Joint Channel Estimation and Data Detection for Time-Varying MIMO Channels in UAV Networks

Sajjad Nassirpour and Duy H. N. Nguyen
Department of Electrical and Computer Engineering, San Diego State University, San Diego, USA
Email: snassirpour@sdsu.edu and duy.nguyen@sdsu.edu

***Abstract*—In the context of wireless communications, channel estimation and data detection are two pivotal tasks that exert considerable influence on system performance, such as enhancing spectral efficiency and improving quality of service (QoS). Nevertheless, these tasks become more challenging as user mobility increases, e.g., in unmanned aerial vehicle (UAV) networks, leading to channel variations over time. Consequently, our focus in this study is on an uplink massive multi-input multi-output (MIMO) system in a UAV network with time-varying channels. We propose an online processing approach for joint estimation and detection (JED) in these channels. Specifically, we introduce a method that utilizes variational Bayes (VB) inference to approximate the true posterior distributions. Our assessment of the VB method encompasses its performance in terms of symbol error rate (SER) and computational complexity. Additionally, we conduct an analysis of how time correlation and communication time impact its effectiveness.**

***Index Terms*—Bayesian inference, detection, estimation, massive MIMO, variational Bayesian, Unmanned aerial vehicles.**

I. Introduction

Unmanned aerial vehicles (UAVs) have become invaluable in the current and future wireless communication networks due to their efficient and automatic execution of various crucial tasks. The incorporation of UAVs into wireless communication networks has emerged as a prominent area of research, with two primary approaches [1], [2]. The first approach centers on utilizing UAVs as base stations (BSs) to enhance network capacity, expand coverage, and swiftly establish a mobile network architecture during catastrophic events [1]. The second approach concentrates on exploring communication services that wireless networks can offer to UAVs [2].

Within the latter approach, the authors in [3] have delved into the application of massive multiple-input multiple-output (MIMO) technology to support cellular communications for UAVs, aiming to enhance reliability in UAV command and control (C&C) links. Furthermore, [4] introduced a 3D channel model tailored for network-connected UAVs and presented a coverage analysis applicable to different network deployments.

The adoption of massive MIMO results in large channel matrices, which necessitate the development of effective algorithms for channel estimation and data detection. In uplink scenarios, channel estimation is conventionally achieved using known pilot signals. Nevertheless, this method encounters scalability challenges, as it requires having a number of orthogonal pilot signals equal to or greater than the number of users. This limitation becomes more pronounced as the number of users grows. To address this issue, literature has introduced blind channel estimation algorithms [5], [6], which solely depend on the received signals. However, these methods are susceptible to phase ambiguities in the demodulated symbols.

Semi-blind channel estimation is another technique to reduce dependency on orthogonal pilot signals, which incorporates a limited number of known pilot signals alongside received signals to enhance the accuracy of the estimation procedure. In [7], the focus was on semi-blind channel estimation for multi-user MIMO systems in which two methods based on the expectation-maximization framework were introduced for channel estimation. Then, in [8], the investigation centered on channel estimation within a multi-cell multi-user massive MIMO network. It proposed a method that estimates the uplink data from the target cell and then obtains the least square channel estimate by treating the detected uplink data as pilot symbols. The work in [9] studied the joint estimation and detection (JED) problem in hybrid massive MIMO systems and proposed two iterative algorithms via a low-rank matrix completion formulation. Furthermore, the authors of [10] introduced an iterative algorithm based on nonlinear optimization, tailored to address a relaxed version of the maximum a-posteriori JED problem in cell-free massive MIMO systems, providing point estimates for the data symbols.

In the aforementioned algorithms, it is assumed that the channels remain time-invariant during the communication period. This assumption is only valid for static channels (i.e., block fading channel model), where users are either static or have low mobility. However, UAVs are high-speed users, and in this case, the channels corresponding to UAVs experience time variations due to Doppler spread. Therefore, channel information needs to be updated instantaneously throughout the communication period [11].

Motivated by the above, in this paper, we investigate an uplink massive MIMO system in UAV networks and propose an online processing technique based on variational Bayes (VB) inference, enabling JED for time-varying channels. VB inference is a robust statistical inference framework derived from machine learning, which addresses the challenge of approximating the posterior distribution of latent variables. The VB inference achieves this by optimizing simpler distributions from a known family to replace the intractable true posterior distributions. In this study, we employ the mean-field variational family to serve as an approximation for the true posterior distributions. Additionally, we assume that the receiver (i.e., BS) lacks knowledge of the noise variance to highlight the influence of residual inter-user interference in our computations.

We compare the performance of our proposed VB method with the conventional linear minimum mean squared error (LMMSE) and maximum likelihood (ML) detection methods. Our results demonstrate that VB surpasses LMMSE and ML in terms of symbol error rate (SER) and computational complexity. Moreover, we analyze how time correlation and communication time impact the performance of the VB method.

Paper organization: Section II includes the system model. The proposed VB method for JED is explained in Section III. The numerical analysis is demonstrated in Section IV, and Section V concludes the paper.

