Decentralized Joint Channel Estimation and Hierarchical Data Detection in Cell-Free Massive MIMO under Bethe Free Energy Optimization Framework

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

***Abstract*—Cell-Free (CF) Massive Multiple-Input Multiple-Output (MaMIMO) system is a promising architecture for next-generation networks. Without the cellular boundary, every User Equipment (UE) in a region is jointly served by all the Access Points (APs) scattered in the same region. In those CF scenarios, the number of users may exceed the length of pilot sequence. As a result, the information obtained from pilot is insufficient to treat the channel as deterministic. In Bayesian estimation, constrained Bethe Free Energy (BEF) optimization leads to a system of equations describing the approximate marginal posteriors. Although the system of equations cannot be solved analytically, a message-passing–like method can be used to solve it iteratively. If the constraints of the BFE optimization are strict consistency constraints, solving the optimization problem iteratively leads to Belief Propagation (BP), which is intractable due to the product of Gaussian Mixtures. On the other hand, with relaxed moments constraints, the BFE optimization leads to Expectation Propagation (EP). Although EP avoid the evaluation of product of Gaussian Mixtures, the iterative method will sometimes contain invalid operations (computing the moments of a function that has infinite integral value). In this paper, besides the strict consistency constraints, we also propose an additional constraint that leads to a Variational-Bayes-(VB)-like operation to avoid the intractable operations. Furthermore, we also exploit the hierarchical UE symbol structure to prevent the invalid operations.**

I. Introduction

Unlike conventional mobile networks, Cell-Free (CF) networks no longer have the cellular topology. As a result, all the user equipments (UEs) will be served by all the access points (APs) in an area. A huge throughput gain can be obtained by transforming into the new CF network topology [1]. One of the challenges introduced by CF network is pilot contamination, which occurs when the number of UEs exceeds the length of the pilot sequence. As a result, multiple UEs may use the same pilot sequence.

Semi-Blind approaches [2], [2], [3] have been explored to mitigate the effects of pilot contamination by jointly estimate the channel and user data. However, Bayesian inference is intractable due to the high dimensional integrals.

A variety of Message Passing based algorithms have been proposed to simplify the high dimensional Bayesian inference problem into local low dimensional inference problems [4]–[9].

Different iterative and message passing algorithms can be unified under the constrained Bethe Free Energy (BFE) minimization [10], [11]. BFE is the variational energy between Bethe approximation and the factored joint probability density function (PDF). A factor graph can be obtained by connecting all the factors and variables in the factored joint PDF if the variable is a parameter of the factor. If the underlying factor graph is cyclic, the BFE degrades to an approximation of the variational energy. It has been shown that Belief Propagation (BP) [12] can be obtained by optimizing BFE under the strict marginal consistency constraints, and Expectation Propagation (EP) can be derived based on BFE optimization with relaxed moment consistency constraints. Although BP simplifies the global inference problem by decomposing it to multiple local inference problems, it still contains intractable operations during joint channel estimation and data detection due to the evaluation of product of Gaussian mixtures. EP works around this problem by projecting the approximated marginal posteriors to Guassian family. However, since the messages are obtained by the PDF division, it may look like a Gaussian with "negative" variance which may lead to invalid operations such as computing the moments of a function that has infinite integral value. Actually, Gaussian messages with "negative" variances are common problems in message passing algorithms [13]–[19].

A. Main Contributions

In this paper, we propose an iterative joint channel estimation and data detection base on constrained BFE for uplink CF systems. Besides the strict consistency constraints, we also introduce a novel constraint to decompose the bilinear combination in the semi-blind estimation. Due to this constraint, our algorithm deviates from the traditional BP algorithm and contains a VB-like update. Despite the simplification, the augmentation may also lead to invalid operations like EP due to the PDF division. However, by exploiting the hierarchical data symbol structure, we can not only avoid messages with "negative" variances, but also lower the complexity.

II. System Model

Consider a semi-blind uplink model

$$\begin{bmatrix}\mathbf{Y}_p & \mathbf{Y}\end{bmatrix} = \boldsymbol{H}\begin{bmatrix}\mathbf{X}_p & \mathbf{X}\end{bmatrix} + \begin{bmatrix}\mathbf{V}_p & \mathbf{V},\end{bmatrix} \tag{1}$$

where $\mathbf{V}_p, \mathbf{V}$ denote the AWGN with power σ_v^2, $\boldsymbol{H} \in \mathbb{C}^{N\times K}$ is the channel matrix, $\mathbf{X}_p \in \mathbb{C}^{K\times P}$ is the transmitted pilot matrix, and $\mathbf{X} \in \mathbb{C}^{K\times L}$ is the transmitted data matrix. Each transmitted data symbol is assumed to be drawn from a finite alphabet. We define the symbol power to be $\mathbb{E}[|x_{kl}|^2] = \sigma_x^2$.

