Variational Bayesian Inference for Time-Varying Massive MIMO Channels: Estimation and Detection

Sajjad Nassirpour, *Member, IEEE*, Toan-Van Nguyen, *Member, IEEE*, and Duy H. N. Nguyen, *Senior Member, IEEE*

***Abstract*—Massive multiple-input multiple-output (MIMO) stands as a key technology for advancing performance metrics such as data rate, reliability, and spectrum efficiency in the fifth generation (5G) and beyond of wireless networks. However, its efficiency depends greatly on obtaining accurate channel state information (CSI). This task becomes particularly challenging with increasing user mobility. In this paper, we focus on an uplink scenario in which a massive MIMO base station serves multiple high-mobility users. We leverage variational Bayesian (VB) inference for joint channel estimation and data detection (JED), tailored for time-varying channels. In particular, we use the VB framework to provide approximations of the true posterior distributions. To cover more real-world scenarios, we assume the time correlation coefficients associated with the channels are unknown. Our simulations demonstrate the efficacy of our proposed VB-based approach in tracking these unknown time correlation coefficients. We present two processing strategies within the VB framework: online and block processing strategies. The online strategy offers a low-complexity solution for a given time slot, requiring only the knowledge of the parameters/statistics within that time slot. In contrast, the block processing strategy focuses on the entire communication block and processes all received signals together to reduce channel estimation errors. Additionally, we introduce an interleaved structure for the online processing strategy to further enhance its performance. Finally, we conduct a comparative analysis of our VB approach against the linear minimum mean squared error (LMMSE), the Kalman Filter (KF), and the expectation propagation (EP) methods in terms of symbol error rate (SER) and channel normalized mean squared error (NMSE). Our findings reveal that our VB framework surpasses these benchmarks across the performance metrics.**

***Index Terms*—Bayesian inference, detection, estimation, massive MIMO, time-varying channels, variational Bayesian.**

I. Introduction

The International Telecommunication Union (ITU) estimates that by 2030, over 17 billion devices, such as mobile phones, unmanned aerial vehicles (UAVs), Internet of Things (IoT) devices, etc., will be supported via the fifth generation (5G) and beyond of wireless networks [2]. Consequently, there is a pressing need to enhance key performance indicators like data rates, reliability, and spectrum efficiency in 5G networks. To meet these requirements, a widely recognized solution is massive multiple-input multiple-output (MIMO) [3]–[6]. This involves deploying an array of antennas at the base station (BS), enabling the simultaneous transmission of multiple data streams to multiple users, thereby vastly increasing spectral efficiency and network capacity.

To achieve the aforementioned potential gains, accurate channel state information (CSI) is essential for both users, aiding in signal detection, and the BS, facilitating effective beamforming. This imperative underscores the need for effective algorithms for channel estimation and data detection. However, the adoption of massive MIMO results in large channel matrices, posing a significant challenge in obtaining CSI. In uplink scenarios, CSI acquisition is typically performed via known pilot signals, but this method faces scalability issues as it requires numerous orthogonal pilot signals equal to or greater than the number of users. Such constraints can lead to excessive overhead, consequently diminishing spectral efficiency as the number of users increases.

A. State-of-the-art

To tackle the scalability challenge of channel estimation, previous studies have proposed blind channel estimation algorithms [7], [8], which rely solely on received signals. Nevertheless, these approaches are vulnerable to phase ambiguities present in the demodulated symbols. Semi-blind channel estimation offers an alternative method to mitigate the reliance on orthogonal pilot signals. This approach integrates a limited number of known pilot signals with received signals to refine the accuracy of the estimation process. In [9], the research presented Cramer-Rao bounds (CRBs) for channel estimation methods using semi-blind, blind, and training sequence approaches in a single-input multiple-output (SIMO) network. Next, in [10], the authors investigated two semi-blind channel estimation techniques based on maximum likelihood (ML) estimation under deterministic and Gaussian models. Thereafter, in [11], the emphasis was on semi-blind channel estimation for multi-user MIMO systems and the authors proposed two methods based on the expectation maximization (EM) framework to estimate the channels. Subsequently, in [12], the focus was placed on channel estimation within an uplink multi-cell massive MIMO network. The authors presented a method that first detects the data from the target cell and then computes the least square channel estimate by treating the detected data as pilot signals.

While individual methods for channel estimation and data detection can yield promising results, joint channel estimation and data detection (JED) is often preferred in practical settings. The JED approach achieves lower channel estimation error and

S. Nassirpour, T. V. Nguyen, and D. H. N. Nguyen are with the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA 92182, USA. Emails: snassirpour@sdsu.edu, tnguyen58@sdsu.edu, and duy.nguyen@sdsu.edu.

This work was supported in part by the U.S. Army Research Office under grant W911NF2310226, and in part by the U.S. National Science Foundation under grants ECCS-2146436, CCF-2225576, and CCF-2322190. Part of this work was presented at the IEEE International Conference on Communications (ICC) Workshops, Denver, CO, USA, June 2024 [1].

improved spectral efficiency compared to individual channel estimation methods. Traditional approaches rely solely on pilot signals, where longer pilot signals are desirable to reduce estimation error. However, the long sequence of known pilot signals degrades spectral efficiency. The JED approach overcomes this by generating long semi-pilot signals through the combination of short pilot signals with detected unknown data symbols. The JED approach leverages the interdependency between channels and data to continuously refine the performance of the channel estimation and data detection processes, improving overall system performance while maintaining high spectral efficiency. Thus, the research conducted in [13] delved into the challenge of JED in hybrid massive MIMO systems. It introduced two iterative algorithms employing a low-rank matrix completion method. Next, the study in [14] focused on a massive MIMO system with generalized spatial modulation (GSM). It assumed the angular channel model with double sparsity in both the channels and signals, and developed two methods, blind and semi-blind, based on message passing to jointly perform user activity estimation and JED. Then, the authors in [15] studied the JED problem in an orthogonal frequency division multiplexing (OFDM) system. They considered the extended vehicular A (EVA) channel model and proposed a method based on the EM technique. This approach identifies reliable data tones from both desired and interfering users, which are then used as pilots for channel re-estimation.

The above algorithms assume that channels remain unchanged over the communication time, which holds true only if users are stationary or have low mobility. However, 5G networks are anticipated to support high-mobility users, e.g., UAVs and high-speed trains, causing the communication channels to fluctuate over time due to Doppler spread. Therefore, efficient channel estimation protocols are essential for promptly updating channel coefficients throughout the communication time. In [16]–[18], the authors focused on time-varying channels as a model that reflects channel aging in massive MIMO networks. They introduced various channel estimation techniques based on linear minimum mean squared error (LMMSE). Then, the work in [19] introduced an EM-based sparse Bayesian learning (SBL) method for time-varying channels, tailored for both uplink and downlink communications. It reformulated time-varying channels as a sparse signal model using the virtual channel representation and then followed the SBL method to estimate the time-varying channels. Subsequently, [20], [21] presented interpolation-based strategies for channel estimation. The effectiveness of these methods heavily depends on the number of missing channel coefficients between pilot signals as well as the length of the pilot signals. Hence, in [22], the authors proposed a first-order Taylor expansion channel prediction (FIT-CP) method, which utilizes the first-order Taylor expansion to predict the missing channel coefficients in the interpolation-based framework. The authors demonstrated in [22] that their approach outperforms interpolation methods as well as traditional channel estimation approaches. Additionally, [23], [24] used Kalman Filter (KF) for channel estimation in time-varying channels. Later, [25], [26] proposed deep learning-based approaches for training neural networks tailored to the dynamics of time-varying channels. Recently, [27] proposed a reinforcement learning (RL)-based data-aided channel estimator for time-varying channels by formulating the channel estimation problem as a Markov decision process (MDP) and developing an RL algorithm to optimize its policy.

In the above algorithms, channel estimation relies solely on pilot signals, while JED techniques enhance performance by leveraging both pilots and data symbols. Therefore, [28] focused on time-varying channels in the uplink of multi-cell massive MIMO systems and proposed a semi-blind JED technique, based on the expectation propagation (EP) algorithm. The EP algorithm is an iterative approach that uses the factorization structure of the true distributions in order to approximate the joint a-posteriori distribution of the unknown channel matrix and data symbols [29], [30]. The authors in [28] assumed that the time correlation coefficients associated with the channels are known and fixed. However, in practice, these coefficients could be unknown and may even vary slowly during the communication time. As a result, in this paper, we study an uplink scenario in massive MIMO systems under time-varying channels with unknown time correlation coefficients. We introduce a technique based on variational Bayesian (VB) inference, enabling JED for time-varying channels. The VB method serves as a statistical inference framework, tackling the challenge of approximating the posterior distribution of latent variables by optimizing simpler distributions from a known family to replace the intractable true posterior distributions. VB inference, originating from machine learning, has also gained attention in wireless communications [31]–[36]. The research in [33] applied the VB method for performing joint carrier frequency offset (CFO) estimation and JED in an underwater acoustic OFDM system. Then, in [34], the authors introduced a quantized version of VB to tackle the channel estimation problem in millimeter-wave (mmWave) MIMO systems, where each antenna element is equipped with a low-resolution analog-to-digital converter (ADC). Furthermore, in [35], the authors studied two-way relaying time-invariant mmWave channels and proposed a VB-based approach for channel estimation. Additionally, the study in [36] investigated massive MIMO systems with orthogonal time-frequency space (OTFS) modulation and adopted an EM-based VB framework to estimate uplink sparse channel parameters.

In this manuscript, we leverage the VB framework to address the JED problem in time-varying channels and show via simulations that our VB-based method outperforms state-of-the-art benchmarks.

B. Contribution

We provide more details about our contributions below:

- We devise a JED technique using VB inference for uplink scenarios in massive MIMO systems, suited for high-mobility users experiencing time-varying channels.
- We consider that the BS does not possess knowledge of the noise variance, emphasizing the impact of residual inter-user interference in our calculations. Additionally, to better reflect real-world conditions, we assume the time

correlation coefficients corresponding to the communication channels are unknown and let the VB framework estimate them. Our simulations demonstrate that the proposed VB method yields nearly identical performance in terms of symbol error rate (SER) compared to a scenario where the time correlation coefficients are known.
- As the computational complexity could lead to a delay in the channel estimation process, and thus the risk of outdated CSI, we design an online VB processing strategy using only the parameters/statistics of each time slot. We also develop an interleaved structure to further improve its performance.
- Since CSI accuracy is essential in massive MIMO systems, we offer a block processing strategy based on VB inference, which analyzes all received signals together to enhance the channel estimation performance.
- We compare the performance of our proposed VB method with the LMMSE, KF, and EP methods in terms of SER and channel normalized mean squared error (NMSE). We show that our VB method outperforms the benchmarks in all performance metrics.

The rest of this paper is organized as follows. In Section II, we provide the system model. Then, in Sections III and IV, we explain the details of our proposed online and block processing strategies. Section V includes the simulation results of this work, and Section VI concludes this paper.

C. Notation

In this paper, we use italic, bold lowercase, and bold uppercase font styles to signify scalars, vectors, and matrices, respectively. We utilize the notation $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote a complex Gaussian random vector with a mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$. $\Gamma(a, b)$ implies a Gamma distribution parametrized by a and b. The space of $x \times y$ dimensional complex-valued matrices is denoted by $\mathbb{C}^{x\times y}$. We use $\sim$ and $\propto$ to convey the notations of "distributed according to" and "proportional to," respectively. $\text{diag}\{\mathbf{a}\}$ is a diagonal matrix based on $\mathbf{a}$. $|\mathbf{A}|$, $\|\mathbf{A}\|$, $\|\mathbf{A}\|_{\text{F}}$, $\mathbf{A}^{\top}$, and $\mathbf{A}^{H}$ represent the determinant, Euclidean norm, Frobenius norm, transpose, and conjugate transpose of $\mathbf{A}$, respectively. The notation $[\mathbf{X}]_{ij}$ denotes the element located at the i^{th} row and j^{th} column of $\mathbf{X}$. $\text{Tr}\,(\cdot)$ is the trace function, $\exp\{\cdot\}$ denotes the exponential function, $\log_{10}(\cdot)$ is the logarithmic function with base 10, and $\mathbf{I}_M$ stands for the identity matrix of size M. We use $x^{\star}$ and $\Re\{x\}$ to indicate the complex conjugate and the real part of x, respectively. The probability density function (PDF) of a length-K complex-valued random vector $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by: $\mathcal{CN}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\pi^K|\boldsymbol{\Sigma}|}\exp\big(-(\mathbf{x}-\boldsymbol{\mu})^H\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\big)$. $\mathbb{E}_{p(x)}[x]$ and $\text{Var}_{p(x)}[x]$ represent the mean and variance of x according to its distribution $p(x)$. Similarly, $\langle x\rangle$, $\langle |x|^2\rangle$, and τ^x represent the mean, second moment, and variance of x based on a variational distribution $q(x)$. The subscript $t|t-1$ denotes predicted statistics at time t utilizing data from time $t-1$, whereas $t|t$ denotes the posterior estimated statistics at time t derived from observations made at time t.