Notations: Throughout this paper, italic, bold-face lowercase, and bold-face uppercase letters are utilized to denote scalars, vectors, and matrices, respectively; $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$ represents a complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$; $\mathbb{C}^{a\times b}$ is the space of an $a\times b$ dimensional complex-valued matrix; The symbols $\sim$ and $\propto$ represent the concepts of being "distributed according to" and "proportional to," respectively; $x \sim \Gamma(a, b)$ signifies that x follows a Gamma distribution with parameters a and b; $\mathrm{diag}\{\mathbf{a}\}$ is a diagonal matrix based on $\mathbf{a}$; $|\mathbf{A}|$, $\|\mathbf{A}\|$, $\mathrm{Tr}\,(.)$, $\mathbf{A}^\top$, and $\mathbf{A}^H$ indicate determinant of $\mathbf{A}$, Euclidean norm of $\mathbf{A}$, the trace function, the transpose of $\mathbf{A}$, and the conjugate transpose of $\mathbf{A}$, respectively; $\mathbf{I}_M$ is an identity matrix with size M; $x^\star$ and $\Re\{x\}$ represent the conjugate of x and the real part of x, respectively; $\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{\pi^K|\mathbf{\Sigma}|}\exp\big(-(\mathbf{x}-\boldsymbol{\mu})^H\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\big)$ is the probability density function (PDF) of a length-K random vector $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{\Sigma})$; $\mathbb{E}_{p(x)}[x]$ and $\mathrm{Var}_{p(x)}[x]$ denote the mean and variance of x with respect to its distribution $p(x)$; $\langle x\rangle$, $\langle|x|^2\rangle$, and τ^x denote the mean, the second moment, and the variance of x with respect to a variational distribution $q(x)$; $[\mathbf{X}]_{ij}$ represent the element at the ith row and jth column of a matrix $\mathbf{X}$; Index $t|t-1$ denotes the predicted statistics at time t using the statistics at time $t-1$; Index $t|t$ represents the a posteriori estimated statistics at time t using the observation at time t.

II. System Model

We consider an uplink massive MIMO system with K single-antenna UAVs and a BS with M receive antennas, as shown in Fig. 1. In this scenario, the linear uplink MIMO system in time slot t can be modeled as:

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{n}_t, \tag{1}$$

where $\mathbf{y}_t \in \mathbb{C}^{M\times 1}$ is the received signal vector, $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \ldots, x_{K,t}]^\top \in \mathbb{C}^{K\times 1}$, where $x_{i,t}$ is the transmitted signal from UAV$_i$, $1 \le i \le K$, $\mathbf{H}_t = [\mathbf{h}_{1,t}, \mathbf{h}_{2,t}, \ldots, \mathbf{h}_{K,t}] \in \mathbb{C}^{M\times K}$ represents the uplink channels where $\mathbf{h}_{i,t} \in \mathbb{C}^{M\times 1}$, denotes the channel between UAV$_i$ and the BS, and $\mathbf{n}_t \sim \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}_M)$ models the independent and identically distributed (i.i.d.) additive Gaussian noise at the BS. It is assumed that channel vector $\mathbf{h}_{i,t}$ is Gaussian distributed with $p(\mathbf{h}_{i,t}) = \mathcal{CN}(\mathbf{h}_{i,t}; \mathbf{0}, \mathbf{R}_i)$, where $\mathbf{R}_i \triangleq \mathbb{E}[\mathbf{h}_{i,t}\mathbf{h}_{i,t}^H]$ is the covariance matrix. Without loss of generality, in this paper, we consider that $\mathbf{R}_i = \mathbf{I}_M$. We also assume that the channels from different UAVs are independent of each other, i.e., $\mathbb{E}[\mathbf{h}_i\mathbf{h}_j^H] = \mathbf{0}$, if $1 \le i, j \le K, i \ne j$.

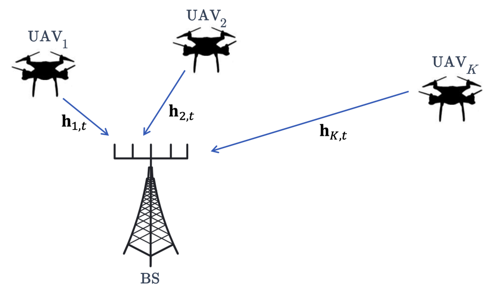

Fig. 1. An uplink massive MIMO serving K single-antenna UAVs.

The transmitted symbol vector $\mathbf{x}_t$ are drawn *independently* from a discrete constellation $\mathcal{S}$, e.g., 16-quadrature amplitude modulation (16QAM) and quadrature phase-shift keying (QPSK), such that the prior distribution $p(\mathbf{x}_t)$ is given by $p(\mathbf{x}_t) = \prod_{i=1}^K p(x_{i,t})$. The symbol $x_{i,t}$ can be a *known pilot* symbol or an *unknown data* symbol. If it is a pilot symbol, then the prior distribution of $x_{i,t}$ is given by $p(x_{i,t}) = \delta(x_{i,t} - \bar{x}_{i,t})$, where $\bar{x}_{i,t}$ is *known* at the BS. If it is a data symbol, the prior distribution of $x_{i,t}$ is *discrete* with $p(x_{i,t}) = \sum_{a\in\mathcal{S}} p_a\delta(x_{i,t} - a)$, where p_a corresponds to a known prior probability of the constellation point $a \in \mathcal{S}$.

A. Modeling Time-Varying Channel

During the data transmission phase, it is assumed that channel $\mathbf{h}_{i,t}, t = 1, \ldots, T$, follows the first-order Gauss-Markov model as below.

$$\mathbf{h}_{i,t} = \eta_i\mathbf{h}_{i,t-1} + \sqrt{1-\eta_i^2}\,\mathbf{g}_{i,t}, \tag{2}$$

where $\mathbf{h}_{i,0} = \mathbf{g}_{i,0}$, $\mathbf{g}_{i,t} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ denotes the innovation process, and $0 \le \eta_i \le 1$ represents the time correlation coefficient corresponding to UAV$_i$. In this paper, in order to gain insight into the online processing in time-varying channels, it is assumed that η_i to be known. The acquisition of η_i would be an interesting future research direction. It is also considered that $\mathbf{g}_{i,t}$ is uncorrelated with $\mathbf{h}_{i,t-1}$, such that $\mathbb{E}\left[\mathbf{h}_{i,t-1}\mathbf{h}_{i,t}^H\right] = \eta_i$ and

$$p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}; \eta_i) = \mathcal{CN}\left(\mathbf{h}_{i,t}; \eta_i\mathbf{h}_{i,t-1}, 1-\eta_i^2\right). \tag{3}$$

Notice that if $\eta_i = 1$, the Gauss-Markov channel model effectively becomes the block fading model. In this work, we focus on time-varying channels ($\eta_i \ne 1$) and propose a VB-based method suited for the JED problem. A background on VB is presented below.