A. Hierarchical Data Symbol

In this paper, we assume that the data symbols can be decomposed into subsymbols:

$$x_{kt} = \sum_{m=1}^{M} x_{m,kt}, \tag{2}$$

where $|x_{m,kt}|^2 = \sigma_{xm}^2$. A common example of this constellation scheme is 4^M-QAM. If we assume the transmission power of 4^M-QAM to be σ_x^2, then we have

$$\sigma_{xm}^2 = \frac{3\,\sigma_x^2\,\frac{1}{4}^{m-1}}{4\left(1-\frac{1}{4}^M\right)}. \tag{3}$$

III. Orthogonal Pilot Sequence

In this paper, we consider orthogonal pilot sequences, i.e., every user uses a pilot sequence drawn from an orthogonal codebook. Therefore, the sequences of any two users are either orthogonal or identical.

Denote $\tilde{\boldsymbol{x}}_{p,g}$ as the g-th pilot sequence from the pilot codebook. We define $\widetilde{y}_{ng}$ as the correlation result of the received signal sequence of the n-th antenna with the g-th pilot sequence,

$$\widetilde{y}_{ng} \triangleq \frac{1}{P\sigma_x^2}\boldsymbol{y}_{p,n}^{\mathsf{H}}\widetilde{\boldsymbol{x}}_{p,g} = \sum_{k\in\mathcal{G}_g} h_{nk} + \widetilde{v}_{ng}, \tag{4}$$

where $\widetilde{v}_{ng}$ is white Gaussian noise with power $\sigma_{\widetilde{v}}^2 = \frac{\sigma_v^2}{P\sigma_x^2}$. Therefore, we have $\widetilde{v}_{ng} \sim \mathcal{N}(0, \sigma_{\widetilde{v}}^2)$.

IV. Factorization

Define $z_{mk,nl} = h_{nk}x_{m,kl}$, the probabilistic model of (1) is

$$\begin{aligned}
&p(\mathbf{Y}_p, \mathbf{Y}, \mathbf{Z}, \boldsymbol{H}, \mathbf{X}) \\
&\propto \prod_{n,g} f_{y_{nl}}(\forall m,k: z_{mk,nl}) \prod_{m,k,n,l} f_{\delta_{mk,nl}}(z_{mk,nl}, h_{nk}, x_{m,kl}) \\
&\cdot \prod_{n,g} f_{\mathbf{h}_{n\mathcal{G}_g}}(\mathbf{h}_{n\mathcal{G}_g}) \prod_{m,k,l} f_{x_{m,kl}}(x_{m,kl}),
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
&f_{y_{nl}}(\forall m,k: z_{mk,nl}) = p\left(y_{nl}|\sum_{m,k} z_{mk,nl}\right) \\
&f_{\delta_{mk,nl}}(z_{mk,nl}, h_{nk}, x_{m,kl}) \propto \delta(z_{mk,nl} - h_{nk}x_{m,kl}) \\
&f_{\mathbf{h}_{n\mathcal{G}_g}}(\mathbf{h}_{n\mathcal{G}_g}) \propto p(\widetilde{y}_{ng}, \mathbf{h}_{n\mathcal{G}_g}) \\
&f_{x_{m,kl}}(x_{m,kl}) = p(x_{m,kl}).
\end{aligned} \tag{6}$$

V. Bethe Free Energy of Bilinear System

We use strict consistency constraints. However, in addition to the consistency constraint, we introduce a structure constraint:

$$\begin{aligned}
&b_{\delta_{mk,nl}}(z_{mk,nl}, h_{nk}, x_{m,kl}) \\
&= b_{\delta^h_{ml,nk}}(h_{nk})b_{\delta^x_{mn,kl}}(x_{m,kl})\delta(z_{mk,nl} - h_{nk}x_{m,kl})
\end{aligned} \tag{7}$$

The BFE of (5) is

$$\begin{aligned}
BFE &= \sum_{n,l} D(b_{f_{y_{nl}}}\|f_{y_{nl}}) + \sum_{m,k,n,l} D(b_{\delta_{mk,nl}}\|f_{\delta_{mk,nl}}) \\
&+ \sum_{n,g} D(b_{f_{\mathbf{h}_{n\mathcal{G}_g}}}\|f_{\mathbf{h}_{n\mathcal{G}_g}}) + \sum_{m,k,l} D(b_{f_{x_{m,kl}}}\|f_{x_{m,kl}}) \\
&+ \sum_{m,k,n,l} H(b_{z_{mk,nl}}) + \sum_{n,k} ML\cdot H(b_{h_{nk}}) + \sum_{m,k,l} N\cdot H(b_{x_{m,kl}}),
\end{aligned} \tag{8}$$

where the factor beliefs ($b_{f_{y_{nl}}}$, $b_{\delta_{mk,nl}}$, $b_{f_{\mathbf{h}_{n\mathcal{G}_g}}}$, $b_{f_{x_{m,kl}}}$) has the parameters the same as (6) and the variable beliefs ($b_{z_{mk,nl}}$, $b_{h_{nk}}$, $b_{x_{m,kl}}$) has the parameters the same as their subscription. Besides the constraints that all the beliefs must be normalized to one, we also need the consistency constraints between factor beliefs and variable beliefs, i.e., the marginalized result of the factor beliefs must be the same as the variable beliefs.

