated as a latent variable and estimated within the variational Bayesian framework [16]. Overall, variational Bayesian inference is a principled approach, admits extensions to diverse scenarios, and is computationally more stable than approximate message passing and expectation

The authors are with the Dept. of ECE, Indian Institute of Science, Bengaluru 560012, India. Emails: rubinpeter@iisc.ac.in, cmurthy@iisc.ac.in.

Table I: Comparison with literature

Reference	Unknown Covariance Information	Low-Resolution ADC	Joint Estimation and Decoding	Analysis of Covariance Matrix Estimation Errors
[5]–[7], [12], [16]	×	✓	✓	×
[7]–[11], [25]	×	✓	×	×
[19]–[23]	✓	×	×	×
This paper	✓	✓	✓	✓

propagation-based algorithms [17].

The previous methods for joint channel estimation and symbol decoding such as [5]–[7], [16] assume perfect knowledge of the channel covariance matrix at the base station (BS); performance degrades when a nominal (or inaccurate) channel covariance matrix is employed. A deep unfolding-based channel estimation and soft symbol decoding under unknown channel statistics was developed in [18]. However, this approach requires computationally expensive retraining of the neural network whenever the statistics change. Algorithms for channel estimation in the presence of unknown channel covariance information have been investigated in [19]–[22]. Batch and online estimators based on the expectation-maximization approach were introduced for covariance matrix estimation [23]. Online and block-processing strategies based on variational Bayesian inference for time-varying channels were developed in [24]. However, these algorithms rely on full-resolution digitized sampling. Moreover, they require separate sets of pilot symbols for channel estimation and covariance matrix estimation.

In low-resolution ADC systems, quantization noise severely degrades channel estimation and symbol decoding performance. Since coarse quantization significantly reduces the information contained in the received measurements, conventional receivers that perform channel estimation and symbol decoding separately suffer substantial performance loss. As a result, joint channel estimation and symbol decoding becomes essential for effectively exploiting the limited information available in the quantized observations. Furthermore, prior channel statistics play a critical role in reliable inference, and inaccuracies in the assumed statistics can significantly affect both channel estimation and data detection. Most existing Bayesian approaches assume that the channel covariance matrix is perfectly known at the receiver. Such an assumption is often unrealistic in practice. Estimating the channel covariance matrix is especially difficult in the face of the quantization distortion introduced by low-resolution ADC architectures, and inaccurate/mismatched channel statistics can result in considerable performance degradation. These limitations motivate receiver algorithms that jointly estimate the channel, transmitted symbols, and channel statistics directly from the quantized measurements, thereby improving robustness against statistical mismatches while retaining the benefits of Bayesian inference. In this paper, we develop novel algorithms for jointly estimating the channel vectors and channel covariance matrices as well as decoding data symbols in an uplink MU-MIMO system, when the BS is equipped with low-resolution ADCs. Our main contributions are:

1) We consider joint channel estimation and soft symbol decoding in the uplink of a MU-MIMO system with low-resolution ADCs in the absence of second-order channel statistics. We develop the following algorithms within a variational Bayesian framework to simultaneously estimate the channel covariance matrices, the channel matrices, and decode the data symbols transmitted from the user equipments (UEs) to the BS.
   - We develop the *seq-QVB* algorithm, in which the channel covariance matrices of the users are estimated recursively within a variational Bayesian framework. The algorithm eliminates the need to store measurements corresponding to data and pilot symbols. In turn, it reduces latency in symbol decoding and is well-suited for practical applications. We also present the *seq-VBI* algorithm for processing full-resolution ADC measurements.
   - We develop the *batch-QVB* algorithm, where the quantized measurements are stored in memory and used to estimate the channel statistics. The estimated statistics are subsequently used for estimating the channel vectors and decoding the data symbols. We also introduce the *batch-VBI* algorithm for processing unquantized data.
2) We derive a marginalized Bayesian CRLB for the channel covariance estimation problem, which provides a tight lower bound on the empirical performance of our estimator. We obtain a closed-form expression for this bound by deriving upper and lower bounds for the marginalized Bayesian Fisher information matrix using Jensen's inequality. To the best of our knowledge, this approach for deriving a bound on a Bayesian estimator has not been reported previously in the literature.
3) We empirically quantify the impact of channel covariance uncertainty on few-bit ADC-based MU-MIMO receivers and demonstrate the benefits of the proposed joint estimation framework. The results show that receiver performance can degrade substantially when inaccurate channel statistics are employed. By learning the channel covariance matrices directly from quantized measurements, the proposed algorithms recover most of the performance loss associated with unknown channel statistics. In a representative scenario with 30 receive antennas at the BS and 6 UEs, the proposed algorithms achieve an SNR gain exceeding 4 dB at a symbol error rate (SER) of 10^{-2} and more than 10 dB reduction in

the normalized mean square error (NMSE) of the channel estimates relative to schemes using mismatched channel statistics. Moreover, the resulting performance closely approaches that of genie-aided receivers with perfect covariance knowledge.

Notation: The symbols a, $\mathbf{a}$, $\mathbf{A}$, and $\mathbb{I}$ denote a scalar, a vector, a matrix, and an identity matrix, respectively. diag$(\mathbf{a})$ is a diagonal matrix with diagonal entries given by $\mathbf{a}$, and blkdiag$(\mathbf{A}, \mathbf{B})$ is a block diagonal matrix with $\mathbf{A}$ and $\mathbf{B}$ as the diagonal blocks. The kth element of $\mathbf{a}$ is denoted by $[\mathbf{a}]_k$. The (i, j)th element of $\mathbf{R}_k$ is denoted by $[\mathbf{R}_k]_{ij}$, $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$, and $\mathbf{A} \succeq \mathbf{B}$ denotes that $\mathbf{A} - \mathbf{B}$ is positive semi-definite. tr$(\mathbf{A})$ denotes the trace of $\mathbf{A}$. $|\cdot|$, $\|\cdot\|$, $\|\cdot\|_F$, $[\cdot]^T$,$[\cdot]^*$, $[\cdot]^H$, $[\cdot]^\dagger$ and $\mathbb{E}(\cdot)$, denote the magnitude (or cardinality of a set or determinant of a square matrix), ℓ_2 norm, Frobenius norm, transpose, conjugate, conjugate transpose, pseudo inverse and the expectation, respectively. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a circularly symmetric complex Gaussian distribution with mean (vector) $\boldsymbol{\mu}$ and variance (covariance matrix) $\boldsymbol{\Sigma}$. The indicator function that outputs one when $z \in (y_1, y_2]$ and zero otherwise is denoted by $\mathbb{1}_{\{y_1 < z \leq y_2\}}$. The real and imaginary parts of a matrix $\mathbf{A}$ are denoted by $\Re\{\mathbf{A}\}$ and $\Im\{\mathbf{A}\}$, respectively.

II. System Model and Preliminaries

We consider uplink communication from K single-antenna UEs to a BS equipped with N_r receive antennas. The analog signals received by the BS antennas are down-converted to baseband and digitized using b-bit ADCs. Each communication frame consists of τ_p pilot and τ_d data symbols. We assume $\min(N_r, \tau_p) \geq K$, i.e., the K UEs can be spatially separated using the N_r antennas and the pilots transmitted by the K UEs can be chosen from an orthogonal set. The received discrete-time pilot and data symbols prior to quantization are modeled as [16], [19]

$$\begin{aligned} \mathbf{Z}_p^{t_n} &= \mathbf{H}^{t_n}\mathbf{X}_p^{t_n} + \mathbf{V}_p^{t_n} \in \mathbb{C}^{N_r \times \tau_p}, \\ \mathbf{Z}_d^{t_n} &= \mathbf{H}^{t_n}\mathbf{X}_d^{t_n} + \mathbf{V}_d^{t_n} \in \mathbb{C}^{N_r \times \tau_d}, \quad n = 1, 2, ..., N. \end{aligned} \tag{1}$$

Here, the unquantized pilot and data symbol matrices are denoted by $\mathbf{Z}_p^{t_n}$ and $\mathbf{Z}_d^{t_n}$, respectively. The uplink channel between UEs and BS during the nth frame is denoted by $\mathbf{H}^{t_n} \in \mathbb{C}^{N_r \times K}$, and N denotes the number of frames considered. The pilot and data symbols matrices in the nth frame are $\mathbf{X}_p^{t_n} = [\mathbf{x}_{p,1}^{t_n}, \ldots, \mathbf{x}_{p,\tau_p}^{t_n}] \in \mathbb{C}^{K \times \tau_p}$ and $\mathbf{X}_d^{t_n} = [\mathbf{x}_{d,1}^{t_n}, \ldots, \mathbf{x}_{d,\tau_d}^{t_n}] \in \mathbb{C}^{K \times \tau_d}$, respectively. The average power of the data and pilot symbols transmitted by the kth UE is denoted by P_k. The data symbols are drawn independently and identically distributed (i.i.d.) from constellation $\mathcal{M}$ of size $M = |\mathcal{M}|$. The matrices $\{\mathbf{V}_p^{t_n}\}_{n=1}^N$ and $\{\mathbf{V}_d^{t_n}\}_{n=1}^N$ represent the additive white Gaussian noise at the receive antennas. The elements of these matrices are i.i.d. as $\mathcal{CN}(0, \sigma_w^2)$.

We assume a correlated quasi-static Rayleigh fading channel between the UEs and the BS. The channel vector from each UE to the BS remains constant within a frame and varies in an i.i.d. fashion across frames, and the channel statistics remain unchanged during the N frames. They are independent but not identically distributed across the K UEs since their path losses and channel covariance matrices could be different. Then, the probability density function (pdf) of the channel matrices between the UEs and the BS factorizes as

$$p(\{\mathbf{H}^{t_n}\}_{n=1}^N) = \prod_{n=1}^N p(\mathbf{H}^{t_n}) = \prod_{n=1}^N \prod_{k=1}^K p(\mathbf{h}_k^{t_n}). \tag{2}$$

The pdf of the channel vector $\mathbf{h}_k^{t_n}$ given its *(unknown)* covariance matrix $\mathbf{R}_k$ is $\mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$, i.e.,

$$p(\mathbf{h}_k^{t_n}) = p(\mathbf{h}_k^{t_n}|\mathbf{R}_k) = \frac{1}{\pi^{N_r}|\mathbf{R}_k|}\exp(-\mathbf{h}_k^{t_n\,H}\mathbf{R}_k^{-1}\mathbf{h}_k^{t_n}). \tag{3}$$

We let $\mathbf{R} = \text{blkdiag}(\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_K) \in \mathbb{C}^{N_r K \times N_r K}$ denote the combined block-diagonal channel covariance matrix. In the above equation, with a slight abuse of notation, we use $p(\mathbf{h}_k^{t_n}|\mathbf{R}_k)$ to denote the pdf of $\mathbf{h}_k^{t_n}$ conditioned on the covariance matrix $\mathbf{R}_k$. Now, to develop a variational Bayesian algorithm for estimating $\mathbf{R}$, we impose a complex inverse Wishart distribution as its prior pdf [26]. The complex inverse Wishart distribution is the conjugate prior for the covariance matrix of a complex Gaussian random vector and enables a tractable update step in variational Bayesian inference. Hence, the prior pdf of $\mathbf{R}$ is given by

$$\begin{aligned} p(\mathbf{R}) &= \prod_{k=1}^K p(\mathbf{R}_k), \\ p(\mathbf{R}_k) &= \frac{|\boldsymbol{\Psi}|^\nu}{\mathcal{C}\Gamma_{N_r}(\nu)}|\mathbf{R}_k|^{-(\nu+N_r)}\exp\big(-\text{tr}(\boldsymbol{\Psi}\mathbf{R}_k^{-1})\big), \\ \mathcal{C}\Gamma_{N_r}(\nu) &= \pi^{\frac{1}{2}N_r(N_r-1)}\prod_{j=1}^{N_r}\Gamma(\nu - j + 1), \end{aligned} \tag{4}$$

where $\nu \geq N_r + 1$ denotes the number of degrees of freedom (DoF), and $\Gamma(k) = (k-1)!$ denotes the Gamma function. The mean matrix of the above distribution is $\mathbb{E}_{p(\mathbf{R}_k)}(\mathbf{R}_k) = \boldsymbol{\Psi}/(\nu - N_r)$, where $\boldsymbol{\Psi} \in \mathbb{C}^{N_r \times N_r}$ denotes the scale matrix of the complex inverse Wishart distribution.[1] Note that, in our simulation studies, the covariance matrices are not generated from the assumed complex inverse Wishart prior. Nevertheless, the proposed algorithms achieve good estimation accuracy, indicating that their performance is not particularly sensitive to the assumed prior pdf. This robustness can be attributed to the fact that the inverse Wishart distribution is used only to model the uncertainty in the covariance matrices. Consequently, the covariance estimates are determined primarily by the information contained in the observed quantized measurements rather than by the specific form of the assumed prior distribution. As the amount of available data increases, the influence of the prior further diminishes, and the posterior distribution becomes increasingly data-driven.

In this work, we address scenarios with and without quantization using low-resolution ADCs. When low-resolution ADCs are employed, the in-phase and quadrature components of the baseband signal in (1) are fed separately to b-bit ADCs, and the quantized data is then used for further processing.

[1] We set the hyperparameters $\nu = 2N_r$ and $\boldsymbol{\Psi} =$ the identity matrix in our simulations, but the performance of the algorithm developed in this work is insensitive to the choice of these parameters. They can be set to any convenient value without significantly affecting the performance.

Let $[-V, V]$ denote the typical voltage range at the input of the ADC. In a b-bit uniform quantizer, the dynamic range of the ADC is divided into $J = 2^b$ intervals of equal width. Therefore, the width of a particular quantization interval is $\Delta = \frac{2V}{J}$, and the corresponding jth quantization interval is $[-V + (j-1)\Delta, -V + j\Delta)$. Suppose a real-valued sample z_i falls within the interval $[-V + (j-1)\Delta, -V + j\Delta)$; then the corresponding quantizer output is $\mathcal{Q}(z_i) = L_j \triangleq -V + (j - 0.5)\Delta$. Thus, the quantization operation using a low-resolution ADC can be expressed as

$$\Re\{\mathbf{Y}_p^{t_n}\} = \mathcal{Q}(\Re\{\mathbf{Z}_p^{t_n}\}), \Im\{\mathbf{Y}_p^{t_n}\} = \mathcal{Q}(\Im\{\mathbf{Z}_p^{t_n}\}),$$
$$\Re\{\mathbf{Y}_d^{t_n}\} = \mathcal{Q}(\Re\{\mathbf{Z}_d^{t_n}\}), \Im\{\mathbf{Y}_d^{t_n}\} = \mathcal{Q}(\Im\{\mathbf{Z}_d^{t_n}\}).$$

Here, the quantization operation $\mathcal{Q}(.)$ is applied separately to every element of its argument.

In joint processing, the MMSE estimation of the covariance matrix $\mathbf{R}$ and the channel matrix $\mathbf{H}^{t_N}$ requires the computation of the posterior pdfs $p(\mathbf{R} \mid \{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^N)$ and $p(\mathbf{H}^{t_N} \mid \{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^N)$, respectively. Furthermore, soft-symbol decoding requires the posterior distribution $p(\mathbf{X}_d^{t_N} \mid \{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^N)$. Generally, the computation of these posterior pdfs is intractable due to the multidimensional integrations required for marginalizing the joint pdf. Hence, we compute an approximate posterior pdf under the variational Bayesian framework to jointly estimate the channel matrices $\{\mathbf{H}^{t_n}\}_{n=1}^N$, the channel covariance matrix $\mathbf{R}$, and decode the data symbols $\{\mathbf{X}_d^{t_n}\}_{n=1}^N$. As a point of departure, we provide a brief description of the variational Bayesian inference principle in the following subsection.