B. Background on VB

This section provides a summary of the variational Bayes (VB) method for approximate inference. Denote the set of all observed variables by $\mathbf{y}$ and set of L latent variables and parameters by $\mathbf{x}$. For detection purposes, it is essential to find the posterior $p(\mathbf{x}|\mathbf{y})$, which is computationally intractable. To tackle this issue, the VB method was introduced to obtain a distribution $q(\mathbf{x})$ with its own setting of variational parameters within a family $\mathcal{Q}$ of densities such that $q(\mathbf{x})$ approximates $p(\mathbf{x}|\mathbf{y})$. To do so, the VB method defines the following optimization problem using the Kullback-Leibler (KL) divergence from $q(\mathbf{x})$ to $p(\mathbf{x}|\mathbf{y})$.

$$q^{\star}(\mathbf{x}) = \underset{q(\mathbf{x})\in\mathcal{Q}}{\arg\min}\ \mathrm{KL}\big(q(\mathbf{x})\|p(\mathbf{x}|\mathbf{y})\big), \tag{4}$$

where $q^{\star}(\mathbf{x})$ is the optimal variational distribution and

$$\mathrm{KL}\big(q(\mathbf{x})\|p(\mathbf{x}|\mathbf{y})\big) = \mathbb{E}_{q(\mathbf{x})}\big[\ln q(\mathbf{x})\big] - \mathbb{E}_{q(\mathbf{x})}\big[\ln p(\mathbf{x}|\mathbf{y})\big]. \tag{5}$$

In [12], the authors showed that the optimization problem in (4) is equivalent to the following maximization problem.

$$q^{\star}(\mathbf{x}) = \underset{q(\mathbf{x})\in\mathcal{Q}}{\arg\max}\ \left\{\mathbb{E}_{q(\mathbf{x})}\left[\ln p(\mathbf{x},\mathbf{y})\right] - \mathbb{E}_{q(\mathbf{x})}\left[\ln q(\mathbf{x})\right]\right\}, \tag{6}$$

where the objective function is referred to as the evidence lower bound (ELBO). Here, (6) is maximized if $q(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$. Since obtaining the true posterior distribution is intractable, it is more practical to consider a restricted family of distribution $q(\mathbf{x})$. Hence, this work focuses on the *mean-field variational family*, which is defined as follows:

$$q(\mathbf{x}) = \prod_{i=1}^{L} q_i(x_i), \tag{7}$$

where the latent variables are mutually independent and each governed by a distinct factor in the variational density. The general expression for the optimal solution of the variational density $q_i(x_i)$ can be obtained as [12]:

$$\begin{aligned} q_i^{\star}(x_i) &\propto \exp\left\{\left\langle \ln p(\mathbf{y},\mathbf{x})\right\rangle\right\} \\ &\propto \exp\left\{\left\langle \ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x})\right\rangle\right\}, \end{aligned} \tag{8}$$

where $\langle\cdot\rangle$ represents the expectation with respect to all latent variables except x_i using the currently fixed variational density $q_{-i}(\mathbf{x}_{-i}) = \prod_{j=1,j\neq i}^{L} q_j(x_j)$. By sequentially updating $q_i^{\star}(x_i)$ in an iterative manner for all j, the objective function of (6) is monotonically enhanced. This forms the fundamental principle of the *Coordinate Ascent Variational Inference (CAVI)* algorithm, which ensures convergence to at least a local optimum solution for the optimization problem in (6) [13].

C. Problem Formulation

According to Section II-B, it is essential to compute joint distribution in order to apply the VB method. Here, the noise variance N_0 is assumed to be unknown *a priori*. We use $\gamma_t = 1/N_0$ to denote the precision of the noise at time slot t. Therefore, we compute joint distribution $p(\mathbf{y}_t, \mathbf{H}_t, \mathbf{H}_{t-1}, \mathbf{x}_t, \gamma_t; \boldsymbol{\eta})$ as follows:

$$\begin{aligned} &p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{H}_t, \mathbf{H}_{t-1}, \gamma_t; \boldsymbol{\eta}) \\ &\quad = p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{H}_t, \gamma_t) p(\mathbf{x}_t) p(\mathbf{H}_t|\mathbf{H}_{t-1}; \boldsymbol{\eta}) p(\gamma_t), \end{aligned} \tag{9}$$

where $p(\mathbf{H}_t|\mathbf{H}_{t-1}; \boldsymbol{\eta}) = \prod_{i=1}^{K} p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}; \eta_i)$ and $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_K]$.

III. VB for JED with Online Processing Strategy

We consider the detection at a particular time slot t. We assume that the channel from UAV$_i$ distributed as $p(\mathbf{h}_{i,t-1}) = \mathcal{CN}(\mathbf{h}_{i|t-1}; \hat{\mathbf{h}}_{i,t-1|t-1}, \boldsymbol{\Sigma}_{i,t-1|t-1}), t = 1, \ldots, T$, where we consider $p(\mathbf{h}_{i,0}) = \mathcal{CN}(\mathbf{h}_0; \mathbf{0}, \mathbf{I}_M)$ and $\mathbf{h}_{i,0}$ was acquired from the pilot transmission phase and the previous data transmission time slots. Our task is to estimate the data symbol $\mathbf{x}_t$ and approximate the posterior distribution of the channel $\mathbf{h}_{i,t}$. To achieve this, we propose a VB-based method, organized into two phases: prediction and estimation. The details of these phases are presented below.