VI. Simplification of BFE

The following relation follows consistency constraints immediately:

$$\begin{aligned}
&b_{\delta^x_{mk,nl}}(x_{m,kl}) = b_{x_{m,kl}}(x_{m,kl}) \\
&b_{\delta^h_{mk,nl}}(h_{nk}) = b_{h_{nk}}(h_{nk}) \\
&b_{f_{x_{m,kl}}}(x_{m,kl}) = b_{x_{m,kl}}(x_{m,kl}).
\end{aligned} \tag{9}$$

Therefore, the Lagrangian function is

$$\begin{aligned}
\mathcal{L} = BFE &+ \sum_{m,k,n,l}\int \lambda_{mk,nl}^{z;f_y}(z_{mk,nl})\left[b_{z_{mk,nl}}(z_{mk,nl})\right. \\
&\left.- b_{f_{y_{nl}}}(z_{mk,nl})\right] dz_{mk,nl} \\
&+ \sum_{m,k,n,l}\int \lambda_{mk,nl}^{z;\delta}(z_{mk,nl})\left[b_{z_{mk,nl}}(z_{mk,nl})\right. \\
&\left.- b_{\delta_{mk,nl}}(z_{mk,nl})\right] d_{z_{mk,nl}} \\
&+ \sum_{n,g}\sum_{k\in\mathcal{G}_g}\int \lambda_{nk}^{h;f_{\mathbf{h}}}(h_{nk})\left[b_{h_{nk}}(h_{nk}) - b_{f_{\mathbf{h}_{n\mathcal{G}_g}}}(h_{nk})\right] dh_{nk}
\end{aligned} \tag{10}$$

We set the derivative of the Lagrangian function w.r.t. each belief to zero,

$$\begin{aligned}
&\forall n,l: \\
&b_{f_{y_{nl}}}(\sum_{m,k} z_{mk,nl}) \propto f_{y_{nl}}(\sum_{m,k} z_{mk,nl}) \prod_{m,k} \mu_{mk,nl}^{z;f_y}(z_{mk,nl})
\end{aligned} \tag{11}$$

$$\forall n,g: \quad b_{f_{\mathbf{h}_{n\mathcal{G}_g}}}(\mathbf{h}_{n\mathcal{G}_g}) \propto f_{\mathbf{h}_{n\mathcal{G}_g}}(\mathbf{h}_{n\mathcal{G}_g}) \prod_{k\in\mathcal{G}_g} \mu_{nk}^{h;f_{\mathbf{h}}}(h_{nk}) \tag{12}$$

On the other hand, the variable level beliefs can be written as

$$\begin{aligned}
&\forall n,k: \\
&b_{z_{mk,nl}}(z_{mk,nl}) \propto \mu_{mk,nl}^{z;f_y}(z_{mk,nl})\mu_{mk,nl}^{z;\delta}(z_{mk,nl})
\end{aligned} \tag{13}$$

$$\begin{aligned}
&\forall n,k: \\
&\mu_{nk}^{h;f_{\mathbf{h}}}(h_{nk}) = \prod_{m,l} \exp\left(\mathbb{E}_{b_{x_{m,kl}}}\left[\ln \mu_{mk,nl}^{z;\delta}(h_{nk}x_{m,kl})\right]\right)
\end{aligned} \tag{14}$$

$$\begin{aligned}
\forall m,k,l: \quad & b_{x_{m,kl}}(x_{m,kl}) \propto f_{x_{m,kl}}(x_{m,kl}) \\
&\cdot \prod_{n} \exp\left(\mathbb{E}_{b_{h_{nk}}}\left[\ln \mu_{mk,nl}^{z;\delta}(h_{nk}x_{m,kl})\right]\right)
\end{aligned} \tag{15}$$

In (11)-(15), we define message-like functions $\mu_\alpha^\Omega(\theta) \propto \exp[\lambda_\alpha^\Omega(\theta)]$.

The constraints are

$$\begin{aligned}
&\forall m,k,n,l: \\
&b_{z_{mk,nl}}(z_{mk,nl}) = \int b_{f_{y_{nl}}}(\forall m,k: z_{mk,nl}) d\mathbf{z}_{\overline{mk},nl}
\end{aligned} \tag{16}$$

$$\forall m,k,n,l: \quad b_{\delta_{mk,nl}}(z_{mk,nl},h_{nk},x_{m,kl}) = \delta(z_{mk,nl}-h_{nk}x_{m,kl})b_{h_{nk}}(h_{nk})b_{x_{m,kl}}(x_{m,kl}) \tag{17}$$

and

$$\forall m,k,n,l: \quad b_{z_{mk,nl}}(z_{mk,nl}) = \int b_{\delta_{mk,nl}}(z_{mk,nl},h_{nk},x_{m,kl})dh_{nk}dx_{m,kl} \tag{18}$$

$$\forall n,g,k \in \mathcal{G}_g: \quad b_{h_{nk}}(h_{nk}) = \int b_{f_{\mathbf{h}_n\mathcal{G}_g}}(\mathbf{h}_{n\mathcal{G}_g}) \prod_{k'\in\mathcal{G}_g/\{k\}} dh_{nk'} \tag{19}$$

