Deliverable 4.1: Intermediate Results on Semi-blind Algorithms and Methods

Christian Forsch, Zilu Zhao, Dirk Slock, Laura Cottatellucci

January 30, 2026

Abstract

This deliverable reports the results achieved within Task 4.1 of the CellFree6G project, with a primary focus on semi-blind signal processing techniques for uplink cell-free massive multiple-input multiple-output (CF-MaMIMO) systems with single-antenna access points (APs). The objective of this task is to develop low-complexity Bayesian algorithms for joint channel estimation and data detection (JCD), while accounting for pilot contamination and scalability constraints anticipated in future 6G networks.

In the time-frame of Task 4.1, we formulated the semi-blind JCD problem as a bilinear Bayesian inference task and addressed it using an expectation propagation (EP)–based framework. A novel factorization of the joint posterior distribution of channel coefficients and transmitted data symbols is introduced, enabling efficient message-passing updates with polynomial computational complexity. The proposed bilinear-EP JCD algorithm allows for iterative refinement of channel and data estimates and effectively captures the coupling between estimation and detection.

The message-passing formulation naturally supports both centralized and distributed implementations. Consequently, the proposed methodology can be evaluated under either processing architecture and enables performance benchmarking in centralized settings while remains compatible with distributed realizations that are further considered in subsequent project activities. Simulation results demonstrate that the proposed approach significantly outperforms state-of-the-art EP-based and linear MMSE baselines in terms of symbol error rate and channel estimation accuracy, particularly in challenging pilot-contaminated scenarios.

Overall, the results presented in this deliverable establish a solid algorithmic foundation for semi-blind Bayesian processing in CF-MaMIMO systems and provide key building blocks for the distributed architectures addressed in later stages of the CellFree6G project.

This document is created as part of the "CellFree6G" project which focuses on reduced complexity techniques and performance analysis towards 6G cell-free massive MIMO. This project was funded by the DFG (German Research Foundation) for FAU and by the ANR for EURECOM – Project CellFree6G CO 1311/1-1, Project ID 491320625.

Contents

1 Introduction **3**

2 System Model **5**

3 Expectation Propagation on Graphs **5**

4 Bilinear-EP JCD [1] **7**
4.1 Factorization and Messages 7
4.2 Initialization and Scheduling 8
4.3 Update Rules 10
4.4 Inference 12
4.5 Algorithm Stability 12
4.6 Analysis of Fronthaul Load 12
4.7 Computational Complexity 13
4.8 Simulation Results 13
4.8.1 Setting 13
4.8.2 Performance evaluation and complexity analysis 13

5 Improving Performance under Pilot Contamination [2] **16**
5.1 Modified Bilinear-EP Algorithm 16
5.2 Quantifying Pilot Contamination 19
5.3 Extended Simulation Results 20

6 Further Performance Analysis **23**
6.1 Minimum User Distance Constraint 24
6.2 Reduced Fronthaul Load 25

7 Conclusion **27**

A Derivation of Factor-to-Variable Messages **27**

B References **30**

1 Introduction

Cell-free massive multiple-input multiple-output (CF-MaMIMO) networks enable primary goals for sixth generation (6G) wireless communication systems, such as ubiquitous coverage with uniform quality of service (QoS) and ultra-high-rate, energy-efficient data transmission [3–6]. In CF-MaMIMO systems, a large number of geographically distributed access points (APs) are jointly serving a much smaller number of user equipments (UEs). The joint processing is coordinated by one or multiple central processing units (CPUs) which are connected to the APs via fronthaul links. The geographically distributed nature of CF-MaMIMO networks enhances the attractive properties of centralized MaMIMO systems by reducing the average minimum distance between APs and UEs. This allows CF-MaMIMO networks to provide uniform high data rates over the coverage area and high energy efficiency. However, in contrast to centralized MaMIMO, channel hardening and favorable propagation do not generally hold [7–10]. Thus, the low-complexity matched filter, which provides near-optimal detection performance in MaMIMO systems [11], is not effective for CF-MaMIMO systems, and optimal joint signal processing at the centralized CPU is not computationally affordable. Similarly, existing pilot decontamination solutions for centralized MaMIMO [12–16] are not effective. Thus, the quest for low-complexity detection techniques with performance approaching that of centralized joint optimal processing schemes is still an open challenge as well as the search of effective pilot decontamination methods.

Distributed processing at the APs can efficiently reduce computational complexity at the CPU, see e.g., the extensive analysis in [17, 18] and [19], whereas semi-blind channel estimation has shown to combat effectively pilot contamination (PC) [14–16]. Therefore, in this work, we propose a distributed semi-blind joint channel estimation and data detection (JCD) algorithm based on expectation propagation (EP) which exhibits a low complexity.

EP is an approximate inference technique which iteratively finds a tractable approximation of factorized probability distributions by projecting each factor onto an exponential family [20, 21]. EP was already applied for data detection in previous works. In [22], the authors proposed a low-complexity EP-based MIMO detector while assuming perfect channel state information (CSI) and a Gaussian approximation for the posterior distribution of data symbols. In [23], the authors extended the EP-detection algorithm proposed in [22] to imperfect CSI by embedding channel estimation errors in the EP formulation.

The EP method was also applied to blind channel estimation and noncoherent detection. In [24], the authors presented a blind channel estimation algorithm for multi-cell MaMIMO systems. In [25], a noncoherent multi-user detection scheme was proposed for the single-input multiple-output (SIMO) multiple access channel (MAC). In both schemes, the approximated channel and symbol distributions were chosen to be multivariate Gaussian and multinoulli distributions, respectively. In [24], the proposed approximate joint posterior distribution of channels and transmitted symbols resulted in two *decoupled* EP-based schemes for channel estimation and data detection but, because of the high complexity of the EP-based symbol detection, only EP-based channel estimation was retained for practical system implementations followed by conventional linear symbol detectors. The decoupling between channel estimation and symbol detection implied that the EP-based channel estimation could only exploit prior statistical knowledge on the transmitted symbols but not their deterministic imperfect knowledge. In [25], the proposed approximate posterior distribution factorization yielded a factor tree with a branch per user. The detection scheme was derived by applying message-passing rules for EP on this tree. The resulting algorithm could be applied to both pilot-assisted and pilot-free communications. However, the complexity of the proposed algorithm was exponential in the product of the number of channel uses and the logarithm of the symbol constellation set cardinality and, hence, it was not affordable for practical high data rate systems.

The EP framework was also used to develop decentralized detection schemes for MaMIMO systems in [19, 26–28]. In these works, the computational complexity was reduced compared to centralized schemes by processing the signals received by antenna sub-arrays locally via EP message passing and then combining the messages from sub-arrays at the CPU. The posterior data symbol distributions were approximated by multivariate Gaussians and perfect CSI knowledge was

assumed. In [19], the authors reduced message sizes by utilizing averaging. In [26], the decentralized MaMIMO receiver embedded both detection and decoding and the extrinsic information from the decoder was utilized as a priori information for the EP-based detector. In [27], the authors introduced a pre-processing based on QR-factorization and a variance compensation scheme in the decentralized detector of [26]. In [28], two decentralized EP detection approaches were proposed, the first based on user grouping and, thus, group-wise joint detection, and the second on a daisy-chain architecture.

EP-based receivers were also applied in CF-MaMIMO systems. The authors of [29] utilized a centralized EP-based detector with Gaussian data approximations which incorporates channel estimation errors as in [23]. The channel estimation error accounted for PC and general estimation errors due to noise. In [30], the authors proposed a distributed EP detector for CF-MaMIMO based on the decentralized subarray-based detector in [19]. The aforementioned approach was extended to an iterative channel estimation and data detection (ICD) scheme in [31]. Here, the data detection was based on EP and the iterative algorithm took into account the channel estimation errors. The channel estimation was based on minimum mean squared error (MMSE) estimation with detected data symbols as additional pilots.

In this work, we propose a new distributed algorithm for JCD in CF-MaMIMO systems. Our contributions can be summarized as follows:

- We develop a novel message-passing algorithm for a bilinear inference problem arising in JCD. The inference is based on the approximate EP method and assumes general multivariate Gaussian and multinoulli distributions for the posterior distributions of the AP channels and data symbols, respectively, as in [25]. In contrast to the approximate posterior distributions of [24,25], the factorization of the proposed approximate posterior joint distribution for channels and data symbols allows for an alternating refinement of channel and data estimates and a distributed implementation of the algorithm with local processing at the APs. Furthermore, the inclusion of single-user soft-input soft-output (SISO) decoders is straightforward. Finally, the algorithm can automatically take into account the effect of PC.

- We show that the proposed bilinear-EP JCD algorithm has polynomial complexity in system parameters and bridges the complexity gap between algorithms based on EP that approximate the posterior distribution of data with a Gaussian distribution such as, e.g., [19, 22, 24, 26–32] and those that assume more precise categorical distributions [25].

- We consider four baseline schemes, namely a centralized linear MMSE detector, the detector in [30] assuming perfect CSI, the ICD algorithm in [31], and a modified version of the proposed JCD algorithm with perfect channel knowledge which provides a lower bound to the symbol error rate (SER). Our simulation results show the superior performance of our approach compared to the first three baseline algorithms.

It is worth noting that the appealing features of the proposed algorithm stem from the choice of the *factorization* of the approximate joint posterior distribution and the way to take into account the interplay between the two sets of variables, i.e., channel coefficients and data symbols. This method can be applied to other bilinear inference problems such as gain calibration, matrix factorization and compressed matrix sensing [33].

This delivarable is organized as follows. In Section 2, we introduce the CF-MaMIMO system model. In Section 3, we review the EP algorithm and its application on graphs. Based on this theoretical framework, we develop the bilinear-EP JCD algorithm in Section 4. Then, in Section 4.8, we present numerical results which show the superior performance of the proposed algorithm. Finally, some conclusions are drawn.

Notation: Lower case, bold lower case, and bold upper case letters, e.g., $x, \boldsymbol{x}, \boldsymbol{X}$, represent scalars, vectors, and matrices, respectively. $\boldsymbol{I}_N$ is the identity matrix of dimension N. $\{x_{l,k,t}\} = \bigcup_{l,k,t} x_{l,k,t}$ denotes the collection of all variables obtained by varying the possible indices. $\mathcal{A} \setminus \mathcal{B}$ is the set complement operator between two sets $\mathcal{A}, \mathcal{B}$. $\delta(\cdot)$ is the Dirac delta function. The indicator function $\mathbb{1}_{x \in \mathcal{S}}$ takes value 1 if the condition in the subscript is satisfied and zero otherwise, e.g.

element x is in the set $\mathcal{S}$. The operator $\text{vec}(\boldsymbol{X})$ maps the matrix $\boldsymbol{X}$ onto the vector obtained by stacking the columns of $\boldsymbol{X}$ on top of one another. $\otimes$ is the Kronecker product between two matrices. $E_{p(\boldsymbol{x})}\{\cdot\}$ stands for the expectation operator with respect to the probability distribution $p(\boldsymbol{x})$ and $\hat{E}\{\cdot\}$ stands for the empirical expectation operator. We denote by $\pi(\cdot)$ and $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{C})$ the categorical distribution and the circularly symmetric Gaussian distribution of complex-valued vectors $\boldsymbol{x}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$, respectively. The notation $x \sim p$ indicates that the random variable x follows the distribution p.

2 System Model

We consider a communication system in the uplink transmission that consists of L APs and K single-antenna UEs. Each AP is equipped with N co-located antennas. Each user sends a signal vector $\overline{\boldsymbol{x}}_k = [\boldsymbol{x}_{p;k}^T, \boldsymbol{x}_k^T]^T$ consisting of a P-dimensional pilot vector $\boldsymbol{x}_{p;k} \in \mathbb{C}^P$ and a T-dimensional data vector $\boldsymbol{x}_k \in \mathcal{S}^T$ where $\mathcal{S}$ is the employed constellation set. We combine pilot and data vectors of all users in the signal matrix $\overline{\boldsymbol{X}} = [\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, ..., \overline{\boldsymbol{x}}_K]^T \in \mathbb{C}^{K\times(P+T)}$. The matrix $\overline{\boldsymbol{X}}$ can also be expressed as $\overline{\boldsymbol{X}} = [\boldsymbol{X}_p, \boldsymbol{X}]$ where $\boldsymbol{X}_p$ and $\boldsymbol{X}$ are matrices consisting of the pilot and data vectors of all users, respectively. The equivalent complex baseband received signal at AP $l \in \{1, ..., L\}$ is given by

$$\overline{\boldsymbol{Y}}_l = [\boldsymbol{Y}_{p;l}, \boldsymbol{Y}_l] = \boldsymbol{H}_l[\boldsymbol{X}_p, \boldsymbol{X}] + \boldsymbol{N}_l, \tag{1}$$

where $\overline{\boldsymbol{Y}}_l \in \mathbb{C}^{N\times(P+T)}$ denotes a matrix whose element $\overline{y}_{l,i,j}$ is the received signal at antenna i of AP l during the j-th channel use. Here, $\boldsymbol{Y}_{p;l}$ and $\boldsymbol{Y}_l$ are the components of the matrix $\overline{\boldsymbol{Y}}_l$ corresponding to the pilot matrix $\boldsymbol{X}_p$ and the data matrix $\boldsymbol{X}$, respectively. $\boldsymbol{H}_l \in \mathbb{C}^{N\times K}$ is the block Rayleigh fading channel matrix whose element $h_{l,n,k}$ is the coefficient of the channel between user k and antenna n at AP l, which is assumed to be constant during $P+T$ consecutive channel uses, and $\boldsymbol{N}_l \in \mathbb{C}^{N\times(P+T)}$ is a matrix whose element $n_{l,i,j}$ is the additive white Gaussian noise (AWGN) at antenna i during the j-th channel use at the l-th AP. Therefore, both channel and additive noise are zero mean complex Gaussian vectors with covariance matrices $\boldsymbol{\Xi}_l \in \mathbb{C}^{NK\times NK}$ and $\sigma^2\boldsymbol{I}_{N(P+T)}$, respectively, i.e., $\text{vec}(\boldsymbol{H}_l) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Xi}_l)$ and $\text{vec}(\boldsymbol{N}_l) \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{N(P+T)})$. Furthermore, we assume that the channels of different users are uncorrelated. Thus, the covariance matrix $\boldsymbol{\Xi}_l$ is a block diagonal matrix of K blocks $\boldsymbol{\Xi}_{kl} \in \mathbb{C}^{N\times N}$.