Prediction Phase: In this phase, we use the (variational) posterior distribution of the unknown random variables at time slot $t-1$ as their prior distribution at time slot t. Here, we make the following assumptions:

Assumption 1: The variational distribution of the channel $\mathbf{h}_{i,t-1}$ is Gaussian $\mathcal{CN}\big(\hat{\mathbf{h}}_{i,t-1|t-1}, \boldsymbol{\Sigma}_{i,t-1|t-1}\big)$;

Assumption 2: The variational distribution of the noise precision γ_t is Gamma distributed as $\Gamma(a_0, b_0)$.

It is noted that the distribution of $\mathbf{h}_{i,t}$ depends on η_i. Based on the Gauss-Markov model in (2), the predictive distribution $p(\mathbf{h}_{i,t}; \eta_i)$ for a given η_i is $\mathcal{CN}(\mathbf{h}_{i,t}; \hat{\mathbf{h}}_{i,t|t-1}, \boldsymbol{\Sigma}_{i,t|t-1})$ where

$$\hat{\mathbf{h}}_{i,t|t-1} = \eta_i \hat{\mathbf{h}}_{i,t-1|t-1}, \tag{10}$$

$$\boldsymbol{\Sigma}_{i,t|t-1} = \eta_i^2 \boldsymbol{\Sigma}_{i,t-1|t-1} + (1-\eta_i^2). \tag{11}$$

The predictive distribution $p(\mathbf{h}_{i,t}; \eta_i) = \mathcal{CN}(\mathbf{h}_{i,t}; \hat{\mathbf{h}}_{i,t|t-1}, \boldsymbol{\Sigma}_{i,t|t-1})$ is now used as the prior distribution for $\mathbf{h}_i; \eta_i$ at time slot t.

Estimation Phase: In the estimation phase, our goal is to attain the Bayes-optimal estimation of $\mathbf{H}_t$ and $\mathbf{x}_t$, which can be found via the posterior distribution $p\big(\mathbf{x}_t, \mathbf{H}_t, \gamma_t|\mathbf{y}_t; \boldsymbol{\eta}\big)$. Since it is intractable to derive $p\big(\mathbf{x}_t, \mathbf{H}_t, \gamma_t|\mathbf{y}_t; \boldsymbol{\eta}\big)$, the mean-field variational distribution $q(\mathbf{x}_t, \mathbf{H}_t, \gamma_t)$ is utilized such that

$$\begin{aligned} p\big(\mathbf{x}_t, \mathbf{H}_t, \gamma_t|\mathbf{y}_t; \boldsymbol{\eta}\big) &\approx q(\mathbf{x}_t, \mathbf{H}_t, \gamma_t) \\ &= \left[\prod_{i=1}^{K} q_i(x_{i,t}) q(\mathbf{h}_{i,t})\right] q(\gamma_t). \end{aligned} \tag{12}$$

Based on (8), the joint distribution $p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{H}_t, \gamma_t; \boldsymbol{\eta})$ is needed to find the optimal solution of the variational densities in (12). To do so, $p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{H}_t, \gamma_t; \boldsymbol{\eta})$ can be factorized as:

$$\begin{aligned} &p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{H}_t, \gamma_t; \boldsymbol{\eta}) \\ &\quad = p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{H}_t, \gamma_t) \left[\prod_{i=1}^{K} p(x_{i,t}) p(\mathbf{h}_{i,t}; \eta_i)\right] p(\gamma_t). \end{aligned} \tag{13}$$

As mentioned earlier in Section II-B, the CAVI algorithm converges to the local optimum solutions $q^{\star}(x_{i,t})$, $q^{\star}(\mathbf{h}_{i,t})$, and $q^{\star}(\gamma_t)$ by iteratively optimizing one latent variable while the others are fixed. The following parts demonstrate how each latent variable can be updated.

1) Updating $\mathbf{h}_{i,t}$*:* By taking the expectation of the conditional (13) with respect to all latent variables except for $\mathbf{h}_{i,t}$, the variational distribution of $\mathbf{h}_{i,t}$ is given in (16). Thus, the variational distribution $q(\mathbf{h}_i)$ is Gaussian with the following covariance matrix and mean:

$$\boldsymbol{\Sigma}_{i,t} = \Big[\langle\gamma_t\rangle\langle|x_{i,t}|^2\rangle\mathbf{I}_M + \boldsymbol{\Sigma}_{i,t|t-1}^{-1}\Big]^{-1}, \tag{14}$$

$$\langle\mathbf{h}_{i,t}\rangle = \boldsymbol{\Sigma}_{i,t}\Bigg[\langle\gamma_t\rangle\Bigg(\mathbf{y}_t - \sum_{j\neq i}^{K}\langle\mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\Bigg)\langle x_{i,t}^*\rangle + \eta_i\boldsymbol{\Sigma}_{i,t|t-1}^{-1}\hat{\mathbf{h}}_{i,t-1|t-1}\Bigg]. \tag{15}$$

The following presents a lemma on the variational posterior mean of multiple random variables that will be applied later to update $x_{i,t}$ and γ_t.