A. Variational Bayesian Inference

Bayesian estimators such as the maximum a posteriori probability (MAP) and minimum mean squared error (MMSE) estimator require the computation of the posterior pdf of the random variables (RVs) to be estimated. Let $\mathcal{Y}$ denote a set of measurements, and $\mathcal{X}$ denote a set of unknown RVs. Then, by Bayes' theorem, the posterior pdf of $\mathcal{X}$ given $\mathcal{Y}$ is

$$p(\mathcal{X}|\mathcal{Y};\omega) = \frac{p(\mathcal{Y}|\mathcal{X};\omega)p(\mathcal{X};\omega)}{\int p(\mathcal{Y}|\mathcal{X};\omega)p(\mathcal{X};\omega)d\mathcal{X}}, \quad (5)$$

where $p(\mathcal{Y}|\mathcal{X};\omega)$ denotes the likelihood function, and $p(\mathcal{X};\omega)$ denotes the prior pdf. All deterministic parameters involved in the pdfs are included in the set ω. The computation of the multidimensional integral in the denominator of (5) is generally intractable. Consequently, a closed-form expression of the posterior pdf is hard to compute. In variational inference, an approximate posterior pdf $q(\mathcal{X})$ is obtained by minimizing the Kullback-Leibler (KL) divergence between the approximate and true posterior distributions. To this end, note that the marginalized log-likelihood function (evidence) of the measurements $\log p(\mathcal{Y};\omega)$ is given by

$$\log p(\mathcal{Y};\omega) = \mathbb{E}_{q(\mathcal{X})}\left[\log\left(\frac{p(\mathcal{Y},\mathcal{X};\omega)}{q(\mathcal{X})}\right)\right] + \mathrm{KL}(q(\mathcal{X})||p(\mathcal{X}|\mathcal{Y};\omega)), \quad (6)$$

where $\mathrm{KL}(q(\mathcal{X})||p(\mathcal{X}|\mathcal{Y};\omega))$ denotes the Kullback-Leibler (KL) divergence between the approximate pdf $q(\mathcal{X})$ and the true posterior pdf $p(\mathcal{X}|\mathcal{Y};\omega)$. Since the left-hand side of (6) does not depend on $q(\mathcal{X})$, minimizing the (non-negative) KL divergence between $q(\mathcal{X})$ and $p(\mathcal{X}|\mathcal{Y};\omega)$ maximizes the evidence lower bound (ELBO) $\mathbb{E}_{q(\mathcal{X})}\big[\log(\frac{p(\mathcal{Y},\mathcal{X};\omega)}{q(\mathcal{X})})\big]$. The computation of $q(\mathcal{X})$ can be made tractable by partitioning the latent variables $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_R\}$ and assuming a factorized form

$$q(\mathcal{X}) = q_1(\mathcal{X}_1)q_2(\mathcal{X}_2)\ldots q_R(\mathcal{X}_R). \quad (7)$$

If the joint pdf $p(\mathcal{Y}, \mathcal{X};\omega)$ of the observed and latent variables belongs to the exponential family of distributions (as is the case here), then the rth factor in (7) can be computed from [27]

$$q_r(\mathcal{X}_r) \propto \exp(\mathbb{E}_{\prod_{j\neq r} q_j(\mathcal{X}_j)} \log(p(\mathcal{Y},\mathcal{X};\omega))), \quad (8)$$

where the expectation of the logarithm of the joint pdf is evaluated using the approximate pdf of all other factors. In particular, the moments of the approximate pdf $q(\mathcal{X}_r)$ are computed using (8), which depends on the moments of other factors. Hence, algorithms based on variational inference are typically iterative: we initialize the approximate pdfs arbitrarily, and update each approximate posterior pdf alternately using the approximate pdfs of the other factors.

With this background, we first present a general method to jointly estimate the channel and soft symbols when the channel statistics are unknown.

B. Joint Channel Estimation and Soft Symbol Decoding in the Absence of Channel Statistics

In joint channel estimation and soft symbol decoding, the goal is to estimate the posterior statistics of the channel and data symbols using the received pilot and data symbols. Further, in the absence of channel statistics, the channel covariance matrix also needs to be estimated using the channel estimates obtained over multiple frames. The channel covariance estimate can be formed either in a sequential fashion or by processing a batch of consecutive frames. In this paper, we explore both options using quantized and unquantized measurements.

In a variational Bayesian framework, the data symbols, $\{\mathbf{X}_d^{t_n}\}_{n=1}^N$, the channel vectors, $\{\mathbf{H}^{t_n}\}_{n=1}^N$, and their covariance matrices, $\mathbf{R}$, are latent variables. Further, since the receiver has access to only quantized data, the unquantized pilot and data measurements $\{\mathbf{Z}_p^{t_n}\}_{n=1}^N$ and $\{\mathbf{Z}_d^{t_n}\}_{n=1}^N$ are also latent variables. The pilot and data samples output by the low-resolution ADCs, namely, $\{\mathbf{Y}_p^{t_n}\}_{n=1}^N$ and $\{\mathbf{Y}_d^{t_n}\}_{n=1}^N$, are observed variables. A Bayesian network graphical model for the MU-MIMO system is shown in Fig. 1, where observed variables are represented using shaded circles and latent variables are represented using unshaded circles. Square boxes denote deterministic parameters such as the noise variance σ_w^2, the transmitted pilot symbols $\{\mathbf{X}_p^{t_n}\}_{n=1}^N$, and the prior parameters of the channel covariance matrix, namely, the scale matrix $\mathbf{\Psi}$ and the DoF ν.

If no constraints are imposed on the posterior pdf, the distribution that maximizes the evidence $\log p(\mathcal{Y};\omega)$ is $q(\mathcal{X}) = p(\mathcal{X} \mid \mathcal{Y};\omega)$. Since the exact posterior $p(\mathcal{X} \mid \mathcal{Y};\omega)$ is analytically intractable, mean-field variational Bayesian inference approximates it using a factorized family of distributions as in (7) and selects the member that minimizes the KL divergence from the true posterior. This factorization also simplifies the computation of the expectation in (8). In this framework, the posterior distribution of the unquantized measurements, data symbols, channel vectors, and their covariance matrices conditioned on all quantized pilot and data observations is approximated by a factorized distribution over these variables. The specific approximations are detailed in the following sections. Further details on mean-field variational Bayesian inference can be found in [27].

In the sequel, in addition to quantized measurements being available at the receiver, we also consider the case where the receiver directly observes unquantized measurements. The Bayesian graphical model for this case is a simplified version of Fig. 1, where $\{\mathbf{Z}_p^{t_n}\}_{n=1}^N$ and $\{\mathbf{Z}_d^{t_n}\}_{n=1}^N$ are observed variables, and $\{\mathbf{Y}_p^{t_n}\}_{n=1}^N$ and $\{\mathbf{Y}_d^{t_n}\}_{n=1}^N$ are eliminated.

In Sec. III, we present the *seq-QVB* and *seq-VBI* algorithms, where the measurements are processed sequentially as they arrive, while in Sec. IV, we present the *batch-QVB* and *batch-VBI* algorithms, which use a batch of measurements for estimating the channel statistics.

III. Sequential Estimation and Symbol Decoding

Here, we describe the *seq-QVB* and *seq-VBI* algorithms for sequentially estimating the channel covariance matrix, channel vectors, and soft symbol decoding. The *seq-QVB* algorithm accounts for the quantization noise and computes an approximate posterior pdf of the latent variables, while the *seq-VBI* algorithm is designed for unquantized data. In the sequel, we consider that the processing of the first $N-1$ communication frames has been completed, and describe the processing of the (current) Nth frame.

In a sequential algorithm, an approximate posterior pdf of all latent variables is computed using the measurements corresponding to the pilot and data symbols of the Nth frame and the approximate posterior pdf of $\mathbf{R}$ from the previous frame. The processing of the Nth frame also involves updating $\mathbf{R}$, estimating the corresponding channels (e.g., as the mean of the posterior distribution), and soft symbol decoding.

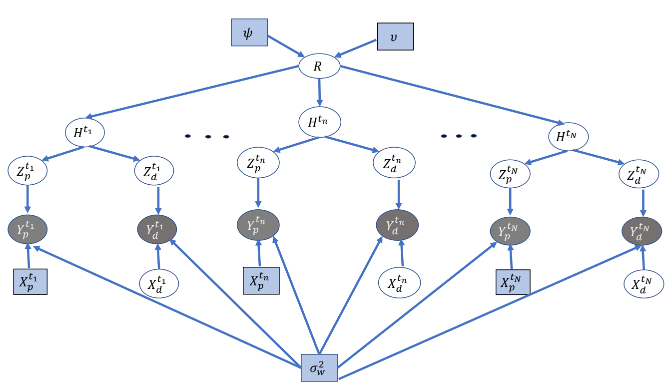

Fig. 1: Bayesian graphical model for our system setup.

A. Sequential Processing using Quantized Measurements

The *seq-QVB* algorithm recursively computes the following approximate factorized posterior pdf of $\{\mathbf{H}^{t_N}, \mathbf{X}_d^{t_N}, \mathbf{R}\}$ conditioned on all quantized measurements received up to and including time frame t_N:

$$\begin{aligned}
&p(\mathbf{Z}_d^{t_N}, \mathbf{Z}_p^{t_N}, \mathbf{X}_d^{t_N}, \mathbf{H}^{t_N}, \mathbf{R} | \boldsymbol{\Xi}; \{\mathbf{X}_p^{t_n}\}_{n=1}^N, \sigma_w^2) \approx \\
&(\prod_{s=1}^{\tau_p}\prod_{r=1}^{N_r} q(z_{p,rs}^{t_N}))(\prod_{s=1}^{\tau_d}\prod_{r=1}^{N_r} q(z_{d,rs}^{t_N}))(\prod_{s=1}^{\tau_d}\prod_{k=1}^{K} q(x_{d,ks}^{t_N})) \\
&\prod_{k=1}^{K} q(\mathbf{h}_k^{t_N}) q(\mathbf{R}_k^{t_N}),
\end{aligned} \tag{9}$$

where the set $\boldsymbol{\Xi} = \{\{\mathbf{Y}_p^{t_n}\}_{n=1}^N, \{\mathbf{Y}_d^{t_n}\}_{n=1}^N\}$ contains the quantized measurements used to compute the approximate pdfs of the latent variables of the Nth frame. The factors $\{q(z_{p,rs}^{t_N})\}_{s=1,r=1}^{s=\tau_p,r=N_r}, \{q(z_{d,rs}^{t_N})\}_{s=1,r=1}^{s=\tau_d,r=N_r}$ denote the approximate posterior pdfs of the unquantized measurements. The approximate posterior pdfs of the data symbols used for the soft symbol decoding are denoted by $\{q(x_{d,kt}^{t_N})\}_{t=1,k=1}^{t=\tau_d,k=K}$, while that of the channel vectors is denoted by $\{q(\mathbf{h}_k^{t_N})\}_{k=1}^K$. The pdf $q(\mathbf{R}^{t_N}) = \prod_{k=1}^K q(\mathbf{R}_k^{t_N})$ approximates the posterior pdf of the channel covariance matrix $\mathbf{R}$ conditioned on $\boldsymbol{\Xi}$, i.e., on the observations available up to time t_N.

Using the conditional independence among random variables as depicted in Fig. 1, the joint pdf of the latent and observed random variables is

$$\begin{aligned}
&p(\mathbf{Z}_d^{t_N}, \mathbf{Z}_p^{t_N}, \mathbf{X}_d^{t_N}, \mathbf{H}^{t_N}, \mathbf{R}, \{\mathbf{Y}_d^{t_n}, \mathbf{Y}_p^{t_n}\}_{n=1}^N) \\
&= p(\mathbf{Y}_p^{t_N}|\mathbf{Z}_p^{t_N}) p(\mathbf{Y}_d^{t_N}|\mathbf{Z}_d^{t_N}) p(\mathbf{Z}_p^{t_N}|\mathbf{H}^{t_N}; \mathbf{X}_p^{t_N}, \sigma_w^2) \\
&\times p(\mathbf{Z}_d^{t_N}|\mathbf{H}^{t_N}, \mathbf{X}_d^{t_N}; \sigma_w^2) p(\mathbf{X}_d^{t_N}) p(\mathbf{H}^{t_N}|\mathbf{R}) \\
&\times p(\mathbf{R}|\{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^{N-1}; \{\mathbf{X}_p^{t_n}\}_{n=1}^{N-1}, \sigma_w^2) \\
&\times p(\{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^{N-1}; \{\mathbf{X}_p^{t_n}\}_{n=1}^{N-1}, \sigma_w^2).
\end{aligned} \tag{10}$$

We wish to use the above and (8) to *sequentially* update the approximate posterior pdfs of the latent variables in (9). However, the posterior pdf of $\mathbf{R}$ based on the previous $N-1$ frames, $p(\mathbf{R}|\{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^{N-1}; \{\mathbf{X}_p^{t_n}\}_{n=1}^{N-1}, \sigma_w^2)$, is not amenable to sequential estimation, as it depends on all the past observations. Therefore, we can modify (10) by replacing $p(\mathbf{R}|\{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^{N-1}; \{\mathbf{X}_p^{t_n}\}_{n=1}^{N-1}, \sigma_w^2)$ with its approximation, $q(\mathbf{R}^{t_{N-1}})$, available from the previous iteration of the *seq-QVB* algorithm.[2] This helps to obtain a tractable sequential algorithm for the estimation of the channel statistics. The joint pdf in (10) is thus approximated as

$$\begin{aligned}
&p(\mathbf{Z}_d^{t_N}, \mathbf{Z}_p^{t_N}, \mathbf{X}_d^{t_N}, \mathbf{H}^{t_N}, \mathbf{R}, \{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^N) \\
&\approx p(\mathbf{Y}_p^{t_N}|\mathbf{Z}_p^{t_N}) p(\mathbf{Y}_d^{t_N}|\mathbf{Z}_d^{t_N}) p(\mathbf{Z}_p^{t_N}|\mathbf{H}^{t_N}; \mathbf{X}_p^{t_N}, \sigma_w^2) \\
&\times p(\mathbf{Z}_d^{t_N}|\mathbf{H}^{t_N}, \mathbf{X}_d^{t_N}; \sigma_w^2) p(\mathbf{X}_d^{t_N}) p(\mathbf{H}^{t_N}|\mathbf{R}) \\
&\times q(\mathbf{R}^{t_{N-1}}) p(\{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^{N-1}; \{\mathbf{X}_p^{t_n}\}_{n=1}^{N-1}, \sigma_w^2).
\end{aligned} \tag{11}$$

Next, we discuss the computation of each factor in (9) by minimizing the KL divergence between the true and

[2] Note that $q(\mathbf{R}^{t_{N-1}})$ is different from the factor $q(\mathbf{R}^{t_N})$ in (9).

approximate posterior distributions. That is, we use (8) and (11) (note that all factors in (11) are known) to compute the factors of the approximate pdf in (9). In the following, we briefly present the derivation of the first factor, $q(\mathbf{R}^{t_N})$. The other factors can be derived following a procedure similar to [12]; we only present the final expressions here.

