Decentralized Message-Passing for Semi-Blind Channel Estimation in Cell-Free Systems Based on Bethe Free Energy Optimization

Zilu Zhao, Dirk Slock
Communication Systems Department, EURECOM, France
zilu.zhao@eurecom.fr, dirk.slock@eurecom.fr

***Abstract*—In this work, we investigate uplink communication in Semi-Blind Cell-Free (CF) Massive Multiple-Input Multiple-Output (MaMIMO) systems. One of the major challenges in CF MaMIMO systems is pilot contamination, where multiple user terminals (UTs) may use the same pilot sequence due to an imbalance between the number of UTs and the length of the pilot sequence. Semi-blind approaches have been proposed to address this issue, where access points (APs) jointly estimate both the channel and user data. This joint estimation leads to a bilinear problem. Channel estimation in bilinear systems with Gaussian input has been studied in prior work, with expectation propagation (EP)-based algorithms, such as variable-level (VL)-EP and hybrid expectation maximization (EM)-EP, being proposed. However, in this paper, the user data follows a categorical distribution. To develop a tractable algorithm that leverages the finite alphabet of the user data, we investigate the Bethe free energy (BFE) of the bilinear system and propose a message-passing algorithm by minimizing the BFE. The resulting algorithm combines variational Bayes (VB), belief propagation (BP), and EP.**

I. Introduction

In Cell-Free (CF) Massive Multiple-Input Multiple-Output (MaMIMO) systems, user terminals (UTs) are simultaneously served by all access points (APs) in a given region. A significant challenge in CF MaMIMO systems is pilot contamination, which occurs when the number of users exceeds the length of the pilot sequences. Consequently, the APs cannot estimate the channel solely based on the pilot sequences. To address this issue, semi-blind channel estimation is employed [1]. In Semi-Blind settings, APs jointly estimate the channel and user data based on received signals and limited pilot sequences.

In [2], it is shown that a Semi-Blind channel estimation problem can be transformed into a Blind estimation problem with an augmented channel prior. We adopt this technique in Section II to simplify the system model. Due to the bilinear relationship between the channel matrix and data sequences, Semi-Blind problems inevitably result in a bilinear system model.

The user data can be categorized as either continuous input or discrete input. A typical continuous input distribution is Gaussian, as discussed in [3], where the Majorization-Minimization (MM) algorithm is used for bilinear detection. In this paper, we primarily focus on finite alphabet data for improved accuracy, though the proposed algorithm can be easily adapted for Gaussian input.

The goal of our method is to jointly estimate channel parameters and data symbols in a cell-free semi-blind uplink network.

A. Prior Work

Bayesian estimation in semi-blind structures holds significant potential [1], but it also presents challenges due to high-dimensional and intractable integrals. Message-passing algorithms, particularly Expectation Propagation (EP) [4] and Belief Propagation (BP) [5], are widely used in Bayesian estimation. Both EP and BP assume a factored joint probability density function and simplify high-complexity global inference problems into manageable local inference tasks. EP further reduces complexity by approximating the factors of the joint pdf with simpler forms, such as Gaussian distributions.

Variable-Level EP (VL-EP) was introduced for Gaussian input data by combining Expectation-Maximization (EM) with EP [6]. To improve its convergence properties, hybrid EM-EP and loop-free EM-EP algorithms were proposed in [3]. However, these approaches are not designed to handle user symbols from finite alphabets.

1) Expectation Propagation for Gaussian Mixture Models: The bilinear combination of a Gaussian distribution (e.g., channel distribution) and a discrete distribution (e.g., input data distribution) leads to a Gaussian Mixture Model. To address the limitations of VL-EP, a distributed bilinear-EP algorithm was proposed in [7], which adopts a brute-force approach to inference over finite alphabets, avoiding high-dimensional computations by considering only one data symbol at a time. Inspired by [7] and [8], the authors of [2] proposed a simplified decentralized bilinear-EP algorithm.

2) Bethe Free Energy: The Bethe Free Energy (BFE) is another powerful tool for Bayesian inference. It represents the variational energy of a factored joint pdf under a specific trial distribution, whose form is determined by the factorization scheme of the joint pdf. It has been demonstrated [9] that various message-passing algorithms, such as EP and BP, can be derived by optimizing BFE under different constraints on the trial distribution.

Hybrid Vector Message Passing (HVMP) [10] was proposed based on BFE optimization, introducing a mean-field constraint for the bilinear factor. However, this method does not account for finite alphabets and entirely neglects the correlation between the channel and data.

B. Main Contributions

We propose a low-complexity algorithm for semi-blind channel and data estimation, leveraging a framework based on BFE-constrained optimization. To handle the posterior interference effectively, we introduce an auxiliary variable. Unlike [7] and [2], our method treats the entire data sequence of a single user as a single atomic variable. Additionally, we introduce mean-field assumptions in the belief factors containing delta functions to avoid non-analytical integrals, simplifying derivations and reducing computational complexity. Our approach also integrates seamlessly with the decentralized scheme from [2].

The proposed algorithm preserves the interference terms as in [2], while simplifying computations, similar to [10]. This balance between accuracy and complexity makes our approach a significant advancement in semi-blind channel and data estimation.

II. System Model

We examine the uplink cell-free semi-blind network containing K single-antenna user terminals (UTs) and L access points (APs). Each AP is equipped with M antennas. The received signals of the l-th AP is

$$\begin{bmatrix}\mathbf{Y}_{p,l} & \mathbf{Y}_l\end{bmatrix} = \mathbf{H}_l \begin{bmatrix}\mathbf{X}_p^\mathsf{T} & \mathbf{X}^\mathsf{T}\end{bmatrix} + \begin{bmatrix}\mathbf{V}_{p,l} & \mathbf{V}_l\end{bmatrix} \in \mathbb{C}^{M\times(P+T)}, \quad (1)$$

where the channel matrix $\mathbf{H}_l \in \mathbb{C}^{M\times K}$ comprises of independent columns. We use $\mathbf{h}_{lk}$ to denote the k-th column which follows $\mathcal{CN}(\mathbf{h}_{lk}|\mathbf{0}, \boldsymbol{\Xi}_{\mathbf{h}_{lk}})$. The matrix $\mathbf{X}_p$ represents the transmitted pilots. We assume that orthogonal pilots are used, i.e., difference columns of $\mathbf{X}_p$ are either the same or orthogonal. We further assume that each pilot sequence has length P and a total power of $P\sigma_x^2$. Similarly, $\mathbf{X}$ represents the data sequence. We denote the data sequence sent by the k-th UT as $\boldsymbol{x}_k$, which is the k-th column of $\mathbf{X}$. Moreover, we assume that each element of $\mathbf{X}$ follows an i.i.d. discrete distribution, i.e., the symbols in $\mathbf{X}$ are drawn from a constellation set $\mathcal{S}$ with power σ_x^2. We define $\boldsymbol{x}_k \sim p_{\boldsymbol{x}_k}(\boldsymbol{x}_k)$. We also assume additive white Gaussian noise $\mathbf{V}_{p,l}$ and $\mathbf{V}_l$, where each entry has a power of σ_v^2. For simplicity, we define $\mathbf{C}_{\boldsymbol{v}} = \sigma_v^2\mathbf{I}$.