3 Expectation Propagation on Graphs

In the following, we associate a factor graph to the factorized true distribution and describe a message-passing scheme that results from the local computation of the approximate factors on the factor graph nodes. More details on the derivation of the message-passing update rules can be found in [25, 34, 35].

We aim at computing an approximation of a joint probability distribution which is assumed to be factorizable as follows

$$p(\mathbf{x}) = \prod_{\alpha} \Psi_\alpha(\mathbf{x}_\alpha), \tag{2}$$

where $\mathbf{x}_\alpha$ is a sub-vector of $\mathbf{x}$, i.e., $\mathbf{x}_\alpha \subseteq \mathbf{x}$, and $\bigcup_\alpha \mathbf{x}_\alpha = \mathbf{x}$. The approximation $\hat{p}(\mathbf{x})$ of the joint distribution reflects the same factorized form,

$$\hat{p}(\mathbf{x}) = \frac{1}{Z} \cdot \prod_{\alpha} \hat{p}_\alpha(\mathbf{x}_\alpha), \tag{3}$$

where Z is a normalization constant and $\hat{p}$ is constrained to lie within the exponential family $\mathcal{F}$. Furthermore, we assume that the approximation can be fully factorized as follows

$$\hat{p}(\mathbf{x}) = \prod_{\beta} \hat{p}_\beta(\mathbf{x}_\beta), \tag{4}$$

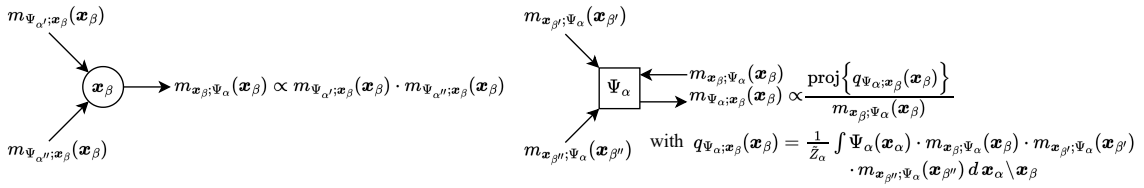


Figure 1: Illustration of the EP message-passing update rules.

where the sub-vectors $\mathbf{x}_\beta$ contain elements which always occur together in a factor. The sub-vectors $\mathbf{x}_\beta$ are not overlapping, i.e., $\mathbf{x}_\beta \cap \mathbf{x}_{\beta'} = \emptyset \quad \forall \beta \neq \beta'$. Furthermore, we have $\bigcup_\beta \boldsymbol{x}_\beta = \boldsymbol{x}$. Therefore, the sub-vectors $\boldsymbol{x}_\beta$ define a partition of $\boldsymbol{x}$. Additionally, it holds for all sub-vectors $\boldsymbol{x}_\beta$ and $\boldsymbol{x}_\alpha$ that the sub-vector $\boldsymbol{x}_\beta$ either lies fully within $\boldsymbol{x}_\alpha$, i.e., $\boldsymbol{x}_\beta \subseteq \boldsymbol{x}_\alpha$, or they are completely disjoint, i.e., $\boldsymbol{x}_\beta \cap \boldsymbol{x}_\alpha = \emptyset$.

The factorization described above induces a representation of the joint distribution by a factor graph with factor nodes associated to functions Ψ_α and variable nodes associated to sub-vectors $\mathbf{x}_\beta$. An edge exists between factor node Ψ_α and variable node $\mathbf{x}_\beta$ if $\boldsymbol{x}_\beta \subseteq \boldsymbol{x}_\alpha$. We define the neighbor set N_α of factor node Ψ_α as the set of indices β of all variable nodes $\mathbf{x}_\beta$ that are connected to Ψ_α, i.e., $N_\alpha = \{\beta \,|\, \mathbf{x}_\beta \subseteq \mathbf{x}_\alpha\}$. Similarly, we define the neighbor set N_β of variable node $\mathbf{x}_\beta$ as the set of indices α of all factor nodes Ψ_α that are connected to $\mathbf{x}_\beta$, i.e., $N_\beta = \{\alpha \,|\, \mathbf{x}_\beta \subseteq \mathbf{x}_\alpha\}$.

Since the approximation $\hat{p}$ lies within the exponential family $\mathcal{F}$, the approximate factors $\hat{p}_\alpha$ and $\hat{p}_\beta$ belong also to $\mathcal{F}$. The assumption of the fully-factorized approximation in (4) yields the following factorizations for the approximate factors,

$$\hat{p}_\alpha(\mathbf{x}_\alpha) = \prod_{\beta \in N_\alpha} m_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta), \tag{5}$$

$$\hat{p}_\beta(\mathbf{x}_\beta) = \frac{1}{Z_\beta} \prod_{\alpha \in N_\beta} m_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta), \tag{6}$$

where Z_β is a normalization constant and $m_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$ is interpreted as a message from the factor node Ψ_α to the variable node $\mathbf{x}_\beta$ also belonging to the specified exponential family. Thus, we can update the approximate factors, and therefore also the approximate joint distribution $\hat{p}(\mathbf{x})$, by exchanging and updating messages on a factor graph.

The EP algorithm is an iterative algorithm. Hence, we denote the messages during iteration i as $m^{(i)}$. At first, we initialize all the factor-to-variable messages $m^{(0)}_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$. For initialization, prior statistical knowledge or, if not available, uninformative distributions can be used. A particular example is discussed in section 4. Then, the variable-to-factor messages and the factor-to-variable messages are updated, respectively, as follows,

$$m^{(i-1)}_{\mathbf{x}_\beta;\Psi_\alpha}(\mathbf{x}_\beta) \propto \prod_{\alpha' \in N_\beta \backslash \alpha} m^{(i-1)}_{\Psi_{\alpha'};\mathbf{x}_\beta}(\mathbf{x}_\beta), \tag{7}$$

$$m^{(i)}_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta) \propto \frac{\text{proj}\left\{q^{(i-1)}_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)\right\}}{m^{(i-1)}_{\mathbf{x}_\beta;\Psi_\alpha}(\mathbf{x}_\beta)}, \tag{8}$$

where the distribution $q^{(i-1)}_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$ in (8) is given by

$$q^{(i-1)}_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta) = \frac{1}{\tilde{Z}_\alpha} \int \Psi_\alpha(\mathbf{x}_\alpha) \prod_{\beta' \in N_\alpha} m^{(i-1)}_{\mathbf{x}_{\beta'};\Psi_\alpha}(\mathbf{x}_{\beta'})\, \mathrm{d}\mathbf{x}_\alpha \backslash \mathbf{x}_\beta. \tag{9}$$

Here, $\tilde{Z}_\alpha$ is a normalization constant and proj$\{\cdot\}$ denotes the projection operator defined as

$$\text{proj}\{f(\mathbf{x})\} = \underset{g(\mathbf{x})\in\mathcal{F}}{\arg\min}\, D_{KL}(f(\mathbf{x})||g(\mathbf{x})), \tag{10}$$

where $D_{KL}(f(\mathbf{x})||g(\mathbf{x}))$ is the Kullback-Leibler (KL) divergence between f and g and $\mathcal{F}$ is an exponential family distribution. The message-passing update rules (7) and (8) are illustrated in Fig. 1.

4 Bilinear-EP JCD [1]

4.1 Factorization and Messages

In the following, we derive an EP-based algorithm for semi-blind JCD which we call bilinear-EP JCD. In this section, we focus on data signals since the received pilot signals are utilized to characterize the a priori channel distribution. Inspired by [25], we propose a novel factorization of the joint posterior distribution of data and channel. We decompose (1) in T vector relations, one for each channel use t, as follows

$$\boldsymbol{y}_{lt} = \sum_{k=1}^{K} x_{kt}\boldsymbol{h}_{lk} + \boldsymbol{n}_{lt}, \tag{11}$$

where $\boldsymbol{h}_{lk}$ is the k-th column, $\boldsymbol{y}_{lt}$ and $\boldsymbol{n}_{lt}$ are the t-th column and x_{kt} is the element in row k and column t of the matrices $\boldsymbol{H}_l$, $\boldsymbol{Y}_l$, $\boldsymbol{N}_l$ and $\boldsymbol{X}$ respectively. Furthermore, we define

$$\boldsymbol{z}_{lkt} := x_{kt}\boldsymbol{h}_{lk}. \tag{12}$$

We can factorize the posterior density function using Bayes theorem as follows

$$\begin{aligned} &p_{\{x_{kt}\},\{\boldsymbol{z}_{lkt}\},\{\boldsymbol{h}_{lk}\}|\{\boldsymbol{y}_{lt}\},\{\boldsymbol{Y}_{p;l}\}} \\ &\propto p_{\{x_{kt}\},\{\boldsymbol{z}_{lkt}\},\{\boldsymbol{h}_{lk}\},\{\boldsymbol{y}_{lt}\},\{\boldsymbol{Y}_{p;l}\}} \\ &= p_{\{\boldsymbol{y}_{lt}\}|\{\boldsymbol{z}_{lkt}\}} p_{\{\boldsymbol{z}_{lkt}\}|\{x_{kt}\},\{\boldsymbol{h}_{lk}\}} p_{\{x_{kt}\}} p_{\{\boldsymbol{h}_{lk}|\boldsymbol{Y}_{p;l}\}}, \end{aligned} \tag{13}$$

where the last equality follows from (11) and (12). We further assume that the a priori channel distributions are independent across users k and APs l. Then, exploiting the independence of the noise across time t and APs l, the independence of the data symbols $\{x_{kt}\}$, and (12), we can factorize (13) as follows

$$\begin{aligned} &p_{\{x_{kt}\},\{\boldsymbol{z}_{lkt}\},\{\boldsymbol{h}_{lk}\}|\{\boldsymbol{y}_{lt}\},\{\boldsymbol{Y}_{p;l}\}} \\ &\propto \prod_{l=1}^{L}\left[\prod_{t=1}^{T}\left(p_{\boldsymbol{y}_{lt}|\boldsymbol{z}_{l1t},\ldots,\boldsymbol{z}_{lKt}} \prod_{k=1}^{K} p_{\boldsymbol{z}_{lkt}|x_{kt},\boldsymbol{h}_{lk}}\right)\right] \\ &\times \prod_{k=1}^{K}\prod_{l=1}^{L} p_{\boldsymbol{h}_{lk}|\boldsymbol{Y}_{p;l}} \prod_{k=1}^{K}\prod_{t=1}^{T} p_{x_{kt}}, \end{aligned} \tag{14}$$

For the probability distributions in (14) we adopt the following compact notation

$$\begin{aligned} \Psi_{0,lt} &:= p_{\boldsymbol{y}_{lt}|\boldsymbol{z}_{l1t},\ldots,\boldsymbol{z}_{lKt}}, \\ \Psi_{1,lkt} &:= p_{\boldsymbol{z}_{lkt}|x_{kt},\boldsymbol{h}_{lk}}, \\ \Psi_{2,lk} &:= p_{\boldsymbol{h}_{lk}|\boldsymbol{Y}_{p;l}}, \\ \Psi_{3,kt} &:= p_{x_{kt}}. \end{aligned}$$