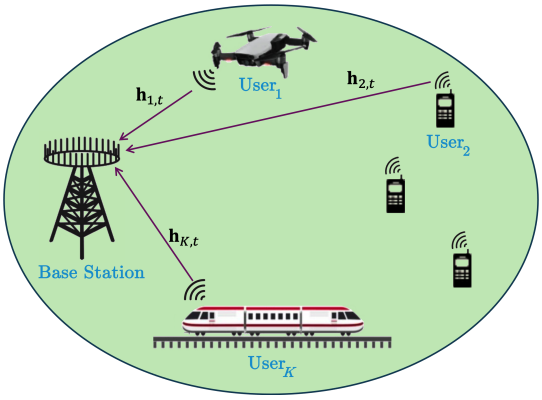

Fig. 1. The uplink scenario in a massive MIMO network supporting K single-antenna users, which could be high-mobility users.

II. System Model

In this work, we focus on an uplink scenario in a massive MIMO network, as illustrated in Fig. 1, where a BS with multiple antennas aims to communicate with multiple users simultaneously. Specifically, we address a JED problem within this network, which supports high-mobility users such as UAVs and high-speed trains that experience time-varying channels. To tackle the JED problem, we propose a method based on VB inference to approximate the true posterior distributions. We begin this section by explaining the channel model, followed by a quick introduction to VB inference, and then proceed to formulate the problem.

A. Channel Model

We use $\mathbf{h}_{i,t} = [h_{i,t}^{[1]}, h_{i,t}^{[2]}, \ldots, h_{i,t}^{[M]}]^{\top} \in \mathbb{C}^{M\times 1}$ to denote the communication channel between the i^{th} user and the BS at time t, where M is the number of antenna elements at the BS, $i \in \{1, 2, \ldots, K\}$, and K is the number of users. We assume each user has a single antenna. In this work, we utilize the first-order Gauss-Markov model to characterize $\mathbf{h}_{i,t}$ as follows [37]:

$$\mathbf{h}_{i,t} = \eta_i \mathbf{h}_{i,t-1} + \sqrt{1-\eta_i^2}\mathbf{R}_i^{\frac{1}{2}}\mathbf{g}_{i,t}, \quad t = 1, \ldots, T, \quad (1)$$

where $\mathbf{h}_{i,0} = \mathbf{R}_i^{\frac{1}{2}}\mathbf{g}_{i,0}$, T is the communication time, and $\mathbf{g}_{i,t}$ models small-scale fading, which is independent of $\mathbf{h}_{i,t-1}$ and assumed to be Rayleigh fading (i.e., $\mathbf{g}_{i,t} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$). Further, η_i denotes the time correlation coefficient of the channels corresponding to the i^{th} user, which is given by [37]:

$$\eta_i = J_0(2\pi f_i^d T_s), \quad (2)$$

in which $J_0(\cdot)$ denotes the zero-order Bessel function of the first kind, f_i^d is the maximum Doppler frequency corresponding to the channels of the i^{th} user, and T_s is the sampling period. Moreover, in (1), $\mathbf{R}_i$ represents the covariance of $\mathbf{h}_{i,t}$ (i.e., $\mathbb{E}\left[\mathbf{h}_{i,t}\mathbf{h}_{i,t}^H\right] = \mathbf{R}_i$), which is equal to:

$$\mathbf{R}_i = \beta_i^{-1}\mathbf{R}, \quad (3)$$

where β_i is the large-scale fading for the propagation channel from the i^{th} user, and $\mathbf{R}$ denotes the spatial correlation at the BS. We assume that $\mathbf{R}_i$ is known at the BS and use it to

compute the autocovariance between $\mathbf{h}_{i,t-1}$ and $\mathbf{h}_{i,t}$, which follows $\mathbb{E}\left[\mathbf{h}_{i,t-1}\mathbf{h}_{i,t}^H\right] = \eta_i \mathbf{R}_i$. Then, we have:

$$p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}, \eta_i; \mathbf{R}_i) = \mathcal{CN}\left(\mathbf{h}_{i,t}; \eta_i \mathbf{h}_{i,t-1}, (1-\eta_i^2)\mathbf{R}_i\right). \tag{4}$$

We also assume a channel independence among users, implying that $\mathbb{E}[\mathbf{h}_i \mathbf{h}_j^H] = \mathbf{0}$, $1 \leq i, j \leq K$ and $i \neq j$.

Received signal: In our massive MIMO network, the BS captures $y_{m,t}, 1 \leq m \leq M$, as the received signal through its m^{th} antenna element at time t, which can be modeled as:

$$y_{m,t} = \sum_{i=1}^{K} h_{i,t}^{[m]} x_{i,t} + n_{m,t}, \tag{5}$$

where $h_{i,t}^{[m]}$ is the m^{th} element of $\mathbf{h}_{i,t}$, $x_{i,t}$ is the transmitted signal from the i^{th} user, and $n_{m,t} \sim \mathcal{CN}(0, N_0)$ denotes the independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) at the m^{th} antenna element during time t. Subsequently, we arrange the received signals, transmitted signals, communication channels, and noise components into a vector format to establish a linear model for the received signals as below:

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{n}_t, \tag{6}$$

where $\mathbf{y}_t = [y_{1,t}, y_{2,t}, \dots y_{M,t}]^\top \in \mathbb{C}^{M\times 1}$, $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \dots, x_{K,t}]^\top \in \mathbb{C}^{K\times 1}$, $\mathbf{H}_t = [\mathbf{h}_{1,t}, \mathbf{h}_{2,t}, \dots, \mathbf{h}_{K,t}] \in \mathbb{C}^{M\times K}$, and $\mathbf{n}_t = [n_{1,t}, n_{2,t}, \dots n_{M,t}]^\top \in \mathbb{C}^{M\times 1}$.

In this paper, we concentrate on time-varying channels with the aim of jointly estimating $\mathbf{H}_t$ and detecting $\mathbf{x}_t$ using the received signal $\mathbf{y}_t$. To accomplish this, we propose a method based on VB inference. Before delving into the details, we provide a brief overview of VB.

B. An Overview of VB inference

Variational Bayesian (VB) inference stands as a robust statistical framework derived from machine learning, devised to address the challenge of approximating the posterior distribution of latent variables. Here, we denote the set of all observed variables as $\mathbf{y}$ and the set of V latent variables as $\mathbf{x}$. To detect $\mathbf{x}$, it is crucial to determine the posterior $p(\mathbf{x}|\mathbf{y})$, which often proves computationally intractable. To tackle this challenge, the VB method endeavors to identify a distribution $q(\mathbf{x})$ characterized by variational parameters within a predefined family $\mathcal{Q}$ of densities, such that $q(\mathbf{x})$ closely approximates $p(\mathbf{x}|\mathbf{y})$. To this end, VB defines an optimization problem leveraging the Kullback-Leibler (KL) divergence from $q(\mathbf{x})$ to $p(\mathbf{x}|\mathbf{y})$:

$$q^\star(\mathbf{x}) = \arg\min_{q(\mathbf{x})\in\mathcal{Q}} \mathrm{KL}\big(q(\mathbf{x})\|p(\mathbf{x}|\mathbf{y})\big), \tag{7}$$

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

$$\mathrm{KL}\big(q(\mathbf{x})\|p(\mathbf{x}|\mathbf{y})\big) = \mathbb{E}_{q(\mathbf{x})}\big[\ln q(\mathbf{x})\big] - \mathbb{E}_{q(\mathbf{x})}\big[\ln p(\mathbf{x}|\mathbf{y})\big]. \tag{8}$$

The minimization of (7) occurs when $q(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$. However, because obtaining the true posterior distribution is often infeasible, it is more practical to employ a restricted family of distributions $q(\mathbf{x})$. Thus, the literature [38] commonly focuses on the mean-field variational family, defined as:

$$q(\mathbf{x}) = \prod_{i=1}^{V} q_i(x_i), \tag{9}$$

where the latent variables are assumed to be mutually independent, each governed by a distinct factor in the variational density. In this case, the optimal value of $q_i(x_i)$ is equal to [31, Chapter 10]:

$$q_i^\star(x_i) \propto \exp\left\{\left\langle \ln p(\mathbf{y}, \mathbf{x})\right\rangle\right\}, \tag{10}$$

where $\langle\cdot\rangle$ is the expectation across all latent variables excluding x_i, utilizing the currently fixed variational density $q_{-i}(\mathbf{x}_{-i}) = \prod_{j=1, j\neq i}^{V} q_j(x_j)$. Then, to optimize (7), we use the Coordinate Ascent Variational Inference (CAVI) algorithm, which is an iterative method that ensures convergence to at least a locally optimal solution [39]. In particular, the CAVI algorithm updates $q_i^\star(x_i)$ sequentially for all latent variables to monotonically enhance the objective function in (7).

C. Problem Formulation

According to (10), in order to apply the VB framework, it is necessary to compute the joint distribution. Unlike previous studies [31], [40], we assume that the time correlation η_i and the noise variance N_0 are i.i.d. random variables and unknown a-priori at the BS. We denote the precision of the noise at time t as $\gamma_t = 1/N_0$. Utilizing these assumptions, we derive the joint distribution $p(\mathbf{y}_t, \mathbf{H}_t, \mathbf{H}_{t-1}, \mathbf{x}_t, \gamma_t, \boldsymbol{\eta}; \bar{\mathbf{R}})$ as follows:

$$\begin{aligned} p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{H}_t, \mathbf{H}_{t-1}, \gamma_t, \boldsymbol{\eta}; \bar{\mathbf{R}}) &= p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{H}_t, \gamma_t) p(\mathbf{x}_t) \\ &\times \underbrace{p(\mathbf{H}_t|\mathbf{H}_{t-1}) p(\mathbf{H}_{t-1}|\boldsymbol{\eta}; \bar{\mathbf{R}})}_{p(\mathbf{H}_t|\boldsymbol{\eta}; \bar{\mathbf{R}})} p(\gamma_t) p(\boldsymbol{\eta}), \end{aligned} \tag{11}$$

where $p(\mathbf{H}_t|\boldsymbol{\eta}; \bar{\mathbf{R}}) = \prod_{i=1}^{K} p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}) p(\mathbf{h}_{i,t-1}|\eta_i; \mathbf{R}_i)$, $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_K]$, $\bar{\mathbf{R}} = [\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_K]$, $p(\mathbf{x}_t) = \prod_{i=1}^{K} p(x_{i,t})$, and $p(\boldsymbol{\eta}) = \prod_{i=1}^{K} p(\eta_i)$.

In this paper, we utilize the VB framework to approximate the posterior distribution $p(\mathbf{x}_t, \mathbf{H}_t, \boldsymbol{\eta}, \gamma_t|\mathbf{y}_t)$. There are two possible strategies for processing time-varying channels: one that utilizes information available at time t, which we refer to as the online processing strategy, and another that processes all observed signals $\mathbf{y}_t$'s together, which we call the block processing strategy. Each of these strategies has its own advantages and drawbacks. We explain the details of these strategies in the next two sections.

III. Online Processing Strategy

In this section, we explain the details of our proposed VB-based online processing strategy designed to address the JED problem in time-varying channels. The primary objective of the online processing strategy is to achieve a low-complexity low-latency solution. To do so, we propose a method that only requires the parameters and statistics corresponding to the given time t. Since estimating the channel and detecting data based on the limited statistics available at time t is challenging, we introduce a two-phase online processing strategy consisting of a prediction phase and an estimation phase.

The prediction phase aims to provide a dynamic prior distribution for $p(\mathbf{h}_{i,t}|\eta_i, \mathbf{R}_i)$ instead of assuming a fixed prior distribution. To achieve this, we continuously update the prior distribution for $(\mathbf{h}_{i,t}|\eta_i, \mathbf{R}_i)$ using the information available at time $t-1$. Following the prediction phase, the estimation phase is employed to complete the online processing strategy by estimating the corresponding channel $\mathbf{H}_t$ and detecting the data symbol $\mathbf{x}_t$ at a given time t. The details of these two phases are presented below."