1) Computation of $q(\mathbf{R}^{t_N})$

Following the variational Bayesian estimation principle in (8) and eliminating terms that do not depend on $\mathbf{R}_k^{t_N}$, the factor $q(\mathbf{R}_k^{t_N})$ that maximizes the ELBO is

$$\log q(\mathbf{R}_k^{t_N}) \propto \langle \log(p(\mathbf{h}_k^{t_N}|\mathbf{R}_k)) + \log(q(\mathbf{R}_k^{t_{N-1}}))\rangle, \quad (12)$$

where $\langle\cdot\rangle$ denotes the expectation of the joint pdf with respect to (w.r.t.) all factors except $q(\mathbf{R}_k^{t_N})$. Expanding the right hand side using (3) and (4) and evaluating the expectations yields

$$\begin{aligned}\log q(\mathbf{R}_k^{t_N}) &\propto -\log|\mathbf{R}_k| - \mathrm{tr}(\mathbf{R}_k^{-1}(\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N\,H} + \boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N})) \\ &- (\nu^{t_{N-1}} + N_r)\log|\mathbf{R}_k| - \mathrm{tr}(\mathbf{R}_k^{-1}\boldsymbol{\Psi}_k^{t_{N-1}}), \end{aligned} \quad (13)$$

where $\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}$ and $\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}$ denote the mean vector and covariance matrix of $q(\mathbf{h}_k^{t_N})$, respectively. Hence, the approximate pdf $q(\mathbf{R}_k^{t_N})$ is a complex inverse Wishart pdf ($q(\mathbf{R}_k^{t_N}) \sim \mathcal{CW}^{-1}(\boldsymbol{\Psi}_k^{t_N}, \nu^{t_N}, N_r)$), with DoF and scale matrix

$$\begin{aligned}\nu^{t_N} &= \nu^{t_{N-1}} + 1, \\ \boldsymbol{\Psi}_k^{t_N} &= \boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N\,H} + \boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N} + \boldsymbol{\Psi}_k^{t_{N-1}},\end{aligned} \quad (14)$$

respectively. Note that the DoF are updated *recursively* according to the first equation in (14). The mean of $\mathbf{R}_k^{t_N}$, denoted by $\mathbf{S}_k^{t_N}$, can also be updated recursively as

$$\mathbf{S}_k^{t_N} = \frac{\boldsymbol{\Psi}_k^{t_N}}{\alpha^{t_N}} = \frac{1}{\alpha^{t_N}}(\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N\,H} + \boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}) + \frac{\alpha^{t_N}-1}{\alpha^{t_N}}\mathbf{S}_k^{t_{N-1}}, \quad (15)$$

where $\alpha^{t_N} = \nu^{t_N} - N_r$. In the first frame ($N = 1$), we set $q(\mathbf{R}^{t_0})$ to be the prior distribution $p(\mathbf{R})$ in (4).

2) Computation of $q(\mathbf{H}^{t_N})$

Using (8), the approximate posterior pdf $q(\mathbf{h}_k^{t_N}) = \mathcal{CN}(\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}, \boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N})$, where

$$\begin{aligned}&\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N} = \\ &\left(\frac{\sum_{t=1}^{\tau_p}|x_{p,kt}^{t_N}|^2 + \sum_{t=1}^{\tau_d}\langle|x_{d,kt}^{t_N}|^2\rangle}{\sigma_w^2}\mathbf{I}_{N_r} + \langle(\mathbf{R}_k^{t_N})^{-1}\rangle\right)^{-1}, \\ &\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N} = \frac{1}{\sigma_w^2}\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}\sum_{t=1}^{\tau_p}(\langle\mathbf{z}_{p,t}^{t_N}\rangle - \sum_{k'\neq k}\boldsymbol{\mu}_{\mathbf{h}_{k'}}^{t_N}x_{p,k't}^{t_N})x_{p,kt}^{t_N\,*} \\ &+ \frac{1}{\sigma_w^2}\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}\sum_{t=1}^{\tau_d}(\langle\mathbf{z}_{d,t}^{t_N}\rangle - \sum_{k'\neq k}\boldsymbol{\mu}_{\mathbf{h}_{k'}}^{t_N}\langle x_{d,k't}^{t_N}\rangle)\langle x_{d,kt}^{t_N\,*}\rangle, \\ &\text{with } \langle(\mathbf{R}_k^{t_N})^{-1}\rangle = ((\mathbf{S}_k^{t_N})^{-1})\frac{\alpha^{t_N}+N_r}{\alpha^{t_N}}.\end{aligned} \quad (16)$$

In (16), we have used the fact that, when $\mathbf{R}_k$ is distributed as a complex inverse Wishart pdf with scale matrix $\boldsymbol{\Psi}$ and DoF ν, the expectation of $\mathbf{R}_k^{-1}$ is $\nu\boldsymbol{\Psi}^{-1}$. Note that the mean vector $\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}$ of $q(\mathbf{h}_k^{t_N})$ can serve as a channel estimate when point estimates are desired.

Algorithm 1: *seq-QVB*

Input: $\mathbf{Y}_d^{t_N}$, $\mathbf{Y}_p^{t_N}$, $\{\mathbf{S}_k^{t_{N-1}}\}_{k=1}^K$, $\alpha^{t_{N-1}}$, $\mathrm{Iter}_{\max}$

Initialize: $\{\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}, \boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}\}_{k=1}^K$, $\{\langle x_{d,kt}^{t_N}\rangle\}_{k=1,t=1}^{k=K,t=\tau_d}$

for $i = 1, 2, \ldots, Iter_{\max}$ **do**

  Compute the moments of $q(\mathbf{Z}_p^{t_N})$ and $q(\mathbf{Z}_d^{t_N})$ using (17) and (18)

  for $k = 1, 2\ldots, K$ **do**

    Compute $\mathbf{S}_k^{t_N}$ and α^{t_N} using (15)

    Compute the moments of q($\mathbf{h}_k^{t_N}$) using (16)

    Compute $q(x_{d,kt}^{t_N})$ using (19)

  end

end

Decode the data symbols using (20)

Output: $\alpha^{t_N}, \{\mathbf{S}_k^{t_N}\}_{k=1}^K, \{\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\}_{k=1}^K, \{\hat{x}_{d,kt}^{t_N}\}_{t=1,k=1}^{t=\tau_d,k=K}$

---

3) Computation of $q(\mathbf{Z}_p^{t_N})$ *and* $q(\mathbf{Z}_d^{t_N})$

In the graphical model shown in Fig. 1, the unquantized measurements corresponding to the pilot symbols, $\mathbf{Z}_p^{t_N}$, and data symbols, $\mathbf{Z}_d^{t_N}$, are latent variables. Using the factorization $q(\mathbf{Z}_p^{t_N}) = \prod_{t=1}^{\tau_p}\prod_{r=1}^{N_r} q(z_{p,rt}^{t_N})$ and (8), the factor $q(z_{p,rt}^{t_N})$ can be obtained as follows

$$\begin{aligned}\log q(z_{p,rt}^{t_N}) &= \langle\log(p(y_{p,rt}^{t_N}|z_{p,rt}^{t_N})p(z_{p,rt}^{t_N}|\mathbf{H}^{t_N};\mathbf{X}_p^{t_N},\sigma_w^2))\rangle \\ &\quad + \text{const.}, \\ &= \log(\mathbb{1}_{\left\{y_{p,rt}^{t_N,\text{lo}} < z_{p,rt}^{t_N} \leq y_{p,rt}^{t_N,\text{hi}}\right\}}) - \frac{1}{\sigma_w^2}\langle|z_{p,rt}^{t_N} - \sum_{k=1}^K h_{rk}^{t_N}x_{p,kt}|^2\rangle \\ &\quad + \text{const.}\end{aligned}$$

We see that $q(z_{p,rt}^{t_N})$ is a truncated complex normal distribution. Its mean $\langle z_{p,rt}^{t_N}\rangle$ is

$$\mu_{p,rt}^{t_N} + \frac{\phi(\frac{\{y_{p,rt}^{t_N,\text{lo}} - \mu_{p,rt}^{t_N}\}}{\sigma_w/\sqrt{2}}) - \phi(\frac{\{y_{p,rt}^{t_N,\text{hi}} - \mu_{p,rt}^{t_N}\}}{\sigma_w/\sqrt{2}})}{\Phi(\frac{\{y_{p,rt}^{t_N,\text{hi}} - \mu_{p,rt}^{t_N}\}}{\sigma_w/\sqrt{2}}) - \Phi(\frac{\{y_{p,rt}^{t_N,\text{lo}} - \mu_{p,rt}^{t_N}\}}{\sigma_w/\sqrt{2}})}, \quad (17)$$

where $\mu_{p,rt}^{t_N} = \sum_{k=1}^K \left[\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\right]_r x_{p,kt}^{t_N}$. The functions $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of a standard complex normal distribution, respectively. The lower and upper limit of the quantization interval corresponding to the measurement $y_{p,rt}^{t_N}$ are denoted by $y_{p,rt}^{t_N,\text{lo}}$ and $y_{p,rt}^{t_N,\text{hi}}$, respectively.

Similarly, we can compute $q(\mathbf{Z}_d^{t_N})$ also. The pdf $q(\mathbf{Z}_d^{t_N})$ is a truncated complex normal, with mean $\langle z_{d,rt}^{t_N}\rangle$ is given by

$$\mu_{d,rt}^{t_N} + \frac{\phi(\frac{\{y_{d,rt}^{t_N,\text{lo}} - \mu_{d,rt}^{t_N}\}}{\sigma_w/\sqrt{2}}) - \phi(\frac{\{y_{d,rt}^{t_N,\text{hi}} - \mu_{d,rt}^{t_N}\}}{\sigma_w/\sqrt{2}})}{\Phi(\frac{\{y_{d,rt}^{t_N,\text{hi}} - \mu_{d,rt}^{t_N}\}}{\sigma_w/\sqrt{2}}) - \Phi(\frac{\{y_{d,rt}^{t_N,\text{lo}} - \mu_{d,rt}^{t_N}\}}{\sigma_w/\sqrt{2}})}, \quad (18)$$

where $\mu_{d,rt}^{t_N} = \sum_{k=1}^K \left[\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\right]_r \langle x_{d,kt}^{t_N}\rangle$. In the above, $y_{d,rt}^{t_N,\text{lo}}$ and $y_{d,rt}^{t_N,\text{hi}}$ denote the lower and upper limits of the quantization interval corresponding to $y_{d,rt}^{t_N}$, respectively.

4) Computation of $q(\mathbf{X}_d^{t_N})$

Soft symbol decoding of the data symbol in the tth symbol interval of the Nth frame is carried out using the approximate posterior pdf $q(x_{d,kt}^{t_N})$ in (9). The data symbols are drawn from the constellation $\mathcal{M}$. It can be shown that $q(x_{d,kt}^{t_N})$ follows a Boltzmann distribution, given by

$$\begin{aligned} &q(x_{d,kt}^{t_N}) = \text{softmax}(f_{kt}^{t_N}(x_{d,kt}^{t_N})), \text{ where } f_{kt}^{t_N}(s_m) = \\ &-\frac{1}{\sigma_w^2}((\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N})^H \boldsymbol{\mu}_{\mathbf{h}_k}^{t_N} + \text{tr}(\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}))|s_m|^2 + \ln p(s_m) \\ &+\frac{2}{\sigma_w^2}\Re\{(\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N})^H(\langle \mathbf{z}_{d,t}^{t_N}\rangle - \sum_{k'\neq k} \boldsymbol{\mu}_{\mathbf{h}_{k'}}^{t_N}\langle x_{d,k't}^{t_N}\rangle)s_m^*\}, \end{aligned} \tag{19}$$

and $p(s_m)$ is the prior probability of the symbol $s_m \in \mathcal{M}$. If hard decisions are required, we may use

$$\hat{x}_{d,kt}^{t_N} = \underset{x_{d,kt}^{t_N}\in\mathcal{M}}{\arg\max}\ q(x_{d,kt}^{t_N}). \tag{20}$$

The *seq-QVB* algorithm is shown in Algorithm 1. The vectors $\{\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\}_{k=1}^K$ can be initialized as $(\mathbf{X}_p^{t_N\,T} \otimes \mathbb{I}_{N_r})^\dagger \text{vec}(\mathbf{Y}_p)$ (maximum likelihood estimates of the channel vectors ignoring quantization noise). The matrices $\{\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}\}_{k=1}^K$ can be initialized as all zeros. For initializing the data symbols, a linear MMSE decoder based on the initial channel estimates can be used.

Note that the computation of $q_{kt}(x_{d,kt}^{t_N})$ in (20) requires evaluating the metric for all constellation symbols. The complexity of computing soft symbol estimates can be significantly reduced by using standard *max-log-MAP* approximations [28]. In addition, for square M-QAM constellations, when hard symbol estimates are required, the final decision can be efficiently implemented using a nearest-neighbor slicer, where the real and imaginary components are decoded separately. This avoids an exhaustive search over all M constellation points and reduces the detection complexity from $\mathcal{O}(M)$ to $\mathcal{O}(\sqrt{M})$. Such complexity reduction methods can be particularly useful for high-order modulation schemes employed in modern wireless systems. Furthermore, since the symbol-wise posterior computations are independent across users and symbol intervals, they can be performed in parallel, further reducing the overall processing latency. Consequently, the proposed receiver architecture remains computationally attractive even for systems employing large constellation sizes and a large number of simultaneously active users.

B. Sequential Processing using Unquantized Measurements

The *seq-QVB* algorithm can be modified to perform sequential estimation using full-resolution ADC outputs. The observed variables in this unquantized setting are denoted by $\{\mathbf{Z}_d^{t_n}, \mathbf{Z}_p^{t_n}\}_{n=1}^N$. The factorized approximate posterior pdf with unquantized measurements is given by

$$\begin{aligned} &p(\mathbf{X}_d^{t_N}, \mathbf{H}^{t_N}, \mathbf{R}|\{\mathbf{Z}_p^{t_n}, \mathbf{Z}_d^{t_n}\}_{n=1}^N; \{\mathbf{X}_p^{t_n}\}_{n=1}^N, \sigma_w^2) \approx \\ &(\prod_{t=1}^{\tau_d}\prod_{k=1}^K q(x_{d,kt}^{t_N}))\prod_{k=1}^K q(\mathbf{h}_k^{t_N}) q(\mathbf{R}_k^{t_N}). \end{aligned} \tag{21}$$

The update equations with the above factorization, called the *seq-VBI* algorithm in the sequel, are similar to *seq-QVB*, with the necessary modifications briefly outlined below.

Computation of $q(\mathbf{R}^{t_N})$: From (10), the factors involving $\mathbf{R}$ do not depend on $\mathbf{Z}_p^{t_N}$ and $\mathbf{Z}_d^{t_N}$. Hence, the update equation in (15) for the channel covariance matrix remains unchanged; it depends only on the moments of $q(\mathbf{H}^{t_N})$ and $q(\mathbf{R}^{t_{N-1}})$.

Computation of $q(\mathbf{H}^{t_N})$: The factor $q(\mathbf{H}^{t_N})$ in the *seq-VBI* algorithm is a circularly symmetric complex normal distribution, similar to *seq-QVB*. The corresponding update equations are also analogous to those in (16), with the expectations $\langle \mathbf{z}_{d,t}^{t_N}\rangle$ and $\langle \mathbf{z}_{p,t}^{t_N}\rangle$ replaced by the unquantized measurements $\mathbf{z}_{d,t}^{t_N}$ and $\mathbf{z}_{p,t}^{t_N}$, respectively.

Computation of $q(\mathbf{X}_d^{t_N})$: The factor $q(\mathbf{X}_d^{t_N})$ in the *seq-VBI* algorithm can be computed by replacing $\langle \mathbf{z}_{d,t}^{t_N}\rangle$ in (19) with the unquantized measurement $\mathbf{z}_{d,t}^{t_N}$; the data symbol can be decoded using (20).

IV. Batch Estimation and Symbol Decoding

We now develop the *batch-QVB* algorithm, which uses a batch of measurements for jointly estimating the channel statistics, channel vectors, and soft symbols. Unlike sequential estimation, the batch estimation algorithm does not require an approximation as in (11) to make the problem tractable. Therefore, for a given number of frames, *batch-QVB* can potentially outperform *seq-QVB* at the cost of latency, as decoding happens only after a batch of frames is received.