Lemma 1. *Let $\mathbf{A}$ and $\mathbf{x}$ of size $m \times n$ and $n \times 1$ be two independent random matrices (vectors) with respect to a variational density $q_{\mathbf{A},\mathbf{x}}(\mathbf{A},\mathbf{x}) = q_{\mathbf{A}}(\mathbf{A})q_{\mathbf{x}}(\mathbf{x})$. It is assumed that $\mathbf{A}$ is column-wise independent and let $\langle\mathbf{a}_i\rangle$ and $\boldsymbol{\Sigma}_{\mathbf{a}_i}$ be the variational mean and covariance matrix of the ith column of $\mathbf{A}$. Let $\langle\mathbf{x}\rangle$ and $\boldsymbol{\Sigma}_{\mathbf{x}}$ be the variational mean and covariance matrix of $\mathbf{x}$. Let $\mathbf{y}$ be an arbitrary $m \times 1$ vector and define $\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle$ as the expectation of $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2$ with respect to $q_{\mathbf{A},\mathbf{x}}(\mathbf{A},\mathbf{x})$, we have:*

$$\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle = \|\mathbf{y}-\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\|^2 + \langle\mathbf{x}\rangle^H\mathbf{D}\langle\mathbf{x}\rangle + \mathrm{Tr}\big\{\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{D}\big\} + \mathrm{Tr}\big\{\boldsymbol{\Sigma}_{\mathbf{x}}\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\big\}, \tag{17}$$

where $\mathbf{D} = \mathrm{diag}\big(\mathrm{Tr}\{\boldsymbol{\Sigma}_{\mathbf{a_1}}\},\ldots,\mathrm{Tr}\{\boldsymbol{\Sigma}_{\mathbf{a_n}}\}\big)$.

Proof: Expanding $\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle$ and taking into account independence between $\mathbf{A}$ and $\mathbf{x}$, we have:

$$\begin{aligned}\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle &= \|\mathbf{y}\|^2 - 2\,\Re\big\{\mathbf{y}^H\langle\mathbf{A}\mathbf{x}\rangle\big\} + \langle\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}\rangle \\ &= \|\mathbf{y}-\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\|^2 - \mathrm{Tr}\big\{\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\langle\mathbf{x}^H\rangle\big\} \\ &\quad + \mathrm{Tr}\big\{\langle\mathbf{A}^H\mathbf{A}\rangle\langle\mathbf{x}\mathbf{x}^H\rangle\big\}.\end{aligned} \tag{18}$$

Note that $\langle\mathbf{x}\mathbf{x}^H\rangle = \langle\mathbf{x}\rangle\langle\mathbf{x}^H\rangle + \boldsymbol{\Sigma}_{\mathbf{x}}$. In addition,

$$\begin{aligned}[\langle\mathbf{A}^H\mathbf{A}\rangle]_{ij} &= \begin{cases}\langle\|\mathbf{a}_i\|^2\rangle, & \text{if } i=j\\ \langle\mathbf{a}_i^H\mathbf{a}_j\rangle, & \text{otherwise}\end{cases}\\ &= \begin{cases}\langle\mathbf{a}_i^H\rangle\langle\mathbf{a}_i\rangle + \mathrm{Tr}\{\boldsymbol{\Sigma}_{\mathbf{a}_i}\}, & \text{if } i=j\\ \langle\mathbf{a}_i^H\rangle\langle\mathbf{a}_j\rangle, & \text{otherwise.}\end{cases}\end{aligned}$$

Thus, we have $\langle\mathbf{A}^H\mathbf{A}\rangle = \langle\mathbf{A}\rangle^H\langle\mathbf{A}\rangle + \mathbf{D}$ and as a result:

$$\mathrm{Tr}\big\{\langle\mathbf{A}^H\mathbf{A}\rangle\langle\mathbf{x}\mathbf{x}^H\rangle\big\} = \mathrm{Tr}\big\{\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\langle\mathbf{x}^H\rangle\big\} + \langle\mathbf{x}\rangle^H\mathbf{D}\langle\mathbf{x}\rangle + \mathrm{Tr}\big\{\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{D}\big\} + \mathrm{Tr}\big\{\boldsymbol{\Sigma}_{\mathbf{x}}\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\big\}.$$

The proof thus follows by removing the duplicated terms in (18). ■

Corollary 1. *If $\mathbf{x}$ is deterministic, the expectation of $\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle$ with respect to a variational density $q_{\mathbf{A}}(\mathbf{A})$ is given by:*

$$\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle = \|\mathbf{y}-\langle\mathbf{A}\rangle\mathbf{x}\|^2 + \sum_{i=1}^{n}|x_i|^2\mathrm{Tr}\{\boldsymbol{\Sigma}_{\mathbf{a}_i}\}. \tag{19}$$

Proof: This is a direct result of Lemma 1 by noting that $\boldsymbol{\Sigma}_{\mathbf{x}} = \mathbf{0}$ and $\mathbf{x}^H\mathbf{D}\mathbf{x} = \sum_{i=1}^{n}|x_i|^2\mathrm{Tr}\{\boldsymbol{\Sigma}_{\mathbf{a}_i}\}$. ■

2) Updating $x_{i,t}$: This update only is applied, if $\mathbf{x}_{i,t}$ is an unknown data symbol. Taking the expectation of the conditional (13) with respect to all latent variables except for $x_{i,t}$, the variational distribution $q_i(x_{i,t})$ can be found as follows:

$$\begin{aligned}q_i(x_{i,t}) &\propto \exp\Big\{\big\langle \ln p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{H}_t,\gamma_t) + \ln p(x_{i,t})\big\rangle\Big\}\\ &\propto p(x_{i,t})\exp\big\{\big\langle -\gamma_t\big\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\big\|^2\big\rangle\big\}\\ &\propto p(x_{i,t})\exp\Bigg\{-\langle\gamma_t\rangle\Bigg\langle\Bigg\|\mathbf{y}_t - \mathbf{h}_{i,t}x_{i,t} - \sum_{j\neq i}^{K}\mathbf{h}_{j,t}x_{j,t}\Bigg\|^2\Bigg\rangle\Bigg\}\\ &\propto p(x_{i,t})\exp\Bigg\{-\langle\gamma_t\rangle\Bigg[\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|x_{i,t}|^2\\ &\qquad - 2\,\Re\Bigg\{\langle\mathbf{h}_{i,t}^H\rangle\Bigg(\mathbf{y}_t - \sum_{j\neq i}^{K}\langle\mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\Bigg)x_{i,t}^*\Bigg\}\Bigg]\Bigg\}\\ &\propto p(x_{i,t})\exp\big\{-\langle\gamma_t\rangle\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|x_{i,t}-z_{i,t}|^2\big\},\end{aligned} \tag{20}$$