The factor graph induced by the factorization in (14) is illustrated in Fig. 2 and contains cycles. According to the system model in (1) the true factors are given by

$$\begin{aligned}\Psi_{0,lt} &= \mathcal{N}(\boldsymbol{y}_{lt}; \sum_{k=1}^{K} \boldsymbol{z}_{lkt}, \sigma^2 \boldsymbol{I}_N),\\ \Psi_{1,lkt} &= \delta(\boldsymbol{z}_{lkt} - x_{kt}\boldsymbol{h}_{lk}),\\ \Psi_{2,lk} &= \mathcal{N}(\boldsymbol{h}_{lk}, \boldsymbol{\mu}_{lk;\mathrm{MMSE}}, \boldsymbol{C}_{lk;\mathrm{MMSE}}),\\ \Psi_{3,kt} &= \frac{1}{|\mathcal{S}|}\mathbb{1}_{x_{kt}\in\mathcal{S}},\end{aligned}$$

where we assume uniform prior distributions $p_{x_{kt}}$ for data symbols and $\Psi_{2,lk}$ is given by the pilot-based Bayesian MMSE channel estimate of the channel $\boldsymbol{H}_l$ as follows

$$p_{\boldsymbol{H}_l|\boldsymbol{Y}_{p;l}} = \mathcal{N}\left(\mathrm{vec}(\boldsymbol{H}_l); \frac{1}{\sigma^2}\boldsymbol{C}\boldsymbol{A}_p^H \mathrm{vec}(\boldsymbol{Y}_{p;l}), \boldsymbol{C}\right)$$

with $\boldsymbol{C} := \left(\boldsymbol{\Xi}_l^{-1} + \frac{1}{\sigma^2}\boldsymbol{A}_p^H\boldsymbol{A}_p\right)^{-1}$ and $\boldsymbol{A}_p := (\boldsymbol{X}_p^T \otimes \boldsymbol{I}_N)$. Due to the independence of the channels among different users, we have $\boldsymbol{\mu}_{lk;\mathrm{MMSE}} = \left[\frac{1}{\sigma^2}\boldsymbol{C}\boldsymbol{A}_p^H\mathrm{vec}(\boldsymbol{Y}_{p;l})\right]_{(k-1)N+1:kN}$ and $\boldsymbol{C}_{lk;\mathrm{MMSE}} = \boldsymbol{C}_{(k-1)N+1:kN,(k-1)N+1:kN}$. We choose the approximate factor-to-variable messages to be distributed as

$$\begin{aligned}m_{\Psi_{0,lt};\boldsymbol{z}_{lkt}} &= \mathcal{N}(\boldsymbol{z}_{lkt}; \boldsymbol{\mu}_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}, \boldsymbol{C}_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}),\\ m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} &= \mathcal{N}(\boldsymbol{z}_{lkt}; \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}, \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}),\\ m_{\Psi_{1,lkt};x_{kt}} &= \pi_{1lkt}(x_{kt}),\\ m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} &= \mathcal{N}(\boldsymbol{h}_{lk}, \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}, \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}),\\ m_{\Psi_{2,lk};\boldsymbol{h}_{lk}} &= \mathcal{N}(\boldsymbol{h}_{lk}, \boldsymbol{\mu}_{\Psi_{2,lk};\boldsymbol{h}_{lk}}, \boldsymbol{C}_{\Psi_{2,lk};\boldsymbol{h}_{lk}}),\\ m_{\Psi_{3,kt};x_{kt}} &= \pi_{3kt}(x_{kt}).\end{aligned}$$

Please note that the dimensions of the parameters of all factor-to-variable messages are given by the type of distribution and the dimensions of the variable represented. The variable-to-factor messages are just multiplications of factor-to-variable messages and are used to compute the update of the factor-to-variable messages. Due to space limitation, we present the derivation only for some of the messages in the Appendix. Other messages that are not explicitly derived can be obtained by applying the rules presented in Section 3.

4.2 Initialization and Scheduling

All messages with the exception of the constant messages representing a priori knowledge are initialized as uninformative distributions, i.e., for Gaussian messages the diagonal entries of the covariance matrices are set to infinity and uniform probability mass functions (PMFs) are utilized for categorical messages. In our implementation, all Gaussian messages are parameterized by the precision matrix $\boldsymbol{\Lambda} = \boldsymbol{C}^{-1}$ and the mean vector $\boldsymbol{\gamma} = \boldsymbol{\Lambda}\boldsymbol{\mu}$. Uninformative Gaussian messages are initialized by all-zero precision matrices.

The scheduling is given in Algorithm 1. The first message to be computed in each iteration is $m_{\Psi_{1,lt};\boldsymbol{z}_{lkt}}$. In the first iteration, it is computed based on the CSI from the Bayesian MMSE estimation together with the constant data prior to generate a first estimate of z_{lkt}. During this step, the initialized uninformed message $m_{\Psi_{1,lt};\boldsymbol{z}_{lkt}}$ is updated to contain information. Note that the messages $m_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \equiv \Psi_{2,lk}$ and $m_{\Psi_{3,kt};x_{kt}} \equiv \Psi_{3,kt}$ represent the a priori distributions on data and the Bayesian MMSE channel estimates, respectively, and are not modified by the message passing algorithm. Hence, they are not included in the scheduling.

For our simulations, we used *parallel* scheduling, i.e., the updates of the messages in the innermost for-loops are done independently suitable for efficient implementations.

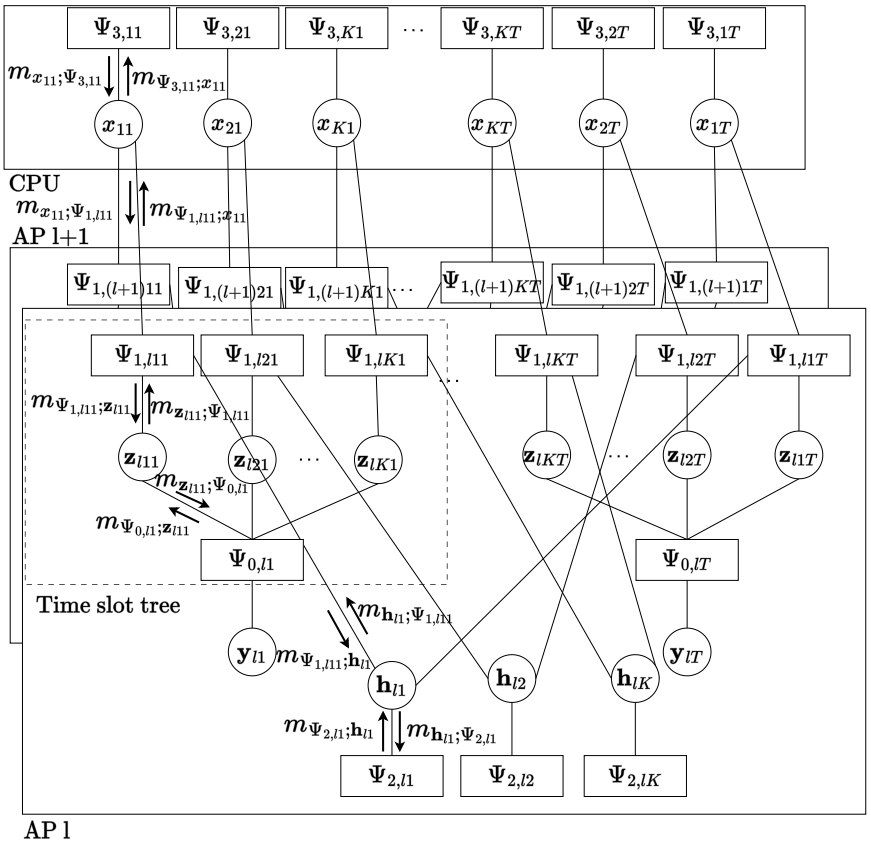


Figure 2: Factor graph of bilinear-EP JCD for distributed CF-MaMIMO.

Algorithm 1 bilinear-EP JCD Scheduling

Require: Initialized Messages
  for number of iterations I **do**
    for all APs l, UEs k and symbol times t **do**
      Update $m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$
    end for
    for all APs l, UEs k and symbol times t **do**
      Update $m_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}$
    end for
    for all APs l, UEs k and symbol times t **do**
      Update $m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}$
    end for
    for all APs l, UEs k and symbol times t **do**
      Update $m_{\Psi_{1,lkt};x_{kt}}$
    end for
  end for

4.3 Update Rules

In this section, we provide expressions to update the messages that appear in Algorithm 1. According to (8), the inputs to update a factor-to-variable message are always variable-to-factor messages. Each variable-to-factor message is computed by applying (7). As an example, we derive the variable-to-factor message forwarded from variable node $\boldsymbol{h}_{lk}$ to factor node $\Psi_{1,lkt}$ as follows

$$\begin{aligned} m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} &= \mathcal{N}(\boldsymbol{h}_{lk}; \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}, \boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}) \\ &\propto m_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \prod_{\tilde{t}\neq t} m_{\Psi_{1,lk\tilde{t}};\boldsymbol{h}_{lk}}. \end{aligned}$$

Since $m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}$ is a Gaussian distribution, its computation reduces to determine its covariance and mean. By applying the Gaussian multiplication lemma [25], we obtain

$$\boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} = \left(\boldsymbol{C}^{-1}_{\Psi_{2,lk};\boldsymbol{h}_{lk}} + \sum_{\tilde{t}\neq t} \boldsymbol{C}^{-1}_{\Psi_{1,lk\tilde{t}};\boldsymbol{h}_{lk}} \right)^{-1}, \tag{15}$$

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} = \boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \Bigg(& \boldsymbol{C}^{-1}_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \boldsymbol{\mu}_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \\ & + \sum_{\tilde{t}\neq t} \boldsymbol{C}^{-1}_{\Psi_{1,lk\tilde{t}};\boldsymbol{h}_{lk}} \boldsymbol{\mu}_{\Psi_{1,lk\tilde{t}};\boldsymbol{h}_{lk}} \Bigg). \end{aligned} \tag{16}$$

All other variable-to-factor messages can be computed analogously.
In the following we provide the rules to update factor-to-variable messages based on the incoming variable-to-factor messages.
Update of $m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} = \mathcal{N}(\boldsymbol{z}_{lkt}; \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}, \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}})$. The covariance and mean are given by

$$\boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} = \left(\boldsymbol{\Sigma}^{-1}_{lkt} - \boldsymbol{C}^{-1}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \right)^{-1}, \tag{17}$$

$$\begin{aligned} \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} = \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} \Big(& \boldsymbol{\Sigma}^{-1}_{lkt} \hat{\boldsymbol{z}}_{lkt} \\ & - \boldsymbol{C}^{-1}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \Big), \end{aligned} \tag{18}$$

where

$$\hat{\boldsymbol{z}}_{lkt} = \sum_{x'_{kt}\in\mathcal{S}} \omega(x'_{kt}) \tilde{\boldsymbol{\mu}}(x'_{kt}), \tag{19}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{lkt} = & \sum_{x'_{kt}\in\mathcal{S}} \omega(x'_{kt}) \left(\tilde{\boldsymbol{\mu}}(x'_{kt}) \tilde{\boldsymbol{\mu}}(x'_{kt})^H + \tilde{\boldsymbol{C}}(x'_{kt}) \right) \\ & - \hat{\boldsymbol{z}}_{lkt} \hat{\boldsymbol{z}}^H_{lkt}, \end{aligned} \tag{20}$$

$$\omega(x_{kt}) = \frac{\tilde{\omega}(x_{kt})}{\sum_{x'_{kt}\in\mathcal{S}} \tilde{\omega}(x'_{kt})}, \tag{21}$$

with $\tilde{\omega}(x_{kt})$ defined in (37). Finally,

$$\tilde{\boldsymbol{C}}(x_{kt}) = \left(\boldsymbol{C}^{-1}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} + \frac{\boldsymbol{C}^{-1}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}}{|x_{kt}|^2} \right)^{-1}, \tag{22}$$

$$\begin{aligned} \tilde{\boldsymbol{\mu}}(x_{kt}) = \tilde{\boldsymbol{C}}(x_{kt}) \Bigg(& \boldsymbol{C}^{-1}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \\ & + \frac{\boldsymbol{C}^{-1}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}}{|x_{kt}|^2} x_{kt} \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \Bigg). \end{aligned} \tag{23}$$

Update of $m_{\Psi_{0,lt};\boldsymbol{z}_{lkt}} = \mathcal{N}(\boldsymbol{z}_{lkt}; \boldsymbol{\mu}_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}, \boldsymbol{C}_{\Psi_{0,lt};\boldsymbol{z}_{lkt}})$. The covariance and mean are given by

$$\begin{aligned} \boldsymbol{C}_{\Psi_{0,lt};\boldsymbol{z}_{lkt}} &= \sigma^2 \boldsymbol{I}_N + \sum_{j \neq k} \boldsymbol{C}_{\Psi_{1,ljt};\boldsymbol{z}_{ljt}}, \\ \boldsymbol{\mu}_{\Psi_{0,lt};\boldsymbol{z}_{lkt}} &= \boldsymbol{y}_{lt} - \sum_{j \neq k} \boldsymbol{\mu}_{\Psi_{1,ljt};\boldsymbol{z}_{ljt}}. \end{aligned}$$

Update of $m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} = \mathcal{N}(\boldsymbol{h}_{lk}; \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}, \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}})$. The covariance and mean are given by

$$\boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} = \left(\hat{\boldsymbol{\Sigma}}^{-1}_{lkt} - \boldsymbol{C}^{-1}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \right)^{-1}, \tag{24}$$

$$\boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} = \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} \left(\hat{\boldsymbol{\Sigma}}^{-1}_{lkt} \hat{\boldsymbol{h}}_{lk} - \boldsymbol{C}^{-1}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \right), \tag{25}$$

where

$$\hat{\boldsymbol{h}}_{lk} = \sum_{x'_{kt} \in \mathcal{S}} \omega(x'_{kt}) \bar{\boldsymbol{\mu}}(x'_{kt}), \tag{26}$$

$$\hat{\boldsymbol{\Sigma}}_{lkt} = \sum_{x'_{kt} \in \mathcal{S}} \omega(x'_{kt}) \left(\bar{\boldsymbol{\mu}}(x'_{kt}) \bar{\boldsymbol{\mu}}(x'_{kt})^H + \bar{\boldsymbol{C}}(x'_{kt}) \right) - \hat{\boldsymbol{h}}_{lk} \hat{\boldsymbol{h}}^H_{lk}. \tag{27}$$