Phase I - Prediction: In (11), we need to find the probability $p(\mathbf{h}_{i,t}|\eta_i; \mathbf{R}_i) = p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1})(\mathbf{h}_{i,t-1}|\eta_i; \mathbf{R}_i)$ in order to compute the desired joint distribution. In this phase, our goal is to find a predictive distribution for $p(\mathbf{h}_{i,t}|\eta_i; \mathbf{R}_i)$ at time t, using the statistics available at time $t-1$, which will later be used in Phase II. To this end, we calculate the posterior distribution of $p(\mathbf{h}_{i,t-1}|\eta_i; \mathbf{R}_i)$ at time slot $t-1$ using the Gauss-Markov model in (1), and then consider it as the prior distribution for $p(\mathbf{h}_{i,t}|\eta_i; \mathbf{R}_i)$ at time slot t. Moreover, according to the VB framework, the prior distributions of the channels are necessary to obtain their posterior distributions. In this case, we assume the prior distribution $p(\mathbf{h}_{i,t-1}) = \mathcal{CN}\big(\mathbf{h}_{i,t-1}; \hat{\mathbf{h}}_{i,t-1|t-1}, \hat{\mathbf{\Sigma}}_{i,t-1|t-1}\big)$. If η_i is known, we use the prior distribution $p(\mathbf{h}_{i,t-1})$ and the Gauss-Markov model in (1) to derive the predictive distribution $p(\mathbf{h}_{i,t}|\eta_i, \mathbf{R}_i)$ as below:

$$p(\mathbf{h}_{i,t}|\eta_i, \mathbf{R}_i) = \mathcal{CN}(\mathbf{h}_{i,t}; \hat{\mathbf{h}}_{i,t|t-1}, \hat{\mathbf{\Sigma}}_{i,t|t-1}), \tag{12}$$

where

$$\hat{\mathbf{h}}_{i,t|t-1} = \eta_i \hat{\mathbf{h}}_{i,t-1|t-1}. \tag{13}$$

$$\hat{\mathbf{\Sigma}}_{i,t|t-1} = \eta_i^2 \hat{\mathbf{\Sigma}}_{i,t-1|t-1} + (1-\eta_i^2)\mathbf{R}_i. \tag{14}$$

Remark 1. *In this work, we assume η_i is an unknown random variable, and the prior distribution $p(\eta_i)$ is a conjugate probability. To achieve this, we assume $p(\eta_i) = \mathcal{N}\big(\eta_i; \hat{\eta}_{i,t-1}, \tau^{\eta}_{i,t-1}\big)$. However, this assumption may result in η_i being less than 0 or exceeding 1, even though it must always lie within the range $0 \leq \eta_i \leq 1$. On the other hand, after a few iterations of the CAVI algorithm, we expect the variance $\tau^{\eta}_{i,t-1}$ to decrease progressively, causing the algorithm to concentrate around the mean $\hat{\eta}_{i,t-1}$. Therefore, it is crucial to select the initial value of $\hat{\eta}_{i,t-1}$ carefully, ensuring it is close to the true value. As a result, we reset the initial value of $\hat{\eta}_{i,t-1}$ if it is less than 0 or exceeds 1, which ensures that η_i always remains within its acceptable range.*

In statistical models, when it is challenging to obtain an accurate closed-form distribution of a random variable, that variable is typically approximated by its mean value. This is based on the expectation that its variance will become progressively smaller throughout the statistical process. Therefore, in this paper, we update (14) by approximating the predictive covariance $\hat{\mathbf{\Sigma}}_{i,t|t-1}$ using the second moment of η_i at time slot $t-1$. In particular, we replace η_i^2 with $\mathbb{E}(\eta_i^2) = \hat{\eta}^2_{i,t-1} + \tau^{\eta}_{i,t-1}$. Here, we have:

$$\begin{aligned}\hat{\mathbf{\Sigma}}_{i,t|t-1} \approx (\hat{\eta}^2_{i,t-1} + \tau^{\eta}_{i,t-1})\hat{\mathbf{\Sigma}}_{i,t-1|t-1} \\ + (1 - \hat{\eta}^2_{i,t-1} - \tau^{\eta}_{i,t-1})\mathbf{R}_i.\end{aligned} \tag{15}$$

TABLE I
THE PRIOR DISTRIBUTION ASSUMPTIONS FOR THE UNKNOWN RANDOM VARIABLES IN THE ONLINE PROCESSING VB STRATEGY

Probability	Prior Distribution Assumption
$p(\mathbf{h}_{i,t-1})$	$\mathcal{CN}\big(\mathbf{h}_{i,t-1}; \hat{\mathbf{h}}_{i,t-1|t-1}, \hat{\mathbf{\Sigma}}_{i,t-1|t-1}\big)$
$p(\eta_i)$	$\mathcal{N}\big(\eta_i; \hat{\eta}_{i,t-1}, \tau^{\eta}_{i,t-1}\big)$
$p(\gamma_t)$	$\Gamma(a_0, b_0)$
$p(x_{i,t})$(pilot transmission)	$\delta(x_{i,t} - \bar{x}_{i,t})$
$p(x_{i,t})$(data transmission)	$\sum_{a\in\mathcal{S}} p_a \delta(x_{i,t} - a)$

Phase II - Estimation: In this phase, our objective is to derive Bayesian optimal estimates for both the data symbol $\mathbf{x}_t$ and the channel matrix $\mathbf{H}_t$. To accomplish this, we require the posterior distribution $p(\mathbf{x}_t, \mathbf{H}_t, \boldsymbol{\eta}, \gamma_t|\mathbf{y}_t)$, which could be challenging to obtain. Consequently, we employ the mean-field variational distribution $q(\mathbf{x}_t, \mathbf{H}_t, \boldsymbol{\eta}, \gamma_t)$ proposed within the VB framework in the following manner:

$$\begin{aligned}p\big(\mathbf{x}_t, \mathbf{H}_t, \boldsymbol{\eta}, \gamma_t|\mathbf{y}_t\big) &\approx q(\mathbf{x}_t, \mathbf{H}_t, \boldsymbol{\eta}, \gamma_t) \\ &= \left[\prod_{i=1}^{K} q_i(x_{i,t}) q(\mathbf{h}_{i,t}) q(\eta_i)\right] q(\gamma_t).\end{aligned} \tag{16}$$

According to (10), in order to get the optimal solution of the variational densities in (16), we need the joint distribution $p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{H}_t, \boldsymbol{\eta}, \gamma_t)$, which is equal to:

$$\begin{aligned}p(\mathbf{y}_t, \mathbf{x}_t, \mathbf{H}_t, \gamma_t) &= p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{H}_t, \boldsymbol{\eta}, \gamma_t) \\ &\times \left[\prod_{i=1}^{K} p(x_{i,t}) p(\mathbf{h}_{i,t}|\eta_i; \mathbf{R}_i) p(\eta_i)\right] p(\gamma_t).\end{aligned} \tag{17}$$

In (17), besides considering $p(\mathbf{h}_{i,t}|\eta_i; \mathbf{R}_i)$ and $p(\eta_i)$, we also need to specify the prior distributions for γ_t and $x_{i,t}$.

Remark 2. *In MIMO systems, γ_t is typically assumed to be a known and fixed parameter [23], [28]. However, recent studies [32], [33] have considered γ_t as an unknown parameter and applied statistical methods to estimate it. Specifically, [33] utilized the EM method to approximate γ_t, while [32] assumed it follows a Gamma distribution and leveraged the VB method to obtain γ_t. In this paper, motivated by these works, we assume $\gamma_t \sim \Gamma(a_0, b_0)$ to capture both the impact of noise and the estimation error in our process, and we then use the CAVI algorithm to find its local optimum solution.*

Assumptions about the prior distribution of $x_{i,t}$ vary depending on whether we are in the pilot transmission phase or data transmission phase. In this study, we divide the total communication time T into two parts: the pilot transmission phase with duration T_p followed by the data transmission phase with duration T_d, where $T_p + T_d = T$. If $1 \leq t \leq T_p$, the prior distribution of $x_{i,t}$ is simply $p(x_{i,t}) = \delta(x_{i,t} - \bar{x}_{i,t})$, where $\bar{x}_{i,t}$ represents the pilot data symbol from the i^{th} user at time t and $\delta(\cdot)$ is the Dirac delta function. For $T_p < t \leq T$, the prior distribution of $x_{i,t}$ is expressed as $p(x_{i,t}) = \sum_{a\in\mathcal{S}} p_a \delta(x_{i,t} - a)$, where p_a denotes the probability of the constellation point $a \in \mathcal{S}$ and $\mathcal{S}$ represents the signal constellation.

In Table I, we summarize the prior distribution assumptions for the desired unknown random variables $p(\mathbf{h}_{t-1})$, $p(\eta_i)$, $p(\gamma_t)$, and $p(x_{i,t})$.

As previously discussed in Section II, the CAVI algorithm iteratively converges to local optimum solutions $q^\star(\mathbf{h}_{i,t})$, $q^\star(\eta_i)$, $q^\star(x_{i,t})$, and $q^\star(\gamma_t)$ by optimizing one latent variable at a time while keeping the others fixed. In the subsequent parts, we illustrate the update process for each latent variable.

1) Updating $\mathbf{h}_{i,t}$: By computing the expectation of (17) for all latent variables except for $\mathbf{h}_{i,t}$, we express the variational distribution $q(\mathbf{h}_{i,t})$ as (18) (at the top of the next page), which shows that $q(\mathbf{h}_{i,t})$ is Gaussian with the subsequent covariance matrix and mean:

$$\mathbf{\Sigma}_{i,t} = \Big[\langle\gamma_t\rangle\langle|x_{i,t}|^2\rangle\mathbf{I}_M + \hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\Big]^{-1}, \tag{19}$$

$$\langle\mathbf{h}_{i,t}\rangle = \mathbf{\Sigma}_{i,t}\Bigg[\langle\gamma_t\rangle\bigg(\mathbf{y}_t - \sum_{j\neq i}^{K}\langle\mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\bigg)\langle x^*_{i,t}\rangle + \langle\eta_i\rangle\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\hat{\mathbf{h}}_{i,t-1|t-1}\Bigg]. \tag{20}$$

2) Updating η_i: If we take the expectation of (17) with respect to all latent variables except for η_i, we obtain the variational distribution $q(\eta_i)$ as (21) (on the next page), which is Gaussian. In this case, the variance and mean of $q(\eta_i)$ are given by:

$$\tau^{\eta}_{i,t} = \Big(\hat{\mathbf{h}}^H_{i,t-1|t-1}\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\hat{\mathbf{h}}_{i,t-1|t-1} + 1/\tau^{\eta}_{i,t-1}\Big)^{-1}, \tag{22}$$

$$\langle\eta_i\rangle = \tau^{\eta}_{i,t}\left(\Re\Big\{\hat{\mathbf{h}}^H_{i,t-1|t-1}\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\langle\mathbf{h}_{i,t}\rangle\Big\} + \frac{\hat{\eta}_{i,t-1}}{\tau^{\eta}_{i,t-1}}\right). \tag{23}$$

We utilize the following lemma to compute the variational posterior mean of several random variables. We employ this lemma later in this section to update $x_{i,t}$ and γ_t.

Lemma 1. *[32] Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ and $\mathbf{x} \in \mathbb{C}^{n\times 1}$ be two independent random matrices (vectors) with respect to a variational density $q_{\mathbf{A},\mathbf{x}}(\mathbf{A},\mathbf{x}) = q_{\mathbf{A}}(\mathbf{A})q_{\mathbf{x}}(\mathbf{x})$. Suppose $\mathbf{A}$ is column-wise independent and $\langle\mathbf{a}_i\rangle$ and $\hat{\mathbf{\Sigma}}_{\mathbf{a}_i}$ are the variational mean and covariance matrix of the i^{th} column of $\mathbf{A}$. Let $\langle\mathbf{x}\rangle$ and $\hat{\mathbf{\Sigma}}_{\mathbf{x}}$ be the variational mean and covariance matrix of $\mathbf{x}$. Consider $\mathbf{y} \in \mathbb{C}^{m\times 1}$ is an arbitrary vector. Here, $\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle$, the expectation of $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2$ with respect to $q_{\mathbf{A},\mathbf{x}}(\mathbf{A},\mathbf{x})$, is given by:*

$$\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle = \|\mathbf{y}-\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\|^2 + \langle\mathbf{x}\rangle^H\mathbf{D}\langle\mathbf{x}\rangle + \mathrm{Tr}\big\{\hat{\mathbf{\Sigma}}_{\mathbf{x}}\mathbf{D}\big\} + \mathrm{Tr}\big\{\hat{\mathbf{\Sigma}}_{\mathbf{x}}\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\big\}, \tag{24}$$

where $\mathbf{D} = \mathrm{diag}\big(\mathrm{Tr}\{\mathbf{\Sigma}_{\mathbf{a}_1}\},\ldots,\mathrm{Tr}\{\mathbf{\Sigma}_{\mathbf{a}_n}\}\big)$.

Proof: We first expand $\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle$ as follows:

$$\begin{aligned}\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle &= \|\mathbf{y}\|^2 - 2\,\Re\big\{\mathbf{y}^H\langle\mathbf{A}\mathbf{x}\rangle\big\} + \langle\mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x}\rangle\\ &= \|\mathbf{y}-\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\|^2 - \mathrm{Tr}\big\{\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\langle\mathbf{x}^H\rangle\big\}\\ &\quad + \mathrm{Tr}\big\{\langle\mathbf{A}^H\mathbf{A}\rangle\langle\mathbf{x}\mathbf{x}^H\rangle\big\}.\end{aligned} \tag{25}$$

We know that $\langle\mathbf{x}\mathbf{x}^H\rangle = \langle\mathbf{x}\rangle\langle\mathbf{x}^H\rangle + \hat{\mathbf{\Sigma}}_{\mathbf{x}}$. Moreover, we have:

$$[\langle\mathbf{A}^H\mathbf{A}\rangle]_{ij} = \begin{cases}\langle\|\mathbf{a}_i\|^2\rangle, & i=j\\ \langle\mathbf{a}_i^H\mathbf{a}_j\rangle, & i\neq j\end{cases} = \begin{cases}\langle\mathbf{a}_i^H\rangle\langle\mathbf{a}_i\rangle + \mathrm{Tr}\{\mathbf{\Sigma}_{\mathbf{a}_i}\}, & i=j\\ \langle\mathbf{a}_i^H\rangle\langle\mathbf{a}_j\rangle, & i\neq j,\end{cases} \tag{26}$$

which results in:

$$\langle\mathbf{A}^H\mathbf{A}\rangle = \langle\mathbf{A}\rangle^H\langle\mathbf{A}\rangle + \mathbf{D}. \tag{27}$$