Since the channels of different APs are independent and all the noise symbols are independent, the received signals $\mathbf{Y}_{p,l}$ and $\mathbf{Y}_l$ of different APs are conditionally independent given $\mathbf{X}$. This paper aims to estimate $\mathbf{X}$ and $\forall l, \mathbf{H}_l$ jointly.

A. Orthogonal Pilots

When orthogonal pilots are used, we correlate the received pilot signals $Y_{p,l}$ with the g-th pilot sequence $\widetilde{\boldsymbol{x}}_{p,g}$ (not to confuse with the pilot sequence of the g-th user) to obtain the correlated version of the received pilot signals $\widetilde{\boldsymbol{y}}_{p,lg}$:

$$\widetilde{\boldsymbol{y}}_{p,lg} = \mathbf{Y}_{p,l}\widetilde{\boldsymbol{x}}_{p,g}^* = P\sigma_x^2\mathbf{H}_{lG_g}\mathbf{1}_{|G_g|} + \widetilde{\boldsymbol{v}}_{p,lg}, \tag{2}$$

where we use G_g to denote the UTs groups using the g-th pilot sequence. The columns of $\mathbf{H}_{lG_g}$ are composed of the channel coefficients corresponding to the users using the g-th pilot, i.e., $\mathbf{h}_{l,k}$ is a column of $\mathbf{H}_{l,g}$ if $\boldsymbol{x}_{p,k} = \boldsymbol{x}_{p,g}$. We denote $\widetilde{\boldsymbol{v}}_{p,lg} = \mathbf{V}_l\widetilde{\boldsymbol{x}}_{p,g}^*$ which is the transformed noise following a distribution $\mathcal{CN}(\mathbf{0}, \sigma_x^2\sigma_v^2 P\mathbf{I}_M)$.

B. Factored Joint Distribution

We introduce an auxiliary variable $\mathbf{Z}_{lk} = \mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T}$ and its vectorization $\mathbf{z}_{lk} = \text{vec}(\mathbf{Z}_{lk})$. Therefore, the likelihood of $\mathbf{Z}_{lk}$ is captured by Dirac function $p(\mathbf{Z}_{lk}|\mathbf{h}_{lk}, \boldsymbol{x}_k) = \delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T})$. The joint probability density function (PDF) can be derived as

$$\begin{aligned} &p(\mathbf{Y}_p, \mathbf{Y}, \mathbf{Z}_{11}, \ldots, \mathbf{Z}_{LK}, \mathbf{H}_1, \ldots \mathbf{H}_L, \mathbf{X}) \\ &= \prod_l p(\mathbf{Y}_l|\mathbf{Z}_{l1}, \ldots, \mathbf{Z}_{lK}) \prod_l \prod_k p(\mathbf{Z}_{lk}|\mathbf{h}_{lk}, \boldsymbol{x}_k) \\ &\prod_l \prod_g p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{H}_{lg}) \prod_k p(\boldsymbol{x}_k). \end{aligned} \tag{3}$$

For simplicity, we define

$$\begin{aligned} &f_{\mathbf{z}_l}(\mathbf{z}_{l1}, \ldots, \mathbf{z}_{lK}) \propto p(\mathbf{Y}_l|\mathbf{Z}_{l1}, \ldots, \mathbf{Z}_{lK}) \\ &f_{\mathbf{h}_{lG_g}}(\mathbf{h}_{lG_g}) \propto p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{H}_{lg}) \\ &f_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = p(\boldsymbol{x}_k) \\ &f_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k) \propto p(\mathbf{Z}_{lk}|\mathbf{h}_{lk}, \boldsymbol{x}_k). \end{aligned} \tag{4}$$

The factorization given by (3) admits a factor graph [5]. We denote $\mathbb{F} = \{f_{\mathbf{z}_l}, f_{\mathbf{h}_{lG_g}}, f_{\boldsymbol{x}_k}, \delta_{lk}\}$ as the set of all factor nodes and $\mathbb{V} = \{\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k\}$ as the set of all variable nodes. From the joint pdf (3), the Semi-Blind system model (1) is simplified into a Blind model

$$\mathbf{Y}_l = \mathbf{H}_l\mathbf{X}^\mathsf{T} + \mathbf{V}_l, \tag{5}$$

with equivalent Gaussian channel prior $\mathbf{H}_l \sim \prod_g f_{\mathbf{h}_{lG_g}}(\mathbf{h}_{lG_g})$.

C. Notations

Throughout the context, we will use bold uppercase letters to denote matrices and bold lowercase letters to denote vectors. Furthermore, we use lowercase letters to denote the vectorization of uppercase letters. For example, $\mathbf{z}_{lk} = \text{vec}(\mathbf{Z}_{lk})$.

III. Bethe Free Energy Optimization Framework

Bethe free energy is the approximated variational free energy between the true probability (3) and a constrained Bethe approximation trial function. For a given factored pdf p, its trial pdf b is obtained by:

$$p(\boldsymbol{\theta}) \propto \prod_\alpha f_\alpha(\boldsymbol{\theta}_\alpha) \Rightarrow b(\boldsymbol{\theta}) = \frac{\prod_\alpha b_{f_\alpha}(\boldsymbol{\theta}_\alpha)}{b_{\theta_i}(\theta_i)^{|N_i|-1}}, \tag{6}$$

s.t.

$$\forall \alpha, \theta_i \in \boldsymbol{\theta}_\alpha \int b_{f_\alpha}(\boldsymbol{\theta}_\alpha) d\boldsymbol{\theta}_{\bar{i}} = b_{\theta_i}(\theta_i) \tag{7}$$

where $|N(i)|$ denotes the number of factors f_α that contain θ_i and $\boldsymbol{\theta}_{\bar{i}}$ denotes all the variables except θ_i.