The factor $\omega(x_{kt})$ is defined in (21). Finally,

$$\bar{\boldsymbol{C}}(x_{kt}) = \left(\boldsymbol{C}^{-1}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} + |x_{kt}|^2 \boldsymbol{C}^{-1}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \right)^{-1}, \tag{28}$$

$$\begin{aligned} \bar{\boldsymbol{\mu}}(x_{kt}) = \bar{\boldsymbol{C}}(x_{kt}) \Bigg(& |x_{kt}|^2 \frac{\boldsymbol{C}^{-1}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}}{x_{kt}} \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \\ & + \boldsymbol{C}^{-1}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \Bigg). \end{aligned} \tag{29}$$

Update of $m_{\Psi_{1,lkt};\boldsymbol{x}_{kt}} = \pi_{1,lkt}(x_{kt})$. The probabilities of the categorical distribution are given by

$$m_{\Psi_{1,lkt};x_{kt}} = \frac{\tilde{\gamma}(x_{kt})}{\sum_{x'_{kt} \in \mathcal{S}} \tilde{\gamma}(x'_{kt})} \tag{30}$$

with

$$\begin{aligned} \tilde{\gamma}(x_{kt}) := \mathcal{N}(\boldsymbol{0}; & \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} - x_{kt} \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}, \\ & \boldsymbol{C}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} + |x_{kt}|^2 \boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}). \end{aligned}$$

4.4 Inference

The desired estimates can be obtained computing the distributions $\hat{p}_{x_{kt}}$ and $\hat{p}_{\boldsymbol{h}_{lk}}$ as follows

$$\hat{x}_{kt} = \underset{x_{kt}\in\mathcal{S}}{\arg\max}\, \hat{p}_{x_{kt}} = \underset{x_{kt}\in\mathcal{S}}{\arg\max}\, m_{\Psi_{3,kt};x_{kt}} \prod_{l=1}^{L} m_{\Psi_{1,lkt};x_{kt}}$$

and

$$\begin{aligned}\hat{\boldsymbol{h}}_{lk} = \underset{\boldsymbol{h}_{lk}\in\mathbb{C}^N}{\arg\max}\, \hat{p}_{\boldsymbol{h}_{lk}} &= \underset{\boldsymbol{h}_{lk}\in\mathbb{C}^N}{\arg\max}\, m_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \prod_{t=1}^{T} m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} \\ &= \underset{\boldsymbol{h}_{lk}\in\mathbb{C}^N}{\arg\max}\, \mathcal{N}(\boldsymbol{h}_{lk}; \boldsymbol{\mu}_{\text{tot};\boldsymbol{h}}, \boldsymbol{C}_{\text{tot};\boldsymbol{h}}) \\ &= \boldsymbol{\mu}_{\text{tot};\boldsymbol{h}},\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{C}_{\text{tot};\boldsymbol{h}} &= \left(\boldsymbol{C}^{-1}_{\Psi_{2,lk};\boldsymbol{h}_{lk}} + \sum_{t=1}^{T} \boldsymbol{C}^{-1}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}\right)^{-1}, \\ \boldsymbol{\mu}_{\text{tot};\boldsymbol{h}} &= \boldsymbol{C}_{\text{tot};\boldsymbol{h}} \Bigg(\boldsymbol{C}^{-1}_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \boldsymbol{\mu}_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \\ &\qquad + \sum_{t=1}^{T} \boldsymbol{C}^{-1}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} \Bigg)\end{aligned}$$

are obtained by applying the Gaussian multiplication lemma [25]. Interestingly, the availability of the posterior distributions allows us to determine the reliability of the estimates $\hat{x}_{kt}$ and $\hat{\boldsymbol{h}}_{lk}$.

4.5 Algorithm Stability

Each parameter of a distribution is iteratively computed by a *soft update* [22], i.e., given the old parameters $\boldsymbol{\theta}^{(i-1)}$, the new parameters $\boldsymbol{\theta}^{(i)}$ are computed according to the message-passing rules and then updated as

$$\boldsymbol{\theta}^{(i)} \leftarrow \eta\, \boldsymbol{\theta}^{(i)} + (1-\eta)\, \boldsymbol{\theta}^{(i-1)},$$

where $\eta \in [0, 1]$ is the soft update parameter. Additionally, the computation of some messages, e.g., (17) may lead to precision matrices with negative eigenvalues. Inspired by [25], we set to zero the negative eigenvalues to assure positive semi-definiteness as required for Gaussian distributions.

4.6 Analysis of Fronthaul Load

In this section we discuss a distributed implementation of the bilinear-EP JCD with special attention to the messages exchanged through the fronthaul. As apparent from (49) in the Appendix, AP l requires the messages $m_{\Psi_{1,\tilde{l}kt};x_{kt}}$ from the other APs $\tilde{l} \neq l$. The data prior $m_{\Psi_{3,kt};x_{kt}} \equiv \Psi_{3,kt}$ is assumed to be uninformative and thus not needed. The messages are exchanged via the CPU which computes, based on the incoming messages from all APs, the following message

$$m_{x_{kt};\text{tot}} \propto \prod_{l=1}^{L} m_{\Psi_{1,lkt};x_{kt}}, \tag{31}$$

and then forwards them to the APs. Then, AP l can remove its own message from the incoming message from the CPU to obtain the desired message as in (49) , i.e.,

$$m_{x_{kt};\Psi_{1,lkt}} \propto \frac{m_{x_{kt};\text{tot}}}{m_{\Psi_{1,lkt};x_{kt}}}.$$

In the following, we quantify the fronthaul load for the communication between APs and CPU in terms of number of messages per iteration. Each AP transmits KT messages $m_{\Psi_{1,lkt};x_{kt}}$ to the CPU. Additionally, the message $m_{x_{kt};\text{tot}}$ is transmitted to each AP totalling LKT messages to be transmitted.

4.7 Computational Complexity

The order of computational complexity at the APs is mainly determined by the computation of inverse matrices of dimension N. Then, we focus on the messages which require matrix inversion and present the highest computational complexity, namely, messages $m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$ and $m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}$. Their order of complexity can be obtained from (22) and (37). We can observe that $KT|\mathcal{A}|$ weighted sums of covariance matrices per AP need to be inverted where $\mathcal{A} := \{a \in \mathbb{R}^+ \,|\, a = |x|^2, x \in \mathcal{S}\}$ denotes the set of distinct amplitudes in the signal constellation set $\mathcal{S}$. Therefore, at the AP, the complexity order per iteration is $\mathcal{O}(KT|\mathcal{A}|N^3)$. The computation of the message $m_{x_{kt};\text{tot}}$ in (31) at the CPU requires the multiplication of real-valued scalars of order $\mathcal{O}(LKT|\mathcal{S}|)$. Therefore, the order of the computational complexity at the CPU is $\mathcal{O}(LKT|\mathcal{S}|)$ per iteration of the bilinear-EP JCD algorithm. The total computational complexity that takes into account the processing in each AP and at the CPU is then given by $\mathcal{O}\left(LKT(|\mathcal{S}| + |\mathcal{A}|N^3)\right)$.

4.8 Simulation Results

4.8.1 Setting

In this section we present the setting that we utilized for our simulation. We consider a square surface with 400m side length and position $L = 16$ APs on a rectangular grid, i.e., on the points of set $\{(i \times \frac{400}{3}\text{m}, j \times \frac{400}{3}\text{m}) \,|\, i, j \in \{0, 1, 2, 3\}\}$ as in [17]. Each AP is equipped with $N = 1$ antenna. The $K = 8$ UEs are uniformly randomly distributed over the square surface. Each UE transmits with a power of $p = 14\,\text{dBm}$ and at each AP the received signal is impaired by an additive white noise with variance $v = -96\,\text{dBm}$. The diagonal elements of the channel covariance matrix $\boldsymbol{\Xi}_l$ are determined by the distance between the UE k and AP l using the fading model in [17], i.e., for UE k and for all co-located antennas n at AP l

$$[\boldsymbol{\Xi}_l]_{nk,nk}[\text{dB}] = -30.5 - 36.7 \log_{10}\left(\frac{d_{kl}}{1m}\right),$$

where d_{kl} is the distance between UE k and AP l. The pilot sequences are orthogonal, specifically, $\boldsymbol{X}_p = \boldsymbol{I}$, $P = K$. The variability of the scenario is captured by sampling 300 realizations of the positions of the UEs. In each position we perform 10^4 transmissions with different small-scale fading realizations. In each transmission, $T \in \{10, 100\}$ symbols are sent per UE. We use 4-QAM as modulation scheme. Our numerical results are obtained with $I = 10$ iterations and we use $\eta = 0.7$ as soft update parameter.

4.8.2 Performance evaluation and complexity analysis

In this section we present our simulation results and discuss the performance-complexity trade-off. As usual in the investigation of CF-MaMIMO to analyze the QoS distribution, we study the performance of the proposed algorithm in terms of the empirical cumulative distribution functions (CDFs) of the symbol error rate SER $:= \hat{E}\{\mathbb{1}_{\hat{x}\neq x}\}$ for detection and normalized mean squared error (NMSE) defined as NMSE $:= \hat{E}\left\{\frac{\|\hat{\boldsymbol{h}}-\boldsymbol{h}\|^2}{\|\boldsymbol{h}\|^2}\right\}$ for channel estimation.

Table 1: Computational complexity of distributed and centralized algorithms

Algorithm	Complexity at AP	Complexity at CPU	Total Complexity
bilinear-EP JCD	$\mathcal{O}(KT\|\mathcal{A}\|N^3)$	$\mathcal{O}(LKT\|\mathcal{S}\|)$	$\mathcal{O}\left(LKT(\|\mathcal{S}\|+\|\mathcal{A}\|N^3)\right)$
He et al. ICD	$\mathcal{O}\left(TKN^2+(T+P)^3N^3\right)$	$\mathcal{O}(TK^2)$	$\mathcal{O}\left(L(TKN^2+(T+P)^3N^3)+TK^2\right)$
Ngo et al.	-	-	$\mathcal{O}\left(K^6\|\mathcal{S}\|^T\right)$
Centralized MMSE	-	-	$\mathcal{O}\left(K\left(LN^3+K^2+LNT+\|\mathcal{S}\|T\right)\right)$

The samples for the empirical CDFs of the SER and NMSE are obtained per each large-scale channel fading and UE realization and per each large-scale channel fading, AP and UE realization, respectively, by averaging the instantaneous values corresponding to different small-scale realizations. Additionally, in the CDF of the NMSE we omit weak channels, i.e., channels where the received power is lower than the noise variance to avoid to take into account errors on weak and insignificant channels. In our simulation, we consider four baseline schemes, namely, the detector in [30] assuming perfect CSI, the ICD algorithm in [31], which is also a semi-blind algorithm based on EP with polynomial complexity order, a modified version of our proposed bilinear-EP JCD algorithm with perfect channel knowledge which provides a lower bound to the SER, and a centralized conventional receiver based on a Bayesian MMSE channel estimation algorithm and a subsequent linear MMSE filter for data detection.

Next, we compare the computational complexity order of the proposed bilinear-EP algorithm with the baseline EP algorithm in [31], a non-coherent centralized EP algorithm proposed in [25], and the baseline centralized approach based on linear MMSE filters for both channel estimation and detection. The results are summarized in Table 1. Notably, the proposed bilinear-EP JCD algorithm and the algorithm in [31] have linear and cubic (quadratic) complexity order in $T(K)$, respectively. Thus, the proposed algorithm is more efficient for long data sequences and a large number of UEs. However in our approach, the highest order term N^3 scales linearly with the number of UEs K. As the bilinear-EP JCD algorithm, the non-coherent centralized EP algorithm in [25] adopts an exact categorical distribution for data. When it is used for uncoded transmission, as in the proposed approach, its complexity order is given by $\mathcal{O}\left(K^6|\mathcal{S}|^T\right)$. Due to its exponential complexity in the length of the symbol sequence length T, the required computational efforts are prohibitive for a comparative study of the algorithm performance. Essentially, the algorithm in [25] approximates a maximum a posteriori (MAP) decoder whereas our algorithm approximates a MAP detector. It is worth noting that the extension of proposed bilinear-EP JCD algorithm with a decoder is straightforward through the use of loopy belief propagation which is known to have polynomial complexity as well. In contrast to the other algorithms, the conventional centralized MMSE algorithm is not iterative in nature and, thus, a fair comparison with the other receivers which require multiple iterations is not straightforward.

In the following, we analyze the performance. Fig. 3 shows the empirical CDF of the SER of the proposed bilinear-EP JCD for $T = \{10, 100\}$ and the above mentioned baselines. For both values of T the proposed bilinear-EP algorithm outperforms the ICD algorithm in [31] by more than one order of magnitude, while the complexity of bilinear-EP is only linear in T and not cubic. Our approach also outperforms the EP-based detector in [30] with perfect CSI and the centralized MMSE approach. The gap in performance with the second baseline approach indicates that the Gaussian approximation of the data symbol distribution and the averaging of the messages in [30, 31] determine a degradation of performance. We observe an improvement of the CDF of the SER obtained with our bilinear-EP JCD algorithm when the length of the transmitted data symbols increases from $T = 10$ to $T = 100$. The performance of our modified algorithm obtained assuming perfect CSI shows that there is room for further improvement. Fig. 4 shows the CDF of the NMSE. We observe an improvement of the NMSE by more than an order of magnitude of our proposed algorithm compared to the ICD scheme in [31]. Additionally, the performance of our channel estimation also improves when T increases.