Then, we use (26) and (27) to obtain the following expression:

$$\begin{aligned}\mathrm{Tr}\big\{\langle\mathbf{A}^H\mathbf{A}\rangle\langle\mathbf{x}\mathbf{x}^H\rangle\big\} &= \mathrm{Tr}\big\{\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\langle\mathbf{x}\rangle\langle\mathbf{x}^H\rangle\big\} + \langle\mathbf{x}\rangle^H\mathbf{D}\langle\mathbf{x}\rangle\\ &\quad + \mathrm{Tr}\big\{\hat{\mathbf{\Sigma}}_{\mathbf{x}}\mathbf{D}\big\} + \mathrm{Tr}\big\{\hat{\mathbf{\Sigma}}_{\mathbf{x}}\langle\mathbf{A}^H\rangle\langle\mathbf{A}\rangle\big\}.\end{aligned} \tag{28}$$

Finally, we apply (28) to (25) and remove the duplicated terms to prove (24). ∎

Corollary 1. *Given $\mathbf{x}$ is deterministic, $\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle$ is simplified to:*

$$\big\langle\|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2\big\rangle = \|\mathbf{y}-\langle\mathbf{A}\rangle\mathbf{x}\|^2 + \sum_{i=1}^{n}|x_i|^2\mathrm{Tr}\{\hat{\mathbf{\Sigma}}_{\mathbf{a}_i}\}. \tag{29}$$

Proof: We derive (29) by setting $\hat{\mathbf{\Sigma}}_{\mathbf{x}} = \mathbf{0}$ and $\mathbf{x}^H\mathbf{D}\mathbf{x} = \sum_{i=1}^{n}|x_i|^2\mathrm{Tr}\{\hat{\mathbf{\Sigma}}_{\mathbf{a}_i}\}$ in Lemma 1. ∎

3) Updating $x_{i,t}$: We only use this update when $T_p < t \leq T$. In this part, we take the expectation of (17) with respect to all latent variables except for $x_{i,t}$ to find the variational distribution $q_i(x_{i,t})$ as below:

$$\begin{aligned}&q_i(x_{i,t})\\ &\propto \exp\Big\{\big\langle\ln p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{H}_t,\gamma_t) + \ln p(x_{i,t})\big\rangle\Big\}\\ &\propto p(x_{i,t})\exp\big\{\big\langle-\gamma_t\big\|\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t\big\|^2\big\rangle\big\}\\ &\propto p(x_{i,t})\exp\Big\{-\langle\gamma_t\rangle\Big\langle\Big\|\mathbf{y}_t-\mathbf{h}_{i,t}x_{i,t}-\sum_{j\neq i}^{K}\mathbf{h}_{j,t}x_{j,t}\Big\|^2\Big\rangle\Big\}\\ &\propto p(x_{i,t})\exp\Bigg\{-\langle\gamma_t\rangle\Bigg[\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|x_{i,t}|^2\\ &\qquad - 2\,\Re\Big\{\big\langle\mathbf{h}^H_{i,t}\big\rangle\bigg(\mathbf{y}_t-\sum_{j\neq i}^{K}\langle\mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\bigg)x^*_{i,t}\Big\}\Bigg]\Bigg\}\\ &\propto p(x_{i,t})\exp\big\{-\langle\gamma_t\rangle\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|x_{i,t}-z_{i,t}|^2\big\},\end{aligned} \tag{30}$$

where

$$z_{i,t} \triangleq \frac{\langle\mathbf{h}^H_{i,t}\rangle}{\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle}\bigg(\mathbf{y}_t-\sum_{j\neq i}^{K}\langle\mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\bigg), \tag{31}$$

acts as a linear estimate of $x_{i,t}$. In (31), we use Corollary 1 to compute $\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle$ as follows:

$$\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle = \|\langle\mathbf{h}_{i,t}\rangle\|^2 + \mathrm{Tr}\{\hat{\mathbf{\Sigma}}_{i,t}\}. \tag{32}$$

Since the prior distribution $p(x_{i,t})$ is discrete, the variational distribution $q_i(x_{i,t})$ is also discrete. Therefore, we need to normalize it such that:

$$q_i(a) = \frac{p_a\exp\big\{-\langle\gamma_t\rangle\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|a-z_{i,t}|^2\big\}}{\sum_{b\in\mathcal{S}}p_b\exp\big\{-\langle\gamma_t\rangle\big\langle\|\mathbf{h}_{i,t}\|^2\big\rangle|a-z_{i,t}|^2\big\}}, \forall a\in\mathcal{S}. \tag{33}$$

As a result, the variational mean and variance of $x_{i,t}$ are equal to:

$$\langle x_{i,t}\rangle = \sum_{a\in\mathcal{S}}aq_i(a), \tag{34}$$

$$\tau^x_{i,t} = \sum_{a\in\mathcal{S}}|a|^2q_i(a) - |\langle x_{i,t}\rangle|^2. \tag{35}$$

$$
\begin{aligned}
q(\mathbf{h}_{i,t}) &\propto \exp\Big\{\big\langle \ln p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{H}_t,\gamma_t)+\ln p(\mathbf{h}_{i,t}|\eta_i;\mathbf{R}_i)\big\rangle\Big\}\\
&\propto \exp\Big\{-\big\langle \gamma_t\|\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t\|^2\big\rangle-\Big\langle(\mathbf{h}_{i,t}-\eta_i\hat{\mathbf{h}}_{i,t-1|t-1})^H\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}(\mathbf{h}_{i,t}-\eta_i\hat{\mathbf{h}}_{i,t-1|t-1})\Big\rangle\Big\}\\
&\propto \exp\Big\{-\Big\langle\gamma_t\Big\|\mathbf{y}_t-\mathbf{h}_{i,t}x_{i,t}-\sum_{j\neq i}^{K}\mathbf{h}_{j,t}x_{j,t}\Big\|^2\Big\rangle-\Big\langle(\mathbf{h}_{i,t}-\eta_i\hat{\mathbf{h}}_{i,t-1|t-1})^H\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}(\mathbf{h}_{i,t}-\eta_i\hat{\mathbf{h}}_{i,t-1|t-1})\Big\rangle\Big\}\\
&\propto \exp\Big\{-\mathbf{h}_{i,t}^H\Big[\langle\gamma_t\rangle\langle|x_{i,t}|^2\rangle\mathbf{I}_M+\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\Big]\mathbf{h}_{i,t}\\
&\quad+2\,\Re\Big\{\langle\gamma_t\rangle\mathbf{h}_{i,t}^H\Big(\mathbf{y}_t-\sum_{j\neq i}^{K}\langle\mathbf{h}_{j,t}\rangle\langle x_{j,t}\rangle\Big)\langle x_{i,t}^*\rangle+\langle\eta_i\rangle\mathbf{h}_{i,t}^H\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\hat{\mathbf{h}}_{i,t-1|t-1}\Big\}\Big\}. \qquad (18)
\end{aligned}
$$

$$
\begin{aligned}
q(\eta_i) &\propto \exp\Big\{\big\langle \ln p\big(\mathbf{h}_{i,t}|\eta_i;\mathbf{R}_i\big)+\ln p\big(\eta_i\big)\big\rangle\Big\}\\
&\propto \exp\Big\{-\Big\langle\big(\mathbf{h}_{i,t}-\eta_i\hat{\mathbf{h}}_{i,t-1|t-1}\big)^H\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\big(\mathbf{h}_{i,t}-\eta_i\hat{\mathbf{h}}_{i,t-1|t-1}\big)\Big\rangle-\|\eta_i-\hat{\eta}_{i,t-1}\|^2/\tau^{\eta}_{i,t-1}\Big\}\\
&\propto \exp\Big\{-\eta_i^2\Big(\hat{\mathbf{h}}^H_{i,t-1|t-1}\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\hat{\mathbf{h}}_{i,t-1|t-1}+1/\tau^{\eta}_{i,t-1}\Big)+2\eta_i\Big(\Re\big\{\hat{\mathbf{h}}^H_{i,t-1|t-1}\hat{\mathbf{\Sigma}}^{-1}_{i,t|t-1}\langle\mathbf{h}_{i,t}\rangle\big\}+\frac{\hat{\eta}_{i,t-1}}{\tau^{\eta}_{i,t-1}}\Big)\Big\}, \qquad (21)
\end{aligned}
$$

4) Updating γ_t: In the last part of the CAVI algorithm, we compute $q(\gamma_t)$. To achieve this, we take the expectation of (17) with respect to all latent variables except for γ_t, to derive the variational distribution $q(\gamma_t)$ as follows:

$$
\begin{aligned}
q(\gamma_t) &\propto \exp\Big\{\big\langle \ln p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{H}_t,\gamma_t)+\ln p(\gamma_t)\big\rangle\Big\}\\
&\propto \exp\Big\{M\ln\gamma_t-\gamma_t\big\langle\|\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t\|^2\big\rangle\\
&\qquad+(a_0-1)\ln\gamma_t-b_0\gamma_t\Big\}. \qquad (36)
\end{aligned}
$$

Based on (36), $q(\gamma_t)$ is Gamma distribution with mean

$$
\langle\gamma_t\rangle=\frac{a_0+M}{b_0+\big\langle\|\mathbf{y}_t-\mathbf{H}_t\mathbf{x}_t\|^2\big\rangle}, \qquad (37)
$$

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

The block diagram of the online processing strategy is shown in Fig. 2(a), where we use solid gray circles to denote variables (parameters) that are either known or can be obtained through prior distribution. Additionally, we use dashed orange circles to denote variables that need to be estimated.

Note that the Gaussian variational distributions of $\mathbf{h}_{i,t}$ and η_i, along with the Gamma variational distribution of γ_t, justify the three assumptions made earlier in Table I. In Algorithm 1, we demonstrate the details of our online processing strategy using the CAVI algorithm. Here, we use $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_T]$, $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_T]$, and $\mathbf{H}=[\mathbf{H}_1,\mathbf{H}_2,\ldots,\mathbf{H}_T]$. We define the convergence condition to occur when the number of iterations reaches its maximum value I_{tr}.

As discussed, the online processing strategy focuses only on the information available at time t, which leads to a low-complexity low-latency JED approach. However, due to its limited information, it cannot fully eliminate errors in channel estimation and data detection, resulting in error propagation in the JED process. To mitigate JED errors, we propose block processing, which is explained in detail in the next section.

Algorithm 1: The Online Processing Strategy

Input: $\mathbf{Y}$, $\hat{\mathbf{h}}_{i,0|0}$, $\hat{\mathbf{\Sigma}}_{i,0|0}$, prior distribution $p_a, \forall a\in\mathcal{S}$, prior distribution $p(\eta_i)$ (i.e., $\hat{\eta}_{i,0}$ and $\tau^{\eta}_{i,0}$), prior distribution $p(\gamma_t)$ (i.e., a_0 and b_0), and I_{tr};
Output: $\mathbf{X}$, $\mathbf{H}$, $\boldsymbol{\eta}$, and γ_t;
for $t=1,2,\ldots,T$ **do**
  if $T_p<t\leq T$ **then**
    Initialize $\langle\mathbf{x}_t\rangle=\mathbf{0}$;
  for $i=1,2,\ldots,K$ **do**
    Use (13) and (14) to predict the prior distribution of $\mathbf{h}_{i,t}$;
    Set $\langle\mathbf{h}_{i,t}\rangle=\hat{\mathbf{h}}_{i,t|t-1}$, $\mathbf{\Sigma}_{i,t}=\hat{\mathbf{\Sigma}}_{i,t|t-1}$;
    Compute $\langle\eta_i\rangle$ and $\langle\gamma_t\rangle$ based on the prior distributions of η_i and γ_t;
  repeat
    for $i=1,2,\ldots,K$ **do**
      Find the distribution $q(\mathbf{h}_{i,t})$ using (18);
      Obtain $\mathbf{\Sigma}_{i,t}$ and $\langle\mathbf{h}_{i,t}\rangle$ using (19) and (20);
    for $i=1,2,\ldots,K$ **do**
      Attain the distribution $q(\eta_i)$ as in (21);
      Determine $\tau^{\eta}_{i,t}$ and $\langle\eta_i\rangle$ using (22) and (23);
    if $T_p<t\leq T$ **then**
      for $i=1,2,\ldots,K$ **do**
        Get the distribution $q_i(x_{i,t})$ as in (30);
        Calculate $\langle x_{i,t}\rangle$ and $\tau^x_{i,t}$ based on (34) and (35);
    Compute $\langle\gamma_t\rangle$ as in (37);
  until *convergence*;
Set $\hat{\mathbf{h}}_{i,t|t}=\langle\mathbf{h}_{i,t}\rangle$, $\hat{\mathbf{\Sigma}}_{i,t|t}=\mathbf{\Sigma}_{i,t}$, $\hat{\eta}_{i,t}=\langle\eta_i\rangle$, and $\gamma_t=\langle\gamma_t\rangle$;
Calculate $\hat{x}_{i,t}=\arg\max_{a\in\mathcal{S}}q_i(a)$.