VII. Detailed Derivation

From the previous discussion, (11)-(19) form a system of equations. The solution to it is the inference result. An alternating optimization can be organized in the form of messages passing. The consistency constraint in (19) can be used to obtain $b_{h_{nk}}(h_{nk})$ by substituting (14) into (12). Therefore, we can see that every belief is uniquely determined if the messages and $b_{x_{m,kl}}$ and $b_{h_{nk}}$ are determined.
We consider the consistency constraints in an iterative manner. By considering the first constraint in (16) with (11) and (13), we can update $\mu^{z;\delta}_{mk,nl}$. By considering the consistency (18) with (13) and (17), we can update $\mu^{z;f_y}_{mk,nl}$.

A. Update of $\mu^{z;\delta}_{mk,nl}$

We consider the consistency (16) with (11) and (13). From (11), we marginalize out $(m',k') \neq (m,k)$:

$$\begin{aligned}&\forall n,l: \\ &b_{f_{y_{nl}}}(z_{mk,nl}) \propto \mu^{z;f_y}_{mk,nl}(z_{mk,nl}) \\ &\cdot \int f_{y_{nl}}(\forall m,k: z_{mk,nl}) \prod_{(m',k')\neq(m,k)} \mu^{z;f_y}_{m'k',nl}(z_{m'k',nl})dz_{m'k',nl}.\end{aligned} \tag{20}$$

Substitute (20) and (13) into (16), we have

$$\begin{aligned}\mu^{z;\delta}_{mk,nl}(z_{mk,nl}) &= \int f_{y_{nl}}(\forall m,k: z_{mk,nl}) \\ &\cdot \prod_{(m',k')\neq(m,k)} \mu^{z;f_y}_{m'k',nl}(z_{m'k',nl})dz_{m'k',nl}.\end{aligned} \tag{21}$$

We use central limit theory and approximate $\sum_{(m',k')\neq(m,k)} z_{m'k',nl}$ to Gaussian based on distribution $\mu^{z;f_y}_{m'k',nl}(z_{m'k',nl})$. Therefore, the message $\mu^{z;\delta}_{mk,nl}(z_{mk,nl})$ can be approximated to Gaussian

$$\mu^{z;\delta}_{mk,nl}(z_{mk,nl}) \propto \mathcal{CN}(z_{mk,nl}|m^{z;\delta}_{mk,nl},\tau^{z;\delta}_{mk,nl}), \tag{22}$$

where

$$\begin{aligned}m^{z;\delta}_{mk,nl} &= y_{nl} - \sum_{(m',k')\neq(m,k)} m^{z;f_y}_{m'k',nl} \\ \tau^{z;\delta}_{mk,nl} &= \sigma_v^2 + \sum_{(m',k')\neq(m,k)} \tau^{z;f_y}_{m'k',nl}.\end{aligned} \tag{23}$$

B. Update of $\mu^{h;f_\mathbf{h}}_{nk}(h_{nk})$

This message is directly updated by (14). Due to the previous discussion, we have Gaussian message $\mu^{z;\delta}_{mk,nl}(z_{mk,nl})$. Therefore, we have

$$\mu^{h;f_\mathbf{h}}_{nk}(h_{nk}) = \mathcal{CN}(h_{nk}|m^{h;f_\mathbf{h}}_{nk},\tau^{h;f_\mathbf{h}}_{nk}), \tag{24}$$

where

$$\begin{aligned}\tau^{h;f_\mathbf{h}}_{nk} &= \left(\sum_{m,l}\frac{\sigma^2_{x_m}}{\tau^{z;\delta}_{mk,nl}}\right)^{-1} \\ m^{h;f_\mathbf{h}}_{nk} &= \tau^{h;f_\mathbf{h}}_{nk}\left(\sum_{m,l}\frac{m^{z;\delta}_{mk,nl}(m^x_{m,kl})^*}{\tau^{z;\delta}_{mk,nl}}\right).\end{aligned} \tag{25}$$

C. Update of $b_{h_{nk}}$

In order to derive the update of $b_{h_{nk}}$ in (19), we substitute $b_{f_{\mathbf{h}_n\mathcal{G}_g}}$ computed in (12) into (19). For simplicity, we combine the message with prior and define

$$\breve{\tau}^h_{nk} = \left[\left(\tau^{h;f_\mathbf{h}}_{nk}\right)^{-1} + \sigma^{-2}_{h_{nk}}\right]^{-1}; \quad \breve{m}^h_{nk} = \breve{\tau}^h_{nk}\frac{m^{h;f_\mathbf{h}}_{nk}}{\tau^{h;f_\mathbf{h}}_{nk}}. \tag{26}$$