where

$$z_{i,t} \triangleq \frac{\langle\mathbf{h}_{i,t}^H\rangle}{\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle}\Bigg(\mathbf{y}_t - \sum_{j\neq i}^{K}\langle\mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\Bigg), \tag{21}$$

as a linear estimate of $x_{i,t}$. Note that $\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle = \|\langle\mathbf{h}_{i,t}\rangle\|^2 + \mathrm{Tr}\{\boldsymbol{\Sigma}_{i,t}\}$ by following Corollary 1. Since prior $p(x_{i,t})$ is discrete, the variational distribution $q_i(x_{i,t})$ is also discrete and can be found easily by normalizing so that:

$$q_i(a) = \frac{p_a\exp\big\{-\langle\gamma_t\rangle\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|a-z_{i,t}|^2\big\}}{\sum_{b\in\mathcal{S}}p_b\exp\big\{-\langle\gamma_t\rangle\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|a-z_{i,t}|^2\big\}}, \forall a\in\mathcal{S}. \tag{22}$$

Hence, the variational mean of $x_{i,t}$, $\langle x_{i,t}\rangle$, and its variance, $\tau_{i,t}^x$, are given by:

$$\langle x_{i,t}\rangle = \sum_{a\in\mathcal{S}}a q_i(a);\;\; \tau_{i,t}^x = \sum_{a\in\mathcal{S}}|a|^2 q_i(a) - |\langle x_{i,t}\rangle|^2. \tag{23}$$

3) Updating γ_t: Taking the expectation of the conditional (13) with respect to all latent variables except for γ_t, the variational distribution $q(\gamma_t)$ is given by:

$$\begin{aligned}q(\gamma_t) &\propto \exp\Big\{\big\langle \ln p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{H}_t,\gamma_t) + \ln p(\gamma_t)\big\rangle\Big\}\\ &\propto \exp\Big\{M\ln\gamma_t - \gamma_t\big\langle\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\|^2\big\rangle\\ &\qquad + (a_0-1)\ln\gamma_t - b_0\gamma_t\Big\}.\end{aligned} \tag{24}$$

The variational distribution $q(\gamma_t)$ is thus Gamma with mean

$$\langle\gamma_t\rangle = \frac{a_0+M}{b_0 + \big\langle\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\|^2\big\rangle}, \tag{25}$$

where $\big\langle\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\|^2\big\rangle = \|\mathbf{y}_t - \langle\mathbf{H}_t\rangle\langle\mathbf{x}_t\rangle\|^2 + \sum_{i=1}^{K}\big[\tau_{i,t}^x\|\langle\mathbf{h}_{i,t}\rangle\|^2 + \langle|x_{i,t}|^2\rangle\mathrm{Tr}\{\boldsymbol{\Sigma}_{i,t}\}\big]$ by following Lemma 1.

It is worth noting that the Gaussian variational distribution of $\mathbf{h}_{i,t}$ and the Gamma variational distribution of γ_t justify the two assumptions in the estimation phase.

$$
\begin{aligned}
q(\mathbf{h}_{i,t}) &\propto \exp\Big\{\big\langle \ln p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{H}_t, \gamma_t) + \ln p(\mathbf{h}_{i,t}; \eta_i)\big\rangle\Big\} \\
&\propto \exp\Big\{ - \big\langle \gamma_t \big\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\big\|^2\big\rangle - \big\langle (\mathbf{h}_{i,t} - \eta_i \hat{\mathbf{h}}_{i,t-1|t-1})^H \boldsymbol{\Sigma}_{i,t|t-1}^{-1} (\mathbf{h}_{i,t} - \eta_i \hat{\mathbf{h}}_{i,t-1|t-1})\big\rangle\Big\} \\
&\propto \exp\bigg\{ - \bigg\langle \gamma_t \bigg\|\mathbf{y}_t - \mathbf{h}_{i,t}x_{i,t} - \sum_{j\neq i}^{K} \mathbf{h}_{j,t}x_{j,t}\bigg\|^2 \bigg\rangle - \bigg\langle (\mathbf{h}_{i,t} - \eta_i \hat{\mathbf{h}}_{i,t-1|t-1})^H \boldsymbol{\Sigma}_{i,t|t-1}^{-1} (\mathbf{h}_i - \eta_i \hat{\mathbf{h}}_{i,t-|t-1})\bigg\rangle\bigg\} \\
&\propto \exp\bigg\{ - \mathbf{h}_{i,t}^H \Big[\langle\gamma_t\rangle\langle|x_{i,t}|^2\rangle \mathbf{I}_M + \boldsymbol{\Sigma}_{i,t|t-1}^{-1}\Big]\mathbf{h}_{i,t} \\
&\qquad + 2\,\Re\bigg\{\langle\gamma_t\rangle \mathbf{h}_{i,t}^H \bigg(\mathbf{y}_t - \sum_{j\neq i}^{K} \langle \mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\bigg)\langle x_{i,t}^*\rangle + \eta_i \mathbf{h}_{i,t}^H \boldsymbol{\Sigma}_{i,t|t-1}^{-1} \hat{\mathbf{h}}_{i,t-1|t-1}\bigg\}\bigg\}. \qquad (16)
\end{aligned}
$$