IV. Block Processing Strategy

In real-world massive MIMO systems, the BS typically estimates uplink channels and uses them for downlink beamforming to enhance data rates and quality of service (QoS), making accurate channel estimation essential. In this section, we propose a block processing strategy to reduce channel estimation errors. The concept of block processing originates

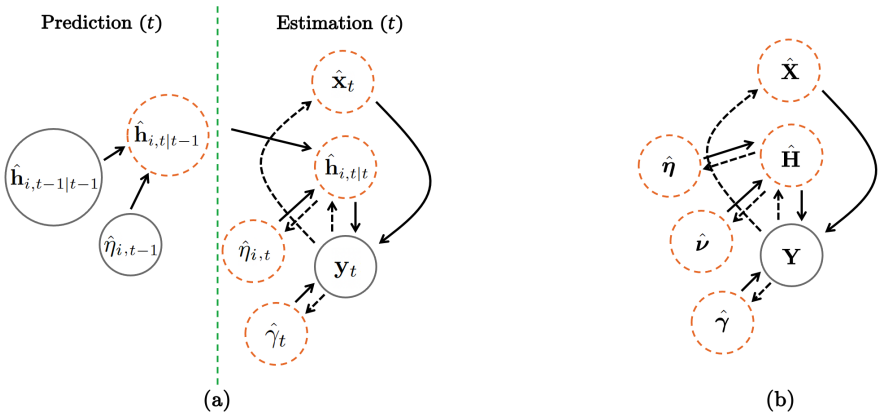

Fig. 2. Block diagram of (a) online processing and (b) block processing strategies.

from estimation theory, where the estimation error decreases as more observed signals become available. We first formulate the problem, where the communication block spans the entire communication time, stacking all received signals from time 1 to T and processing them together. Next, we introduce a scheme for acquiring Bayesian optimal estimates of $\mathbf{H}_t$ and $\mathbf{x}_t$, and then apply the CAVI algorithm to sequentially update variables to obtain a local optimal solution.

A. Problem Formulation

Similar to the online strategy in Section III, we assume η_i and γ_t are i.i.d. unknown variables. Then, we compute joint distribution $p(\mathbf{Y}, \mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}; \bar{\mathbf{R}})$ as below:

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}; \bar{\mathbf{R}}) = \left[\prod_{t=1}^{T} p(\mathbf{H}_t|\mathbf{H}_{t-1}, \boldsymbol{\eta}; \mathbf{R}_i)\right] \times \left[\prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{H}_t, \gamma_t) p(\mathbf{x}_t) p(\gamma_t)\right] p(\boldsymbol{\eta}) p(\mathbf{H}_0), \quad (38)$$

where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \ldots, \gamma_T]$ and $p(\mathbf{H}_0) = \prod_{i=1}^{K} p(\mathbf{h}_{i,0})$ is the prior channel distribution. In this case, to compute the conditional probability $p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}, \eta_i; \mathbf{R}_i)$, we use (4) and rewrite it as:

$$p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}, \eta_i; \mathbf{R}_i) = \mathcal{CN}\left(\mathbf{h}_{i,t}; \eta_i \mathbf{h}_{i,t-1}, \nu_i^{-1} \mathbf{R}_i\right), \quad (39)$$

where $\nu_i \triangleq (1-\eta_i^2)^{-1}$. In the following subsection, we apply the VB framework to estimate ν_i.

B. Estimation phase

In this part, we focus on the entire communication block for JED. Unlike online processing, we simplify the process by bypassing the prediction phase, and then we solely rely on the estimation phase to derive the Bayesian optimal estimates for $\mathbf{H}_t$ and $\mathbf{x}_t$. We achieve this through the posterior distribution $p(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}|\mathbf{Y})$. Although (39) shows that ν_i and η_i are dependent, in the estimation process, we assume these two variables are independent and hope to estimate ν_i such that $\langle \nu_i \rangle = \left(1 - \langle \eta_i \rangle^2\right)^{-1}$. Our simulation results in Section V will validate this assumption. Therefore, we replace the posterior distribution $p(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}|\mathbf{Y})$ by $p(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}|\mathbf{Y})$, where $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_K]$. Then, we use the mean-field variational distribution $q(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu})$ to approximate $p(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}|\mathbf{Y})$, which is given by:

$$p(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}|\mathbf{Y}) \approx q(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}) = \prod_{i=1}^{K} \prod_{t=1}^{T} q_i(x_{i,t}) q(\mathbf{h}_{i,t}) q(\eta_i) q(\nu_i) q(\gamma_t). \quad (40)$$

Further, referring to (10), the optimal solution for the variational densities in (40) requires the joint distribution $p(\mathbf{Y}, \mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}; \bar{\mathbf{R}})$. To obtain this, one can factorize $p(\mathbf{Y}, \mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}; \bar{\mathbf{R}})$ as follows:

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}; \bar{\mathbf{R}}) = \left[\prod_{i=1}^{K} p(\eta_i) p(\mathbf{h}_{i,0})\right] \times \left[\prod_{i=1}^{K} p(\nu_i)\right] \left[\prod_{i=1}^{K} \prod_{t=1}^{T} p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}, \eta_i, \nu_i; \mathbf{R}_i)\right] \times \left[\prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{H}_t, \gamma_t) \left[\prod_{i=1}^{K} p(x_{i,t})\right] p(\gamma_t)\right]. \quad (41)$$

In this strategy, we need the prior probabilities $p(\eta_i)$, $p(\gamma_t)$, $p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}, \eta_i, \nu_i; \mathbf{R}_i)$, $p(x_{i,t})$, and $p(\nu_i)$ to attain the variational distributions in (40), which are outlined in Table II. We show the block diagram of the block processing strategy in Fig. 2(b).

TABLE II
THE PRIOR DISTRIBUTION ASSUMPTIONS FOR THE UNKNOWN RANDOM VARIABLES INVOLVED IN $q(\mathbf{X}, \mathbf{H}_1, \ldots, \mathbf{H}_T, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu})$

Probability	Prior Distribution Assumption
$p(\mathbf{h}_{i,t}\|\mathbf{h}_{i,t-1}, \eta_i, \nu_i; \mathbf{R}_i)$	$\mathcal{CN}\left(\mathbf{h}_{i,t}; \eta_i \mathbf{h}_{i,t-1}, \nu_i^{-1} \mathbf{R}_i\right)$
$p(\gamma_t)$	$\Gamma(a_{0,b}, b_{0,b})$
$p(\eta_i)$	$\mathcal{N}(\eta_i; \hat{\eta}_{i,b}, \tau_{i,b}^{\eta})$
$p(\nu_i)$	$\Gamma(\bar{a}_{i,b}, \bar{b}_{i,b})$
$p(x_{i,t})$(pilot transmission)	$\delta(x_{i,t} - \bar{x}_{i,t})$
$p(x_{i,t})$(data transmission)	$\sum_{a \in \mathcal{S}} p_a \delta(x_{i,t} - a)$

C. The CAVI algorithm

Here, akin to the approach used in online processing, we apply the iterative CAVI algorithm to determine the local optimal solutions for $q^{\star}(x_{i,t})$, $q^{\star}(\mathbf{h}_{i,t})$, $q^{\star}(\eta_i)$, $q^{\star}(\gamma_t)$, and $q^{\star}(\nu_i)$. It is important to note that the updating process for

$x_{i,t}$ and γ_t closely resembles the procedure employed in the online processing strategy. Thus, we only provide the specifics of updating $\mathbf{h}_{i,t}$, η_i, and ν_i in this subsection.

1) Updating $\mathbf{h}_{i,t}$: In Section III, we compute $q(\mathbf{h}_{i,t})$ based on the predictive distribution and the conditional probability $p(\mathbf{y}_t|\mathbf{H}_t, \mathbf{x}_t, \gamma_t)$ at time t. However, as shown in (41), the block processing strategy requires three conditional probabilities that have $\mathbf{h}_{i,t}$ as a factor, i.e., $p(\mathbf{y}_t|\mathbf{H}_t, \mathbf{x}_t, \gamma_t)$, $p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}, \eta_i, \nu_i; \mathbf{R}_i)$, and $p(\mathbf{h}_{i,t+1}|\mathbf{h}_{i,t}, \eta_i, \nu_i; \mathbf{R}_i)$ to derive $q(\mathbf{h}_{i,t})$. By taking the expectation of the conditional probability in (41) with respect to all latent variables except for $\mathbf{h}_{i,t}$, the variational distribution $q(\mathbf{h}_{i,t})$ is given in (44) (at the top of the next page), which is Gaussian with covariance $\mathbf{\Sigma}_{i,t}$ and mean $\langle \mathbf{h}_{i,t} \rangle$, as outlined below:

$$\mathbf{\Sigma}_{i,t} = \Big[\langle \gamma_t \rangle \langle |x_{i,t}|^2 \rangle \mathbf{I}_M + \big(1 + \langle \eta_i^2 \rangle\big) \langle \nu_i \rangle \mathbf{R}_i^{-1} \Big]^{-1}, \quad (42)$$

$$\langle \mathbf{h}_{i,t} \rangle = \mathbf{\Sigma}_{i,t} \Bigg[\langle \gamma_t \rangle \Big(\mathbf{y}_t - \sum_{j \neq i}^{K} \langle \mathbf{h}_{j,t} \rangle \langle x_{j,t} \rangle \Big) \langle x_{i,t}^* \rangle + \langle \eta_i \rangle \langle \nu_i \rangle \mathbf{R}_i^{-1} \left(\mathbf{h}_{i,t+1} + \mathbf{h}_{i,t-1} \right) \Bigg]. \quad (43)$$

2) Updating η_i: Taking the expectation of the conditional probability in (41) with respect to all latent variables except for η_i, we obtain the variational distribution $q(\eta_i)$ as presented in (45) (on the next page). Specifically, to derive $q(\eta_i)$, we require the conditional channel probabilities throughout the entire block and the prior probability of $p\big(\eta_i; \hat{\eta}_{i,b}, \tau_{i,b}^{\eta}\big)$. Notice $\hat{\eta}_{i,b} = \hat{\eta}_{i,0}$ and $\tau_{i,b}^{\eta} = \tau_{i,0}^{\eta}$ in (45). Here, the variational distribution $q(\eta_i)$ is Gaussian with the following variance and mean:

$$\tau_{i,t}^{\eta} = \Big(\langle \nu_i \rangle \sum_{t=1}^{T} \langle \mathbf{h}_{i,t-1} \rangle^H \mathbf{R}_i^{-1} \langle \mathbf{h}_{i,t-1} \rangle + 1/\tau_{i,t-1}^{\eta} \Big)^{-1}, \quad (46)$$

$$\langle \eta_{i,t} \rangle = \tau_{i,t}^{\eta} \Big(\Re\big\{ \langle \nu_i \rangle \sum_{t=1}^{T} \langle \mathbf{h}_{i,t-1} \rangle^H \mathbf{R}_i^{-1} \langle \mathbf{h}_{i,t} \rangle \big\} + \frac{\hat{\eta}_{i,t-1}}{\tau_{i,t-1}^{\eta}} \Big). \quad (47)$$

3) Updating ν_i: In this case, we assume ν_i is independent of η_i and subsequently employ the VB framework to estimate the variational distribution $q(\nu_i)$. To derive $q(\nu_i)$, the following lemma is essential:

Lemma 2. *[41] Let $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{x} \in \mathbb{C}^{n \times 1}$, and $\mathbf{y} \in \mathbb{C}^{m \times 1}$ be three independent random matrices (vectors) with respect to a variational density $q_{\mathbf{A},\mathbf{x},\mathbf{y}}(\mathbf{A}, \mathbf{x}, \mathbf{y}) = q_{\mathbf{A}}(\mathbf{A}) q_{\mathbf{x}}(\mathbf{x}) q_{\mathbf{y}}(\mathbf{y})$. Suppose $\mathbf{A}$ is column-wise independent and let $\langle \mathbf{a}_i \rangle$ and $\mathbf{\Sigma}_{\mathbf{a}_i}$ denote the variational mean and covariance matrix of the i^{th} column of $\mathbf{A}$. Let $\langle \mathbf{x} \rangle$ and $\mathbf{\Sigma}_{\mathbf{x}}$ represent the variational mean and covariance matrix of $\mathbf{x}$, respectively. Further, assume $\langle \mathbf{y} \rangle$ and $\mathbf{\Sigma}_{\mathbf{y}}$ are the variational mean and covariance matrix of $\mathbf{y}$, respectively. Then, for an arbitrary Hermitian matrix $\mathbf{R}$, we define $\big\langle (\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{R} (\mathbf{y} - \mathbf{A}\mathbf{x}) \big\rangle$ as the expectation of $(\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{R} (\mathbf{y} - \mathbf{A}\mathbf{x})$ with respect to $q_{\mathbf{A},\mathbf{x},\mathbf{y}}(\mathbf{A}, \mathbf{x}, \mathbf{y})$. Here, we have:*

$$\begin{aligned} \big\langle (\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{R} (\mathbf{y} - \mathbf{A}\mathbf{x}) \big\rangle &= \langle \mathbf{x} \rangle^H \mathbf{D} \langle \mathbf{x} \rangle + \mathrm{Tr}\big\{ \mathbf{\Sigma}_{\mathbf{x}} \mathbf{D} \big\} \\ &+ \big(\langle \mathbf{y} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle \big)^H \mathbf{R} \big(\langle \mathbf{y} \rangle - \langle \mathbf{A} \rangle \langle \mathbf{x} \rangle \big) \\ &+ \mathrm{Tr}\big\{ \mathbf{R} \mathbf{\Sigma}_{\mathbf{y}} \big\} + \mathrm{Tr}\big\{ \mathbf{\Sigma}_{\mathbf{x}} \langle \mathbf{A}^H \rangle \mathbf{R} \langle \mathbf{A} \rangle \big\}, \end{aligned} \quad (48)$$

where $\mathbf{D} = \mathrm{diag}\big(\mathrm{Tr}\{\mathbf{R}\mathbf{\Sigma}_{\mathbf{a}_1}\}, \ldots, \mathrm{Tr}\{\mathbf{R}\mathbf{\Sigma}_{\mathbf{a}_n}\}\big)$.