By Gaussian reproduction lemma, the mean and variance of $b_{h_{nk}}$ in (19) are

$$\tau^h_{nk}=\left[\left(\breve{\tau}^h_{nk}\right)^{-1}+\left(\tau^{h|\tilde{y}}_{nk}\right)^{-1}\right]^{-1}; \quad m^h_{nk} = \tau^h_{nk}\left[\frac{\breve{m}^h_{nk}}{\breve{\tau}^h_{nk}} + \frac{m^{h|\tilde{y}}_{nk}}{\tau^{h|\tilde{y}}_{nk}}\right], \tag{27}$$

where we define

$$m^{h|\tilde{y}}_{nk} = \tilde{y}_{ng} - \sum_{k'\neq k}\breve{m}^h_{nk'}; \quad \tau^{h|\tilde{y}}_{nk} = \sigma^2_{\tilde{v}} + \sum_{k'\neq k}\breve{\tau}^h_{nk'}. \tag{28}$$

For future discussion, we also define the autocorrelation of (19) as

$$\gamma^h_{nk} \triangleq \tau^h_{nk} + |m^h_{nk}|^2 \tag{29}$$

D. Update of $b_{x_{m,kl}}(x_{m,kl})$

From (15), the product factor can be interpreted as an approximated likelihood of $x_{m,kl}$. Therefore, if we denote

$$\begin{aligned}\tau^{\mathbf{Y}|x}_{m,kl} &= \left(\sum_n \frac{\gamma^h_{nk}}{\tau^{z;\delta}_{mk,nl}}\right)^{-1} \\ m^{\mathbf{Y}|x}_{m,kl} &= \tau^{\mathbf{Y}|x}_{m,kl}\left[\sum_n \frac{m^{z;\delta}_{mk,nl}(m^h_{nk})^*}{\tau^{z;\delta}_{mk,nl}}\right],\end{aligned} \tag{30}$$

then (15) can be written as

$$b_{x_{m,kl}}(x_{m,kl}) \propto f_{x_{m,kl}}(x_{m,kl})\mathcal{CN}\left(x|m^{\mathbf{Y}|x}_{m,kl},\tau^{\mathbf{Y}|x}_{m,kl}\right). \tag{31}$$

The mean of (15) is obtained as

$$m^x_{m,kl} = \mathbb{E}_{b_{x_{m,kl}}}[x]. \tag{32}$$

1) Discrete Symbols: Assume discrete input data:

$$f_{x_{m,kl}}(x_{m,kl}) = \sum_s p_s \delta(x_{m,kl} - x_s). \tag{33}$$

The mean can be obtained by

$$m^x_{m,kl} = \frac{1}{Z} \sum_s x_s p_s \mathcal{CN}\left(x_s | m^{\mathbf{Y}|x}_{m,kl}, \tau^{\mathbf{Y}|x}_{m,kl}\right), \tag{34}$$

where

$$Z = \sum_s p_s \mathcal{CN}\left(x_s | m^{\mathbf{Y}|x}_{m,kl}, \tau^{\mathbf{Y}|x}_{m,kl}\right). \tag{35}$$

To make the computation robust, we try to avoid unregularized exponential. Define the log unnormalized probability of data symbol $x_{m,kl}$ as

$$\forall s : \overline{p}^x_{m,kl}(x_s) = \log p_s + \log \mathcal{CN}\left(x_s | m^{\mathbf{Y}|x}_{m,kl}, \tau^{\mathbf{Y}|x}_{m,kl}\right), \tag{36}$$

the regularization constant can be obtained as

$$c_{m,kl} = \max_s \overline{p}^x_{m,kl}(x_s). \tag{37}$$

Finally, the normalized approximated data posterior can be obtained as

$$p^x_{m,kl}(x_s) = \frac{\exp[\overline{p}^x_{m,kl}(x_s) - c_{m,kl}]}{\sum_s \exp[\overline{p}^x_{m,kl}(x_s) - c_{m,kl}]}. \tag{38}$$

As a result the belief mean of (15) is

$$m^x_{m,kl} = \sum_s x_s p^x_{m,kl}(x_s). \tag{39}$$

E. Update of $\mu^{z;f_y}_{mk,nl}$

We consider the consistency constraint (18) between belief (13) and belief (17). The message is obtained by

$$\mu^{z;f_y}_{mk,nl}(z_{mk,nl}) \propto \frac{b_{\delta_{mk,nl}}(z_{mk,nl})}{\mu^{z;\delta}_{mk,nl}(z_{mk,nl})}. \tag{40}$$

The numerator is the marginal belief in (17), which can be expanded as

$$b_{\delta_{mk,nl}}(z_{mk,nl}) = \sum_s p^x_{m,kl}(x_s) \mathcal{CN}(z_{mk,nl} | x_s m^h_{nk}, \sigma^2_{x_m} \tau^h_{nk}). \tag{41}$$

We notice the following property for the message.