A. Batch Processing using Quantized Measurements

The approximate factorized posterior pdf for *batch-QVB* is

$$\begin{aligned} &p(\{\mathbf{Z}_d^{t_n}, \mathbf{Z}_p^{t_n}, \mathbf{X}_d^{t_n}, \mathbf{H}^{t_n}\}_{n=1}^N, \mathbf{R}|\boldsymbol{\Xi}; \{\mathbf{X}_p^{t_n}\}_{n=1}^N, \sigma_w^2) \approx \\ &\left(\prod_{n=1}^N\prod_{s=1}^{\tau_p}\prod_{r=1}^{N_r} q(z_{p,rs}^{t_n})\right)\left(\prod_{n=1}^N\prod_{s=1}^{\tau_d}\prod_{r=1}^{N_r} q(z_{d,rs}^{t_n})\right) \\ &\times\left(\prod_{n=1}^N\prod_{s=1}^{\tau_d}\prod_{k=1}^{K} q(x_{d,ks}^{t_n})\right)\prod_{k=1}^K\left(\prod_{n=1}^N q(\mathbf{h}_k^{t_n})\right)q(\mathbf{R}_k), \end{aligned} \tag{22}$$

where $q(\mathbf{R}) = \prod_{k=1}^K q(\mathbf{R}_k)$ approximates the posterior pdf of the channel covariance matrix $\mathbf{R}$ using N frames. The joint pdf of the latent and observed random variables is given by

$$\begin{aligned} &p(\{\mathbf{Z}_d^{t_n}, \mathbf{Z}_p^{t_n}, \mathbf{X}_d^{t_n}, \mathbf{H}^{t_n}, \mathbf{Y}_d^{t_n}, \mathbf{Y}_p^{t_n}\}_{n=1}^N, \mathbf{R}) \\ &= \prod_{n=1}^N\Big(p(\mathbf{Y}_p^{t_n}|\mathbf{Z}_p^{t_n})p(\mathbf{Y}_d^{t_n}|\mathbf{Z}_d^{t_n})p(\mathbf{Z}_p^{t_n}|\mathbf{H}^{t_n}; \mathbf{X}_p^{t_n}, \sigma_w^2) \\ &\quad\times p(\mathbf{Z}_d^{t_n}|\mathbf{H}^{t_n}, \mathbf{X}_d^{t_n}; \sigma_w^2)p(\mathbf{X}_d^{t_n})p(\mathbf{H}^{t_n}|\mathbf{R})\Big)\, p(\mathbf{R}). \end{aligned} \tag{23}$$

Each factor in (22) can be computed from (23) using (8). A key difference between (10) used in *seq-QVB* and (23) used in *batch-QVB* is as follows. In *seq-QVB*, while processing a given frame, the joint pdf is expressed in terms of the posterior pdf of $\mathbf{R}$ given all observations up to the previous frame. In contrast, in *batch-QVB*, the prior pdf of the covariance matrix is used directly in (23), and all N frames are processed together.

*1) **Computation of*** $q(\mathbf{R})$

Using (8) for maximizing the ELBO, the factor $q(\mathbf{R}_k)$ follows a complex inverse Wishart pdf with parameters

$$\nu^{t_N} = \nu + N,\ \ \boldsymbol{\Psi}_k^{t_N} = \sum_{n=1}^N(\boldsymbol{\mu}_{\mathbf{h}_k}^{t_n}\boldsymbol{\mu}_{\mathbf{h}_k}^{t_n\,H} + \boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_n}) + \boldsymbol{\Psi}, \tag{24}$$

and hence the mean matrix is $\mathbf{S}_k^{t_N} = \boldsymbol{\Psi}_k^{t_N}/(\nu^{t_N} - N_r)$.

*2) **Computation of** $q(\mathbf{H}^{t_n})$, $q(\mathbf{Z}_p^{t_n})$, $q(\mathbf{Z}_d^{t_n})$, $q(\mathbf{X}_d^{t_n})$*

The update equations for $q(\mathbf{H}^{t_n})$, $q(\mathbf{Z}_p^{t_n})$, $q(\mathbf{Z}_d^{t_n})$, and $q(\mathbf{X}_d^{t_n})$ remain unchanged: (16), (17), (18) and (19) with frame numbers going from 1 to N can be used to compute them, respectively. A difference is that $\langle(\mathbf{R}_k^{t_N})^{-1}\rangle$ used in (16) remains unchanged across all frames. From (24), it is given by $\langle(\mathbf{R}_k^{t_N})^{-1}\rangle = (\mathbf{S}_k^{t_N})^{-1}\cdot \nu^{t_N}/(\nu^{t_N}-N_r)$.

B. Batch Processing using Unquantized Measurements

The *batch-QVB* can be modified to obtain *batch-VBI*, an algorithm that considers unquantized measurements, by taking $\{\{\mathbf{X}_d^{t_n},\mathbf{H}^{t_n}\}_{n=1}^N,\mathbf{R}\}$ as the latent variables across N frames, along with the approximate factorized posterior pdf:

$$p(\{\mathbf{X}_d^{t_n},\mathbf{H}^{t_n}\}_{n=1}^N,\mathbf{R}|\{\mathbf{Z}_p^{t_n},\mathbf{Z}_d^{t_n}\}_{n=1}^N;\{\mathbf{X}_p^{t_n}\}_{n=1}^N,\sigma_w^2)\approx \Big(\prod_{n=1}^N\prod_{s=1}^{\tau_d}\prod_{k=1}^K q(x_{d,ks}^{t_n})\Big)\prod_{n=1}^N\prod_{k=1}^K q(\mathbf{h}_k^{t_n})q(\mathbf{R}). \quad (25)$$

Similarly, the joint pdf of the latent and observed random variables is given by

$$p(\{\mathbf{X}_d^{t_n},\mathbf{H}^{t_n},\mathbf{Z}_d^{t_n},\mathbf{Z}_p^{t_n}\}_{n=1}^N,\mathbf{R}) = \prod_{n=1}^N p(\mathbf{Z}_p^{t_n}|\mathbf{H}^{t_n};\mathbf{X}_p^{t_n},\sigma_w^2) \times\prod_{n=1}^N p(\mathbf{Z}_d^{t_n}|\mathbf{H}^{t_n};\mathbf{X}_d^{t_N},\sigma_w^2)p(\mathbf{H}^{t_n}|\mathbf{R})p(\mathbf{X}_d^{t_N})p(\mathbf{R}). \quad (26)$$

Now, using (8) and (26), all the factors in (25) can be computed. The update equations for $q(\mathbf{R})$ remain unchanged.

The computations of $q(\mathbf{H}^{t_n})$ and $q(\mathbf{X}_d^{t_n})$ in the *batch-VBI* algorithm which uses unquantized samples as inputs are similar to the *batch-QVB* algorithm, with the expectations $\langle\mathbf{z}_{d,t}^{t_n}\rangle$ and $\langle\mathbf{z}_{p,t}^{t_n}\rangle$ replaced by the unquantized measurements $\mathbf{z}_{d,t}^{t_n}$ and $\mathbf{z}_{p,t}^{t_n}$, respectively.

V. Performance Analysis

In this section, we analytically evaluate the performance of the proposed algorithms by deriving a marginalized Bayesian Cramér-Rao lower bound (MB-CRLB) on the MSE in the channel covariance matrix estimation. We consider the sensor data without quantization noise in our analysis. The presence of quantization noise increases the MSE of the estimated channel statistics. Hence, the lower bound on the MSE derived without quantization noise is also valid for the estimation problem with quantized measurements. In deriving the lower bound, we assume that the data symbols $\{\mathbf{X}_d^{t_n}\}_{n=1}^N$ are decoded correctly; this also further lower-bounds the MSE in estimating the channel statistics.

Consider the signal model in (1). Let $\mathbf{z}_p^{t_n}=\text{vec}(\mathbf{Z}_p^{t_n})\in\mathbb{C}^{N_r\tau_p}$ and $\mathbf{z}_d^{t_n}=\text{vec}(\mathbf{Z}_d^{t_n})\in\mathbb{C}^{N_r\tau_d}$ be the measurement vectors corresponding to pilot and data symbols. The channel vector during the nth frame is $\mathbf{h}^{t_n}=\text{vec}(\mathbf{H}^{t_n})\in\mathbf{C}^{N_rK}$. The measurement model in (1) can be rewritten as

$$\begin{aligned}\mathbf{z}_p^{t_n} &= \mathbf{A}_p^{t_n}\mathbf{h}^{t_n}+\mathbf{v}_p^{t_n},\\ \mathbf{z}_d^{t_n} &= \mathbf{A}_d^{t_n}\mathbf{h}^{t_n}+\mathbf{v}_d^{t_n},\quad n=1,...,N\end{aligned} \quad (27)$$

where $\mathbf{A}_p^{t_n}=[\mathbf{A}_{p1}^{t_n},\mathbf{A}_{p2}^{t_n},\dots,\mathbf{A}_{pK}^{t_n}]={\mathbf{X}_p^{t_n}}^T\otimes\mathbf{I}\in\mathbb{C}^{N_r\tau_p\times N_rK}$ and $\mathbf{A}_d^{t_n}=[\mathbf{A}_{d_1}^{t_n},\mathbf{A}_{d_2}^{t_n},\dots,\mathbf{A}_{d_K}^{t_n}]={\mathbf{X}_d^{t_n}}^T\otimes\mathbf{I}\in\mathbb{C}^{N_r\tau_d\times N_rK}$. Let $\mathcal{Z}\triangleq\big\{\{\mathbf{z}_p^{t_n}\}_{n=1}^N,\{\mathbf{z}_d^{t_n}\}_{n=1}^N\big\}$ represent the set of measurements. The logarithm of the joint pdf after marginalizing over the channels $\{\mathbf{h}_k^{t_n}\}_{k=1,n=1}^{K,N}$ is given by

$$\begin{aligned}\log p(\mathcal{Z}) = &-\sum_{n=1}^N\log|\mathbf{C}_p^{t_n}|-\sum_{n=1}^N\log|\mathbf{C}_d^{t_n}|\\ &-\sum_{n=1}^N{\mathbf{z}_p^{t_n}}^H(\mathbf{C}_p^{t_n})^{-1}\mathbf{z}_p^{t_n}-\sum_{n=1}^N{\mathbf{z}_d^{t_n}}^H(\mathbf{C}_d^{t_n})^{-1}\mathbf{z}_d^{t_n}\\ &-(\nu+N_r)\sum_{k=1}^K\log|\mathbf{R}_k|-\sum_{k=1}^K\text{tr}(\mathbf{R}_k^{-1}\mathbf{\Psi})+\text{const.}\end{aligned} \quad (28)$$

where $\mathbf{C}_p^{t_n}=\mathbf{A}_p^{t_n}\mathbf{R}{\mathbf{A}_p^{t_n}}^H+\sigma_w^2\mathbb{I}$, $\mathbf{C}_d^{t_n}=\mathbf{A}_d^{t_n}\mathbf{R}{\mathbf{A}_d^{t_n}}^H+\sigma_w^2\mathbb{I}$.

The unknown random variables are denoted by $\boldsymbol{\theta}=[\boldsymbol{\theta}_{c1}^T,\boldsymbol{\theta}_{c2}^T,\dots,\boldsymbol{\theta}_{cK}^T]^T\in\mathbb{C}^{KN_r^2}$ and $\boldsymbol{\theta}_{ck}=\text{vec}\{\mathbf{R}_k\}$. In the MB-CRLB, the error covariance matrix of unknown random variables is lower bounded by [29]

$$\mathbf{E}(\boldsymbol{\theta})\succeq\mathbf{J}_B^{-1}, \quad (29)$$

where $\mathbf{E}(\boldsymbol{\theta})=\mathbb{E}_{p(\mathcal{Z})}\big\{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^H\big\}$ is the covariance of the estimation error, and $\mathbf{J}_B$ is the marginalized Bayesian Fisher information matrix (M-FIM) with (i,j)th element

$$[\mathbf{J}_B]_{ij}=-\mathbb{E}_{p(\mathcal{Z})}\left\{\frac{\partial^2\ln p(\mathcal{Z})}{\partial\theta_i\,\partial\theta_j^*}\right\} \quad (30)$$

The matrix $\mathbf{J}_B=\text{blkdiag}\{\mathbf{J}_{B_1},\mathbf{J}_{B_2},\dots,\mathbf{J}_{B_K}\}$ is block diagonal because the channel vectors are uncorrelated across UEs. Further, the columns of the M-FIM of $\mathbf{J}_{B_k}$ are

$$\begin{aligned}\mathbf{J}_{B_k}&=[\mathbf{j}_{B_{k1}},\mathbf{j}_{B_{k2}},\dots,\mathbf{j}_{B_{kN_r^2}}]\in\mathbb{C}^{N_r^2\times N_r^2},\\ \mathbf{j}_{B_{k(i+(j-1)N_r)}}&=\text{vec}\left(-\mathbb{E}_{p(\mathcal{Z})}\left\{\frac{\partial^2\ln p(\mathcal{Z})}{\partial\,[\mathbf{R}_k]_{ij}\,\partial\mathbf{R}_k^*}\right\}\right).\end{aligned}$$

Computing the partial derivative of the joint pdf in (28) w.r.t. $\mathbf{R}_k^*$, we get the matrix $\mathbf{G}^k(\mathbf{R})\in\mathbb{C}^{N_r\times N_r}$ given by

$$\begin{aligned}\mathbf{G}^k(\mathbf{R})=&-\sum_{n=1}^N\Big({\mathbf{A}_{pk}^{t_n}}^H(\mathbf{C}_p^{t_n})^{-1}\mathbf{A}_{pk}^{t_n}+{\mathbf{A}_{dk}^{t_n}}^H(\mathbf{C}_d^{t_n})^{-1}\mathbf{A}_{dk}^{t_n}\Big)\\ &+\sum_{n=1}^N{\mathbf{A}_{pk}^{t_n}}^H(\mathbf{C}_p^{t_n})^{-1}\mathbf{U}_p^{t_n}(\mathbf{C}_p^{t_n})^{-1}\mathbf{A}_{pk}^{t_n}\\ &+\sum_{n=1}^N{\mathbf{A}_{dk}^{t_n}}^H(\mathbf{C}_d^{t_n})^{-1}\mathbf{U}_d^{t_n}(\mathbf{C}_d^{t_n})^{-1}\mathbf{A}_{dk}^{t_n}\\ &-(\nu+N_r)\mathbf{R}_k^{-1}+\mathbf{R}_k^{-1}\mathbf{\Psi}\mathbf{R}_k^{-1}\end{aligned}$$

where $\mathbf{U}_p^{t_n}\triangleq\mathbf{z}_p^{t_n}{\mathbf{z}_p^{t_n}}^H$ and $\mathbf{U}_d^{t_n}\triangleq\mathbf{z}_d^{t_n}{\mathbf{z}_d^{t_n}}^H$. Computing the partial derivative of $\mathbf{G}^k(\mathbf{R})$ w.r.t. the (i,j)th element of $\mathbf{R}_k$ and evaluating the expectation w.r.t. $p(\mathcal{Z}|\mathbf{R})$,[3] we get

[3]Note that $\log p(\mathcal{Z}|\mathbf{R})$ equals the first four terms of (28) plus a constant.

$$\mathbf{G}_{ij}^{k}(\mathbf{R})=-\sum_{n=1}^{N}\Big({\mathbf{A}_{pk}^{t_n}}^H\mathbf{Q}_{p,ij}^{t_n}\mathbf{A}_{pk}^{t_n}+{\mathbf{A}_{dk}^{t_n}}^H\mathbf{Q}_{d,ij}^{t_n}\mathbf{A}_{dk}^{t_n}\Big) - \mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1}\boldsymbol{\Psi}\mathbf{R}_k^{-1} - \mathbf{R}_k^{-1}\boldsymbol{\Psi}\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1} + (\nu+N_r)\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1} \tag{31}$$

where $\mathbf{Q}_{p,ij}^{t_n} \triangleq (\mathbf{C}_p^{t_n})^{-1}\mathbf{A}_{pk}^{t_n}\mathbf{E}_{ij}{\mathbf{A}_{pk}^{t_n}}^H(\mathbf{C}_p^{t_n})^{-1}$, $\mathbf{Q}_{d,ij}^{t_n} \triangleq (\mathbf{C}_d^{t_n})^{-1}\mathbf{A}_{dk}^{t_n}\mathbf{E}_{ij}{\mathbf{A}_{dk}^{t_n}}^H(\mathbf{C}_d^{t_n})^{-1}$, and $\mathbf{E}_{ij}$ is an $N_r\times N_r$ matrix whose (i,j)th element is 1 and the remaining elements are 0. To compute the marginalized CRLB, we need to find the expectation of vec$(\mathbf{G}_{ij}^{k}(\mathbf{R}))$ w.r.t. the prior distribution of $\mathbf{R}_k$. The resultant vector output becomes the $(i+(j-1)N_r)$th column of the M-FIM of $\mathbf{R}_k$, that is,

$$\mathbf{j}_{B_{k(i+(j-1)N_r)}} = \text{vec}\left(\mathbb{E}_{\mathbf{R}}\left\{\mathbf{G}_{ij}^{k}(\mathbf{R})\right\}\right) \tag{32}$$

The following theorem is helpful in computing the M-FIM.