Proof: We omit the proof of this lemma since it is similar to the proof of Lemma 1. ∎

By taking the expectation of the probability in (41) with respect to all latent variables except for ν_i, we have:

$$\begin{aligned} q(\nu_i) &\propto \exp\Big\{ \big\langle \sum_{t=1}^{T} \ln p\big(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1}, \eta_i, \nu_i; \mathbf{R}_i\big) \\ &\qquad + \ln p\big(\nu_i; \bar{a}_{i,t-1}, \bar{b}_{i,t-1}\big) \big\rangle \Big\} \\ &\propto \exp\Big\{ TM \ln \nu_i + (\bar{a}_{i,t-1} - 1) \ln \nu_i - \bar{b}_{i,t-1} \nu_i \\ &\quad - \nu_i \Big\langle \sum_{t=1}^{T} \big(\mathbf{h}_{i,t} - \eta_i \mathbf{h}_{i,t-1}\big)^H \mathbf{R}_i^{-1} \big(\mathbf{h}_{i,t} - \eta_i \mathbf{h}_{i,t-1}\big) \Big\rangle \Big\}. \end{aligned} \quad (49)$$

Thus, the variational distribution $q(\nu_i)$ is $\Gamma(\bar{a}_{i,t}, \bar{b}_{i,t})$, where:

$$\bar{a}_{i,t} = \bar{a}_{i,t-1} + TM, \quad (50)$$

and $\bar{b}_{i,t}$ is given in (51) (on the next page). Here, $\langle \nu_i \rangle$ is equal to $\bar{a}_{i,t}/\bar{b}_{i,t}$. In Algorithm 2, we explain the details of how our block processing strategy utilizes the CAVI algorithm to optimize $q(\mathbf{h}_{i,t})$, $q_i(x_{i,t})$, $q(\eta_i)$, $q(\nu_i)$, and $q(\gamma_t)$.

V. Simulation Results

In this section, we evaluate the performance of our proposed VB-based approaches in time-varying channels. We consider a JED problem in a massive MIMO network supporting $K = 4$ high-mobility users. Initially, we focus on the online processing strategy and compare its performance with three baselines (i.e., the LMMSE, KF, and EP methods) in terms of SER under two cases: (i) when η_i is fixed and (ii) when it is a random variable. Furthermore, we compare the computational complexity and channel NMSE of our proposed online strategy with the benchmarks, where NMSE is given by:

$$\mathrm{NMSE\ (dB)} = 10 \log_{10} \left(\frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_{\mathrm{F}}^2}{\|\mathbf{H}\|_{\mathrm{F}}^2} \right). \quad (52)$$

Subsequently, we use an interleaved structure to enhance the SER performance of the VB-based online strategy. Finally, we assess the performance of both our online and block processing strategies in terms of SER and channel NMSE.

Throughout this section, we use $I_{\mathrm{tr}} = 50$, $T_p = 8$, and $T_d = 128$ to represent the maximum number of iterations for the CAVI algorithm and the KF and EP methods, the pilot transmission time, and the data transmission time, respectively.

We normalize the covariance matrix $\mathbf{R}_i$ $\forall i$ such that its diagonal elements are all set to $1/M$, resulting in $\mathbb{E}[|\mathbf{h}_i|^2] = 1$.

$$
\begin{aligned}
q(\mathbf{h}_{i,t}) &\propto \exp\Big\{ \big\langle \ln p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{H}_t,\gamma_t) + \ln p(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1},\eta_i,\nu_i;\mathbf{R}_i) + \ln p(\mathbf{h}_{i,t+1}|\mathbf{h}_{i,t},\eta_i,\nu_i;\mathbf{R}_i)\big\rangle \Big\} \\
&\propto \exp\Big\{ -\big\langle \gamma_t \big\| \mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t \big\|^2 \big\rangle - \Big\langle (\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1})^H \nu_i \mathbf{R}_i^{-1} (\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1}) \Big\rangle \\
&\quad - \Big\langle (\mathbf{h}_{i,t+1}-\eta_i\mathbf{h}_{i,t})^H \nu_i \mathbf{R}_i^{-1} (\mathbf{h}_{i,t+1}-\eta_i\mathbf{h}_{i,t}) \Big\rangle \Big\} \\
&\propto \exp\Big\{ -\Big\langle \gamma_t \Big\| \mathbf{y}_t - \mathbf{h}_{i,t}x_{i,t} - \sum_{j\neq i}^{K} \mathbf{h}_{j,t}x_{j,t} \Big\|^2 \Big\rangle - \Big\langle (\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1})^H \nu_i \mathbf{R}_i^{-1} (\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1}) \Big\rangle \\
&\quad - \Big\langle (\mathbf{h}_{i,t+1}-\eta_i\mathbf{h}_{i,t})^H \nu_i \mathbf{R}_i^{-1} (\mathbf{h}_{i,t+1}-\eta_i\mathbf{h}_{i,t}) \Big\rangle \Big\} \\
&\propto \exp\Big\{ -\mathbf{h}_{i,t}^H \Big[\langle\gamma_t\rangle \langle |x_{i,t}|^2 \rangle \mathbf{I}_M + \big(1+\langle\eta_i^2\rangle\big) \langle\nu_i\rangle \mathbf{R}_i^{-1} \Big] \mathbf{h}_{i,t} \\
&\quad + 2\,\Re\Big\{ \langle\gamma_t\rangle \mathbf{h}_{i,t}^H \Big(\mathbf{y}_t - \sum_{j\neq i}^{K} \langle \mathbf{h}_{j,t}\rangle \langle x_{j,t}\rangle \Big) \langle x_{i,t}^*\rangle + \langle\eta_i\rangle\langle\nu_i\rangle \mathbf{h}_{i,t}^H \mathbf{R}_i^{-1} \left(\mathbf{h}_{i,t+1} + \mathbf{h}_{i,t-1} \right) \Big\} \Big\}.
\end{aligned} \tag{44}
$$

$$
\begin{aligned}
q(\eta_i) &\propto \exp\Big\{ \big\langle \sum_{t=1}^{T} \ln p\big(\mathbf{h}_{i,t}|\mathbf{h}_{i,t-1},\eta_i,\nu_i;\mathbf{R}_i\big) + \ln p\big(\eta_i;\hat{\eta}_{i,t-1},\tau_{i,t-1}^{\eta}\big) \big\rangle \Big\} \\
&\propto \exp\Big\{ -\Big\langle \sum_{t=1}^{T} \big(\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1}\big)^H \nu_i \mathbf{R}_i^{-1} \big(\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1}\big) \Big\rangle - \|\eta_i - \hat{\eta}_{i,t-1}\|^2 / \tau_{i,t-1}^{\eta} \Big\} \\
&\propto \exp\Big\{ -\eta_i^2 \Big(\langle\nu_i\rangle \sum_{t=1}^{T} \langle \mathbf{h}_{i,t-1}\rangle^H \mathbf{R}_i^{-1} \langle \mathbf{h}_{i,t-1}\rangle + 1/\tau_{i,t-1}^{\eta} \Big) + 2\eta_i \Big(\Re\big\{ \langle\nu_i\rangle \sum_{t=1}^{T} \langle \mathbf{h}_{i,t-1}\rangle^H \mathbf{R}_i^{-1} \langle \mathbf{h}_{i,t}\rangle \big\} + \frac{\hat{\eta}_{i,t-1}}{\tau_{i,t-1}^{\eta}} \Big) \Big\},
\end{aligned} \tag{45}
$$

$$
\begin{aligned}
\bar{b}_{i,t} &= \bar{b}_{i,t-1} + \Big\langle \sum_{t=1}^{T} \big(\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1}\big)^H \mathbf{R}_i^{-1} \big(\mathbf{h}_{i,t}-\eta_i\mathbf{h}_{i,t-1}\big) \Big\rangle \\
&= \bar{b}_{i,t-1} + \sum_{t=1}^{T} \Big(\langle \mathbf{h}_{i,t}\rangle - \langle\eta_i\rangle\langle \mathbf{h}_{i,t-1}\rangle \Big)^H \mathbf{R}_i^{-1} \Big(\langle \mathbf{h}_{i,t}\rangle - \langle\eta_i\rangle\langle \mathbf{h}_{i,t-1}\rangle \Big) + \sum_{t=1}^{T} \mathrm{Tr}\big\{ \mathbf{R}_i^{-1} \boldsymbol{\Sigma}_{i,t} \big\} \\
&\quad + \sum_{t=1}^{T} \tau_{i,t}^{\eta} \langle \mathbf{h}_{i,t-1}\rangle^H \mathrm{Tr}\{\mathbf{R}_i^{-1}\} \langle \mathbf{h}_{i,t-1}\rangle + \sum_{t=1}^{T} \mathrm{Tr}\big\{ \tau_{i,t}^{\eta} \boldsymbol{\Sigma}_{i,t-1} \mathrm{Tr}\{\mathbf{R}_i^{-1}\} \big\} + \sum_{t=1}^{T} \mathrm{Tr}\big\{ \boldsymbol{\Sigma}_{i,t-1} \langle \eta_{i,t}\rangle \mathbf{R}_i^{-1} \langle \eta_{i,t}\rangle \big\} \big\},
\end{aligned} \tag{51}
$$

Subsequently, the noise variance N_0 is determined using the signal-to-noise ratio (SNR), given by:

$$
\mathrm{SNR} = \frac{\mathbb{E}[\|\mathbf{H}\mathbf{x}\|^2]}{\mathbb{E}[\|\mathbf{n}\|^2]} = \frac{\sum_{i=1}^{K} \mathrm{Tr}(\mathbf{R}_i)}{MN_0} = \frac{K}{MN_0}. \tag{53}
$$

To model the correlated channels, we consider the exponential spatial correlation model [42] for each column of $\mathbf{H}_t$, in which the covariance matrix $\mathbf{R}_i$ is set to:

$$
[\mathbf{R}_i]_{k\ell} = \begin{cases} \frac{1}{M}\alpha^{k-\ell}, & \text{if } k \geq \ell, \\ \frac{1}{M}\left(\alpha^{\ell-k}\right)^*, & \text{if } k < \ell, \end{cases} \tag{54}
$$

where $1 \leq k, \ell \leq M$, and α is the (complex) correlation coefficient between neighboring receive antennas.

Further, we use the LMMSE channel estimation method to find $\hat{\mathbf{h}}_{i,0|0}$ and $\hat{\boldsymbol{\Sigma}}_{i,0|0}$. We also initialize $p_a = 1/|\mathcal{S}|$, $\hat{\eta}_{i,0} = 0.95, \tau_{i,0}^{\eta} = 10^{-3}, \bar{a}_{i,0} = 10^{-4}, \bar{b}_{i,0} = 10^{-4}, a_0 = 10^{-4}$, and $b_0 = 10^{-4}$, where $|\mathcal{S}|$ denotes the cardinality of $\mathcal{S}$. Ultimately, we evaluate the simulation results across 1000 trials.

A. Performance of Online Processing VB

In this part, we evaluate the effectiveness of the proposed online strategy within the VB framework. To accomplish this, we use the LMMSE, KF, and EP methods [28] as benchmarks and compare them with our approach. Also, to gain deeper insights into the performance of the online VB method, we employ another benchmark where the VB method knows η_i.

We first focus on the SER metric and examine two cases: one where η_i remains constant and another where it varies as a random variable with minor fluctuations.

Case I – η_i is fixed: We assume correlated channels with $\alpha = 0.5 + j0.5$ when $M = 32$, $K = 4$, and $\eta_i = 0.985$, where η_i is unknown at the BS. Notice that $\eta_i = 0.985$ signifies the Doppler frequency of the i^{th} high-mobility user traveling at a velocity of 158 Km/h [28]. Then, we let the VB framework estimate and track this parameter. Fig. 3 illustrates the SER versus SNR using quadrature phase-shift keying (QPSK). As depicted in Fig. 3, both VB-based approaches, with known and