With (6)-(7), the BFE can be obtained as

$$\text{BFE}{=}D[b(\boldsymbol{\theta})\| \prod_\alpha f_\alpha(\boldsymbol{\theta}_\alpha)] = D(b_{f_\alpha}\|f_\alpha) + (|N_i|-1)H(b_{\theta_i}), \tag{8}$$

where we define $D(b\|q) = \int b(\boldsymbol{\theta}) \ln \frac{b(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta}$, and $H(\cdot)$ as entropy. It is worth noticing that (8) only holds if the factorization (6) is loop-free and strict constraints (7) are applied. Otherwise, (8) is only an approximation.

A. Bethe Approximation with Constraints

Following [9], the BFE of (3) is:

$$\begin{aligned} \text{BFE} = &\sum_l D[b_{f_{\mathbf{z}_l}}(\mathbf{z}_{l1}, \ldots, \mathbf{z}_{lK})\|p(\mathbf{Y}_l|\mathbf{z}_{l1}, \ldots, \mathbf{z}_{lK})] \\ &+ \sum_{l,g} D[b_{f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lG_g})\|p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{h}_{lG_g})] + \sum_k D[b_{f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k)\|p(\boldsymbol{x}_k)] \\ &+ \sum_{l,k} D[b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k)\|\delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T})] + \sum_{l,k} H[b_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] \\ &+ \sum_{l,k} H[b_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] + \sum_k L \cdot H[b_{\boldsymbol{x}_k}(\boldsymbol{x}_k)]. \end{aligned} \tag{9}$$

where all the factor-level beliefs $b_{f_{\mathbf{z}_l}}$, $b_{\delta_{\mathbf{z}_{lk}}}$, $b_{f_{\mathbf{h}_{lG_g}}}$, $b_{f_{\boldsymbol{x}_k}}$, and variable-level beliefs $b_{\mathbf{h}_{lk}}$, $b_{\mathbf{z}_{lk}}$, $b_{\boldsymbol{x}_k}$ are proper distributions normalized to one. Furthermore, to make all these factors consistent, the variable-level beliefs must be the marginal distribution of the factor-level beliefs. For all $l \in [1, L]$, $k \in [1, K]$, the constraints for the $\boldsymbol{x}_k$ are

$$\int b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k) d\mathbf{z}_{lk} d\mathbf{h}_{lk} = b_{\boldsymbol{x}_k}(\boldsymbol{x}_k) \tag{10}$$

$$b_{f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) = b_{\boldsymbol{x}_k}(\boldsymbol{x}_k). \tag{11}$$

However, satisfying the strict constraints of $\mathbf{h}_{lk}$ and $\mathbf{z}_{lk}$ will lead to an intractable problem. Therefore, we relax the strict constraints to first and second-order moment constraints (specifically, mean and covariance constraints). W.l.o.g., we denote those sufficient statistics as $\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})$, $\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})$

$$\mathbb{E}_{b_{f_{\mathbf{z}_l}}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] = \mathbb{E}_{b_{\mathbf{z}_{lk}}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] \tag{12}$$

$$\mathbb{E}_{\delta_{lk}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] = \mathbb{E}_{b_{\mathbf{z}_{lk}}}[\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})] \tag{13}$$

$$\mathbb{E}_{b_{f_{\mathbf{h}_{lG_g}}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] = \mathbb{E}_{b_{\mathbf{h}_{lk}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] \tag{14}$$

$$\mathbb{E}_{b_{\delta_{lk}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] = \mathbb{E}_{b_{\mathbf{h}_{lk}}}[\phi_{\mathbf{h}_{lk}}(\mathbf{h}_{lk})] \tag{15}$$

Moreover, to make the further derivation tractable with finite input $\mathbf{X}$, we only consider the covariance constraints of elements within every size-M block $\forall t \in [1, T], [\mathbf{z}_{lk}]_{(t-1)M+1:tM}$.

B. Bethe Free Energy Optimization

The optimization criteria can be concluded by

$$\begin{aligned} &\min_b \text{ BFE} \\ &\text{s.t. } (10) \sim (15). \end{aligned} \tag{16}$$

We observe the term $D[b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k)\|\delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T})]$ in (9). Since we need to minimize the BFE, the posterior factor $b_{\delta_{lk}}$ must contain the factor $\delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T})$ to avoid infinity BFE value. In order to have an analytical algorithm, we use the following mean-field approximation for the joint belief $b_{\delta_{lk}}$:

$$b_{\delta_{lk}}(\mathbf{z}_{lk}, \mathbf{h}_{lk}, \boldsymbol{x}_k) = b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k)\delta(\mathbf{Z}_{lk} - \mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T}), \tag{17}$$

where the belief $b_{\delta_{\mathbf{h},lk}}$ and $b_{\delta_{\boldsymbol{x},lk}}$ are beliefs normalized to one. By using Lagrangian methods, we can obtain the following message-passing style system of equations along with (17):

$$b_{f_{\mathbf{z}_l}}(\mathbf{z}_{l1}, \dots, \mathbf{z}_{lK}) = p(\mathbf{Y}_l|\mathbf{z}_{l1}, \dots, \mathbf{z}_{lK}) \prod_k \mu_{\mathbf{z}_{lk};f_{\mathbf{z}_l}}(\mathbf{z}_{lk}) \tag{18}$$

$$b_{f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lG_g}) = p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{h}_{lG_g}) \prod_{k\in G_g} \mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk}) \tag{19}$$

$$b_{f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) = p(\boldsymbol{x}_k)\mu_{\boldsymbol{x}_k;f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) \tag{20}$$

$$b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) = \mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk}) e^{\int b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) \ln \mu_{\mathbf{z}_{lk};\delta_{lk}}(\text{vec}(\mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T}))d\boldsymbol{x}_k} \tag{21}$$

$$b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) = \mu_{\boldsymbol{x}_k;\delta_{lk}}(\boldsymbol{x}_k) e^{\int b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) \ln \mu_{\mathbf{z}_{lk};\delta_{lk}}(\text{vec}(\mathbf{h}_{lk}\boldsymbol{x}_k^\mathsf{T}))d\mathbf{h}_{lk}} \tag{22}$$

$$b_{\mathbf{z}_{lk}}(\mathbf{z}_{lk}) = \mu_{\mathbf{z}_{lk};f_{\mathbf{z}_l}}(\mathbf{z}_{lk})\mu_{\mathbf{z}_{lk};\delta_{lk}}(\mathbf{z}_{lk}) \tag{23}$$

$$b_{\mathbf{h}_{lk}}(\mathbf{h}_{lk}) = \mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk})\mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk}) \tag{24}$$

$$b_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = [\mu_{\boldsymbol{x}_k;f_{\boldsymbol{x}_k}}(\boldsymbol{x}_k) \prod_l \mu_{\boldsymbol{x}_k;\delta_{lk}}(\boldsymbol{x}_k)]^{1/L}, \tag{25}$$