Theorem 1. *Let $\mathbf{R}$ be a complex inverse Wishart distributed random matrix with scale matrix $\boldsymbol{\Psi}_c$ and DoF ν. Then, the expectation of the random matrix $(({\mathbf{A}_p^{t_n}}^*\mathbf{R}^*{\mathbf{A}_p^{t_n}}^T+\sigma_w^2\mathbb{I})^{-1}\otimes(\mathbf{A}_p^{t_n}\mathbf{R}{\mathbf{A}_p^{t_n}}^H+\sigma_w^2\mathbb{I})^{-1})$ w.r.t. the pdf of $\mathbf{R}$ satisfies the following matrix lower and upper bounds:*

$$\begin{aligned}&\left(\frac{1}{\tilde{\nu}}{\mathbf{A}_p^{t_n}}^*\boldsymbol{\Psi}_c^*{\mathbf{A}_p^{t_n}}^T+\sigma_w^2\mathbb{I}\right)^{-1}\otimes\left(\frac{1}{\tilde{\nu}}\mathbf{A}_p^{t_n}\boldsymbol{\Psi}_c{\mathbf{A}_p^{t_n}}^H+\sigma_w^2\mathbb{I}\right)^{-1}\\ &\preceq \mathbb{E}\left(({\mathbf{A}_p^{t_n}}^*\mathbf{R}^*{\mathbf{A}_p^{t_n}}^T+\sigma_w^2\mathbb{I})^{-1}\otimes(\mathbf{A}_p^{t_n}\mathbf{R}{\mathbf{A}_p^{t_n}}^H+\sigma_w^2\mathbb{I})^{-1}\right)\\ &\preceq\left({\mathbf{F}_p^{t_n}}^{-1}+2\sigma_w^2{\mathbf{F}_p^{t_n}}^{-1/2}+\sigma_w^4\mathbb{I}\right)^{-1},\\ &\textit{where, } \mathbf{F}_p^{t_n}=({\mathbf{A}_p^{t_n}}^T\otimes{\mathbf{A}_p^{t_n}}^H)^\dagger\boldsymbol{\Gamma}(({\mathbf{A}_p^{t_n}}^*\otimes\mathbf{A}_p^{t_n})^\dagger,\\ &\boldsymbol{\Gamma}=\nu^2({\boldsymbol{\Psi}_c^{-1}}^T\otimes\boldsymbol{\Psi}_c^{-1})+\nu vec\left(\boldsymbol{\Psi}_c^{-1}\right)vec\left(\boldsymbol{\Psi}_c^{-1}\right)^H,\\ &\textit{and } \tilde{\nu}=\nu-N_r.\end{aligned} \tag{33}$$

Proof. See Appendix A. ∎

The general matrix inequality presented in Theorem 1 could be of independent interest because an expression of the above form frequently appears in problems where a complex inverse Wishart prior distribution is imposed on the covariance matrix, e.g., see [26]. Also, note that the lower and upper bounds become tight as ν becomes large. Using the Theorem, we can compute an upper bound on the M-FIM as follows:

Lemma 1. *The M-FIM of kth user, $\mathbf{J}_{B_k}\in\mathcal{C}^{N_r^2\times N_r^2}$, can be upper bounded as*

$$\begin{aligned}&\mathbf{J}_{B_k}\preceq\\ &\sum_{n=1}^{N}\left(({\mathbf{A}_{dk}^{t_n}}^T\otimes{\mathbf{A}_{dk}^{t_n}}^H)\mathbf{D}_d^{t_n}({\mathbf{A}_{dk}^{t_n}}^T\otimes{\mathbf{A}_{dk}^{t_n}}^H)^H\right)\\ &+\sum_{n=1}^{N}\left(({\mathbf{A}_{pk}^{t_n}}^T\otimes{\mathbf{A}_{pk}^{t_n}}^H)\mathbf{D}_p^{t_n}({\mathbf{A}_{pk}^{t_n}}^T\otimes{\mathbf{A}_{pk}^{t_n}}^H)^H\right)\\ &+b_1\left({\boldsymbol{\Psi}^{-1}}^T\otimes\boldsymbol{\Psi}^{-1}\right)+b_2 vec\left(\boldsymbol{\Psi}^{-1}\right)vec\left(\boldsymbol{\Psi}^{-1}\right)^H,\end{aligned} \tag{34}$$

where $\mathbf{D}_p^{t_n}=({\mathbf{F}_p^{t_n}}^{-1}+2\sigma_w^2{\mathbf{F}_p^{t_n}}^{-1/2}+\sigma_w^4\mathbb{I})^{-1}$, $\mathbf{D}_d^{t_n}=({\mathbf{F}_d^{t_n}}^{-1}+2\sigma_w^2{\mathbf{F}_d^{t_n}}^{-1/2}+\sigma_w^4\mathbb{I})^{-1}$, $\mathbf{F}_p^{t_n}=({\mathbf{A}_p^{t_n}}^T\otimes{\mathbf{A}_p^{t_n}}^H)^\dagger\boldsymbol{\Gamma}({\mathbf{A}_p^{t_n}}^*\otimes\mathbf{A}_p^{t_n})^\dagger$, $\mathbf{F}_d^{t_n}=({\mathbf{A}_d^{t_n}}^T\otimes{\mathbf{A}_d^{t_n}}^H)^\dagger\boldsymbol{\Gamma}({\mathbf{A}_d^{t_n}}^*\otimes\mathbf{A}_d^{t_n})^\dagger$, $\boldsymbol{\Gamma}=\nu^2({\boldsymbol{\Psi}_c^{-1}}^T\otimes\boldsymbol{\Psi}_c^{-1})+\nu\, vec(\boldsymbol{\Psi}_c^{-1})vec(\boldsymbol{\Psi}_c^{-1})^H$, $b_1=N_r\nu^2+2\nu+\nu^3$, $b_2=N_r\nu+3\nu^2$, and $\boldsymbol{\Psi}_c\triangleq$ blkdiag$(\boldsymbol{\Psi},\ldots,\boldsymbol{\Psi})\in\mathbb{C}^{KN_r\times KN_r}$.

Proof. See Appendix B. ∎

Table II: Per-iteration Computational Complexity of *seq-QVB*

Variables	*Computational Complexity*
$\{\mathbf{S}_k^{t_N}\}_{k=1}^{K}$	KN_r^2
$\{\boldsymbol{\mu}_{\mathbf{h}_k}^{t_N}\}_{k=1}^{K}$	$KN_r^2+K^2N_r(\tau_p+\tau_d)$
$\{\hat{x}_{d,kt}^{t_N}\}_{t=1,k=1}^{t=\tau_d,k=K}$	$MK^2N_r\tau_d$
$\{\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}\}_{k=1}^{K}$	$K(N_r^3+\tau_d)$
$\{\langle z_{p,rt}^{t_N}\rangle\}_{t=1,r=1}^{\tau_p,N_r}$	$KN_r\tau_p$
$\{\langle z_{d,rt}^{t_N}\rangle\}_{t=1,r=1}^{\tau_d,N_r}$	$KN_r\tau_d$

With the M-FIM in hand, we can now lower bound the MSE in estimating the channel covariance matrix as follows:

Theorem 2. *[MB-CRLB] The mean squared error in estimating the (i,j)th element of the kth user's channel covariance matrix $\mathbf{R}_k$ is lower bounded by*

$$\begin{aligned}&\Bigg[\Bigg(\sum_{n=1}^{N}\left(({\mathbf{A}_{dk}^{t_n}}^T\otimes{\mathbf{A}_{dk}^{t_n}}^H)\mathbf{D}_d^{t_n}({\mathbf{A}_{dk}^{t_n}}^T\otimes{\mathbf{A}_{dk}^{t_n}}^H)^H\right)\\ &+\sum_{n=1}^{N}\left(({\mathbf{A}_{pk}^{t_n}}^T\otimes{\mathbf{A}_{pk}^{t_n}}^H)\mathbf{D}_p^{t_n}({\mathbf{A}_{pk}^{t_n}}^T\otimes{\mathbf{A}_{pk}^{t_n}}^H)^H\right)\\ &+b_1({\boldsymbol{\Psi}^{-1}}^T\otimes\boldsymbol{\Psi}^{-1})+b_2 vec(\boldsymbol{\Psi}^{-1})vec(\boldsymbol{\Psi}^{-1})^H\Bigg)^{-1}\Bigg]_{l,l},\\ &\textit{where } l=i+(j-1)N_r.\end{aligned} \tag{35}$$

Proof. The bound is obtained using (29) and (34). ∎

A. Computational Complexity Analysis

Compared to the variational Bayesian algorithms in [12], [16], *seq-QVB* and *batch-QVB* additionally require the computation of the mean matrices $\{\mathbf{S}_k^{t_N}\}_{k=1}^{K}$ of the approximate posterior pdf $q(\mathbf{R}^{t_N})$. The order-wise per-iteration computational complexity of *seq-QVB* algorithm is given in Table II. The complexity of computing $\{\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_N}\}_{k=1}^{K}$ scales cubically with the number of antennas, and this dominates the computational complexity of the algorithm in the large number of antennas-regime. It also scales quadratically in the number of users and linearly in the data/pilot durations.

In the *batch-QVB* algorithm, the most recent N quantized measurements $\{\mathbf{Y}_p^{t_n}\}_{n=1}^{N},\{\mathbf{Y}_d^{t_n}\}_{n=1}^{N}$ are stored to compute $\{\mathbf{S}_k^{t_N}\}_{k=1}^{K}$. This requires recomputing all other parameters such as $\{\langle z_{p,rt}^{t_n}\rangle\}_{t=1,r=1}^{\tau_p,N_r}$, $\{\langle z_{d,rt}^{t_n}\rangle\}_{t=1,r=1}^{\tau_d,N_r}$, $\{\boldsymbol{\Sigma}_{\mathbf{h}_k}^{t_n}\}_{k=1}^{K}$, $\{\boldsymbol{\mu}_{\mathbf{h}_k}^{t_n}\}_{k=1}^{K}$ and $\{\hat{x}_{d,kt}^{t_n}\}_{t=1,k=1}^{t=\tau_d,k=K}$ from $n=1$ to N. Consequently, the real-time implementation of the *batch-QVB* algorithm requires an N-fold higher computational complexity than the *seq-QVB* algorithm.

The approximate posterior distributions of the channel matrices using (16) require matrix inversions, resulting in a significant computational burden in the proposed algorithms. This complexity can be mitigated by employing approximate message passing (AMP) techniques for linear mixing models, thereby avoiding explicit matrix inversions in the computation of posterior means and covariance matrices. Similar approaches have been considered in the literature [30]. In particular, AMP-based implementations can significantly reduce the per-iteration computational complexity and memory requirements, making the resulting algorithms more scalable to systems with large antenna arrays and a large number of users. Moreover, AMP-based methods have been shown to provide accurate approximations to Bayesian inference in large-scale linear estimation problems while maintaining low computational complexity [30]. Extending our algorithms in this direction is an interesting topic for future research.

VI. Simulation Results

In this section, we demonstrate the performance of the *seq-QVB*, *batch-QVB*, *seq-VBI*, and *batch-VBI* algorithms in estimating the channel statistics, channel vectors, and decoding the data symbols. We consider uplink communications and evaluate the performance of the new algorithms in the presence and absence of quantization noise, with unknown channel statistics. The channel estimation accuracy is characterized using the normalized MSE (NMSE). The data decoding performance is evaluated by computing the SER as a function of various system parameters.

The covariance matrix of the kth user is modeled as $\mathbf{R}_k = \sum_{l=1}^{L_{p_k}} P_k \beta_{kl} \mathbf{a}_{\phi_{lk}} \mathbf{a}_{\phi_{lk}}^H$, where L_{p_k} denotes the number of propagation paths between the kth UE and the BS. The unit mean exponentially distributed random variable β_{kl} captures the small-scale fading of the lth path, while P_k denotes the transmit power of the kth user [31]. The SNR is defined as $\text{SNR}_k = \frac{P_k}{\sigma_w^2}$. The mth element of the array response vector $\mathbf{a}_{\phi_{lk}} \in \mathbb{C}^{N_r}$ is $\exp(-j\frac{2\pi d}{\lambda}(m-1)\cos(\theta_{kl}))$, where θ_{kl}, λ and $d = \lambda/2$ represent the angle of arrival of the lth path from the kth UE, carrier wavelength, and antenna spacing, respectively. The arrival angles $\{\theta_{kl}\}_{k=1,l=1}^{k=K,l=L_p}$ are distributed uniformly at random on $[0, \pi)$. The number of paths is $L_{p_k} = 16$. In the simulation studies, quadrature phase shift keying modulation is used for the data symbols. The pilot symbol matrix consists of orthogonal sequences and satisfies $\mathbf{X}_p^{t_n} \mathbf{X}_p^{t_n H} = \tau_p \mathbb{I}$.

First, we evaluate the performance of the sequential algorithms described in Section III. We consider the state-of-the-art *conv-QVB-JED* algorithm described in [16] for comparison. The *conv-QVB-JED* algorithm requires knowledge of the channel covariance matrices. Hence, we use the identity matrix as the nominal covariance matrix when running the *conv-QVB-JED* algorithm.

Fig. 2 shows that the performance of the *conv-QVB-JED* is degraded compared to the *seq-QVB* and *seq-VBI* algorithms due to the mismatch between assumed and actual channel covariance matrix. On the other hand, its performance matches with that of the *seq-QVB* algorithm in the genie-aided case where the channel statistics are revealed to the *conv-QVB-JED* decoder. The performance of *seq-QVB* and *seq-VBI* algorithms is comparable to the corresponding

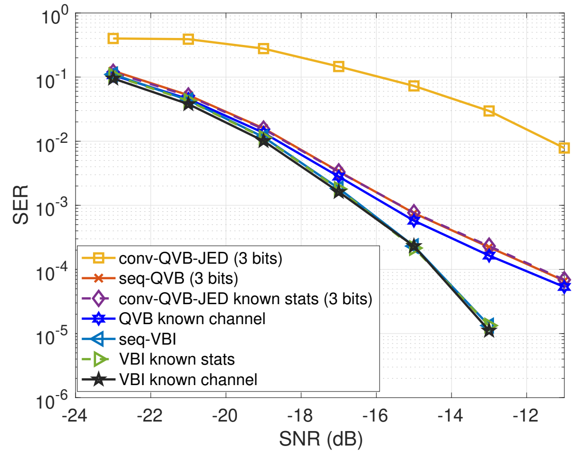

Fig. 2: SER vs. SNR, with $N_r = 30$, $K = 6$, $\tau_d = 250$, $\tau_p = 24$, $N = 500$. Comparison of sequential algorithms. Performance of proposed algorithms are comparable with the genie-aided decoders.

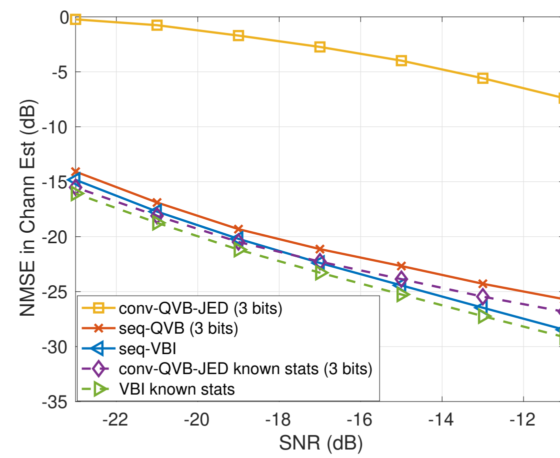

Fig. 3: NMSE in channel matrix estimation as a function of the SNR, with $N_r = 30$, $K = 6$, $\tau_d = 250$, $\tau_p = 24$, $N = 500$. The gap between sequential and the genie-aided decoders is less than 2 dB.

genie-aided decoders that have access to the true covariance matrix (labeled `conv-QVB-JED known stats` and `VBI known stats`, respectively). Finally, we also compare the SER with a genie-aided decoder that has perfect channel state information (CSI). We note that perfect CSI is a strong assumption when the BS employs low resolution ADCs, since the pilot signals are also subject to the quantization noise. Yet, the performance gap is marginal in the quantized case (*seq-QVB*) and negligible in the unquantized case (*seq-VBI*).