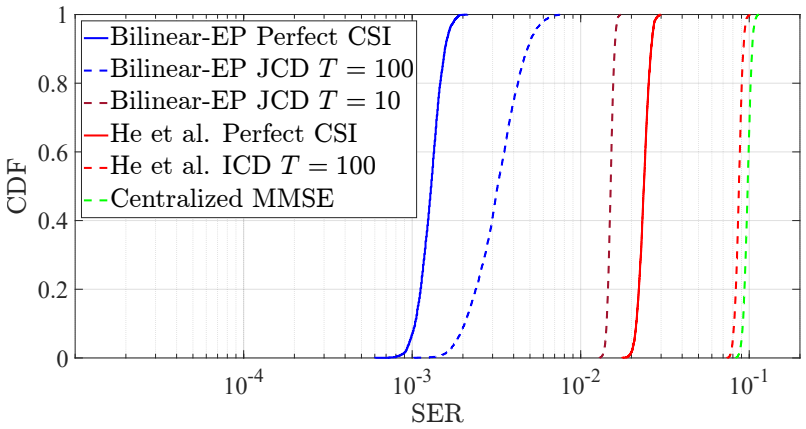


Figure 3: CDF of SER for $L = 16, K = 8, N = 1$.

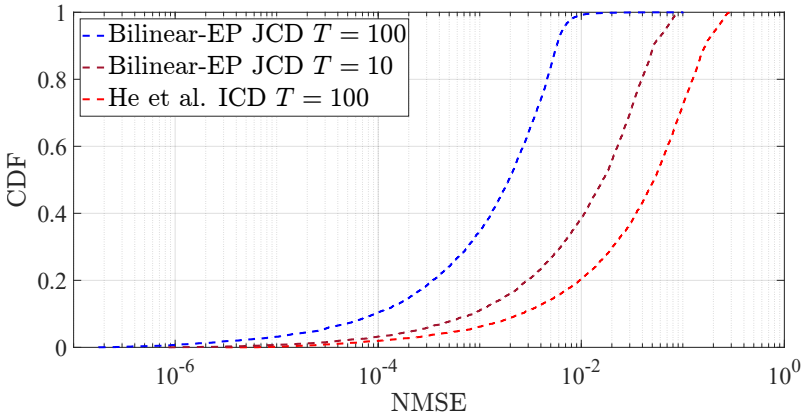


Figure 4: CDF of NMSE for $L = 16, K = 8, N = 1$.

$$m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}\left(\frac{\boldsymbol{z}_{lkt}}{x_{kt}}\right) = \frac{1}{\pi^N \det\left(\boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}\right)} \cdot \exp\left(-\left(\frac{\boldsymbol{z}_{lkt}}{x_{kt}} - \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}\right)^H \boldsymbol{C}^{-1}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}\left(\frac{\boldsymbol{z}_{lkt}}{x_{kt}} - \boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}\right)\right)$$
$$= |x_{kt}|^{2N}\mathcal{N}(\boldsymbol{z}_{lkt}; x_{kt}\boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}, |x_{kt}|^2\boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \tag{36}$$

$$m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(\boldsymbol{z}_{lkt}) m_{x_{kt};\Psi_{1,lkt}}(x_{kt})\mathcal{N}(\boldsymbol{z}_{lkt}; x_{kt}\boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}, |x_{kt}|^2\boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}})$$
$$= \mathcal{N}(\boldsymbol{z}_{lkt}; \tilde{\boldsymbol{\mu}}(x_{kt}), \tilde{\boldsymbol{C}}(x_{kt})) \cdot \underbrace{m_{x_{kt};\Psi_{1,lkt}}(x_{kt})\mathcal{N}(\boldsymbol{0}; \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} - x_{kt}\boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}, \boldsymbol{C}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} + |x_{kt}|^2\boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}})}_{:=\tilde{\omega}(x_{kt})} \tag{37}$$

$$m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(x_{kt}\boldsymbol{h}_{lk}) = \frac{1}{\pi^N \det\left(\boldsymbol{C}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}\right)} \cdot \exp\left(-\left(x_{kt}\boldsymbol{h}_{lkt} - \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}\right)^H \boldsymbol{C}^{-1}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}\left(x_{kt}\boldsymbol{h}_{lk} - \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}\right)\right)$$
$$= \frac{1}{|x_{kt}|^{2N}}\mathcal{N}(\boldsymbol{h}_{lk}; \frac{\boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}}{x_{kt}}, \frac{\boldsymbol{C}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}}{|x_{kt}|^2}) \tag{38}$$

Algorithm 2 Modified Bilinear-EP Algorithm

Input: Pilot matrix $\mathbf{X}^p$, transmit power σ_x^2, received signal $\mathbf{Y}$, noise variance σ_n^2, prior distribution $\tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k}) \equiv (\tilde{\boldsymbol{\mu}}_{h_{l,k}}, \tilde{\mathbf{C}}_{h_{l,k}})$.

Output: Estimated channels $\hat{\mathbf{h}}_{l,k}$ and data $\hat{x}_{kt}^d$.

1: $\forall k, l, t$: Initialize all messages uninformatively except

$$\boldsymbol{\mu}_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}} = \boldsymbol{\mu}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} = \tilde{\boldsymbol{\mu}}_{h_{l,k}},$$

$$\mathbf{C}_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}} = \mathbf{C}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} = \tilde{\mathbf{C}}_{h_{l,k}},$$

$$\boldsymbol{\mu}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \begin{cases} \tilde{\boldsymbol{\mu}}_{h_{l,k}} x_{kt}^p & \text{for } t \leq T_p \\ \mathbf{0} & \text{for } t > T_p \end{cases},$$

$$\mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \begin{cases} \tilde{\mathbf{C}}_{h_{l,k}} |x_{kt}^p|^2 & \text{for } t \leq T_p \\ (\tilde{\mathbf{C}}_{h_{l,k}} + + \tilde{\boldsymbol{\mu}}_{h_{l,k}} \tilde{\boldsymbol{\mu}}_{h_{l,k}}^{\mathsf{H}}) \sigma_x^2 & \text{for } t > T_p \end{cases}.$$

2: **for** $i = 1$ to $i_{\max}$ **do**

3: $\forall k, l, t$: Update $m_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}$ via (32), (33).

4: $\forall k, l, t > T_p$: Update $m_{\Psi_{z_{l,kt}};x_{kt}}$ via (34).

5: $\forall k, l, t > T_p$: Update $m_{x_{kt};\Psi_{z_{l,kt}}}$ via (35).

6: $\forall k, l, t$: Update $m_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}}$ via (36), (37).

7: $\forall k, l, t$: Update $m_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}}$ via (38), (39).

8: $\forall k, l, t$: Update $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ via (40), (41).

9: **end for**

10: **return** $\hat{\mathbf{h}}_{l,k}$ calculated via (42) $\forall k, l$.

11: **return** $\hat{x}_{kt}^d$ calculated via (43) $\forall k, t$.

5 Improving Performance under Pilot Contamination [2]

In this section, we propose a novel JCD algorithm based on EP, specifically designed to enhance robustness against PC in CF-MaMIMO systems. The algorithm builds upon the bilinear-EP algorithm presented in Section 4 and incorporates a modified scheduling and message passing for bilinear structure to effectively exploit the inherent structure of the received data signals and suppress the impact of PC. We evaluate the proposed method against state-of-the-art Bayesian learning algorithms and demonstrate superior detection and estimation performance, particularly under severe PC. This thorough analysis of PC was not conducted in prior work. Our analysis further considers both contamination caused by non-orthogonal sequences and by the reuse of identical orthogonal sequences, revealing that Bayesian methods exhibit greater robustness when non-orthogonal pilots with low correlation are used. This insight motivates the introduction of a novel metric tailored to quantify PC and assess its impact on iterative JCD algorithms.

5.1 Modified Bilinear-EP Algorithm

The modified bilinear-EP algorithm is summarized in Algorithm 2. Compared to the EP algorithm presented in Section 4, it contains additional and augmented message updates. To be more precise, the updates of the messages $m_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}$, $m_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}}$, $m_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}}$, and $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ are now also considered for $t \leq T_p$ i.e., for the pilot part as well. Furthermore, the update of the message $m_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}}$ is enhanced by taking into account the additional information from $t \leq T_p$. To make this more clear, we describe the scheduling, initialization, and message updates in the following. Note that many of the message updates presented below are equal to the message updates presented in the previous section. However, we recapitulate all the message updates to enable a clear presentation of the modified algorithm.

The bilinear-EP algorithm models the variables x_{kt} with categorical distributions, while $\mathbf{z}_{l,kt}$ and $\mathbf{h}_{l,k}$ are modeled as multivariate complex Gaussian distributions. Accordingly, EP message-passing update rules are applied to the factor graph in Fig. 5. We present the message scheduling

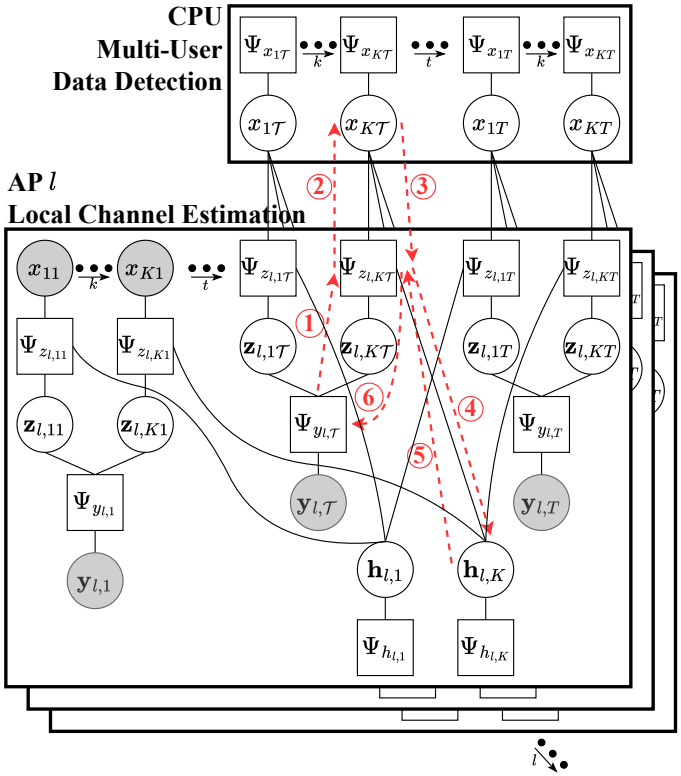


Figure 5: Factor graph for bilinear-EP with $\mathcal{T} \coloneqq T_p + 1$. The numbered red dashed arrows show the message update scheduling according to Algorithm 2.

and the final update rules. The derivations are analog to those presented in Appendix A. Note that the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\mathbf{C}$ of a Gaussian random variable are equivalently represented by the natural parameters $\boldsymbol{\gamma} = \mathbf{C}^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\Lambda} = \mathbf{C}^{-1}$. In the following, we will interchangeably use both these representations without explicitly stating the transformation, i.e., if $\boldsymbol{\mu}_{\Psi_\alpha;\mathbf{x}_\beta}$ and $\mathbf{C}_{\Psi_\alpha;\mathbf{x}_\beta}$ are computed, then $\boldsymbol{\gamma}_{\Psi_\alpha;\mathbf{x}_\beta}$ and $\boldsymbol{\Lambda}_{\Psi_\alpha;\mathbf{x}_\beta}$ are automatically given and vice versa.

The message initialization is performed as follows: parameters describing the messages $m_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}$, $m_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}}$, and $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ $\forall k, l, t$ are initialized per Algorithm 2. All other messages are initialized in an uninformative way.