Proposition 0.1. *With discrete user data and Gaussian channel prior and noise, if $b_{h_{nk}}$ and $b_{x_{m,kl}}$ are obtained by using the same $\mu^{z;\delta}_{mk,nl}$ with $b_{z_{mk,nl}}$, then the message obtained by* (40) *is a Gaussian mixture.*

Proof. Observe (12), which can be rewritten into

$$b_{f_{\mathbf{h}_{n\mathcal{G}_g}}}(\mathbf{h}_{n\mathcal{G}_g}) \propto \mu^{h;f_{\mathbf{h}}}_{nk}(h_{nk}) \cdot f_{\mathbf{h}_{n\mathcal{G}_g}}(\mathbf{h}_{n\mathcal{G}_g}) \prod_{k' \in \mathcal{G}_g / \{k\}} \mu^{h;f_{\mathbf{h}}}_{nk'}(h_{nk'}). \tag{42}$$

On the other hand, according to (14), the message $\mu^{h;f_{\mathbf{h}}}_{nk}(h_{nk})$ can be written as

$$\mu^{h;f_{\mathbf{h}}}_{nk}(h_{nk}) \propto \mathcal{CN}\left(h_{nk} | \frac{m^{z;\delta}_{mk,nl}(m^x_{m,kl})^*}{\sigma^2_{x_m}}, \frac{\tau^{z;\delta}_{mk,nl}}{\sigma^2_{x_m}}\right). \tag{43}$$

From the discussion in Section VII-C, the second line in (42) can be identified as another Gaussian distribution. Since $b_{h_{nk}}$ is updated by computing (19), rewrite (19), and we have

$$\begin{aligned} b_{h_{nk}}(h_{nk}) &\propto \mu^{h;f_{\mathbf{h}}}_{nk}(h_{nk}) \\ &\cdot \int f_{\mathbf{h}_{n\mathcal{G}_g}}(\mathbf{h}_{n\mathcal{G}_g}) \prod_{k' \in \mathcal{G}_g / \{k\}} \mu^{h;f_{\mathbf{h}}}_{nk'}(h_{nk'}) dh_{nk'}, \end{aligned} \tag{44}$$

This observation implies

$$\tau^h_{nk} < \frac{\tau^{z;\delta}_{mk,nl}}{\sigma^2_{x_m}}. \tag{45}$$

Now we focus on the marginal belief $b_{\delta_{mk,nl}}(z_{mk,nl})$. By taking (41) and (45) into consideration, we have

$$\sigma^2_{x_m} \tau^h_{nk} < \tau^{z;\delta}_{mk,nl}. \tag{46}$$

Therefore, the division in (40) will lead to an unnormalized Gaussian mixture. □

We notice the following relation: if $\mathbf{B} - \mathbf{A}$ is positive definite, then

$$\frac{\mathcal{CN}(\boldsymbol{x}|\boldsymbol{a}, \mathbf{A})}{\mathcal{CN}(\boldsymbol{x}|\mathbf{b}, \mathbf{B})} = \frac{\det(\mathbf{B})^2}{\det(\mathbf{B} - \mathbf{A})^2} \frac{\mathcal{CN}(\boldsymbol{x}|\mathbf{c}, \mathbf{C})}{\mathcal{CN}(\mathbf{0}|\mathbf{b} - \boldsymbol{a}, \mathbf{B} - \mathbf{A})}. \tag{47}$$

Therefore, (40) can be obtained as

$$\begin{aligned} &\mu^{z;f_y}_{mk,nl}(z_{mk,nl}) \\ &= \sum_s w^{z;f_y}_{mk,nl}(x_s) \mathcal{CN}(z_{mk,nl} | m^{z;f_y|x_s}_{mk,nl}, \tau^{z;f_y|x_s}_{mk,nl}), \end{aligned} \tag{48}$$

where

$$\begin{aligned} \tau^{z;f_y|x_s}_{mk,nl} &= \left[\left(\sigma^2_{x_m}\tau^h_{nk}\right)^{-1} - \left(\tau^{z;\delta}_{mk,nl}\right)^{-1}\right]^{-1} \\ m^{z;f_y|x_s}_{mk,nl} &= \tau^{z;f_y|x_s}_{mk,nl}\left(\frac{x_s m^h_{nk}}{\sigma^2_{x_m}\tau^h_{nk}} - \frac{m^{z;\delta}_{mk,nl}}{\tau^{z;\delta}_{mk,nl}}\right), \end{aligned} \tag{49}$$

and the weighting factor $w^{z;f_y}_{mk,nl}(x_s)$

$$\begin{aligned} w^{z;f_y}_{mk,nl}(x_s) &\propto p^x_{m,kl}(x_s) \left(\frac{\tau^{z;\delta}_{mk,nl}}{\tau^{z;\delta}_{mk,nl} - \sigma^2_{x_m}\tau^h_{nk}}\right)^2 \cdot \\ &\cdot \left[\mathcal{CN}(0 | m^{z;\delta}_{mk,nl} - x_s m^h_{nk}, \tau^{z;\delta}_{mk,nl} - \sigma^2_{x_m}\tau^h_{nk})\right]^{-1} \end{aligned} \tag{50}$$

must be normalized to one, such that, $\sum_s w^{z;f_y}_{mk,nl}(x_s) = 1$. Therefore, $\mu^{z;f_y}_{mk,nl}(z_{mk,nl})$ in (48) can be identified as a probability density function.