Fig. 3 illustrates the NMSE in channel estimation. We see that estimating channel statistics using *seq-QVB* or *seq-VBI* significantly reduces the NMSE compared to *conv-QVB-JED* with mismatched channel statistics. The performance of the *seq-QVB* and *seq-VBI* is in fact comparable to that of a genie-aided scenario where the exact channel covariance matrix is known. Fig. 4 shows that the algorithms continue to perform well as the number of BS antennas N_r is increased. The SER reduces roughly linearly (in the log scale) with increasing N_r due to the array gain from the antennas. The figure also shows the effect of the number of frames, N, used

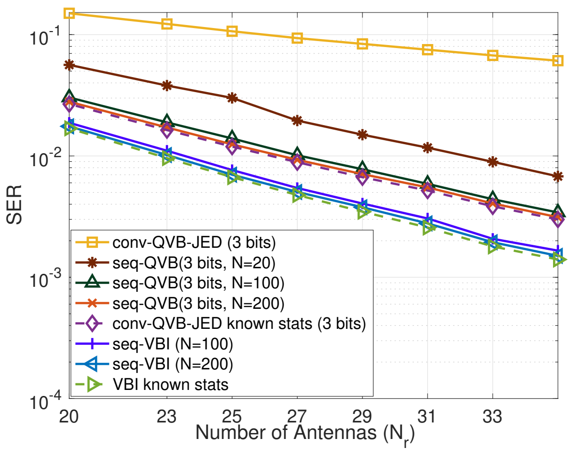

Fig. 4: Comparison of SER with number of antennas at BS with $K = 6$, $\tau_d = 250$, $\tau_p = 24$, and SNR -16 dB.

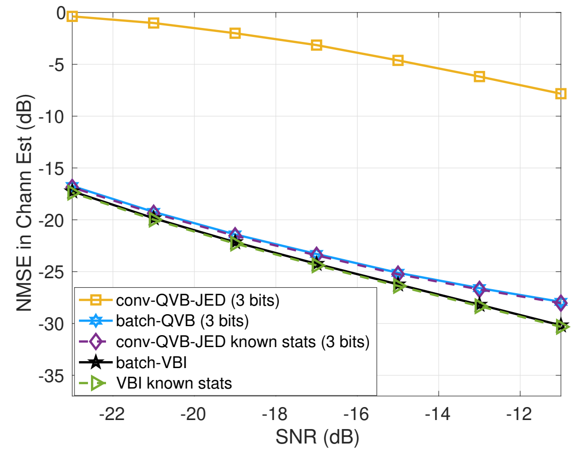

Fig. 6: NMSE in channel matrix estimation vs. the SNR, with $N_r = 42$, $K = 8$, $\tau_d = 250$, $\tau_p = 48$, $N = 500$. The performance of the batch algorithms matches with genie-aided decoders.

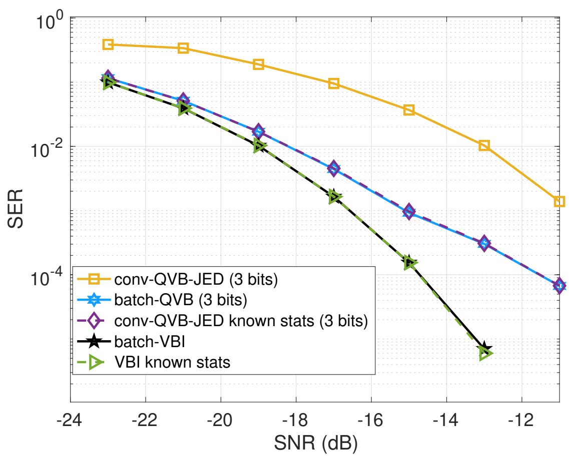

Fig. 5: The performance of the batch algorithms is comparable to that of the genie-aided decoders. Settings: $N_r = 42$, $K = 8$, $\tau_d = 250$, $\tau_p = 48$, $N = 500$.

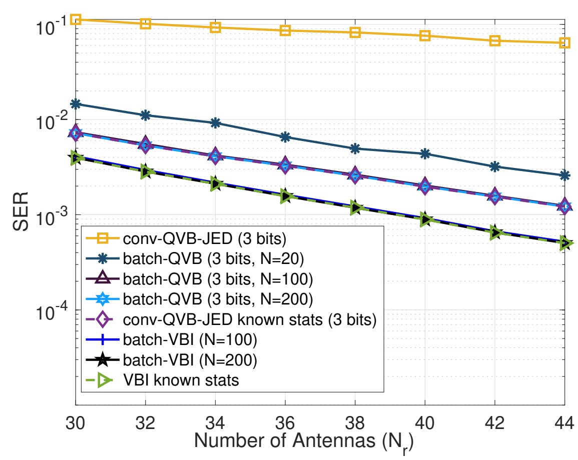

Fig. 7: Comparison of batch algorithms with $K = 8$, $\tau_d = 250$, $\tau_p = 48$, and SNR -16 dB.

in the covariance matrix estimation: at $N \approx 100$, the SER gap between the sequential algorithms and genie-aided method (known channel covariance) becomes negligible.

Next, we evaluate the performance of the *batch algorithms* discussed in Section IV. Fig. 5 compares the SER of *batch-QVB* and *batch-VBI* with the *conv-QVB-JED* and genie aided decoders. The *batch-QVB* algorithm significantly outperforms the *conv-QVB-JED* algorithm since it estimates the channel covariance matrix. Further, the SER of *batch-QVB* and *batch-VBI* overlaps with that of the corresponding genie aided decoders that are aware of the exact channel statistics. The NMSE in channel estimation is illustrated in Fig. 6. Compared to the *conv-QVB-JED* algorithm, the estimation of channel statistics in *batch-QVB* helps to reduce the NMSE in the channel estimates by over 10 dB. As before, the NMSE in estimating the channel vectors using the batch algorithms matches with that of the genie-aided receivers.

Fig. 7 depicts the linear reduction in SER with increasing the number of antennas at the BS, similar to the sequential algorithms. Note that mismatches in the available channel statistics degrade the performance of the *conv-QVB-JED* algorithm; the in-situ estimation of channel statistics using *batch-QVB* recovers this loss and yields a performance that overlaps with that of the corresponding genie-aided decoders.

In Fig. 15, we compare the performance of *seq-QVB* and *batch-QVB* algorithms with a varying number of bits used in analog-to-digital conversion. With 4-bit quantization, the *seq-QVB* and *batch-QVB* algorithms perform on par with *seq-VBI* and *batch-VBI* which use unquantized measurements. Fig. 17 shows the effect of number of ADC bits in the estimation error of the channels. Again, a 4-bit ADC suffices to match with the performance obtained with unquantized measurements, within the SNR range of interest.

In Fig. 16, we evaluate the tightness of the MB-CRLB on the MSE of channel covariance matrix estimation derived in Section V. In simulations, multiple realizations of the covariance matrix $\mathbf{R}$ are used to estimate the MSE in estimating the channel covariance matrix. In each realization, $\mathbf{R}$ is drawn from a complex inverse Wishart pdf and the parameters of the pdf (scale matrix and DoF) are used to initialize the algorithms. The MB-CRLB provides a tight lower bound to the empirical MSE achieved by the proposed

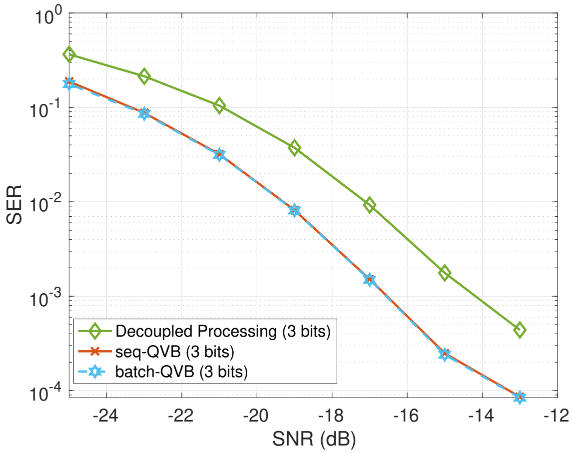

Fig. 8: Comparison of joint processing and decoupled processing with $N_r = 30$, $K = 6$, $\tau_d = 250$, $\tau_p = 6$, and $N = 500$. Joint processing provides a $3\,\text{dB}$ improvement in SNR at an SER of 10^{-3}.

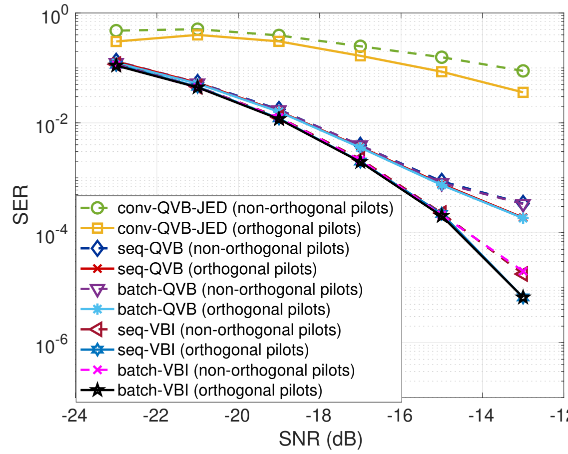

Fig. 10: Comparison of orthogonal and non-orthogonal pilot sequences with $N_r = 30$, $K = 6$, $\tau_d = 250$, $\tau_p = 24$, and $N = 500$.

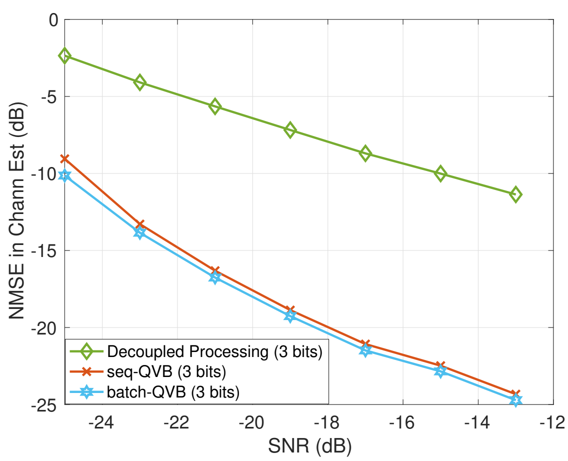

Fig. 9: NMSE in channel estimation with joint and decoupled processing; $N_r = 30$, $K = 6$, $\tau_d = 250$, $\tau_p = 6$, and $N = 500$.

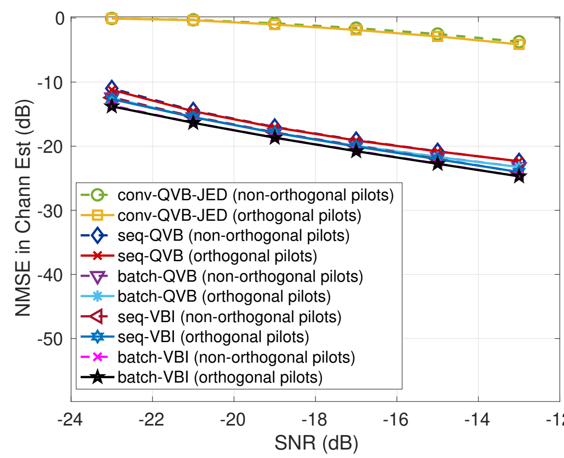

Fig. 11: NMSE in channel estimation for orthogonal and non-orthogonal pilot sequences, with $N_r = 30$, $K = 6$, $\tau_d = 250$, $\tau_p = 24$, and $N = 500$.

algorithms. The effect of the number of frames on the MSE in the covariance matrix estimate is depicted in Fig. 18, at an SNR of -8 dB. The *seq-VBI* and *batch-VBI* algorithms closely follow the MB-CRLB, with a gap of less than 1 dB.

The importance of joint estimation of channel vectors, channel statistics, and soft-symbol decoding is illustrated in Figs. 8 and 9. In the decoupled processing, the estimation of channel vectors, channel statistics, and soft-symbol decoding is carried out separately. Joint processing of the measurements corresponding to pilot symbols and data symbols reduces the NMSE of the channel vector estimate by more than 10 dB at an SNR of -18 dB. Similarly, we observe a significant improvement in SER performance, where an SNR gain of approximately 3 dB is achieved at an SER of 10^{-3}.

The sequential and batch algorithms do not require the pilot sequences to be orthogonal across the UEs. We compare the performance of the proposed algorithms using orthogonal and non-orthogonal pilot sequences in Figs. 10 and 11. The Zadoff–Chu sequences are employed as the non-orthogonal pilot symbols. The simulation results show that the differences in SER and NMSE obtained using orthogonal and Zadoff–Chu sequences are negligible. In the design of Zadoff–Chu sequences, the first six root indices that are relatively prime to $\tau_p = 24$ are selected for $K = 6$, resulting in the set $\{1, 5, 7, 11, 13, 17\}$. Since $\tau_p = 24$ is even, the Zadoff–Chu sequence corresponding to the root index u is generated as $z_u(n) = \exp\left(-j\pi u n^2/\tau_p\right), \quad n = 0, 1, \ldots, \tau_p - 1$. Each sequence is then normalized by $1/\sqrt{\tau_p}$ to obtain unit energy. Finally, the normalized sequences are assigned as the pilot sequences for the six users.

The proposed variational Bayesian algorithms can be viewed as instances of the majorization-minimization framework. Specifically, the KL divergence between the approximate posterior pdf and the true posterior pdf is guaranteed to be non-increasing at each iteration, thereby ensuring convergence to a local optimum. The convergence behavior of the proposed algorithm is illustrated in Fig. 12. The results show that the SER converges within about 10 iterations.

The effect of pilot length is illustrated in Figs. 13 and 14. While increasing τ_p improves channel estimation and data detection, the proposed joint framework substantially

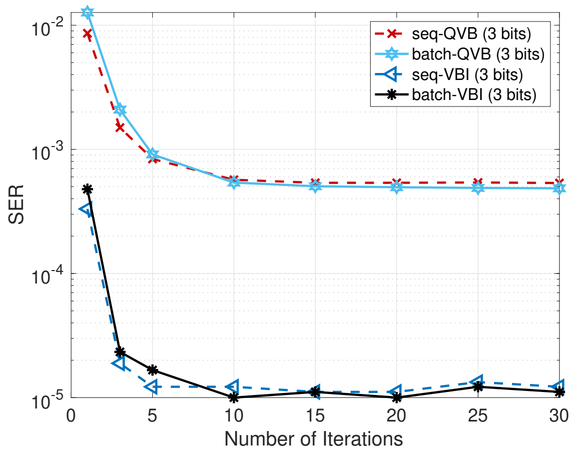

Fig. 12: SER vs. number of iterations, with $N_r = 30$, $K = 6$, $\tau_d = 250$, SNR $= -13$ dB, $\tau_p = 24$, and $N = 500$.

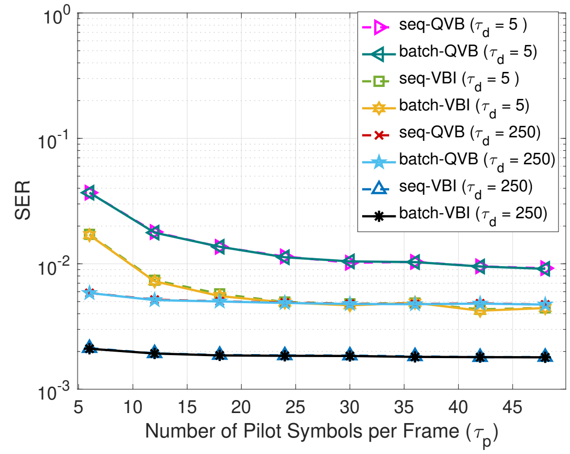

Fig. 13: SER vs. number of pilot symbols (τ_p), with $N_r = 30$, $K = 6$, SNR $= -17$ dB,and $N = 500$.

reduces pilot requirements by exploiting both pilot and data observations. In particular, the channel estimation NMSE falls below -20 dB with only $\tau_p = 6$ pilots and $\tau_d = 250$ data symbols. Further increases in τ_p yield only marginal NMSE and SER improvements, indicating that the proposed algorithms achieve high performance with low pilot overhead.