The equations (17)$\sim$ (22) describes the factor level beliefs while (23)$\sim$(25) are variable level beliefs. For all $f \in \mathbb{F}, \boldsymbol{\theta} \in \mathbb{V}$, we interpret $\mu_{\boldsymbol{\theta};f}$ as the variable to factor message. Furthermore, we can define the factor to variable messages such that the following relation holds [11]

$$\forall f \in N(\boldsymbol{\theta}), \mu_{\boldsymbol{\theta};f}(\boldsymbol{\theta}) = \prod_{f' \in N(\boldsymbol{\theta})/\{f\}} \mu_{f';\boldsymbol{\theta}}(\boldsymbol{\theta}), \tag{26}$$

where $N(\boldsymbol{\theta})$ denotes the neighborhood around $\boldsymbol{\theta}$ in the corresponding factor graph. Thus, (25) can be rewritten into the message passing form

$$b_{\boldsymbol{x}_k}(\boldsymbol{x}_k) = \mu_{f_{\boldsymbol{x}_k};\boldsymbol{x}_k}(\boldsymbol{x}_k) \prod_l \mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k) \tag{27}$$

Since the sufficient statistics we consider here are first and second-order moments, the messages $\mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}$, $\mu_{\delta_{lk};\mathbf{h}_{lk}}$, $\mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}$ and $\mu_{\delta_{lk};\mathbf{z}_{lk}}$ are all (unnormalized) Gaussian distributions. Therefore, in the following, for all $f \in \mathbb{F}, \boldsymbol{\theta} \in \mathbb{V}$, we use $\mathbf{m}_{f;\boldsymbol{\theta}}$, $\mathbf{C}_{f;\boldsymbol{\theta}}$ to denote the mean and covariance of the factor-to-variable (normalized) message distributions $\mu_{f;\boldsymbol{\theta}}$. For convenience, we also denote the mean and covariance of the variable-to-factor (normalized) message $\mu_{\boldsymbol{\theta};f}$ as $\mathbf{m}_{\boldsymbol{\theta};f}$ and $\mathbf{C}_{\boldsymbol{\theta};f}$. We should note here that the factor-to-variable messages fully determine those variable-to-factor messages and beliefs.

Furthermore, since the second-order sufficient statistics of $\mathbf{z}_{lk}$ considered here only include the covariance between the elements within each block sub-vector $\forall t \in [1, T], [\mathbf{z}_{lk}]_{(t-1)M+1:tM}$, the covariance matrices $\mathbf{C}_{\delta_{lk};\mathbf{z}_{lk}} \in \mathbb{C}^{MT \times MT}$ and $\mathbf{C}_{f_{\mathbf{z}_l};\mathbf{z}_{lk}} \in \mathbb{C}^{MT \times MT}$ are block diagonal matrix with block size equals to M. For simplicity, for all block matrix $\mathbf{C}$, we use the notations $\{\mathbf{C}\}_{tt,M}$ to denote the t-th $M \times M$ block matrix on the diagonal of $\mathbf{C}$. Analogously, we use the notation $\{\mathbf{m}\}_{t,M}$ to denote the t-th block vector of size $M \times 1$ in $\mathbf{m}$, i.e., the subvector $[\mathbf{m}]_{(t-1)M+1:tM}$.

The computation of factor-level beliefs (18)$\sim$(20), are composed of two types of factors, the true factors given by the joint pdf model, e.g., $p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{h}_{lG_g})$, and variable-to-factor messages, e.g., $\mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}$. We will use the term "intrinsic" to denote the true factors and use "extrinsic" to denote the variable-to-factor messages. Those messages can be understood as the "rest" part of the approximated posteriors besides the true intrinsic. For example, if we look at (19), the extrinsic $\prod_{k\in G_g} \mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}$ can be interpreted as an approximation of $p(\mathbf{Y}_p, \mathbf{Y}, \mathbf{h}_{lG_g})/p(\widetilde{\boldsymbol{y}}_{p,lg}, \mathbf{h}_{lG_g})$, where $p(\mathbf{Y}_p, \mathbf{Y}, \mathbf{h}_{lG_g})$ is the marginalization result of (3).

Since the optimal point of the BFE can be purely represented by those factor-to-variable messages, we will focus on deriving the update of the factors-to-variable messages in the following context. Meanwhile, the update of all the variable-to-factor messages follows (26).

IV. Detailed Derivations

The messages are updated iteratively (update one message a time while considering the other messages to be known) by satisfying the constraints (10)$\sim$(15), which describe the consistencies between the factor-level beliefs $\forall f \in \mathbb{F}, b_f$ and variable-level beliefs $\forall \boldsymbol{\theta} \in \mathbb{V}, b_{\boldsymbol{\theta}}$. Each pair of the (marginalized) factor-level belief and variable-level belief constrained by (10)$\sim$(15) always has one message different. We will update that different message by considering the consistency constraints. Note, in this paper, we base our discussion on the factor-to-variable messages since the variable-to-factor message is entirely determined by the definition (26). We can consider the variable-to-factor messages as aliases of the corresponding factor-to-variable messages. For example, $\mu_{\mathbf{z}_{lk};\delta_{lk}}$ is considered as the same message as $\mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}$.