Algorithm 2: The Block Processing Strategy

Input: $\mathbf{Y}$, $\hat{\mathbf{h}}_{i,0|0}$, $\hat{\mathbf{\Sigma}}_{i,0|0}$, prior distribution $p_a, \forall a \in \mathcal{S}$, prior distribution $p(\eta_i)$ (i.e., $\hat{\eta}_{i,0}$ and $\tau^{\eta}_{i,0}$), prior distribution $p(\nu_i)$ (i.e., $\bar{a}_{i,0}$ and $\bar{b}_{i,0}$), prior distribution $p(\gamma_t)$ (i.e., a_0 and b_0), and I_{tr};
Output: $\mathbf{X}$, $\mathbf{H}$, $\boldsymbol{\eta}$, $\boldsymbol{\nu}$, and γ_t;
for $t = 1, 2, \ldots, T$ **do**
  if $T_p < t \leq T$ **then**
    Initialize $\langle \mathbf{x}_t \rangle = \mathbf{0}$;
  for $i = 1, 2, \ldots, K$ **do**
    Compute $\langle \eta_i \rangle$, ν_i, and $\langle \gamma_t \rangle$ based on the prior distributions of η_i, ν_i, and γ_t;
repeat
  for $t = 1, 2, \ldots, T$ **do**
    for $i = 1, 2, \ldots, K$ **do**
      if $t = 1$ **then**
        Set $\mathbf{h}_{i,t-1} = \hat{\mathbf{h}}_{i,0|0}$;
        Follow Algorithm 1 to compute $\mathbf{h}_{i,t+1}$;
      else if $t = T$ **then**
        Set $\mathbf{h}_{i,t+1} = \langle \eta_i \rangle \hat{\mathbf{h}}_{i,0|0}$;
      else
        Follow Algorithm 1 to compute $\mathbf{h}_{i,t+1}$;
      Find the distribution $q(\mathbf{h}_{i,t})$ using (44);
      Obtain $\mathbf{\Sigma}_{i,t}$ and $\langle \mathbf{h}_{i,t} \rangle$ using (42) and (43);
  for $t = 1, 2, \ldots, T$ **do**
    for $i = 1, 2, \ldots, K$ **do**
      Attain the distribution $q(\eta_i)$ as in (45);
      Determine $\tau^{\eta}_{i,t}$ and $\langle \eta_i \rangle$ using (46) and (47);
  for $t = 1, 2, \ldots, T$ **do**
    for $i = 1, 2, \ldots, K$ **do**
      Find the distribution $q(\nu_i)$ as in (49);
      Compute $\bar{a}_{i,t}$ and $\bar{b}_{i,t}$ using (50) and (51);
  for $t = 1, 2, \ldots, T$ **do**
    if $T_p < t \leq T$ **then**
      for $i = 1, 2, \ldots, K$ **do**
        Get the distribution $q_i(x_{i,t})$ as in (30);
        Calculate $\langle x_{i,t} \rangle$ and $\tau^{x}_{i,t}$ based on (34) and (35);
    Compute $\langle \gamma_t \rangle$ as in (37);
until *convergence*;
Set $\mathbf{h}_{i,t} = \langle \mathbf{h}_{i,t} \rangle$, $\eta_i = \langle \eta_i \rangle$, $\nu_i = \langle \nu_i \rangle$, and $\gamma_t = \langle \gamma_t \rangle$;
Get $\hat{x}_{i,t} = \arg\max_{a \in \mathcal{S}} q_i(a)$.

unknown η_i, exhibit nearly identical performance. This observation validates the efficacy of the proposed VB-based method in successfully tracking and estimating η_i. Further, despite the VB method with unknown η_i estimating the parameter over the air, its performance surpasses that of the LMMSE, KF, and EP methods, while these benchmarks require knowing η_i before communication. This is due to the fact that LMMSE does not take the time correlation into account, and the KF method is a one-shot approach, which only passes over the frame once, and thus cannot improve its performance during the process. Moreover, in Fig. 3, our proposed online VB method demonstrates superior performance compared to the EP method because it treats γ_t as an unknown variable, whereas the EP method in [28] assumes it is a fixed known component. This variable nature of γ_t grants the VB framework greater flexibility in estimating $\mathbf{\Sigma}_{i,t}$ in (19), leading to its enhanced performance relative to the EP method.

Then, Fig. 4 illustrates the SER performance of our method in comparison to three benchmark approaches, with $M = 32$,

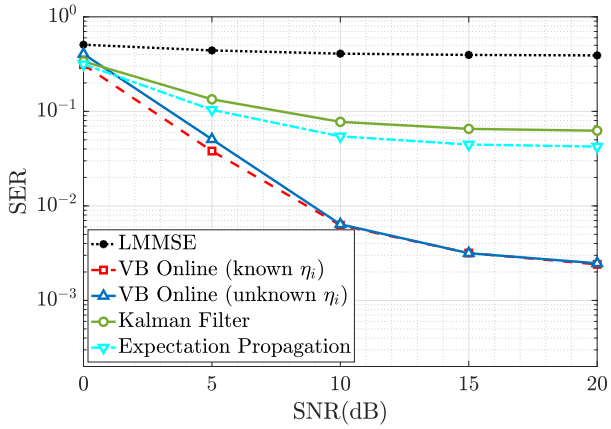

Fig. 3. An SER comparison between LMMSE, KF, EP, and our online VB using QPSK under correlated channels when $\alpha = 0.5 + j0.5, \eta_i = 0.985, T_p = 8, T_d = 128$, and SNR $\in [0, 20]$ dB.

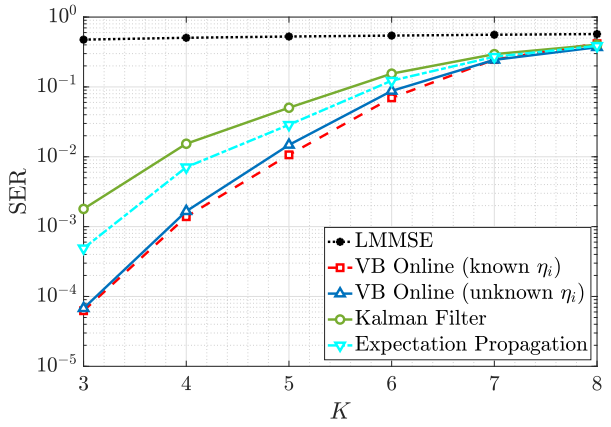

Fig. 4. The SER performance comparison between LMMSE, KF, EP, and our online processing VB-based strategy utilizing QPSK modulation, when $\eta_i = 0.97$, $T_p = 8$, $T_d = 128$, SNR $= 15$ dB, and $K \in [3, 8]$.

SNR $= 15$ dB, $\eta_i = 0.97$, $\mathbf{R}_i = \frac{1}{M}\mathbf{I}_M$, and $K \in [3, 8]$. It is evident that the performance of all methods declines as K increases. This occurs as we can treat a network with K_2 users similarly to one with K_1 users, where $K_1 < K_2$, by setting $K_2 - K_1$ of the K_2 users to be inactive. However, when these $K_2 - K_1$ users become active, they contribute additional interference, further complicating channel estimation and data detection. This is exacerbated by the fact that the VB framework, which tends to converge to local optimums, cannot perfectly eliminate errors during the JED process. As a result, as the number of users increases, the probability of errors in estimation and detection rises, leading to a higher SER. Next, in Fig. 5, we compare the performance of our online VB method with the LMMSE, KF, and EP methods for various values of $\eta_i \in [0.94, 0.98]$ when $M = 32$, $K = 4$, $\mathbf{R}_i = \frac{1}{M}\mathbf{I}_M$, and SNR $= 15$ dB. The results indicate that the SER performance of all methods improves as η_i increases. This happens because, as η_i approaches 1, the variation in the channels in the Gauss-Markov model in (1) between consecutive time slots decreases. This leads to improved accuracy in channel estimation and, consequently, better SER performance.

Case II – η_i is a random variable: To cover a more realistic scenario, we consider a situation where η_i is a random

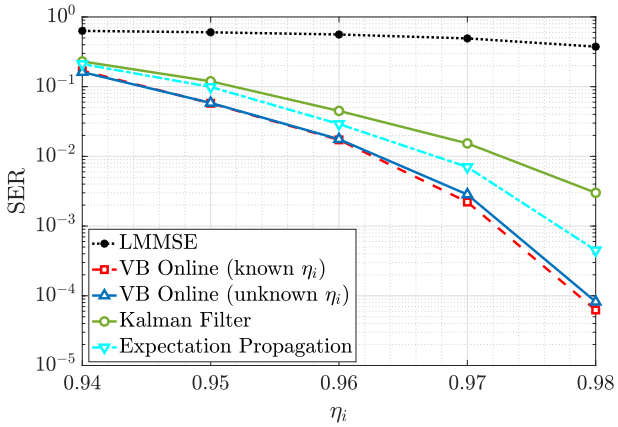

Fig. 5. Comparison of SER performance for LMMSE, KF, EP, and the proposed online VB-based strategy using QPSK modulation, with $\eta_i \in [0.94, 0.98]$, $T_p = 8$, $T_d = 128$, and SNR $= 15$ dB.

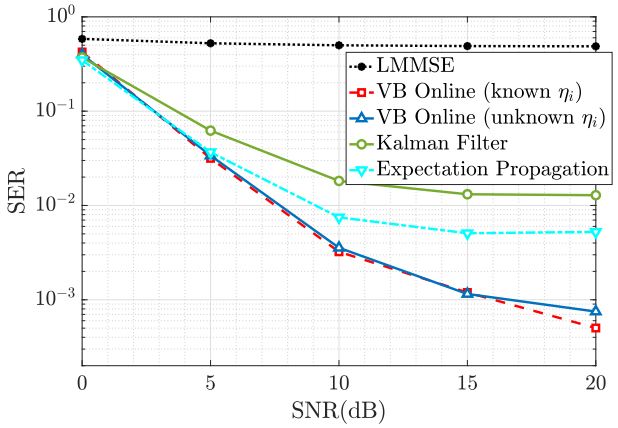

Fig. 6. A comparison of SER among LMMSE, KF, EP, and our online processing VB-based approach employing QPSK modulation where $\eta_i \sim \mathcal{N}(0.97, 5 \times 10^{-5})$, $T_p = 8$, $T_d = 128$, and SNR ranging from 0 to 20 dB.

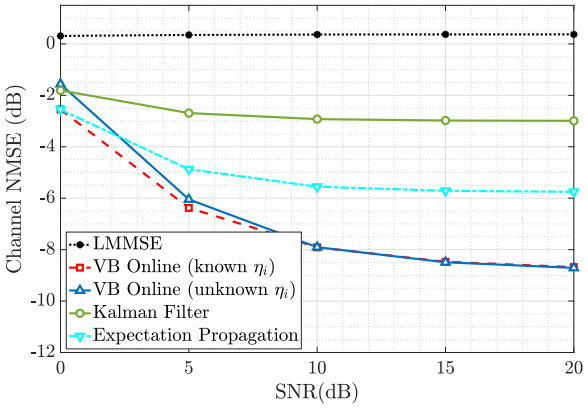

Fig. 7. The channel NMSE performance comparison between LMMSE, KF, EP, and our online VB utilizing QPSK modulation under correlated channels, where the parameters are set to $\alpha = 0.5 + j0.5, \eta_i = 0.985$, $T_p = 8$, $T_d = 128$, and SNR $\in [0, 20]$ dB.

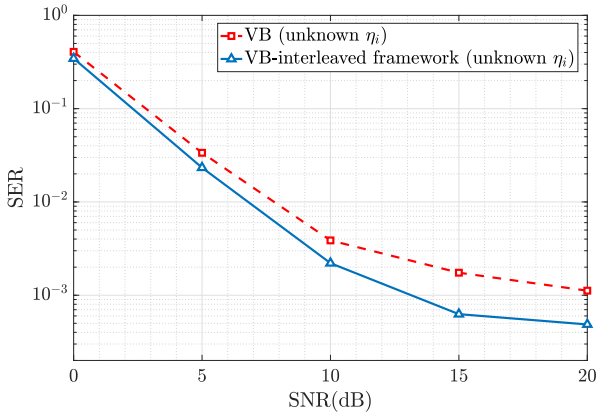

Fig. 8. The SER performance of the proposed online processing using the VB framework with and without interleaved structure under the QPSK signal constellation when $\eta_i = 0.97, T_p = 8, T_d = 128$, and SNR $\in [0, 20]$ dB.

variable with minor fluctuations. This assumption is motivated by the fact that the Doppler frequency of the i^{th} user could change slowly over the communication time due to changes in its velocity and direction. Thus, in this context, we model $\eta_i \sim \mathcal{N}(0.97, 5 \times 10^{-5})$, while $M = 32$, $K = 4$, $\mathbf{R}_i = \frac{1}{M}\mathbf{I}_M$, and $\eta_i = 0.97$. As illustrated in Fig. 6, the observed trend closely mirrors that of the scenario where η_i is fixed, indicating the effectiveness of the proposed online processing approach in accurately estimating and tracking the variable correlation coefficient η_i, consequently showcasing superior performance over the baseline methods in terms of SER.

Secondly, since the channel estimation process plays a crucial role in system performance, we study the channel NMSE performance of our online processing method alongside other benchmarks here. The simulation setup mirrors that depicted in Fig. 3. According to Fig. 7, our proposed online method, based on VB, yields the lowest NMSE results compared to the baselines. This observation elucidates that the superior channel NMSE performance of our method comes from its more precise data detection process compared to the other benchmarks. Furthermore, as shown in Fig. 7, all methods exhibit an error floor. This phenomenon arises from two primary sources of error in the channel estimation process: (1) weak signals relative to noise, which occurs at low SNRs, and (2) the time-varying nature of the channels. The figure demonstrates that increasing SNR (i.e., receiving a stronger signal) improves estimation performance up to SNR $= 10$ dB with the online VB method. Beyond this point, however, further improvement is minimal because the channels continue to vary over time, making it difficult for the approximated statistical model of the channel at one time slot to remain accurate in the next time slot, even with stronger signals.

Finally, it is crucial to assess the computational complexity of our approach in comparison to the benchmarks. The computational complexity of LMMSE is $\mathcal{O}\left(MK^2 + |\mathcal{S}|K\right)$, while for both KF and EP methods, it is $\mathcal{O}\left(I_{\text{tr}}\left(M^3K^3 + M^2K^2 + |\mathcal{S}|K\right)\right)$ [40]. However, our online processing VB strategy has a computational complexity of $\mathcal{O}(I_{\text{tr}}(M^3K + |S|K))$. This complexity corresponds to finding a local optimum solution using the CAVI algorithm within the VB method [32]. As a result, our online processing strategy within the VB framework has lower computational complexity than the KF and EP methods and higher complexity than the LMMSE method.