Algorithm 1: Variational Bayesian for online JED

Input: $\mathbf{Y}$, $\hat{\mathbf{h}}_{i,0|0}$, $\boldsymbol{\Sigma}_{i,0|0}$, prior distribution $p(a), \forall a \in \mathcal{S}$, prior distribution of $p(\gamma_t)$;
Output: $\mathbf{X}$, $\mathbf{H}_t$, γ_t;
for $t = 1, 2, \ldots, T$ **do**
  Compute the prior distribution of $\mathbf{h}_{i,t}$ using (10) and (11);
  Initialize $\langle \mathbf{x}_t\rangle = \mathbf{0}$, $\langle \mathbf{h}_{i,t}\rangle = \hat{\mathbf{h}}_{i,t|t-1}$, $\boldsymbol{\Sigma}_{i,t} = \boldsymbol{\Sigma}_{i,t|t-1}$, $\langle\gamma_t\rangle = 0$;
  repeat
    for $i = 1, 2, \ldots, K$ **do**
      Compute $\boldsymbol{\Sigma}_{i,t}$ as in (14) and $\langle \mathbf{h}_{i,t}\rangle$ as in (15);
    end
    for $i = 1, 2, \ldots, K$ **do**
      Compute and normalize the distribution $q_i(x_{i,t})$ as in (20) ;
      Compute $\langle x_{i,t}\rangle$ and $\tau_{i,t}^x$ with respect to $q_i(x_{i,t})$ as in (23);
    end
    Compute $\langle\gamma_t\rangle$ as in (25);
  until *convergence*;
  Set $\hat{\mathbf{h}}_{i,t|t} = \langle \mathbf{h}_{i,t}\rangle$ and $\boldsymbol{\Sigma}_{i,t|t} = \boldsymbol{\Sigma}_{i,t}$.
  MAP estimate: $\hat{x}_{i,t} = \arg\max_{a\in\mathcal{S}} q_i(a)$.
end

Algorithm 1 explains the details of the CAVI algorithm to iteratively optimizing $q(\mathbf{h}_{i,t})$, $q_i(x_{i,t})$, and $q(\gamma_t)$ in order to estimate $\mathbf{H}_t$, $\mathbf{x}_t$, and γ_t. Here, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T]$ and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]$. At convergence, Algorithm 1 updates the approximate posterior statistics at time slot t as $\hat{\mathbf{h}}_{i,t|t} = \langle \hat{\mathbf{h}}_{i,t}\rangle$ and $\boldsymbol{\Sigma}_{i,t|t} = \boldsymbol{\Sigma}_{i,t}$, which will be used as the prior distribution in the next time slot.

IV. Numerical analysis

This section analyzes the performance of the proposed VB method, which is suitable for JED in a MIMO UAV network.

Simulation setup: It is assumed that $K = 4$ and $M = 32$. We define $I_{\text{tr}} = 50$ to denote the maximum number of iterations of the CAVI algorithm. For simplicity, we consider $\eta_i = \eta, 1 \le i \le K$, where η is a constant.

This section uses LMMSE and ML, two well-known benchmarks for data detection. Note that we use LMMSE channel estimation in the LMMSE and ML data detection methods. Here, the performance of VB is first compared with LMMSE and ML in terms of the SER and computational complexity. Next, the impact of η is studied. Finally, it is investigated how the SER of VB, LMMSE, and ML behaves as communication time T increases.

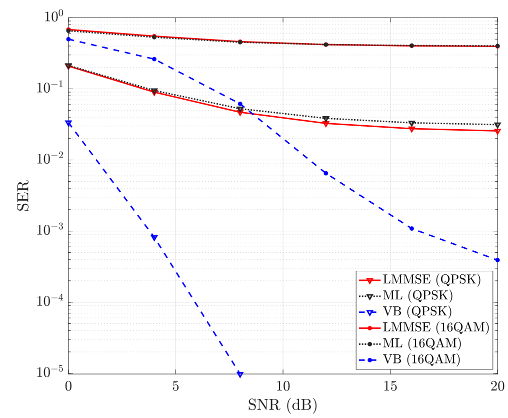

Fig. 2. An SER comparison between LMMSE, ML, and VB using QPSK and 16QAM when $\eta = 0.995$, $T = 128$, and SNR $\in [0, 20]$ dB.

VB vs. benchmarks: In this part, the performance of the proposed VB method is compared with LMMSE and ML in terms of the SER and computational complexity. First, QPSK and 16QAM are considered, and then the SER of LMMSE, ML, and VB when $\eta = 0.995$, $T = 128$, and the signal-to-noise ratio (SNR) varies between 0 and 20 dB is computed. Fig. 2 depicts that VB outperforms LMMSE and ML in terms of SER due to considering the time correlation between channels and the impact of variable precision for the noise. It also indicates that the performance of all methods degrades as the signal constellation size increases.

Second, the computational complexity of the LMMSE, ML, and VB methods is examined. Here, the complexity of LMMSE and ML is given by $\mathcal{O}\left(MK^2 + |\mathcal{S}|K\right)$ and $\mathcal{O}\left(MK|\mathcal{S}|^K\right)$, respectively, where $|\mathcal{S}|$ denotes the size of $\mathcal{S}$. On the other hand, the complexity of finding a local optimum solution based on the CAVI algorithm in the VB method is equal to $\mathcal{O}\left(MK + |\mathcal{S}|K\right)$ [14]. Therefore, the VB method not only provides a lower SER than the LMMSE and ML methods but also has a lower computational complexity.