Discussion: In the batch algorithms, decoding $\{\mathbf{X}_d^{t_n}\}_{n=1}^N$ requires computing $q(\mathbf{R})$ and $\{q(\mathbf{H}^{t_n}), q(\mathbf{z}_p^{t_n}), q(\mathbf{z}_d^{t_n})\}_{n=1}^N$ using all quantized measurements $\{\mathbf{Y}_p^{t_n}, \mathbf{Y}_d^{t_n}\}_{n=1}^N$. Consequently, decoding $\mathbf{X}_d^{t_1}$ requires observations from all N frames, leading to a latency of approximately N communication frames and increased memory requirements. In contrast, the sequential algorithms decode the data symbols in the Nth frame using only the current measurements $\{\mathbf{Y}_p^{t_N}, \mathbf{Y}_d^{t_N}\}$ and the covariance statistics $\{\mathbf{S}_k^{t_{N-1}}\}_{k=1}^K$ from previous frames. Hence, *seq-QVB* and *seq-VBI* achieve a latency of approximately one communication frame while significantly reducing memory and computational requirements. This reduction in latency comes at a slight performance loss. Unlike the batch algorithms, newly received measurements are not used to refine latent variables associated with earlier frames, and the approximation in (11) is required for tractability. By jointly processing all available observations, the batch algorithms can exploit the measurements more effectively and achieve improved channel estimation accuracy. Therefore, for a given number of communication frames, *batch-QVB* can potentially outperform *seq-QVB* at the cost of increased latency. The results in Figs. 9 and 14 confirm this trend, showing that the batch algorithms achieve slightly lower NMSE than the sequential algorithms.

In this simulation study, the *conv-QVB-JED* algorithm is used as a benchmark. A detailed comparison of *conv-QVB-JED* with other state-of-the-art joint channel estimation and symbol decoding algorithms is provided in [16]. Note that these algorithms assume perfect knowledge of the channel covariance matrices. Consequently, their performance degrades when the assumed channel statistics are mismatched with the actual channel statistics. In contrast,

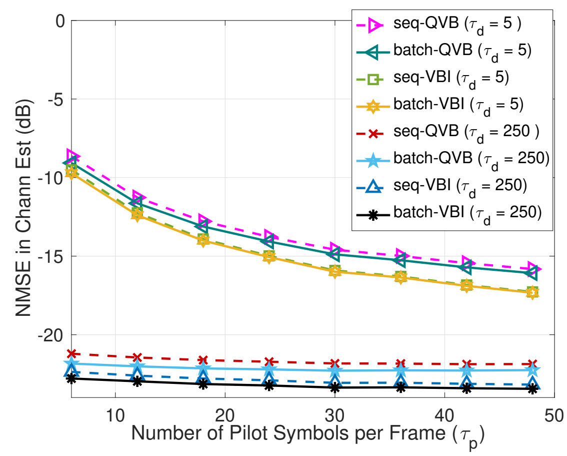

Fig. 14: NMSE vs. number of pilot symbols (τ_p) with $N_r = 30$, $K = 6$, SNR $= -17$ dB, and $N = 500$.

the proposed sequential and batch algorithms estimate the channel statistics in-situ by jointly processing the observations corresponding to both pilot and data symbols. Therefore, the proposed methods exhibit increased robustness to statistical mismatches and can outperform existing approaches that require accurate knowledge of the channel statistics. When perfect channel statistics are available, the proposed algorithms achieve performance close to *conv-QVB-JED*, further confirming that the main performance differences arise from the unknown-statistics setting rather than the underlying inference framework. Overall, the simulation results illustrate that, when the channels are correlated, it is crucial to incorporate the estimation of channel statistics into the receiver chain in order to obtain good performance.

VII. Conclusions

We considered the joint estimation of channel statistics, channel vectors, and soft symbol decoding in the uplink of a MU-MIMO system where low-resolution ADCs are employed at the BS. We introduced sequential and batch algorithms for estimating the channel covariance matrix. The sequential update of the covariance matrix within a variational Bayesian framework reduces latency in joint estimation and symbol decoding. We derived a marginalized Bayesian CRLB for estimating the channel covariance matrix. We empirically evaluated the MSE of the estimated channel covariance matrix and showed that it is close to the marginalized CRLB. We also evaluated the performance of the proposed algorithms using Monte Carlo simulations and demonstrated the utility of accurately estimating the channel covariance matrix in the subsequent channel estimation and data decoding. Future work can consider optimal quantizer design based on the analytical results developed here, theoretical analysis of the impact of imperfect covariance estimation on metrics such as achievable data rate and BER, and extension to multi-cell scenarios accounting for inter-cell interference.

Appendix A: Proof of Theorem 1

A function $\mathbf{G}(\mathbf{R})$ is said to be convex in the matrix inequality sense if the scalar function $f(\mathbf{R}) = \mathbf{w}^H\mathbf{G}(\mathbf{R})\mathbf{w}$ is convex for any choice of the vector $\mathbf{w}$. Moreover, the function $f(\mathbf{R})$ is convex if and only if the function $g(\alpha) = f(\mathbf{R}_0 + \alpha\mathbf{R}_1)$ is convex in the scalar $\alpha \geq 0$ for any choice of non-negative definite matrices $\mathbf{R}_0$ and $\mathbf{R}_1$. Expanding $\mathbf{G}(\mathbf{R}) = (\mathbf{A}_p^{t_n*}\mathbf{R}^*\mathbf{A}_p^{t_n T} + \sigma_w^2\mathbb{I})^{-1} \otimes (\mathbf{A}_p^{t_n}\mathbf{R}\mathbf{A}_p^{t_n H} + \sigma_w^2\mathbb{I})^{-1}$, we get

$$\mathbf{G}(\mathbf{R}) = (\mathbf{A}_p^{t_n *}\mathbf{R}^*\mathbf{A}_p^{t_n T} \otimes \mathbf{A}_p^{t_n}\mathbf{R}\mathbf{A}_p^{t_n H} + \sigma_w^2\mathbb{I} \otimes \mathbf{A}_p^{t_n}\mathbf{R}\mathbf{A}_p^{t_n H} \\ \mathbf{A}_p^{t_n *}\mathbf{R}^*\mathbf{A}_p^{t_n T} \otimes \sigma_w^2\mathbb{I} + \sigma_w^4\mathbb{I})^{-1}$$

Using the eigenvalue decomposition of the positive definite matrix $\mathbf{A}_p^{t_n}\mathbf{R}\mathbf{A}_p^{t_n H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, we have $\mathbf{G}(\mathbf{R}) = (\mathbf{V}(\mathbf{\Lambda} \otimes \mathbf{\Lambda})\mathbf{V}^H + \mathbf{V}(\mathbb{I} \otimes \mathbf{\Lambda} + \mathbf{\Lambda} \otimes \mathbb{I})\mathbf{V}^H\sigma_w^2 + \sigma_w^4\mathbb{I})^{-1}$, where $\mathbf{V} = (\mathbf{U}^* \otimes \mathbf{U})$. Let $(\lambda_i + \lambda_j)\sigma_w^2$ denote a diagonal element of $\sigma_w^2(\mathbb{I} \otimes \mathbf{\Lambda} + \mathbf{\Lambda} \otimes \mathbb{I})$. Using the arithmetic–geometric mean inequality, this term can be lower bounded as $2\sqrt{\lambda_i\lambda_j}\sigma_w^2$. Substituting this into the expression above gives

$$\begin{aligned} &\mathbf{G}(\mathbf{R}) \preceq \mathbf{G}_1(\mathbf{R}) \qquad (36)\\ &= (\mathbf{V}(\mathbf{\Lambda} \otimes \mathbf{\Lambda})\mathbf{V}^H + 2\sigma_w^2\mathbf{V}(\mathbf{\Lambda}^{1/2} \otimes \mathbf{\Lambda}^{1/2})\mathbf{V}^H + \sigma_w^4\mathbb{I})^{-1} \\ &= ((\mathbf{A}_p^{t_n *} \otimes \mathbf{A}_p^{t_n})(\mathbf{R}^* \otimes \mathbf{R})(\mathbf{A}_p^{t_n T} \otimes \mathbf{A}_p^{t_n H}) + \\ &\quad 2\sigma_w^2((\mathbf{A}_p^{t_n *} \otimes \mathbf{A}_p^{t_n})(\mathbf{R}^* \otimes \mathbf{R})(\mathbf{A}_p^{t_n T} \otimes \mathbf{A}_p^{t_n H}))^{\frac{1}{2}} + \sigma_w^4\mathbb{I})^{-1} \end{aligned}$$

Let us define the transformation $\mathbf{Z} = ((\mathbf{A}_p^{t_n *} \otimes \mathbf{A}_p^{t_n})(\mathbf{R}^* \otimes \mathbf{R})(\mathbf{A}_p^{t_n T} \otimes \mathbf{A}_p^{t_n H}))^{-1}$. Then, $\mathbf{G}_1(\mathbf{R}) = ((\mathbf{A}_p^{t_n *} \otimes \mathbf{A}_p^{t_n})(\mathbf{R}^* \otimes \mathbf{R})(\mathbf{A}_p^{t_n T} \otimes \mathbf{A}_p^{t_n H}) + 2\sigma_w^2(\mathbf{A}_p^{t_n *} \otimes \mathbf{A}_p^{t_n})(\mathbf{R}^* \otimes \mathbf{R})(\mathbf{A}_p^{t_n T} \otimes \mathbf{A}_p^{t_n H})^{\frac{1}{2}} + \sigma_w^4\mathbb{I})^{-1} = \mathbf{G}_2(\mathbf{Z})$, where $\mathbf{G}_2(\mathbf{Z}) = (\mathbf{Z}^{-1} + 2\sigma_w^2\mathbf{Z}^{-1/2} + \sigma_w^4\mathbb{I})^{-1}$ with $\mathbf{Z}$ appropriately defined. The function $\mathbf{G}_2(\mathbf{Z})$ is concave w.r.t. the matrix inequality (the proof is in the supplemental material). Thus, by Jensen's inequality, $\mathbb{E}(\mathbf{G}_2(\mathbf{Z})) \preceq \mathbf{G}_2(\mathbb{E}(\mathbf{Z}))$. From (36), the matrix $\mathbf{G}(\mathbf{R})$ is upper bounded by $\mathbf{G}_1(\mathbf{R}) = \mathbf{G}_2(\mathbf{Z})$. Hence, $\mathbb{E}(\mathbf{G}(\mathbf{R}))$ can be upper bounded as

$$\mathbb{E}(\mathbf{G}(\mathbf{R})) \preceq \mathbb{E}(\mathbf{G}_2(\mathbf{Z})) \preceq \mathbf{G}_2(\mathbb{E}(\mathbf{Z})). \qquad (37)$$

The closed-form expression of $\mathbb{E}(\mathbf{Z})$ is

$$\mathbb{E}(\mathbf{Z}) = \mathbf{F}_p^{t_n} = (\mathbf{A}_p^{t_n T} \otimes \mathbf{A}_p^{t_n H})^{\dagger}\mathbf{\Gamma}((\mathbf{A}_p^{t_n *} \otimes \mathbf{A}_p^{t_n})^{\dagger},$$

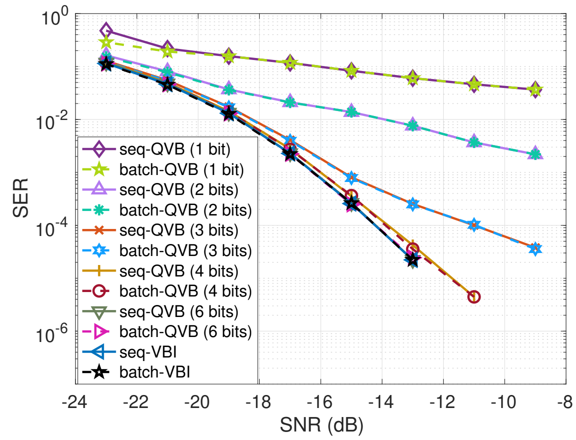

Fig. 15: SER vs. SNR, with $N_r = 30$, $K = 6$, $\tau_d = 250$, $\tau_p = 24$, $N = 500$. The SER with 4-bit ADC measurements is comparable to that with unquantized measurements.

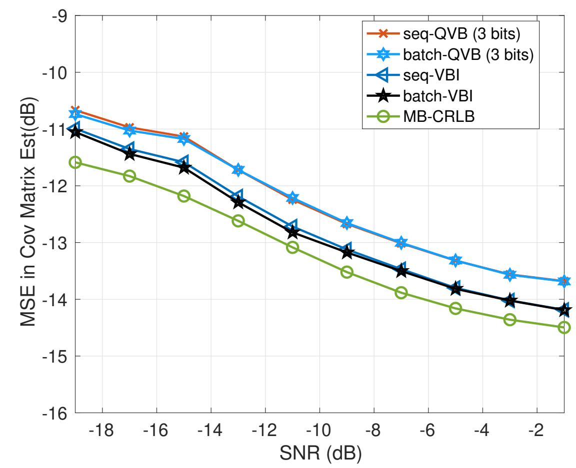

Fig. 16: MSE in channel *covariance matrix estimation* vs. SNR with $N_r = 20$, $K = 4$, $\tau_p = 4$ and $N = 1000$. Comparison of different QVB based algorithms with the MB-CRLB in Theorem 2.

$$\begin{aligned} \text{where, } \mathbf{\Gamma} &= \mathbb{E}\left((\mathbf{R}^*)^{-1} \otimes (\mathbf{R})^{-1}\right) \\ &= \nu^2({\mathbf{\Psi}_c^{-1}}^T \otimes \mathbf{\Psi}_c^{-1}) + \nu \mathrm{vec}\left(\mathbf{\Psi}_c^{-1}\right)\mathrm{vec}\left(\mathbf{\Psi}_c^{-1}\right)^H. \end{aligned} \qquad (38)$$

The derivation of $\mathbb{E}\big((\mathbf{R}^*)^{-1} \otimes (\mathbf{R})^{-1}\big)$ is given in the supplemental material. Using (37) and (38), the final upper bound becomes

$$\mathbb{E}\left((\mathbf{A}_p^{t_n*}\mathbf{R}^*\mathbf{A}_p^{t_n T} + \sigma_w^2\mathbb{I})^{-1} \otimes (\mathbf{A}_p^{t_n}\mathbf{R}\mathbf{A}_p^{t_n H} + \sigma_w^2\mathbb{I})^{-1}\right) \\ \preceq \left({\mathbf{F}_p^{t_n}}^{-1} + 2\sigma_w^2{\mathbf{F}_p^{t_n}}^{-1/2} + \sigma_w^4\mathbb{I}\right)^{-1}.$$

Using a similar approach, we can show that $\mathbb{E}((\mathbf{A}_p^{t_n*}\mathbf{R}^*\mathbf{A}_p^{t_n T} + \sigma_w^2\mathbb{I})^{-1} \otimes (\mathbf{A}_p^{t_n}\mathbf{R}\mathbf{A}_p^{t_n H} + \sigma_w^2\mathbb{I})^{-1}) \succeq (\mathbf{A}_p^{t_n *}\mathbb{E}(\mathbf{R})^*\mathbf{A}_p^{t_n T} + \sigma_w^2\mathbb{I})^{-1} \otimes (\mathbf{A}_p^{t_n}\mathbb{E}(\mathbf{R})\mathbf{A}_p^{t_n H} + \sigma_w^2\mathbb{I})^{-1}$. The details of the derivation of the above lower bound are provided in the supplemental material.