The messages $m_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}$ $\forall k, l, t$ are updated first, in which each AP performs interference cancellation on the received signal using the current knowledge of the auxiliary variables $\mathbf{z}_{l,kt}$,

$$\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} = \mathbf{y}_{l,t} - \sum_{k' \neq k} \boldsymbol{\mu}_{\Psi_{z_{l,k't}};\mathbf{z}_{l,k't}}, \tag{32}$$

$$\mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} = \sigma_n^2 \mathbf{I}_N + \sum_{k' \neq k} \mathbf{C}_{\Psi_{z_{l,k't}};\mathbf{z}_{l,k't}}. \tag{33}$$

The updated information on the variables $\mathbf{z}_{l,kt}$ at each AP l is used to refine the local beliefs on the data symbols x_{kt}, which are subsequently shared with the CPU. This is done by updating the message $m_{\Psi_{z_{l,kt}};x_{kt}}$ $\forall k, l, t > T_p$,

$$\pi_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) \propto \theta(x_{kt}), \tag{34}$$

with

$$\theta(x_{kt}) = \mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_{\Psi_{y_{l,t};\mathbf{z}_{l,kt}}} - \boldsymbol{\mu}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} x_{kt}, \\ \mathbf{C}_{\Psi_{y_{l,t};\mathbf{z}_{l,kt}}} + \mathbf{C}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} |x_{kt}|^2).$$

Next, the messages $m_{x_{kt};\Psi_{z_{l,kt}}}$ $\forall k, l, t > T_p$ are updated by aggregating the data symbol beliefs from all APs at the CPU, which then sends the following refined beliefs back to the APs,

$$\pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \propto \prod_{l' \neq l} \pi_{\Psi_{z_{l',kt}};x_{kt}}(x_{kt}). \tag{35}$$

The updated beliefs on the symbols x_{kt} at each AP are then used to refine the channel estimates $\mathbf{h}_{l,k}$. This refinement is achieved through the update of the message $m_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}}$ $\forall k, l, t$, given by

$$\boldsymbol{\Lambda}_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}} = \hat{\boldsymbol{\Lambda}}_{1_{l,kt}} - \boldsymbol{\Lambda}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}}, \tag{36}$$

$$\boldsymbol{\gamma}_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}} = \hat{\boldsymbol{\gamma}}_{1_{l,kt}} - \boldsymbol{\gamma}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}}, \tag{37}$$

with $\hat{\boldsymbol{\mu}}_{1_{l,kt}} = \frac{\check{\boldsymbol{\mu}}_{l,kt}(x^p_{kt})}{x^p_{kt}}$, $\hat{\mathbf{C}}_{1_{l,kt}} = \frac{\check{\mathbf{C}}_{l,kt}(x^p_{kt})}{|x^p_{kt}|^2}$ for $t \leq T_p$ and

$$\hat{\boldsymbol{\mu}}_{1_{l,kt}} = \frac{1}{Z_{l,kt}} \sum_{x_{kt} \in \mathcal{X}} \pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \frac{\theta(x_{kt})}{x_{kt}} \cdot \check{\boldsymbol{\mu}}_{l,kt}(x_{kt}),$$

$$\hat{\mathbf{C}}_{1_{l,kt}} = \frac{1}{Z_{l,kt}} \sum_{x_{kt} \in \mathcal{X}} \pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \frac{\theta(x_{kt})}{|x_{kt}|^2} \cdot \big(\check{\mathbf{C}}_{l,kt}(x_{kt}) \\ + \check{\boldsymbol{\mu}}_{l,kt}(x_{kt}) \cdot \check{\boldsymbol{\mu}}^{\mathsf{H}}_{l,kt}(x_{kt})\big) - \hat{\boldsymbol{\mu}}_{l,kt} \hat{\boldsymbol{\mu}}^{\mathsf{H}}_{l,kt},$$

for $t > T_p$ with

$$Z_{l,kt} = \sum_{x_{kt} \in \mathcal{X}} \pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}),$$

$$\check{\boldsymbol{\Lambda}}_{l,kt}(x_{kt}) = \boldsymbol{\Lambda}_{\Psi_{y_{l,t};\mathbf{z}_{l,kt}}} + \boldsymbol{\Lambda}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} |x_{kt}|^{-2},$$

$$\check{\boldsymbol{\gamma}}_{l,kt}(x_{kt}) = \boldsymbol{\gamma}_{\Psi_{y_{l,t};\mathbf{z}_{l,kt}}} + \boldsymbol{\gamma}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} \frac{x_{kt}}{|x_{kt}|^2}.$$

Then, the messages $m_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}}$ $\forall k, l, t$ are updated yielding new estimates of $\mathbf{h}_{l,k}$ by combining the information acquired across all time slots with the prior channel information,

$$\boldsymbol{\Lambda}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} = \boldsymbol{\Lambda}_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}} + \sum_{t' \neq t} \boldsymbol{\Lambda}_{\Psi_{z_{l,kt'};\mathbf{h}_{l,k}}}, \tag{38}$$

$$\boldsymbol{\gamma}_{\mathbf{h}_{l,k};\Psi_{z_{l,kt}}} = \boldsymbol{\gamma}_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}} + \sum_{t' \neq t} \boldsymbol{\gamma}_{\Psi_{z_{l,kt'};\mathbf{h}_{l,k}}}. \tag{39}$$

The messages $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ $\forall k, l, t$ are updated last in an EP iteration, generating refined estimates of the variables $\mathbf{z}_{l,kt}$ which are then utilized for interference cancellation in the next iteration,

$$\boldsymbol{\Lambda}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \hat{\boldsymbol{\Lambda}}_{2_{l,kt}} - \boldsymbol{\Lambda}_{\Psi_{y_{l,t};\mathbf{z}_{l,kt}}}, \tag{40}$$

$$\boldsymbol{\gamma}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \hat{\boldsymbol{\gamma}}_{2_{l,kt}} - \boldsymbol{\gamma}_{\Psi_{y_{l,t};\mathbf{z}_{l,kt}}}, \tag{41}$$

with $\hat{\boldsymbol{\mu}}_{2_{l,kt}}=\check{\boldsymbol{\mu}}_{l,kt}(x_{kt}^p)$, $\hat{\mathbf{C}}_{2_{l,kt}}=\check{\mathbf{C}}_{l,kt}(x_{kt}^p)$ for $t \leq T_p$ and

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{2_{l,kt}} &= \frac{1}{Z_{l,kt}} \sum_{x_{kt}\in\mathcal{X}} \pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}) \cdot \check{\boldsymbol{\mu}}_{l,kt}(x_{kt}),\\ \hat{\mathbf{C}}_{2_{l,kt}} &= \frac{1}{Z_{l,kt}} \sum_{x_{kt}\in\mathcal{X}} \pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}) \cdot \big(\check{\mathbf{C}}_{l,kt}(x_{kt})\\ &\qquad + \check{\boldsymbol{\mu}}_{l,kt}(x_{kt}) \cdot \check{\boldsymbol{\mu}}_{l,kt}^{\mathsf{H}}(x_{kt})\big) - \hat{\boldsymbol{\mu}}_{1_{l,kt}} \hat{\boldsymbol{\mu}}_{1_{l,kt}}^{\mathsf{H}},\end{aligned}$$

for $t > T_p$ with $Z_{l,kt}$, $\check{\boldsymbol{\Lambda}}_{l,kt}(x_{kt})$, and $\check{\boldsymbol{\gamma}}_{l,kt}(x_{kt})$ as given before.

After performing the final EP iteration, the channel and data estimates are computed as follows,

$$\hat{\mathbf{h}}_{l,k} = \hat{\boldsymbol{\Lambda}}_{\mathbf{h}_{l,k}}^{-1} \hat{\gamma}_{\mathbf{h}_{l,k}}, \tag{42}$$

$$\hat{x}_{kt}^d = \underset{x_{kt}^d\in\mathcal{X}}{\arg\max}\ \hat{p}_{x_{kt}}(x_{kt}^d), \tag{43}$$

with

$$\hat{\boldsymbol{\Lambda}}_{\mathbf{h}_{l,k}} = \boldsymbol{\Lambda}_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}} + \sum_{t=1}^{T} \boldsymbol{\Lambda}_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}}, \tag{44}$$

$$\hat{\gamma}_{\mathbf{h}_{l,k}} = \boldsymbol{\gamma}_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}} + \sum_{t=1}^{T} \boldsymbol{\gamma}_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}}, \tag{45}$$

and the approximated posterior data distribution

$$\hat{p}_{x_{kt}}(x_{kt}^d) \propto \prod_{l=1}^{L} \pi_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}^d). \tag{46}$$

We note that damping is applied to factor-to-variable messages using a damping parameter $\eta \in [0,1]$ to improve the stability of the bilinear-EP algorithm [1], i.e., each updated parameter is computed as a convex combination of its previous and newly computed values. Furthermore, the parameters of the messages $m_{\Psi_{z_{l,kt}};\mathbf{h}_{l,k}}$ and $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ in line 6 and 8 of Algorithm 2 are updated only if the corresponding covariance/precision matrices obtained by (36) and (40), respectively, are symmetric positive definite. Otherwise, the parameters from the previous iteration are retained.

5.2 Quantifying Pilot Contamination

In this section, we introduce a metric to quantify PC in CF-MaMIMO networks. The level of PC is influenced by the choice of the pilot matrix $\mathbf{X}^p$ and the resulting correlations between pilot sequences. Mutual coherence, which measures the similarity between pilot sequences, was used in [36] for pilot design to mitigate PC. However, due to the distributed architecture of CF-MaMIMO, user signals can often be separated spatially, especially when the users are far apart, resulting in negligible interference. Therefore, the spatial power profiles of all UEs, captured by the large-scale fading coefficients (LSFCs) $\xi_{l,k}$, are also critical for characterizing PC. This motivates us to develop a new PC metric which is particularly suited for JCD in distributed systems. Inspired by the NMSE of the pilot-based MMSE channel estimator, we define the PC metric c_k for UE k as

$$c_k = \min_{l} \frac{\left[\left(\mathrm{diag}(\xi_{l,1},\ldots,\xi_{l,K})^{-1} + \mathbf{X}^p\mathbf{X}^{p\mathsf{H}}\sigma_n^{-2}\right)^{-1}\right]_{k,k}}{\xi_{l,k}}, \tag{47}$$

where $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with its inputs on the main diagonal. The rationale for taking the minimum value over all APs is that a strong, low-contamination link to any single AP

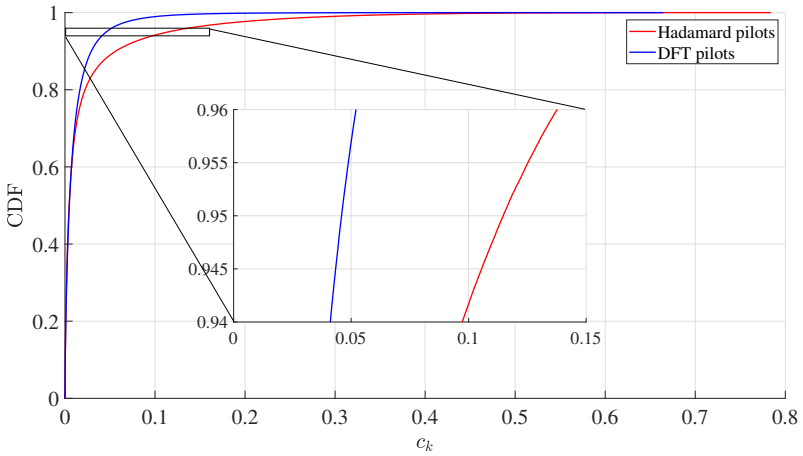


Figure 6: CDF of c_k for different pilot sequences.

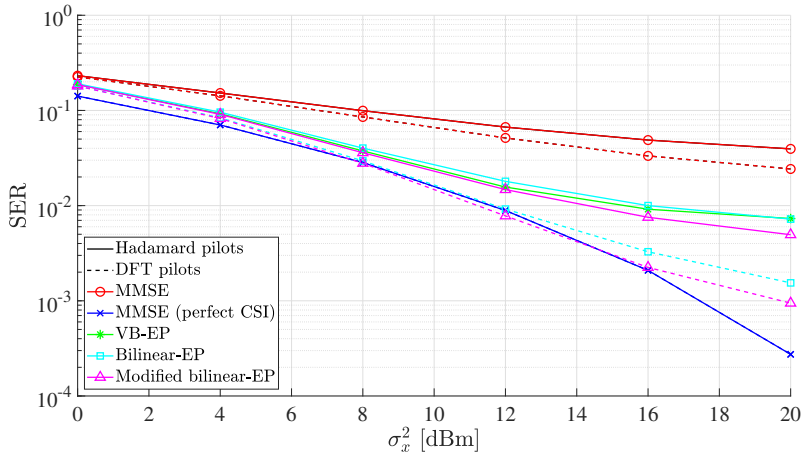
is sufficient for reliable channel estimation and successful data detection. This reliable information can be used to cancel the interference caused by the corresponding UE and, hence, iteratively remove PC. This concept was formalized in [37] where sufficient and necessary conditions for semi-blind identifiability were established.