A. Update of $\mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}$

We first investigate the constraint between the beliefs $b_{f_{\mathbf{z}_l}}$ and $b_{\mathbf{z}_{lk}}$ given by (18) and (23). According to the constraint given by (12), we need to match the marginal mean and covariance matrix of $\mathbf{z}_{lk}$. Since the pdf $p(\boldsymbol{y}|\mathbf{z}_{l1}, \dots, \mathbf{z}_{lK})$ is a Gaussian pdf with block-diagonal covariance matrix, the belief $b_{f_{\mathbf{z}_l}}$ is also a Gaussian with block-diagonal covariance matrix. Because Gaussian pdf is fully determined by mean and covariance matrix, matching the moment of $\mathbf{z}_{lk}$ between (18) and (23) is equivalent to matching the entire distribution between the marginalized version of (18) $b_{f_{\mathbf{z}_l}}(\mathbf{z}_{lk})$ and the belief $b_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})$ given by (23). Therefore, by forcing the equality $b_{f_{\mathbf{z}_l}}(\mathbf{z}_{lk}) = b_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})$, the update equation for the message $\mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}$ can be obtained by Gaussian reproduction lemma [12]:

$$\mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}(\mathbf{z}_{lk}) = \mathcal{CN}(\mathbf{z}_{lk}|\boldsymbol{y}_l - \sum_{k'\neq k} \mathbf{m}_{\mathbf{z}_{lk'};f_{\mathbf{z}_l}}, \mathbf{C}_{\boldsymbol{v}} + \sum_{k'\neq k} \mathbf{C}_{\mathbf{z}_{lk'};f_{\mathbf{z}_l}}).$$

B. Update of $\mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}$

The consistency constraint between (19) and (24) is given by (14). A detailed derivation of the update equation can be found in [2]. The mean and covariance matrices of $\mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}(\mathbf{h}_{lk})$ are given by

$$\mathbf{C}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}} = \left(\boldsymbol{\Xi}_{\mathbf{h}_{lk}}^{-1} + \mathbf{C}_{\boldsymbol{v}+\mathbf{h}_{l\overline{k}}|\boldsymbol{y}}^{-1}\right)^{-1} \tag{28}$$

$$\mathbf{m}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}} = \mathbf{C}_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}} \mathbf{C}_{\boldsymbol{v}+\mathbf{h}_{l\overline{k}}|\boldsymbol{y}}^{-1} \mathbf{m}_{\boldsymbol{v}+\mathbf{h}_{l\overline{k}}|\boldsymbol{y}}, \tag{29}$$

where $\mathbf{C}_{\boldsymbol{v}+\mathbf{h}_{l\overline{k}}|\boldsymbol{y}}$ can be interpreted as the covariance matrix of the interference (estimated from observations $\boldsymbol{y}$ and prior knowledge) plus noise, and $\mathbf{m}_{\boldsymbol{v}+\mathbf{h}_{l\bar{k}}|\boldsymbol{y}}$ can be interpreted as the new observation with interference terms removed, i.e.,

$$\mathbf{C}_{\boldsymbol{v}+\mathbf{h}_{l\bar{k}}|\boldsymbol{y}} = \frac{\sigma_v^2}{\sigma_x^2 P}\mathbf{I} + \sum_{k' \in G_g/\{k\}} \mathbf{C}_{\mathbf{h}_{lk'}|\mathbf{Y}} \tag{30}$$

$$\mathbf{m}_{\boldsymbol{v}+\mathbf{h}_{l\bar{k}}|\boldsymbol{y}} = \frac{1}{\sigma_x^2 P}\widetilde{\boldsymbol{y}}_{p,lg} - \sum_{k' \in G_g/\{k\}} \mathbf{m}_{\mathbf{h}_{lk'}|\mathbf{Y}_d}, \tag{31}$$

where

$$\begin{aligned}\mathbf{C}_{\mathbf{h}_{lk}|\mathbf{Y}_d} &= (\boldsymbol{\Xi}_{\mathbf{h}_{lk}}^{-1} + \mathbf{C}_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}^{-1})^{-1} \\ \mathbf{m}_{\mathbf{h}_{lk}|\mathbf{Y}_d} &= \mathbf{C}_{\mathbf{h}_{lk}|\mathbf{Y}_d}\mathbf{C}_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}^{-1}\mathbf{m}_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}.\end{aligned} \tag{32}$$

C. Update of $\mu_{f_{\boldsymbol{x}_k};\boldsymbol{x}_k}$

The consistency constraint between (20) and (27) is given by (11). Thus, we can immediately get

$$\mu_{f_{\boldsymbol{x}_k};\boldsymbol{x}_k}(\boldsymbol{x}_k) = p(\boldsymbol{x}_k). \tag{33}$$

D. Update of $\mu_{\delta_{lk};\boldsymbol{x}_k}$

The update of the message $\mu_{\delta_{lk};\boldsymbol{x}_k}$ is obtained by satisfying the consistency constraint (10) between the beliefs (22) and (27). Following the definition of $b_{\delta_{lk}}$ in (17), we can immediately obtain $\mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k) = b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k)/\mu_{\boldsymbol{x}_k;\delta_{lk}}(\boldsymbol{x}_k)$. It can be seen from (21) that the belief $b_{\delta_{\mathbf{h},lk}}$ is Gaussian (more details in section IV-E. In fact, we can see $b_{\delta_{\mathbf{h},lk}} = b_{\mathbf{h}_{lk}}$.). Thus, the message $\mu_{\delta_{lk};\boldsymbol{x}_k}$ can be derived as

$$\mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k) \propto \prod_t \mathcal{CN}(x_{kt}|\widehat{m}_{\delta_{lk};x_{kt}}, \widehat{\tau}_{\delta_{lk};x_{kt}}), \tag{34}$$

with

$$\widehat{\tau}_{\delta_{lk};x_{kt}} = \text{tr}\left[\{\mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}}\}_{tt,M}^{-1}\boldsymbol{R}_{b_{\delta_{\mathbf{h},lk}}}\right]^{-1} \tag{35}$$

$$\widehat{m}_{\delta_{lk};x_{kt}} = \widehat{\tau}_{\delta_{lk};x_{kt}}\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}^{\mathsf{H}}\{\mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}}\}_{tt,M}^{-1}\{\mathbf{m}_{\mathbf{z}_{lk};\delta_{lk}}\}_{t,M}, \tag{36}$$

where $\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}$ and $\boldsymbol{R}_{b_{\delta_{\mathbf{h},lk}}} = \mathbf{C}_{b_{\delta_{\mathbf{h},lk}}} + \mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}^{\mathsf{H}}$ denote the mean and correlation matrix of the Gaussian pdf $b_{\delta_{\mathbf{h},lk}}$ calculated in section IV-E. Note here that the (normalized) message $\mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k)$ is a categorical distribution, and thus, the variables $\widehat{m}_{\delta_{lk};x_{kt}}, \widehat{\tau}_{\delta_{lk};x_{kt}}$ are just parameters for computing the message, they do not correspond to the mean and variance of the elements in $\boldsymbol{x}_k \sim \mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k)$. From this point, we can also update the belief by

$$b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) = \mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k)\mu_{\boldsymbol{x}_k;\delta_{lk}}(\boldsymbol{x}_k). \tag{37}$$