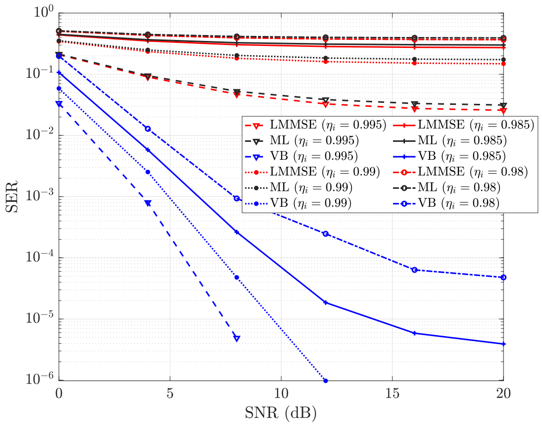

Fig. 3. SER against SNR using the LMMSE, ML, and VB methods when $K = 4, M = 32$, and $T = 128$ under QPSK modulation.

Impact of time correlation coefficient η: Here, the SER of LMMSE, ML, and VB is calculated, as shown in Fig. 3, while $\eta \in \{0.995, 0.99, 0.985, 0.98\}$, $T = 128$, SNR $\in [0, 20]$ dB, and QPSK is used. As anticipated, Fig. 3 illustrates that the SER performance of the VB method deteriorates as η decreases.

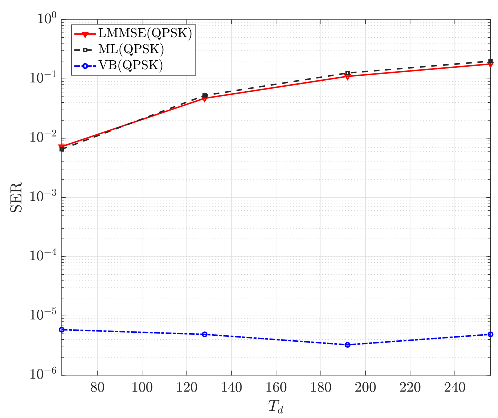

Fig. 4. SER versus SNR using the LMMSE, ML, and VB methods under QPSK assuming $\eta = 0.995$, $T \in \{64, 128, 192, 256\}$, and SNR $= 8$ dB.

Impact of communication time T: Finally, the SER of LMMSE, ML, and VB under QPSK modulation is studied when $\eta = 0.995$, $T \in \{64, 128, 192, 256\}$, and SNR $= 8$ dB. Based on Fig. 4, the performance of LMMSE and ML degrades as T increases, while VB's performance remains relatively stable with increasing T. This occurs because the channel estimation error of the LMMSE and ML approaches propagates during the estimation time. However, VB updates the variational posterior distributions of the time-varying channels and data symbols during the communication time.

V. Conclusion

In this paper, we explored an uplink massive MIMO system in a UAV network characterized by time-varying channels. Our approach involved an online processing strategy based on VB inference for JED. This approach aims to approximate true posterior distributions via variational distributions. Moreover, we assumed the noise variance is a variable to emphasize the impact of inter-user interference. Finally, our investigation delved into the performance of our VB method and compared the results with those of LMMSE and ML. The results indicated that our proposed VB method surpasses the LMMSE and ML methods in terms of SER and computational complexity. One potential area for further study could involve expanding these findings to situations where time correlation is unknown. Additionally, exploring the JED problem within a network setting, incorporating reconfigurable intelligent surfaces (RISs) alongside massive MIMO, could represent another promising avenue for future investigation, as RIS technology has garnered significant interest recently as a potential cornerstone in the next generation of wireless networks [15].

VI. Acknowledgment

This work was supported by the U.S. Department of Defense (DoD) under Grant W911NF-23-1-0226 and by the National Science Foundation (NSF) under Grant CCF-2225576.

References

[1] J. Lyu, Y. Zeng, R. Zhang, and T. J. Lim, "Placement optimization of UAV-mounted mobile base stations," *IEEE Communications Letters*, vol. 21, no. 3, pp. 604–607, 2016.

[2] M. M. Azari, F. Rosas, and S. Pollin, "Reshaping cellular networks for the sky: Major factors and feasibility," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–7.

[3] G. Geraci, A. Garcia-Rodriguez, L. G. Giordano, D. Lopez-Perez, and E. Björnson, "Supporting UAV cellular communications through massive MIMO," in *2018 IEEE international conference on communications workshops (ICC workshops)*. IEEE, 2018, pp. 1–6.

[4] J. Lyu and R. Zhang, "Network-connected UAV: 3-D system modeling and coverage performance analysis," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7048–7060, 2019.

[5] K. Ghavami and M. Naraghi-Pour, "Blind channel estimation and symbol detection for multi-cell massive MIMO systems by expectation propagation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 943–954, 2017.

[6] L. Chen and X. Yuan, "Blind multiuser detection in massive MIMO channels with clustered sparsity," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1052–1055, 2019.

[7] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 540–553, 2017.

[8] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 525–536, 2015.

[9] S. Liang, X. Wang, and L. Ping, "Semi-blind detection in hybrid massive MIMO systems via low-rank matrix completion," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5242–5254, 2019.

[10] H. Song, T. Goldstein, X. You, C. Zhang, O. Tirkkonen, and C. Studer, "Joint channel estimation and data detection in cell-free massive MU-MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 4068–4084, 2021.

[11] M. Naraghi-Pour, M. Rashid, and C. Vargas-Rosales, "Semi-blind channel estimation and data detection for multi-cell massive MIMO systems on time-varying channels," *IEEE Access*, vol. 9, pp. 161 709–161 722, 2021.

[12] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[13] M. J. Wainwright, M. I. Jordan *et al.*, "Graphical models, exponential families, and variational inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[14] D. H. Nguyen, I. Atzeni, A. Tölli, and A. L. Swindlehurst, "A variational bayesian perspective on massive MIMO detection," *arXiv preprint arXiv:2205.11649*, 2022.

[15] S. Nassirpour, A. Vahid, D.-T. Do, and D. Bharadia, "Beamforming design in reconfigurable intelligent surface-assisted IoT networks based on discrete phase shifters and imperfect CSI," *IEEE Internet of Things Journal*, 2023.