From this point of view, the mean and vairance of $\mu^{z;f_y}_{mk,nl}(z_{mk,nl})$ used in Section VII-A are

$$\begin{aligned} m^{z;f_y}_{mk,nl} &= \sum_s w^{z;f_y}_{mk,nl}(x_s) m^{z;f_y|x_s}_{mk,nl} \\ \tau^{z;f_y}_{mk,nl} &= \sum_s \left[w^{z;f_y}_{mk,nl}(x_s)\left(\tau^{z;f_y|x_s}_{mk,nl} + |m^{z;f_y|x_s}_{mk,nl}|^2\right)\right] \\ &\quad - |m^{z;f_y}_{mk,nl}|^2. \end{aligned} \tag{51}$$

We conclude the algorithm in the following table.

Algorithm 1 Proposed VB-BiGaBP

Require: $\widetilde{y}_{ng}$, $\mathbf{Y}$, σ_{x_m}, $\mathcal{G}_g$, σ_v^2

1: **repeat** $[\forall m, k, n, l]$
2: Update $m_{mk,nl}^{z;\delta}$ and $\tau_{mk,nl}^{z;\delta}$ by (23)
3: Update $m_{nk}^{h;f_\mathbf{h}}$ and $\tau_{nk}^{h;f_\mathbf{h}}$ by (25)
4: Update $\breve{m}_{nk}^{h}$ and $\breve{\tau}_{nk}^{h}$ by (26)
5: Update $m_{nk}^{h|\widetilde{y}}$ and $\tau_{nk}^{h|\widetilde{y}}$ by (28)
6: Update m_{nk}^{h} and τ_{nk}^{h} by (27)
7: Update $m_{m,kl}^{\mathbf{Y}|x}$ and $\tau_{m,kl}^{\mathbf{Y}|x}$ by (30)
8: Update $m_{m,kl}^{x}$ by (34) and (35)
9: Update $m_{mk,nl}^{z;f_y|x_s}$ and $\tau_{mk,nl}^{z;f_y|x_s}$ by (49)
10: Update $w_{mk,nl}^{z;f_y}(x_s)$ by (50), s.t., $\sum_s w_{mk,nl}^{z;f_y}(x_s) = 1$
11: Update $m_{mk,nl}^{z;f_y}$ and $\tau_{mk,nl}^{z;f_y}$ by (51)
12: **until** Convergence
13: Output m_{nk}^{h} and $m_{m,kl}^{x}$

VIII. Decentralization Method

Until this point, we have developed a distributed BFE-based message-passing algorithm since a CPU is needed to compute $b_{x_{m,kl}}$. We use the backhaul message-passing scheme (physical message passed from AP to AP) proposed in [16] to decentralize the computing of $b_{x_{m,kl}}$. Assume single antenna APs, and we define auxiliary function based on (15):

$$\mu_{mk,nl}^{\delta;x}(x_{m,kl}) = \mathcal{CN}(x_{m,kl}|m_{m,kl}^{\delta;x}, \tau_{m,kl}^{\delta;x}), \tag{52}$$

$$\text{where } \tau_{m,kl}^{\delta;x} = \frac{\tau_{mk,nl}^{z;\delta}}{\gamma_{nk}^{h}}; \quad m_{m,kl}^{\delta;x} = \frac{m_{mk,nl}^{z;\delta}(m_{nk}^{h})^*}{\gamma_{nk}^{h}}. \tag{53}$$

We define the update rule of the physical backhaul message from AP n to AP n' to be

$$\nu_{m,kl}^{n\to n'}(x_{m,kl}) = \mu_{mk,nl}^{\delta;x}(x_{m,kl}) \prod_{n''\in N(n)/\{n'\}} \nu_{m,kl}^{n''\to n}(x_{m,kl}),$$

where we exploit the notations and use $N(n)$ to denote the neighborhood of AP n in the AP network. At each AP, the approximated version of belief $b_{x_{m,kl}}(x_{m,kl})$ is recovered by

$$\widehat{b}_{x_{m,kl}}(x_{m,kl}) = p(x_{m,kl})\mu_{mk,nl}^{\delta;x}(x_{m,kl}) \prod_{l'\in N(l)} \nu_{m,kl}^{n'\to n}(x_{m,kl}).$$

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.

IX. Simulation Results

In this section, we will verify the algorithm using numerical simulations. We consider a $400m \times 400m$ area with $M = 16$ APs and $K = 8$ UEs. The APs are located at the coordinates $(\frac{400}{3}i, \frac{400}{3}j)$, where $i, j \in \{0, \dots, 3\}$. The UEs are uniformly randomly distributed over this area. The fading model we use is [14] $\sigma_{l,k}^2[\text{dB}] = -30.5 - 36.7\log_{10}(d_{lk})$, where d_{lk} is the distance between AP l and UE 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 default pilot sequence length is $P = 6$. The transmitted data sequence spans a length of $L = 10$. We use 4-QAM constellation to generate the i.i.d. input symbols $\mathbf{X}^T$. We maintain consistent positions for all APs and UEs for different realizations and conduct simulations across 50 unique scenarios with varying $\boldsymbol{H}$, $\mathbf{V}$, and user data. The metric for evaluating performance is channel normalized mean squared error (NMSE). It is calculated as $\frac{\sum \text{tr}(\widetilde{\boldsymbol{H}}\widetilde{\boldsymbol{H}}^H)}{\sum \text{tr}(\boldsymbol{H}\boldsymbol{H}^H)}$, where $\widetilde{\boldsymbol{H}}$ represents the estimation error. The simulation results are concluded in Fig. 1. For comparison, we also plot the results of the VL-EP algorithm [20], which assumes Gaussian inputs, and the EP-based decentralized algorithm [16], which assumes discrete inputs. In the Genie-Aided scenario, we assume the data to be known.