E. Update of $\mu_{\delta_{lk};\mathbf{h}_{lk}}$

The beliefs given by (21) and (24) should satisfy the consistency constraint (15). The exponential factor in (21) can be verified as an (unnormalized) Gaussian. Therefore, the belief $b_{\delta_{\mathbf{h},lk}}$ is Gaussian, and the outbound message is computed by $\mu_{\delta_{lk};\mathbf{h}_{lk}}(\mathbf{h}_{lk}) = b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk})/\mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk})$. The mean and covariance matrix of the Gaussian message $\mu_{\delta_{lk};\mathbf{h}_{lk}}$ are

$$\mathbf{C}_{\delta_{lk};\mathbf{h}_{lk}} = \left(\sum_t [\mathbf{r}_{b_{\delta_{\boldsymbol{x},lk}}}]_t \{\mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}}\}_{tt,M}^{-1}\right)^{-1}$$

$$\mathbf{m}_{\delta_{lk};\mathbf{h}_{l\bar{k}}} = \mathbf{C}_{\delta_{lk};\mathbf{h}_{lk}}\left(\sum_t [\mathbf{m}_{b_{\delta_{\boldsymbol{x},lk}}}]_t^* \{\mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}}\}_{tt,M}\{\mathbf{m}_{\mathbf{z}_{lk};\delta_{lk}}\}_{t,M}\right),$$

where $\mathbf{r}_{b_{\delta_{\boldsymbol{x},lk}}} = \mathbb{E}_{b_{\delta_{\boldsymbol{x},lk}}}[\boldsymbol{x}_k . \boldsymbol{x}_k^*]$, $\mathbf{m}_{b_{\delta_{\boldsymbol{x},lk}}} = \mathbb{E}_{b_{\delta_{\boldsymbol{x},lk}}}[\boldsymbol{x}_k]$ with "." denoting element-wise product.
Thus, we update the belief by

$$\begin{aligned} b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) &= \mathcal{CN}(\mathbf{h}_{lk}|\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}, \mathbf{C}_{b_{\delta_{\mathbf{h},lk}}}) \\ &= \mu_{\delta_{lk};\mathbf{h}_{lk}}(\mathbf{h}_{lk})\mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk}).\end{aligned} \tag{38}$$

F. Update of $\mu_{\delta_{lk};\mathbf{z}_{lk}}$

We examine the moments consistency between (17) and (23) based on (13). Note here the message $\mu_{\mathbf{z}_{lk};\delta_{lk}}$ is implicitly included in $b_{\delta_{lk}}$ due to the definition (21) and (21). Therefore, we will update $\mu_{\delta_{lk};\mathbf{z}_{lk}}$ to make (17) and (23) consistent. The update equation of $\mu_{\delta_{lk};\mathbf{z}_{lk}}$ can be derived to be

$$\mu_{\delta_{lk};\mathbf{z}_{lk}}(\mathbf{z}_{lk}) = \frac{\text{proj}[b_{\delta_{lk}}(\mathbf{z}_{lk})]}{\mu_{\mathbf{z}_{lk};\delta_{lk}}(\mathbf{z}_{lk})}, \tag{39}$$

where the operation $q(\mathbf{z}_{lk}) = \text{proj}[p(\mathbf{z}_{lk})]$ projects the distribution p to Gaussian family q with block covariance matrices, such that the sufficient statistics $\phi_{\mathbf{z}_{lk}}(\mathbf{z}_{lk})$ of p and q are the same. Thus, the message $\mu_{\delta_{lk};\mathbf{z}_{lk}}$ is updated by

$$\begin{aligned}\mu_{\delta_{lk};\mathbf{z}_{lk}}(\mathbf{z}_{lk}) &= \mathcal{CN}(\mathbf{z}_{lk}|\mathbf{m}_{\delta_{lk};\mathbf{z}_{lk}}, \mathbf{C}_{\delta_{lk};\mathbf{z}_{lk}}) \\ &= \frac{\mathcal{CN}(\mathbf{z}_{lk}|\mathbf{m}_{b_{\delta_{\mathbf{z},lk}}}, \mathbf{C}_{b_{\delta_{\mathbf{z},lk}}})}{\mathcal{CN}(\mathbf{z}_{lk}|\mathbf{m}_{\mathbf{z}_{lk};\delta_{lk}}, \mathbf{C}_{\mathbf{z}_{lk};\delta_{lk}})},\end{aligned} \tag{40}$$

with

$$\begin{aligned}\mathbf{m}_{b_{\delta_{\mathbf{z},lk}}} &= \mathbf{m}_{b_{\delta_{\boldsymbol{x},lk}}} \otimes \mathbf{m}_{b_{\delta_{\mathbf{h},lk}}} \\ \mathbf{C}_{b_{\delta_{\mathbf{z},lk}}} &= \text{diag}(\mathbf{r}_{b_{\delta_{\boldsymbol{x},lk}}}) \otimes \mathbf{C}_{b_{\delta_{\mathbf{h},lk}}} + \mathbf{C}_{b_{\delta_{\boldsymbol{x},lk}}} \otimes \mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}\mathbf{m}_{b_{\delta_{\mathbf{h},lk}}}^{\mathsf{H}},\end{aligned}$$