Appendix B: Proof of Lemma 1

From (32), the expectations of $\{\mathbf{G}_{ij}^{k}(\mathbf{R})\}_{i=1,j=1}^{i=N_r,j=N_r}$ w.r.t. the pdf of $\mathbf{R}$ are required for computing the M-FIM. Here, $\mathbf{G}_{ij}^{k}(\mathbf{R})$ is given by (31), which consists of five terms. We evaluate the expectations of these terms one by one. Considering the first term, we note that $\mathrm{vec}({\mathbf{A}_{pk}^{t_n}}^H \mathbf{Q}_{p,ij}^{t_n} \mathbf{A}_{pk}^{t_n}) = \left({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H\right) \times \left(\left({\mathbf{C}_p^{t_n}}^T\right)^{-1} \otimes \left(\mathbf{C}_p^{t_n}\right)^{-1}\right) \times \left({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H\right)^H \mathrm{vec}(\mathbf{E}_{ij})$. Collecting $\{\mathrm{vec}(\sum_{n=1}^{N} {\mathbf{A}_{pk}^{t_n}}^H \mathbf{Q}_{p,ij}^{t_n} \mathbf{A}_{pk}^{t_n})\}_{i=1,j=1}^{i=N_r,j=N_r}$ together, we obtain the $N_r^2 \times N_r^2$ matrix $\mathbf{S}_p^k(\mathbf{R})$ given by

$$\mathbf{S}_p^k(\mathbf{R}) = \sum_{n=1}^{N} \big(({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H)(({\mathbf{A}_p^{t_n}}^* \mathbf{R}^* {\mathbf{A}_p^{t_n}}^T + \sigma_w^2 \mathbb{I})^{-1} \otimes (\mathbf{A}_p^{t_n} \mathbf{R} \mathbf{A}_p^{t_n H} + \sigma_w^2 \mathbb{I})^{-1})({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H)^H\big)$$

We compute the expectation of $\mathbf{S}_p^k(\mathbf{R})$ w.r.t. $p(\mathbf{R})$:

$$\mathbb{E}(\mathbf{S}_p^k(\mathbf{R})) = \sum_{n=1}^{N} \Bigg(({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H)\mathbb{E}\left\{({\mathbf{A}_p^{t_n}}^* \mathbf{R}^* {\mathbf{A}_p^{t_n}}^T + \sigma_w^2 \mathbb{I})^{-1} \otimes (\mathbf{A}_p^{t_n} \mathbf{R} \mathbf{A}_p^{t_n H} + \sigma_w^2 \mathbb{I})^{-1}\right\}({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H)^H\Bigg).$$

The exact computation of $\mathbb{E}(({\mathbf{A}_p^{t_n}}^* \mathbf{R}^* {\mathbf{A}_p^{t_n}}^T + \sigma_w^2 \mathbb{I})^{-1} \otimes (\mathbf{A}_p^{t_n} \mathbf{R} \mathbf{A}_p^{t_n H} + \sigma_w^2 \mathbb{I})^{-1})$ is difficult. Hence, we upper bound it using Theorem 1 to get a lower bound on the MSE. Thus, an upper bound of $\mathbb{E}(\mathbf{S}_p^k(\mathbf{R}))$ is given by

$$\sum_{n=1}^{N} \big(({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H)\mathbf{D}_p^{t_n}({\mathbf{A}_{pk}^{t_n}}^T \otimes {\mathbf{A}_{pk}^{t_n}}^H)^H\big). \tag{39}$$

Similarly, we upper bound the second term, $\mathbb{E}(\mathbf{S}_d^k(\mathbf{R}))$, as

$$\sum_{n=1}^{N} \big(({\mathbf{A}_{dk}^{t_n}}^T \otimes {\mathbf{A}_{dk}^{t_n}}^H)\mathbf{D}_d^{t_n}({\mathbf{A}_{dk}^{t_n}}^T \otimes {\mathbf{A}_{dk}^{t_n}}^H)^H\big),$$

where $\mathbf{D}_p^{t_n}$ and $\mathbf{D}_d^{t_n}$ are $({\mathbf{F}_p^{t_n}}^{-1} + 2\sigma_w^2 {\mathbf{F}_p^{t_n}}^{-1/2} + \sigma_w^4 \mathbb{I})^{-1}$ and $({\mathbf{F}_d^{t_n}}^{-1} + 2\sigma_w^2 {\mathbf{F}_d^{t_n}}^{-1/2} + \sigma_w^4 \mathbb{I})^{-1}$, respectively. The matrices $\mathbf{F}_p^{t_n}$ and $\mathbf{F}_d^{t_n}$ are $({\mathbf{A}_p^{t_n}}^T \otimes {\mathbf{A}_p^{t_n}}^H)^{\dagger} \mathbf{\Gamma} (({\mathbf{A}_p^{t_n}}^* \otimes \mathbf{A}_p^{t_n})^{\dagger}$, $({\mathbf{A}_d^{t_n}}^T \otimes {\mathbf{A}_d^{t_n}}^H)^{\dagger} \mathbf{\Gamma} (({\mathbf{A}_d^{t_n}}^* \otimes \mathbf{A}_d^{t_n})^{\dagger}$, respectively. The matrix $\mathbf{\Gamma} = \nu^2({\mathbf{\Psi}_c^{-1}}^T \otimes \mathbf{\Psi}_c^{-1}) + \nu \mathrm{vec}\left(\mathbf{\Psi}_c^{-1}\right) \mathrm{vec}\left(\mathbf{\Psi}_c^{-1}\right)^H$, where $\mathbf{\Psi}_c = \mathrm{blkdiag}(\mathbf{\Psi}, \mathbf{\Psi}, \ldots, \mathbf{\Psi}) \in \mathbb{C}^{KN_r \times KN_r}$. The computation of the expectations of the remaining terms, namely, $\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1}$, $\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1}\mathbf{\Psi}\mathbf{R}_k^{-1}$ and $\mathbf{R}_k^{-1}\mathbf{\Psi}\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1}$, w.r.t. $p(\mathbf{R})$ are detailed in supplemental material. The expression for $\mathbb{E}(\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1})$ is given by

$$\mathbb{E}(\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1}) = \nu \mathbf{\Psi}^{-1}\mathbf{e}_j^T \mathbf{\Psi}^{-1}\mathbf{e}_i + \nu^2 \mathbf{\Psi}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{\Psi}^{-1}.$$

Let $\mathbf{S}_{w_1}^k(\mathbf{R}) \in \mathbb{C}^{N_r^2 \times N_r^2}$ be the matrix with columns $\{\mathrm{vec}(\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1})\}_{i=1,j=1}^{i=N_r,j=N_r}$. Then, $\mathbb{E}(\mathbf{S}_{w_1}^k(\mathbf{R}))$ is given by $\mathbb{E}(\mathbf{S}_{w_1}^k(\mathbf{R})) = \nu \mathrm{vec}(\mathbf{\Psi}^{-1})\mathrm{vec}(\mathbf{\Psi}^{-1})^H + \nu^2({\mathbf{\Psi}^{-1}}^T \otimes \mathbf{\Psi}^{-1})$.

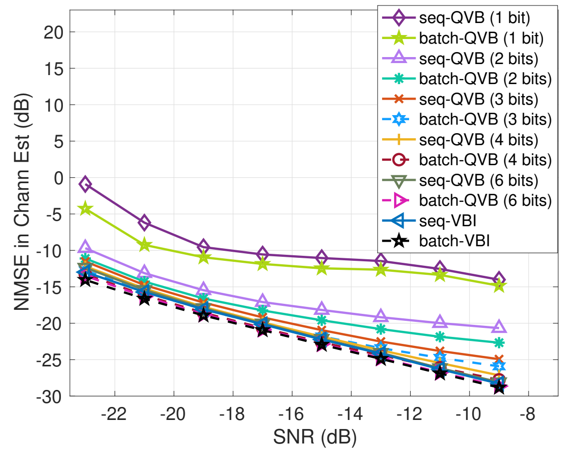

Fig. 17: NMSE vs. SNR for varying ADC resolutions ($N_r = 30, K = 6, \tau_d = 250, \tau_p = 24, N = 500$). QVB performance improves with resolution, matching the unquantized seq-VBI and batch-VBI baselines at 4-bits.

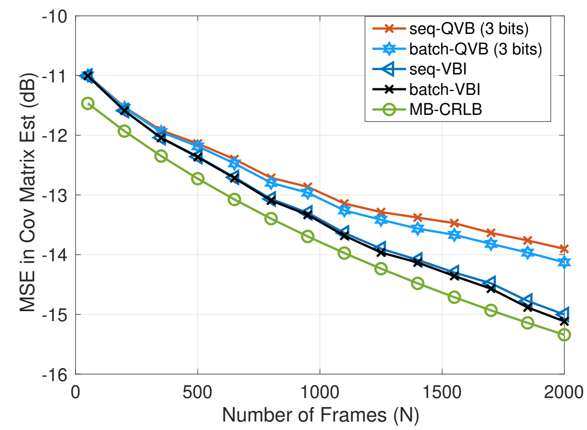

Fig. 18: MSE in channel covariance matrix estimation vs. number of frames, N, with $N_r = 20$, $K = 4$. The MSE (in dB) decreases linearly with N, particularly in the absence of quantization noise.

The expression for $\mathbb{E}(\mathbf{R}_k^{-1}\mathbf{\Psi}\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1})$ is given by

$$\mathbb{E}(\mathbf{R}_k^{-1}\mathbf{\Psi}\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1}) = (\nu N_r + 2\nu^2)\mathbf{\Psi}^{-1}\mathbf{e}_j^T\mathbf{\Psi}^{-1}\mathbf{e}_i + (\nu + N_r\nu^2 + \nu^3)\mathbf{\Psi}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{\Psi}^{-1}.$$

Arranging $\{\mathrm{vec}(\mathbf{R}_k^{-1}\mathbf{\Psi}\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1})\}_{i=1,j=1}^{i=N_r,j=N_r}$ as columns of $\mathbf{S}_{w_2}^k(\mathbf{R}) \in \mathbb{C}^{N_r^2 \times N_r^2}$ and taking the expectation, we get

$$\mathbb{E}(\mathbf{S}_{w_2}^k(\mathbf{R})) = (\nu N_r + 2\nu^2)\mathrm{vec}(\mathbf{\Psi}^{-1})\mathrm{vec}(\mathbf{\Psi}^{-1})^H + (\nu + N_r\nu^2 + \nu^3){\mathbf{\Psi}^{-1}}^T \otimes \mathbf{\Psi}^{-1}.$$

Similarly, we construct the matrix $\mathbf{S}_{w_3}^k(\mathbf{R})$ with columns $\{\mathrm{vec}(\mathbf{R}_k^{-1}\mathbf{E}_{ij}\mathbf{R}_k^{-1}\mathbf{\Psi}\mathbf{R}_k^{-1})\}_{i=1,j=1}^{i=N_r,j=N_r}$, and it is easy to show that $\mathbb{E}(\mathbf{S}_{w_3}^k(\mathbf{R})) = \mathbb{E}(\mathbf{S}_{w_2}^k(\mathbf{R}))$.

Using the above upper bounds, the M-FIM of the kth user, $\mathbf{J}_{B_k} \in \mathbb{C}^{N_r^2 \times N_r^2}$, can be upper bounded as

$$\mathbf{J}_{B_k} \preceq \sum_{n=1}^{N} \big(({\mathbf{A}_{dk}^{t_n}}^T \otimes {\mathbf{A}_{dk}^{t_n}}^H)\mathbf{D}_d^{t_n}({\mathbf{A}_{dk}^{t_n}}^T \otimes {\mathbf{A}_{dk}^{t_n}}^H)^H\big) \tag{40}$$

$$+\sum_{n=1}^{N}\big((\mathbf{A}_{pk}^{t_n\,T}\otimes\mathbf{A}_{pk}^{t_n\,H})\mathbf{D}_p^{t_n}(\mathbf{A}_{pk}^{t_n\,T}\otimes\mathbf{A}_{pk}^{t_n\,H})^H\big)$$
$$+b_1(\mathbf{\Psi}^{-1\,T}\otimes\mathbf{\Psi}^{-1})+b_2\,\mathrm{vec}(\mathbf{\Psi}^{-1})\,\mathrm{vec}(\mathbf{\Psi}^{-1})^H,$$

where the scalar constants b_1 and b_2 are $N_r\nu^2+2\nu+\nu^3$, and $N_r\nu+3\nu^2$, respectively.

REFERENCES

[1] S. Buzzi and C. D'Andrea, "Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmwave systems," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 385–396, Jun. 2018.

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

[3] K. Safa, R. Combes, R. de Lacerda, and S. Yang, "Data detection in 1-bit quantized MIMO systems," *IEEE Trans. Commun.*, vol. 72, no. 9, pp. 5396–5410, Sep. 2024.

[4] 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.

[5] 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.

[6] P. Sun, Z. Wang, R. W. Heath, and P. Schniter, "Joint channel-estimation/decoding with frequency-selective channels and few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 899–914, Feb. 2019.

[7] Y. Xiong, L. Qin, S. Sun, L. Liu, S. Mao, Z. Zhang, and N. Wei, "Joint effective channel estimation and data detection for RIS-aided massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 721–725, Feb. 2023.

[8] 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.

[9] M. Esfandiari, S. A. Vorobyov, and R. W. Heath, "AdaBoost-based efficient channel estimation and data detection in one-bit massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13 935–13 945, Oct. 2024.

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

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

[15] T.-V. Nguyen, S. Nassirpour, I. Atzeni, A. Tölli, A. L. Swindlehurst, and D. H. N. Nguyen, "MIMO detection with spatial Sigma-Delta ADCs: A variational Bayesian approach," *IEEE Trans. Signal Process.*, 2025.

[16] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "Variational Bayes for joint channel estimation and data detection in few-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 72, pp. 3408–3423, 2024.

[17] M.-A. Badiu, G. E. Kirkelund, C. N. Manchón, E. Riegler, and B. H. Fleury, "Message-passing algorithms for channel estimation and decoding using approximate inference," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 2376–2380.

[12] S. S. Thoota and C. R. Murthy, "Variational Bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low-resolution ADCs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3467–3481, May 2021.

[13] X. Cheng, B. Xia, K. Xu, and S. Li, "Bayesian channel estimation and data detection in oversampled OFDM receiver with low-resolution ADC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5558–5571, Sep. 2021.

[14] S. Nassirpour, T.-V. Nguyen, H. Q. Ngo, L.-N. Tran, T. Ratnarajah, and D. H. N. Nguyen, "Variational Bayesian channel estimation and data detection for cell-free massive MIMO with low-resolution quantized fronthaul links," *IEEE J. Sel. Topics Signal Process.*, 2025.

[18] S. Katyal, S. B. Hayavadana, and C. R. Murthy, "Deep unfolding-based channel estimation and soft symbol decoding with low-resolution ADCs," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2024, pp. 862–866.

[19] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, Jun. 2018.

[20] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *Proc. 50th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2016, pp. 974–978.

[21] K. Upadhya and S. A. Vorobyov, "Covariance matrix estimation for massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 546–550, Apr. 2018.

[22] A. K. Kocharlakota, K. Upadhya, and S. A. Vorobyov, "Impact of pilot overhead and channel estimation on the performance of massive MIMO," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8242–8255, Dec. 2021.

[23] A. Decurninge and M. Guillaud, "Covariance estimation with projected data: Applications to CSI covariance acquisition and tracking," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 628–632.

[24] S. Nassirpour, T.-V. Nguyen, and D. H. N. Nguyen, "Variational Bayesian inference for time-varying massive MIMO channels: Estimation and detection," *IEEE Trans. Wireless Commun.*, 2025.

[25] N. Kolomvakis, T. Eriksson, M. Coldrey, and M. Viberg, "Quantized uplink massive MIMO systems with linear receivers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.

[26] H. Djelouat, M. J. Sillanpää, M. Leinonen, and M. Juntti, "Hierarchical MTC user activity detection and channel estimation with unknown spatial covariance," *IEEE Trans. Wireless Commun.*, 2024.

[27] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[28] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proceedings of the IEEE International Conference on Communications (ICC)*, vol. 2. IEEE, Jun 1995, pp. 1009–1013.

[29] H. L. Van Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Hoboken, NJ, USA: Wiley-IEEE Press, 2007.

[30] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[31] H. He, R. Wang, W. Jin, S. Jin, C.-K. Wen, and G. Y. Li, "Beamspace channel estimation for wideband millimeter-wave MIMO: A model-driven unsupervised learning approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1808–1822, Mar. 2023.