5.3 Extended Simulation Results

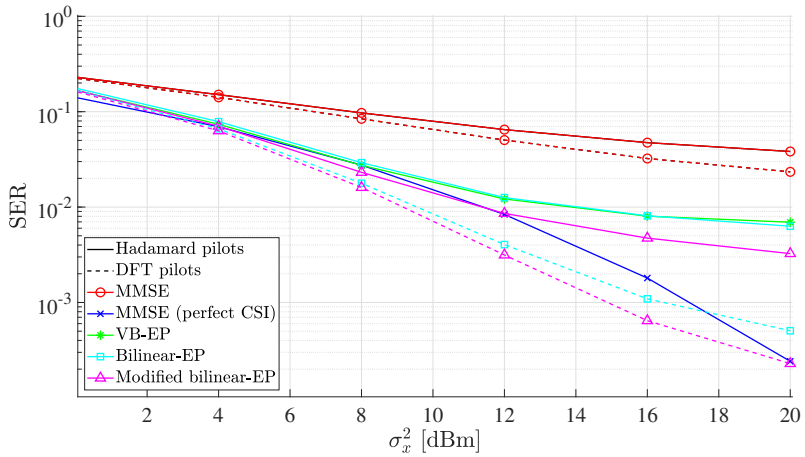
In this section, we present Monte Carlo simulation results for the modified bilinear-EP algorithm and several state-of-the-art benchmark algorithms. We consider a network spanning an area of $400 \times 400\,\mathrm{m}^2$ comprising $L = 16$ single-antenna APs, i.e., $N = 1$, placed on a regular grid at coordinates $\{(50 + i{\cdot}100, 50 + j{\cdot}100)\,\mathrm{m} \,|\, i, j \in \{0, 1, 2, 3\}\}$ and placed at a height of 10 m. A total of $K = 8$ UEs are placed uniformly at random ground locations and transmit $T_p = 4$ pilot symbols and $T_d \in \{10, 30\}$ 4-quadrature amplitude modulation (QAM) uncoded data symbols. The values of K and T_p are chosen such that PC occurs and the complexity of the simulations is not too high. In practice, longer channel coherence times allow for an increase in the number of pilot symbols T_p but also for an increase in the number of UEs K, especially when the number of AP antennas LN is increased as well, such that PC still occurs in these more practical cases and needs to be mitigated. We consider two different choices of pilot sequences, referred to as Hadamard and discrete Fourier transform (DFT) pilots. For Hadamard pilots, T_p orthogonal Hadamard pilot sequences are considered and shared among the K UEs. For DFT pilots, the pilot matrix $\mathbf{X}^p$ is obtained by truncating a $K \times K$ DFT matrix to the first T_p columns, resulting in non-orthogonal sequences. Hence, in our simulations, the set of Hadamard pilots consists of four orthogonal pilot sequences, each of which is shared between two users. In contrast, the set of DFT pilots consists of eight unique but non-orthogonal pilot sequences. The receiver noise power at each AP is set to $\sigma_n^2 = -96\,\mathrm{dBm}$. The LSFCs are obtained using the 3GPP urban microcell model which incorporates correlated shadow fading [17].

The first set of results pertains to the PC metric introduced in (47) and evaluated over 10^5 independent large-scale fading realizations. The CDF of the PC metric c_k is illustrated in Fig. 6. It can be observed that the non-orthogonal DFT pilots result in lower PC than the orthogonal Hadamard pilots which are reused among the UEs.

In the following, we show the channel estimation and data detection performance in terms of the NMSE of the channel estimates and the SER. We compare the proposed modified bilinear-EP algorithm with the bilinear-EP algorithm in [1], the variational Bayes, belief propagation, and expectation propagation (VB-EP) algorithm presented in [38], and different MMSE estimators. For channel estimation, we consider the pilot-based MMSE estimator and the genie-aided MMSE estimator with perfect knowledge of the transmitted data symbols. For symbol detection, we employ



(a) $T_d = 10$.



(b) $T_d = 30$.

Figure 7: SER versus transmit power.

the centralized MMSE MIMO detector with CSI obtained by the pilot-based MMSE estimator and with perfect CSI. Note that the MMSE detector with pilot-based CSI serves as an upper performance bound for any linear detection scheme with imperfect CSI, including those proposed in [39] and [17]. Both of the bilinear-EP algorithms execute 20 iterations for JCD, whereas the VB-EP algorithm runs for 40 iterations. All iterative algorithms use a damping parameter of $\eta = 0.5$. In the following figures, solid and dashed lines distinguish between different pilot sequences or data lengths, while colors and markers indicate the respective algorithms.

The next set of results present the NMSE and the SER as functions of the UE transmit power σ_x^2. The NMSE of the channel matrix estimate $\hat{\mathbf{H}}$ is defined as NMSE $\coloneqq E\left\{\frac{||\mathbf{H}-\hat{\mathbf{H}}||_F^2}{||\mathbf{H}||_F^2}\right\}$, and the SER is obtained by averaging across all UEs, i.e., SER $\coloneqq E\left\{\sum_k \sum_t \mathbb{1}_{x_{kt}^d \neq \hat{x}_{kt}^d}/(KT_d)\right\}$. The expectation operator in both definitions is computed with respect to the channel realizations. In our simulation results, the performance is obtained by averaging 10^4 block transmissions where each block transmission corresponds to an independent realization of the UE positions. The SER results are shown in Fig. 7. It can be observed that the proposed bilinear-EP algorithm outperforms the linear MMSE detector as well as the bilinear-EP algorithm in [1] and the VB-EP algorithm in [38]. Systems adopting non-orthogonal DFT pilots show significantly better performance compared to those using orthogonal Hadamard pilots. Additionally, the performance gain from increasing the number of data symbols from $T_d = 10$ to $T_d = 30$ is more pronounced for DFT pilots. The

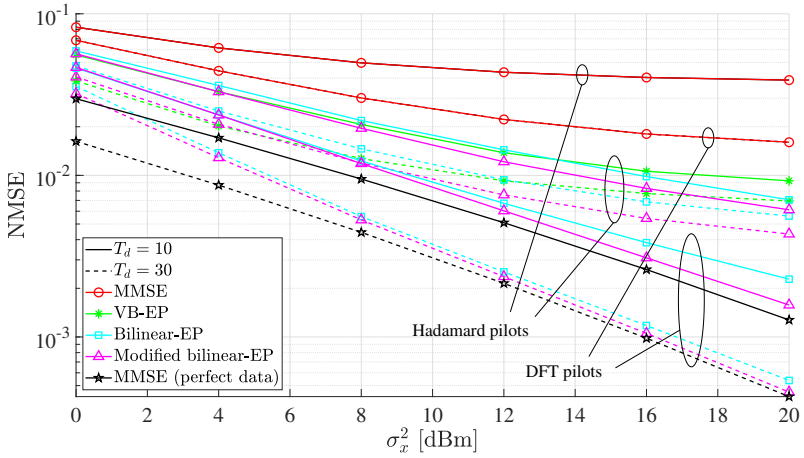


Figure 8: NMSE versus transmit power.

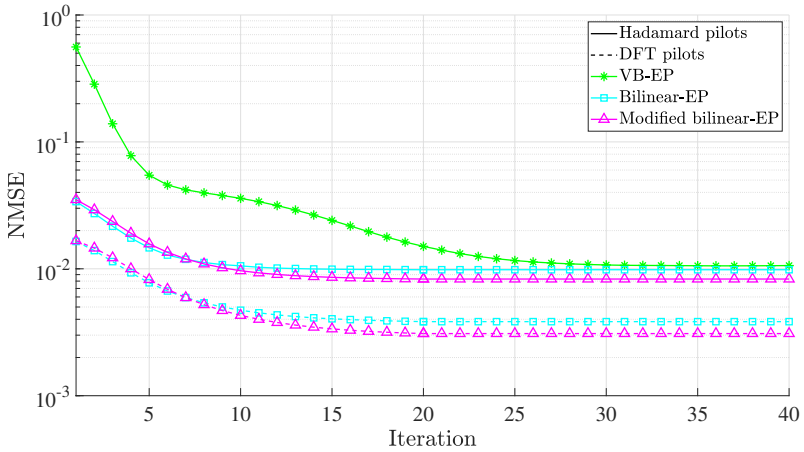


Figure 9: NMSE versus iterations for $T_d = 10$ and $\sigma_x^2 = 16\,\text{dBm}$.

VB-EP algorithm, being designed for orthogonal pilots, is not applicable to systems adopting the proposed DFT pilots. The proposed bilinear-EP algorithm offers greater flexibility in the design of pilot sequences, which represents an additional advantage. Fig. 8 illustrates the NMSE performance as a function of the transmit power for the same settings discussed above. The proposed bilinear-EP algorithm, when used with non-orthogonal DFT pilot sequences, closely approaches the performance of the genie-aided MMSE estimator, a bound which is not attained with orthogonal Hadamard pilot sequences. Fig. 9 shows the convergence behavior of the iterative algorithms. The bilinear-EP algorithm converges significantly faster than the VB-EP algorithm. As the bilinear-EP algorithm is executed for only 20 iterations, the result for iteration number 20 is extended to all following iterations to enable direct comparison with the benchmark algorithms.

For the following set of results, the UE transmit power is fixed to $\sigma_x^2 = 16\,\text{dBm}$ and the performance is assessed in terms of the CDFs of the NMSE and the SER per user, i.e., $\text{NMSE}_k \coloneqq E\left\{\frac{||\mathbf{h}_k - \hat{\mathbf{h}}_k||^2}{||\mathbf{h}_k||^2}\right\}$ and $\text{SER}_k \coloneqq E\left\{\sum_t \mathbb{1}_{x_{kt}^d \neq \hat{x}_{kt}^d}/T_d\right\}$. Here, 1000 independent realizations of the UE positions are considered which results in $1000 \cdot K = 8000$ data points for the CDF. For each UE positioning realization, the performance is averaged over 1000 independent block transmissions, accounting for small-scale fading and noise realizations. The corresponding results are illustrated in Figs. 10 and 11. The CDFs show that the proposed modified bilinear-EP algorithm outperforms the benchmark schemes. Furthermore, the algorithms applied to systems utilizing non-orthogonal DFT pilots show greater performance gains from increased data lengths, especially when considering the

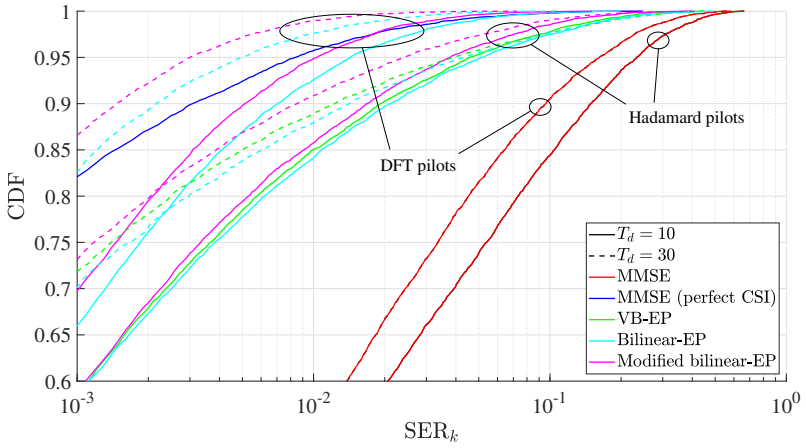


Figure 10: CDF of SER_k for $\sigma_x^2 = 16\,\text{dBm}$.

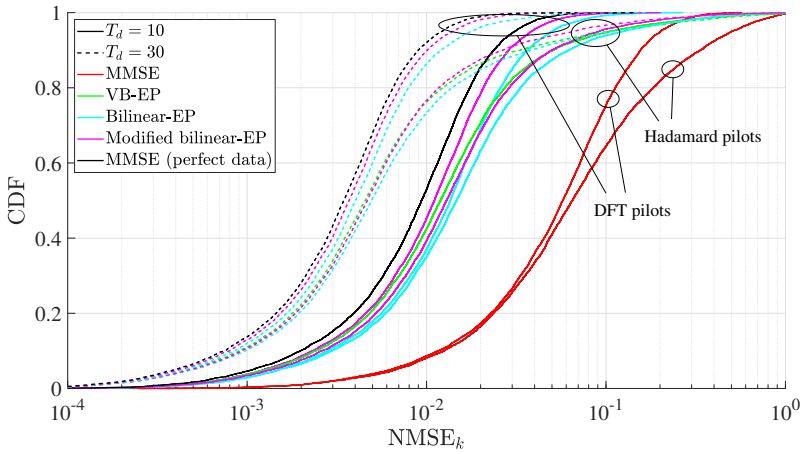


Figure 11: CDF of NMSE_k for $\sigma_x^2 = 16\,\text{dBm}$.

95%-likely performance.

For the final set of results, the exact same setup as before is used with the UE-based performance metrics now evaluated as a function of the PC metric c_k. The results are averaged over UEs experiencing a similar level of PC quantified by c_k. Fig. 12 presents the corresponding results for the SER. It can be observed that as c_k increases, the performance degrades which validates c_k as an appropriate PC metric. Combined with the observations from Fig. 6, it explains the superior average performance of DFT pilots over Hadamard pilots in previous results. Furthermore, the proposed algorithm consistently achieves the best performance for a given level of PC. These results further show that the JCD schemes applied to systems using DFT pilots are more effective at mitigating PC for a given level of PC c_k.

6 Further Performance Analysis

In this section, we further investigate the performance under PC and how to reduce the fronthaul load for the bilinear-EP algorithm. In order to control the level of PC, we introduce a minimum distance constraint between UEs. The fronthaul load is reduced by performing multiple local iterations at the APs before sharing information with the CPU via the fronthaul. More details on these approaches are provided in the following.

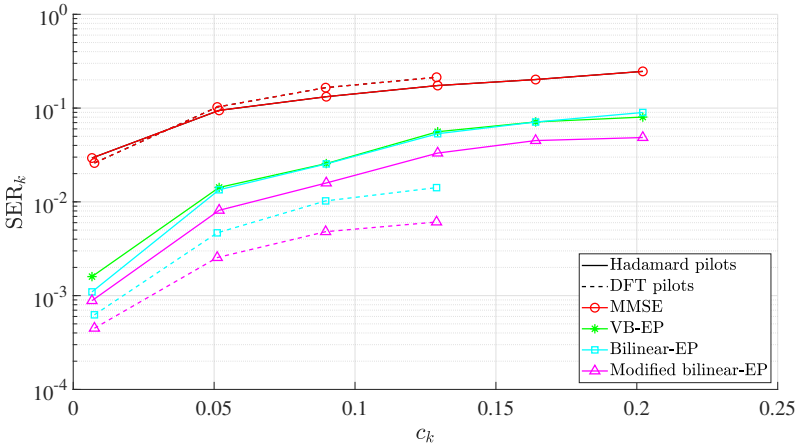


Figure 12: SER_k versus PC metric for $T_d = 30$ and $\sigma_x^2 = 16\,\mathrm{dBm}$.