where $\mathbf{C}_{b_{\delta_{\boldsymbol{x},lk}}} = \mathbb{E}_{b_{\delta_{\boldsymbol{x},lk}}}[(\boldsymbol{x}_k - \mathbf{m}_{b_{\delta_{\boldsymbol{x},lk}}})(\boldsymbol{x}_k - \mathbf{m}_{b_{\delta_{\boldsymbol{x},lk}}})^{\mathsf{H}}]$ is the covariance of the belief $b_{\delta_{\boldsymbol{x},lk}}$. It can be verified that this covariance matrix is diagonal. According to (37), the belief $b_{\delta_{\boldsymbol{x},lk}}$ is entirely determined by the messages $\mu_{f_{\boldsymbol{x}_k};\boldsymbol{x}_k}$ and $\forall l, \mu_{\delta_{lk},\boldsymbol{x}_k}$, which are all independent according to (33) and (34). Therefore, the covariance matrix $\mathbf{C}_{b_{\delta_{\boldsymbol{x},lk}}}$ is a diagonal matrix.
Note that the belief distribution $b_{\delta_{lk}}(\mathbf{z}_{lk})$ is a Gaussian mixture model. Thus, the resulting covariance matrix $\mathbf{C}_{\delta_{lk};\mathbf{z}_{lk}}$ may not be positive semi-definite.
Define the eigenvalue matrix $\boldsymbol{\Lambda}_{\delta_{lk};\mathbf{z}_{lk}}$ and unitary eigenvector $\mathbf{U}_{\delta_{lk};\mathbf{z}_{lk}}$ such that $\mathbf{C}_{\delta_{lk};\mathbf{z}_{lk}} = \mathbf{U}_{\delta_{lk};\mathbf{z}_{lk}}\boldsymbol{\Lambda}_{\delta_{lk};\mathbf{z}_{lk}}\mathbf{U}_{\delta_{lk};\mathbf{z}_{lk}}^{\mathsf{H}}$. We propose the following correction: for all $\lambda \in \text{diag}[\boldsymbol{\Lambda}_{\delta_{lk};\mathbf{z}_{lk}}]$, we clip λ^{-1} to the range $[10^{-8}, 10^8]$. Since we are using the iterative algorithm to find the fixed point of the BFE, resetting the value will not change the final result.

V. Decentralization Method

Until this point, we have developed a distributed BFE-based message-passing algorithm since a CPU is needed to compute the messages $\mu_{\boldsymbol{x}_k;\delta_{lk}}$. These messages are only used to compute the beliefs $b_{\delta_{\boldsymbol{x},lk}}$ which are then used to update the messages $\mu_{\delta_{lk};\mathbf{h}_{lk}}$ and $\mu_{\delta_{lk};\mathbf{z}_{lk}}$. Based on this observation, we use the backhaul message-passing scheme (physical message passed from AP to AP) proposed in [2] to decentralize the computing of $b_{\delta_{\boldsymbol{x},lk}}$. We define the update rule of the backhaul message from AP l to AP l' to be

$$\nu_{l\to l'}(\boldsymbol{x}_k) = \mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k)\prod_{l'' \in N(l)/\{l'\}} \nu_{l''\to l}(\boldsymbol{x}_k), \tag{41}$$

where we exploit the notations and use $N(l)$ to denote the neighborhood of AP l in the AP network. At each AP, the approximated version of belief $b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k)$ is recovered by

$$\widehat{b}_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) = p(\boldsymbol{x}_k)\mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k)\prod_{l' \in N(l)} \nu_{l'\to l}(\boldsymbol{x}_k). \tag{42}$$

This approximated belief is exact when the two conditions hold: 1), the backhaul messages converge to the steady point; 2), the AP network is acyclic. Nevertheless, we will use (42) to replace the exact update in (37). A suggested update order is concluded in Algorithm 1.

Algorithm 1 Proposed Method in one iteration

Require: $\forall l, g, k, \boldsymbol{y}_{l,pg}, \boldsymbol{y}_l, p(\boldsymbol{x}_k), p(\mathbf{h}_{lk}), p(\boldsymbol{y}_l|\mathbf{z}_{l1}, \ldots, \mathbf{z}_{lK})$
Initialize: All the factor-to-variable messages and $b_{\delta_{\mathbf{h},lk}}$ s.t.: 1), If Gaussian, then zero mean and unit covariance matrices. 2), If categorical distribution, then uniform.
[$\forall l$, At AP l, execute the following loop]
repeat [$\forall l' \in N(l)k, g$]
    $\mu_{\mathbf{z}_{lk};f_{\mathbf{z}_l}}(\mathbf{z}_{lk}) = \mu_{\delta_{lk};\mathbf{z}_{lk}}(\mathbf{z}_{lk})$
    Update $\mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}$ based on Section IV-A
    $\mu_{\mathbf{h}_{lk};f_{\mathbf{h}_{lG_g}}}(\mathbf{h}_{lk}) = \mu_{\delta_{lk};\mathbf{h}_{lk}}(\mathbf{h}_{lk})$
    Update $\mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}$ based on Section IV-B
    $\mu_{f_{\boldsymbol{x}_k};\boldsymbol{x}_k}(\boldsymbol{x}) = p(\boldsymbol{x}_k)$ due to Section IV-C
    $\mu_{\mathbf{z}_{lk};\delta_{lk}}(\mathbf{z}_{lk}) = \mu_{f_{\mathbf{z}_l};\mathbf{z}_{lk}}(\mathbf{z}_{lk})$
    Update $\mu_{\delta_{lk};\boldsymbol{x}_k}$ based on Section IV-D
    $\nu_{l\to l'}(\boldsymbol{x}_k) = \mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k)\prod_{l''\in N(l)/\{l'\}}\nu_{l''\to l}(\boldsymbol{x}_k)$
    $b_{\delta_{\boldsymbol{x},lk}}(\boldsymbol{x}_k) = p(\boldsymbol{x}_k)\mu_{\delta_{lk};\boldsymbol{x}_k}(\boldsymbol{x}_k)\prod_{l'\in N(l)}\nu_{l'\to l}(\boldsymbol{x}_k)$
    $\mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk}) = \mu_{f_{\mathbf{h}_{lG_g}};\mathbf{h}_{lk}}(\mathbf{h}_{lk})$
    Update $\mu_{\delta_{lk};\mathbf{h}_{lk}}$ based on Section IV-E
    $b_{\delta_{\mathbf{h},lk}}(\mathbf{h}_{lk}) = \mu_{\delta_{lk};\mathbf{h}_{lk}}(\mathbf{h}_{lk})\mu_{\mathbf{h}_{lk};\delta_{lk}}(\mathbf{h}_{lk})$ based on (38)
    Update $\mu_{\delta_{lk};\mathbf{z}_{lk}}$ based on Section IV-F
until Convergence

VI. Simulation Results

In this section, we verify the algorithm using numerical simulations. We consider a $400m \times 400m$ area with $M = 16$ APs and $K = 8$ UTs. The APs are located at the coordinates $(\frac{400}{3}i, \frac{400}{3}j)$, where $i, j \in \{0, \ldots, 3\}$. The UTs are uniformly randomly distributed over this area. The fading model we use is [7],

$$\sigma_{l,k}^2[\text{dB}] = -30.5 - 36.7\log_{10}(d_{lk}), \tag{43}$$

where d_{lk} is the distance between AP l and UT k. All the neighboring APs within $\frac{400}{3}$ meters are connected and can exchange information of the estimated data symbols. Furthermore, as illustrated in Algorithm 1, a synchronized message-exchanging scheme is used. To induce pilot contamination, the pilot sequence length is set to $P = 4$. Furthermore, the pilots are randomly assigned to the users. We use 4-QAM constellation to generate the input symbols $\mathbf{X}^T$. To make the simulations fair, we use power control for each UT and ensure that the total received power from all the users is the same. In default setting, the transmitted data sequence spans a length of $L = 12$ and the default SNR is set to 19 dB.