B. Enhancing the Performance of the Online VB Strategy

As mentioned earlier in Section III, the online processing strategy involves two phases of prediction and estimation,

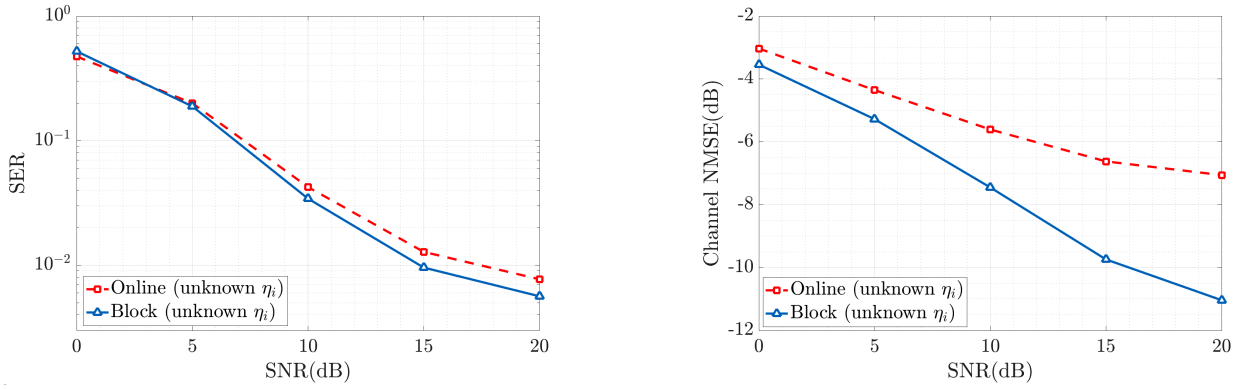

Fig. 9. The SER and channel NMSE performance of the proposed online and block processing strategies using the VB framework with 16QAM when $M = 64$, $\eta_i = 0.985, T_p = 8, T_d = 128$, and SNR $\in [0, 20]$ dB.

where the channel information at time $t-1$ serves as prior knowledge for time t. However, if the predicted channels at time $t-1$ fail to provide accurate prior knowledge for estimating the channel at time t, it can lead to error propagation in the estimation process. More precisely, by increasing T_p and T_d, although we expect to receive more signals and thus improve the SER performance, the predicted channels from the prediction phase become more and more deviated from the actual channel due to the time correlation between the channels. Therefore, increasing T_p and T_d alone cannot lead to performance improvement. To address this issue, we propose an interleaved structure that divides the entire communication time into L sections, each with its own pilot and data signals. This way, our VB-based method leverages the interleaved structure to stop the error propagation and retune the estimated parameters based on the new pilot signals. Here, we have:

$$\sum_{\ell=1}^{L} T_p^{[\ell]} = T_p, \quad \sum_{\ell=1}^{L} T_d^{[\ell]} = T_d, \tag{55}$$

where $T_p^{[\ell]}$ and $T_d^{[\ell]}$ represent the pilot and data communication times for the ℓ^{th} section, respectively.

In Fig. 8, we incorporate an interleaved structure with $L = 2$ sections into the online processing strategy and compare its performance with a scenario where the structure is not utilized. We employ the same assumptions as in Fig. 3. The results show that the interleaved structure enhances the SER compared to the non-interleaved scenario. This improvement arises because the interleaved structure prevents error propagation and enables the BS to adjust parameters using new pilot signals at the beginning of each interleaving section.

C. Online Strategy versus Block Processing Strategy

In Fig. 9, we compare the performance of the block processing strategy with the online processing strategy in terms of SER and channel NMSE. We assume $M = 64$ and $\eta_i = 0.985$ using the 16-quadrature amplitude modulation (16QAM) signal constellation. Fig. 9 depicts that the block processing strategy slightly enhances SER performance. Moreover, it significantly reduces channel NMSE, especially at high SNRs. These improvements stem from stacking the entire received signals together and processing them together. On the other hand, the block processing strategy increases delay and computational complexity compared to the online processing strategy. In particular, the complexity of the block processing VB method is expressed as $\mathcal{O}(I_{\text{tr}}\big[T(3M^2+M^3)K+M^3K+|\mathcal{S}|K\big])$, due to the matrix inversion in (42) and the matrix multiplications required to compute $\tau_{i,t}^{\eta}$, $\langle \eta_{i,t} \rangle$, and $\bar{b}_{i,t}$ in (46), (47), and (51), respectively. As we can see, this complexity exceeds that of the online VB method, which explains the trade-off between higher complexity and enhanced performance in the block processing strategy.

To highlight the applications of the online and block processing strategies, we note that the online strategy is useful in latency-sensitive scenarios, such as voice and video calls, as well as in environments with limited computational resources, such as IoT networks. In contrast, block processing is preferable when lower error rates are required (e.g., for control signaling communication in cellular networks) and sufficient computational power is available.

VI. Conclusion

In this study, we focused on the uplink scenario within a massive MIMO network. We introduced approaches based on VB inference to tackle the JED problem, particularly for high-mobility users experiencing time-varying channels. We developed two processing strategies using VB: (i) a low-complexity low-latency online processing strategy suitable for scenarios with unknown time correlation coefficients at the BS; (ii) a block processing strategy that further enhances performance by reducing channel NMSE. Moreover, we incorporated an interleaved structure into the online processing strategy to prevent error propagation during the JED process, thereby improving its effectiveness. We numerically compared the performance of the VB-based online strategy against three well-known benchmarks (LMMSE, KF, and EP) in terms of SER and channel NMSE. Our simulation results consistently demonstrated the superiority of our VB framework across the performance metrics. One potential direction for future exploration involves extending these findings to networks integrating reconfigurable intelligent surfaces (RISs) with massive MIMO systems, considering RIS's emerging role as a promising, low-complexity, and power-efficient solution in 5G and beyond networks [43], [44].

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful and constructive comments, which helped to improve the content and presentation of the paper.

REFERENCES

[1] S. Nassirpour and D. H. Nguyen, "Joint channel estimation and data detection for time-varying MIMO channels in UAV networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*. IEEE, 2024, pp. 1592–1597.

[2] E. A. Kadir, R. Shubair, S. K. A. Rahim, M. Himdi, M. R. Kamarudin, and S. L. Rosa, "B5G and 6G: Next generation wireless communications technologies, demand and challenges," in *Proc. Int. Congr. Adv. Technol. and Eng. (ICOTEN)*. IEEE, 2021, pp. 1–6.

[3] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, 2014.

[4] W. Tan, M. Matthaiou, S. Jin, and X. Li, "Spectral efficiency of DFT-based processing hybrid architectures in massive MIMO," *IEEE Wireless Commun. Letters*, vol. 6, no. 5, pp. 586–589, 2017.

[5] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Select. Topics in Signal Process.*, vol. 8, no. 5, pp. 742–758, 2014.

[6] A. Gupta, S. Nassirpour, M. Dunna, E. Patamasing, A. Vahid, and D. Bharadia, "GreenMO: Enabling virtualized, sustainable massive MIMO with a single RF chain," in *Proc. Annual Int. Conf. Mobile Comput. and Netw. (MobiCom)*, 2023, pp. 1–17.

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

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

[9] E. De Carvalho and D. T. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. IEEE Signal Process. Workshop on Signal Process. Advances in Wireless Commun.* IEEE, 1997, pp. 129–132.

[10] E. de Carvalho and D. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Proc. Asilomar Conf. Signals, Systems, and Comp.*, vol. 2. IEEE, 1997, pp. 1624–1628.

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

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

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

[14] X. Kuai, X. Yuan, W. Yan, H. Liu, and Y. J. Zhang, "Double-sparsity learning-based channel-and-signal estimation in massive MIMO with generalized spatial modulation," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2863–2877, 2020.

[15] S. Park, J. W. Choi, J.-Y. Seol, and B. Shim, "Expectation-maximization-based channel estimation for multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2397–2410, 2017.

[16] R. Chopra, C. R. Murthy, H. A. Suraweera, and E. G. Larsson, "Performance analysis of FDD massive MIMO systems under channel aging," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1094–1108, 2017.

[17] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1428–1442, 2016.

[18] K. T. Truong and R. W. Heath, "Effects of channel aging in massive MIMO systems," *J. Commun. and Netw.*, vol. 15, no. 4, pp. 338–351, 2013.

[19] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse Bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, 2018.

[20] A. Osinsky, A. Ivanov, and D. Yarotsky, "Bayesian approach to channel interpolation in massive MIMO receiver," *IEEE Commun. Letters*, vol. 24, no. 12, pp. 2751–2755, 2020.

[21] M. Lerch, "Experimental comparison of fast-fading channel interpolation methods for the LTE uplink," in *Proc. Int. Symp. ELMAR*. IEEE, 2015, pp. 5–8.

[22] W. Peng, M. Zou, and T. Jiang, "Channel prediction in time-varying massive MIMO environments," *IEEE Access*, vol. 5, pp. 23 938–23 946, 2017.

[23] S. Kashyap, C. Mollén, E. Björnson, and E. G. Larsson, "Performance analysis of (TDD) massive MIMO with Kalman channel prediction," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.* IEEE, 2017, pp. 3554–3558.

[24] S. Srivastava, A. Mishra, A. Rajoriya, A. K. Jagannatham, and G. Ascheid, "Quasi-static and time-selective channel estimation for block-sparse millimeter wave hybrid MIMO systems: Sparse Bayesian learning (SBL) based approaches," *IEEE Trans. Signal Process.*, vol. 67, no. 5, pp. 1251–1266, 2018.

[25] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, "Deep CNN-based channel estimation for mmWave massive MIMO systems," *IEEE J. Select. Topics in Signal Process.*, vol. 13, no. 5, pp. 989–1000, 2019.

[26] Y. Liao, Y. Hua, and Y. Cai, "Deep learning based channel estimation algorithm for fast time-varying MIMO-OFDM systems," *IEEE Commun. Letters*, vol. 24, no. 3, pp. 572–576, 2019.

[27] T.-K. Kim, Y.-S. Jeon, J. Li, N. Tavangaran, and H. V. Poor, "Semi-data-aided channel estimation for MIMO systems via reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4565–4579, 2022.

[28] M. Naraghi-Pour, M. Rashid, and C. Vargas-Rosales, "Semi-blind channel estimation and data detection for time-varying massive MIMO system," in *Proc. IEEE Int. Conf. Commun. (ICC)*. IEEE, 2021, pp. 1–6.

[29] T. P. Minka, "Expectation propagation for approximate Bayesian inference," *arXiv preprint arXiv:1301.2294*, 2013.

[30] Y. Qi and T. P. Minka, "Window-based expectation propagation for adaptive signal detection in flat-fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 348–355, 2007.

[31] 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, 2021.

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

[33] L. Wan, J. Zhu, E. Cheng, and Z. Xu, "Joint CFO, gridless channel estimation and data detection for underwater acoustic OFDM systems," *IEEE J. Oceanic Eng.*, vol. 47, no. 4, pp. 1215–1230, 2022.

[34] J. Zhu, C.-k. Wen, J. Tong, C. Xu, and S. Jin, "Grid-less variational Bayesian channel estimation for antenna array systems with low resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1549–1562, 2019.

[35] S. Bera, S. Chakraborty, D. Sen, and A. K. Dutta, "Iterative variational Bayesian inference based channel estimation for TWR mmWave systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10 330–10 344, 2023.

[36] Y. Liu, S. Zhang, F. Gao, J. Ma, and X. Wang, "Uplink-aided high mobility downlink channel estimation over massive MIMO-OTFS system," *IEEE J. on Select. Areas in Comm.*, vol. 38, no. 9, pp. 1994–2009, 2020.

[37] R. Chopra, C. R. Murthy, and H. A. Suraweera, "On the throughput of large MIMO beamforming systems with channel aging," *IEEE Signal Process. Letters*, vol. 23, no. 11, pp. 1523–1527, 2016.

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

[39] 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.

[40] 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.

[41] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "A variational Bayesian perspective on MIMO detection with low-resolution ADCs," in *Proc. Asilomar Conf. Signals, Systems, and Comp.* IEEE, 2022, pp. 22–26.

[42] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Letters*, vol. 5, no. 9, pp. 369–371, 2001.

[43] W. Mei and R. Zhang, "Multi-beam multi-hop routing for intelligent reflecting surfaces aided massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1897–1912, 2021.

[44] 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 Things J.*, vol. 11, no. 3, pp. 5301–5315, 2023.

Sajjad Nassirpour (Member, IEEE) received the B.Sc. degree in Electrical Engineering from Shiraz University, Shiraz, Iran, in 2011, his M.Sc. degree in Electrical Engineering from Iran University of Science and Technology, Tehran, Iran, in 2014, and his Ph.D. degree in Electrical Engineering from the University of Colorado Denver, CO, in 2023. He is currently a Post-doctoral Fellow at San Diego State University, CA. His research interests include wireless communications, signal processing, optimization, machine learning, and applications of coding theory in high-performance computer memory systems. He received the Best Paper Award at the IEEE Computing and Communication Workshop and Conference (CCWC) in 2020.

Toan-Van Nguyen (Member, IEEE) received the Ph.D. degree in Electronics and Computer Engineering from Hongik University, South Korea, in 2021. From September 2021 to August 2022, he was a Postdoctoral Researcher in the Department of Electrical and Computer Engineering at Utah State University, USA. He is currently a Postdoctoral Researcher in the Department of Electrical and Computer Engineering at San Diego State University, CA, USA. His current research focuses on signal processing for wireless communications.

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