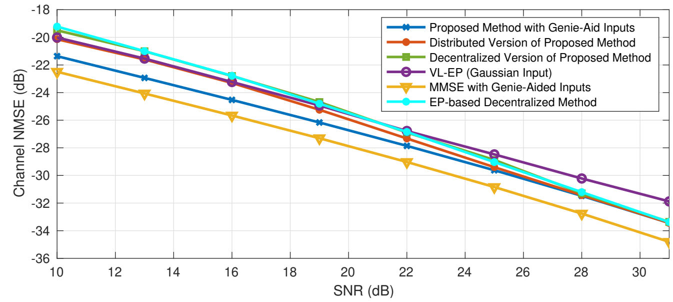

Fig. 1. Channel NMSE (dB) versus SNR (dB)

Due to the additional BFE constraints introduced in the proposed algorithm, VB-BiGaBP experiences a negligible performance loss compared to the EP-based algorithm in [16]. However, in the distributed version of the proposed algorithm, as illustrated in Algorithm 1, the computational complexity at each AP is $O(MKL)$. The CPU is responsible for executing lines 7 and 8 in Algorithm 1, resulting in a computational complexity of $O(NMKL)$.

In contrast, the EP-based algorithm in [16] has the same complexity as that in [14], with an AP-side complexity of $O(KL\cdot 4^M)$ and a CPU-side complexity of $O(NKL \cdot 4^M)$.

X. Conclusions

In this paper, we use constrained BFE optimization framework for joint channel estimation and data detection. The original BFE optimization leads to BP which contains intractable computation. We propose an additional customized constraint to work around the problem. However, the additional constraint may introduce invalid operations if iterative method is used. To cope with this problem, we exploit the hierarchical symbol structure. We show in our paper that with hierarchical structure, the messages always have a finite integral value, as well as the beliefs. Finally, due to the VB-like updates, only the beliefs need to be propagated to every AP. Therefore, we proposed a decentralized scheme, which is exact if every AP is connected in a tree network.

Acknowledgements EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, by the French PEPR-5G projects PERSEUS and YACARI, the EU H2030 project CONVERGE, and by a Huawei France funded Chair towards Future Wireless Networks.

References

[1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-Free Massive MIMO: Uniformly Great Service for Everyone," in *SPAWC*, 2015, pp. 201–205.

[2] 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 ISIT*, 2021.

[3] Z. Zhao and D. Slock, "Semi-Blind Sparse Channel Learning in Cell-Free Massive MIMO: A CRB Analysis," in *ICC*, 2023.

[4] ——, "Bethe Free Energy and Extrinsics in Approximate Message Passing," Asilomar Conference, 2023.

[5] Z. Zhao, F. Xiao, and D. Slock, "Vector Approximate Message Passing for Not-So-Large NIID Generalized I/O Linear Models," in *ICASSP*, 2024.

[6] ——, "Approximate Message Passing for Not-So-Large NIID Generalized Linear Models," in *SPAWC*, 2023.

[7] Z. Zhao and D. Slock, "Variance Predictions in VAMP/UAMP with Right Rotationally Invariant Measurement Matrices for NIID Generalized Linear Models," in *EUSIPCO*, 2023.

[8] ——, "Improved Variance Predictions in Approximate Message Passing," in *MLSP*, 2023.

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

[10] 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.

[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] M. J. Wainwright and M. I. Jordan, "Graphical Models, Exponential Families, and Variational Inference," *Foundations and Trends in Machine Learning*, 2008.

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

[14] 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.

[15] K.-H. Ngo, M. Guillaud, A. Decurninge, S. Yang, and P. Schniter, "Multi-User Detection Based on Expectation Propagation for the Non-Coherent SIMO Multiple Access Channel," *IEEE Transactions on Wireless Communications*, 2020.

[16] Z. Zhao and D. Slock, "Decentralized Expectation Propagation for Semi-Blind Channel Estimation in Cell-Free Networks," in *https://www.eurecom.fr/publication/7816*, 2024.

[17] S. Rangan, "Generalized Approximate Message Passing for Estimation with Random Linear Mixing," in *IEEE ISIT*, 2011, pp. 2168–2172.

[18] R. F. H. Fischer, C. Sippel, and N. Goertz, "VAMP with Vector-Valued Diagonalization," in *ICASSP*, 2020, pp. 9110–9114.

[19] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector Approximate Message Passing," *IEEE Transactions on Information Theory*, 2019.

[20] 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*, 2021.