We maintain consistent positions for all APs and UTs and conduct simulations across 50 unique scenarios with varying $\mathbf{H}$, $\mathbf{V}$, and user data. The metric for evaluating performance is channel normalized mean squared error (NMSE). The simulation results are concluded in Fig. 1. For comparison, we also plot the results of the EP-based decentralized algorithm [2] with the same input and results of VL-EP [6] with Gaussian input of the same power. In the Genie-Aided scenario, we assume the data to be known. The performance curve of our proposed method is worse than the EP-based decentralized method. However, EP-based decentralized method has a higher complexity, i.e., $O[(|\mathcal{A}|M^3 + |\mathcal{S}|)KT]$ at each AP, where $|\mathcal{S}|$ denotes the size of the $\mathcal{S}$ and $\mathcal{A} = \{x^2 | x \in \mathcal{S}\}$. Meanwhile, the proposed algorithm has a complexity of $O[(M^3 + |\mathcal{S}|)KT]$. To evaluate the robustness of our method against pilot contamination, we plot NMSE versus T in Fig. 2. The results show that the estimation error decreases significantly as T increases from 0 to 5. Notably, since the pilot length is 4, at least $T = 4$ data length is required to estimate the channel effectively.

VII. Conclusions

In this paper, we derive a low-complexity message-passing algorithm for semi-blind estimation based on BFE optimization. Simulations

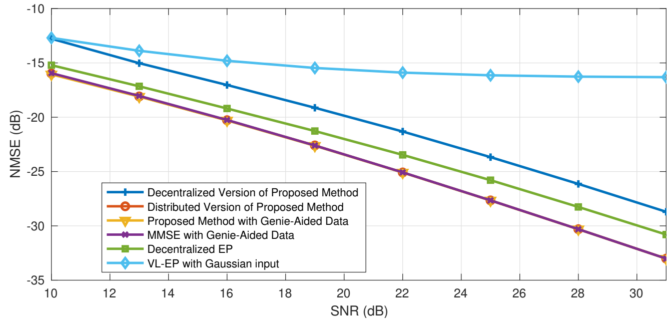

Fig. 1. NMSE vs SNR, Fixed Data Length $T = 12$

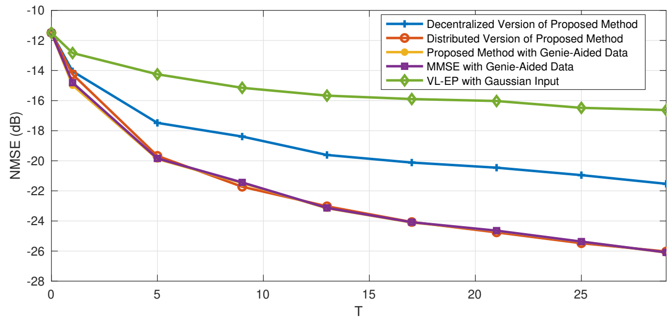

Fig. 2. NMSE vs T, Fixed SNR 19 dB

reveal that the distributed version of our algorithm achieves the same performance as MMSE, even with unknown data. Since the message updates depend solely on the belief $b_{\delta_{\boldsymbol{x},lk}}$, our algorithm integrates seamlessly into a decentralized scheme. However, the decentralized implementation experiences performance degradation due to loops in the backhaul network.

Acknowledgements EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, and by the Franco-German project CellFree6G.

References

[1] R. Gholami, L. Cottatellucci, and D. Slock, "Tackling Pilot Contamination in Cell-Free Massive MIMO by Joint Channel Estimation and Linear Multi-User Detection," in *IEEE International Symposium on Information Theory (ISIT)*, 2021.

[2] Z. Zhao and D. Slock, "Decentralized Expectation Propagation for Semi-Blind Channel Estimation in Cell-Free Networks," 2024. [Online]. Available: https://arxiv.org/abs/2410.01303

[3] ——, "Bilinear Hybrid Expectation Naximization and Expectation Propagation for Semi-Blind Channel Estimation," in *European Signal Processing Conference (EUSIPCO)*, 2024.

[4] T. P. Minka, "A Family of Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[5] 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.

[6] R. Gholami, L. Cottatellucci, and D. Slock, "Message Passing for a Bayesian Semi-Blind Approach to Cell-Free Massive MIMO," in *Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021.

[7] A. Karataev, C. Forsch, and L. Cottatellucci, "Bilinear Expectation Propagation for Distributed Semi-Blind Joint Channel Estimation and Data Detection in Cell-Free Massive MIMO," *IEEE Open Journal of Signal Processing*, 2024.

[8] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear Generalized Approximate Message Passing—Part I: Derivation," *IEEE Transactions on Signal Processing*, 2014.

[9] D. Zhang, X. Song, W. Wang, G. Fettweis, and X. Gao, "Unifying Message Passing Algorithms Under the Framework of Constrained Bethe Free Energy Minimization," *IEEE Transactions on Wireless Communications*, 2021.

[10] H. Jiang, X. Yuan, and Q. Guo, "Hybrid Vector Message Passing for Generalized Bilinear Factorization," *arXiv preprint arXiv:2401.03626*, 2024.

[11] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, "Approximate Inference Techniques with Expectation Constraints," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, 2005.

[12] Q. Zou, H. Zhang, C.-K. Wen, S. Jin, and R. Yu, "Concise Derivation for Generalized Approximate Message Passing Using Expectation Propagation," *IEEE Signal Processing Letters*, 2018.