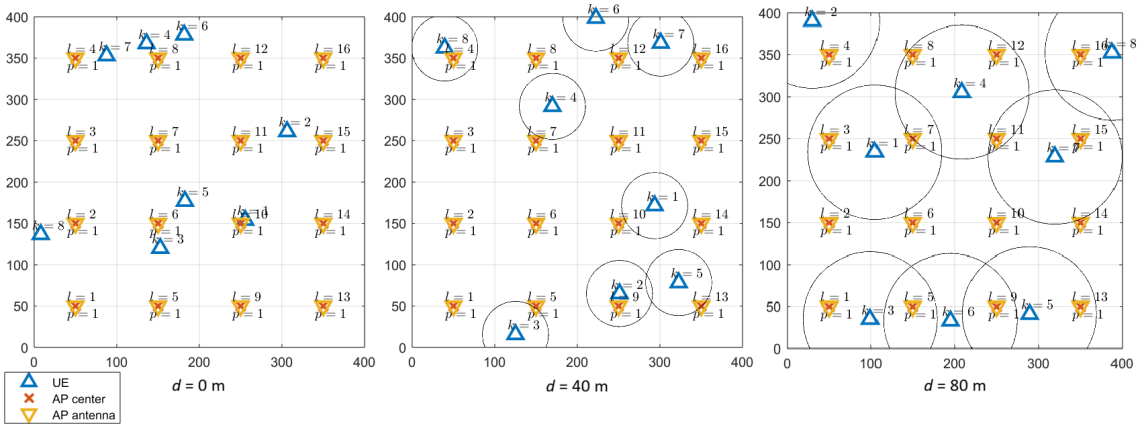


Figure 13: Network realization with different minimum UE distance constraints d.

6.1 Minimum User Distance Constraint

PC occurs due to the use of non-orthogonal pilot sequences. The level of PC is determined by the correlation between the different pilot sequences. However, in a distributed architecture as in CF-MaMIMO networks, the level of PC also depends on the distance between APs and UEs. Two UEs which are close to each other will cause strong interference, resulting in strong PC. This was partly analyzed in Section 5.2. In the following, we investigate how the geographical distance between UEs affect the system performance. For this, we introduce a minimum UE separation distance d, i.e., two UEs are always at least d meters apart from each other. This is illustrated in Fig. 13 for three different values of d. Here, the network area of $400 \times 400\,\mathrm{m}$ is shown with the AP and UE positions. The circles around the UEs visualizes the minimum UE distance constraint.

In the following, the same simulation parameters are considered as in Section 5.3 except that $T = 18$ and $T_p \in \{4, 8\}$ depending on the scenario with and without PC, respectively. For $T_p = 4$, random binary phase-shift-keying (BPSK) pilot sequences are transmitted by the UEs, whereas for $T_p = 8$ orthogonal pilot sequences are used. The minimum UE distance constraint is set to $d \in \{0, 40, 80\}\,\mathrm{m}$. Figs. 14 and 15 show the effect of the user distance constraints on the system performance. In scenarios without PC, the performance remains consistent regardless of the user distance. However, in systems affected by PC, increasing the distance between users leads to an improvement in performance for both algorithms. This improvement is more pronounced for the bilinear-EP algorithm compared to the MMSE data detector which shows the effective exploitation of geographical UE separation for pilot decontamination.

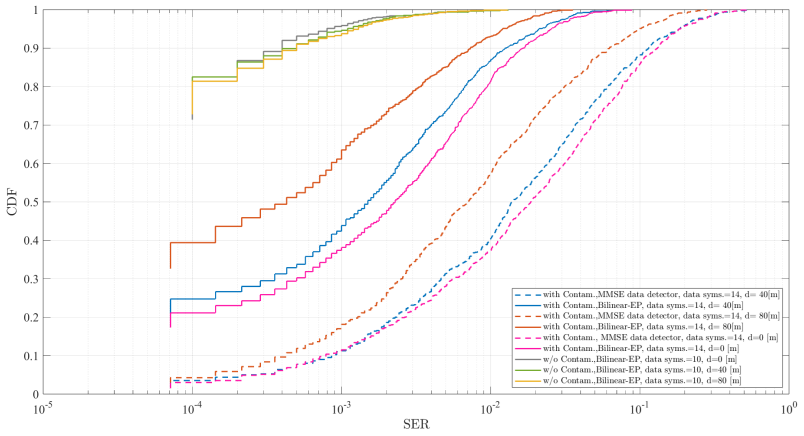


Figure 14: CDF of SER with minimum distance constraint between UEs.

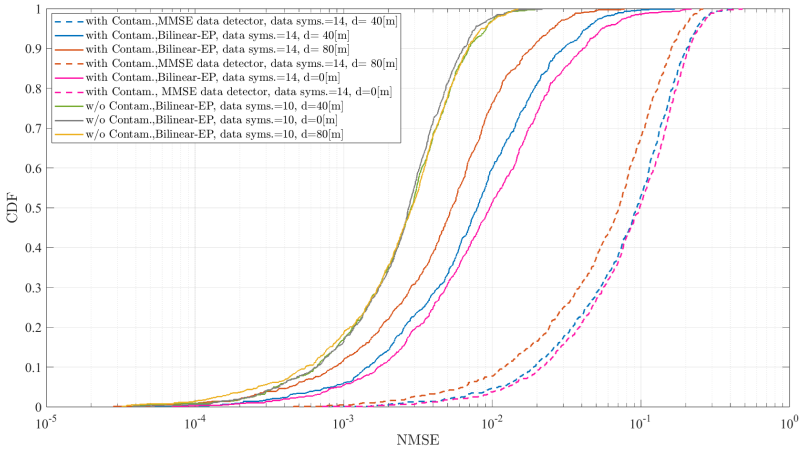


Figure 15: CDF of NMSE with minimum distance constraint between UEs.

6.2 Reduced Fronthaul Load

Distributed signal processing algorithms in CF-MaMIMO systems aim to enable scalability and reduce the computational load on the CPU by allowing APs to locally process received signals before forwarding information to the CPU. However, this approach presents significant challenges. A major issue is the frequent usage of fronthaul links, as messages are iteratively exchanged between the APs and the CPU for JCD via the bilinear-EP algorithm. This iterative communication increases fronthaul load, particularly when many iterations are needed for improved performance. Hence, developing efficient scheduling strategies that optimize the balance between local and centralized processing is crucial for addressing these issues and ensuring the scalability and practicality of CF-MaMIMO for future wireless networks.

The proposed modified scheduling for the bilinear-EP algorithm introduces an alternative fronthaul communication strategy to reduce the fronthaul load by selectively enabling message exchanges between APs and the CPU. In the original bilinear-EP algorithm, fronthaul communication occurs during every EP iteration, i.e., the number of fronthaul uses n_{fh} is equal to the number of EP iterations $i_{\max}$, which can be demanding in terms of fronthaul capacity. The proposed strategy reduces the fronthaul load by allowing fronthaul communication only at specific EP iterations, determined by the scheduling rule

$$\text{round}\left(i \cdot \frac{i_{\max}}{n_{\text{fh}}}\right), \quad i \in \{1, 2, \ldots, n_{\text{fh}}\}. \tag{48}$$

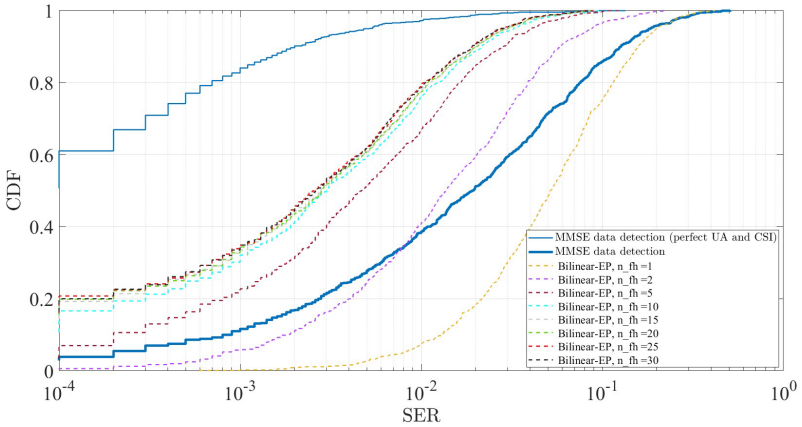


Figure 16: CDF of SER with reduced fronthaul load.

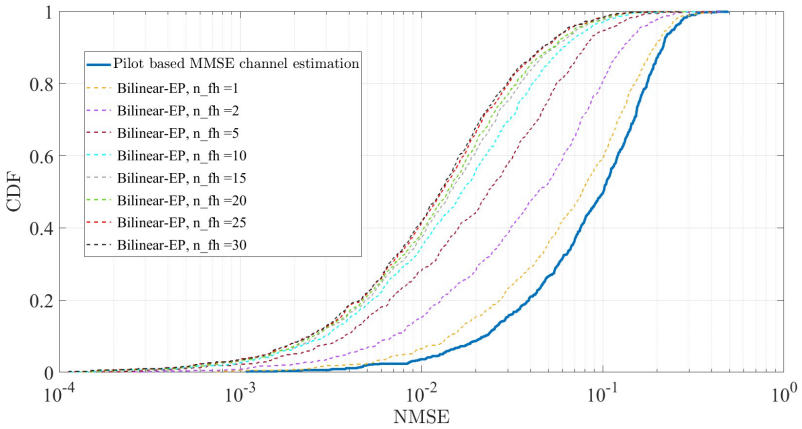


Figure 17: CDF of NMSE with reduced fronthaul load.

In this modified framework, the scheduling of messages the $m_{\Psi_{z_{l,kt}};x_{kt}}$ (from APs to CPU) and $m_{x_{kt};\Psi_{z_{l,kt}}}$ (from CPU to APs) is adjusted. When the fronthaul communication is disabled, updated information is not exchanged; instead, the algorithm relies on values from the previous iteration. This selective communication reduces computational complexity at the CPU and fronthaul usage while maintaining effective iterative refinement of channel and data estimates, ensuring a balance between system performance and efficiency.

For the following simulations, the same parameters are used as in Section 5.3 with altered $T_d = 14$ and random BPSK pilot sequences of length $T_p = 4$. Furthermore, we set the number of EP iterations to $i_{\max} = 30$ and vary the number of fronthaul uses n_{fh}. The simulation results are shown in Figs. 16 and 17. The bilinear-EP algorithm outperforms the centralized MMSE data detector, even with a small number of fronthaul uses since the MMSE data detector relies solely on the initial channel estimates for data detection, while bilinear-EP refines estimates through iterative processing. Furthermore, it can be observed that the bilinear-EP algorithm with $n_{\text{fh}} = 10$ fronthaul uses already performs as good as with exhaustive fronthaul communication $n_{\text{fh}} = i_{\max} = 30$, i.e., reducing the fronthaul load by 67 % with negligible performance degradation. This shows the efficacy of the proposed fronthaul scheduling scheme and the general possibility to reduce the fronthaul load for iterative distributed algorithms.

7 Conclusion

In this work, we considered a CF-MaMIMO system and tackled the quest of low-complexity JCD with near-optimal performance and robustness to PC. We derived a blind or semi-blind distributed JCD algorithm by formulating the problem in the framework of bilinear inference and obtaining the solution as an unfolding of a message passing algorithm incorporating EP-rules over a factor graph. The appealing features of the proposed algorithm stem from our choice of the approximate posterior joint distribution of data symbols and channels. Our simulation results show that the proposed scheme significantly outperforms the selected baseline schemes based on the detector in [30], which assumes perfect CSI, and the ICD algorithm in [31]. Additionally, bilinear-EP JCD has polynomial computational complexity and allows for straightforward embedding of state-of-the art SISO decoders, such that the decoding complexity can also be kept of polynomial order. This enables further investigations on the rate that can be achieved using our proposed JCD algorithm. Furthermore, we extensively investigated the performance under PC and proposed corresponding modifications. It consistently outperformed optimal linear detectors in [39] and [17] and state-of-the-art JCD algorithms. We also compared systems employing orthogonal and non-orthogonal pilots and showed that non-orthogonal sequences provided significant performance gains. Additionally, we introduced a new metric to quantify the impact of PC on iterative JCD algorithms and demonstrated its relevance and consistency. Then, we further analyzed the performance of the proposed bilinear-EP algorithm under minimum UE distance constraints, which effects the level of PC in the network, and under reduced fronthaul capabilities. The results showed that the algorithm can effectively exploit the geographical UE separation for pilot decontamination and that a reduction of the fronthaul load is feasible with negligible performance degradation.

A Derivation of Factor-to-Variable Messages

In this section, we derive some of the factor-to-variable messages. When it does not cause ambiguity, we adopt the following abbreviated notation $m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} = m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}(\boldsymbol{z}_{lkt})$.

Derivation of message $m_{\Psi_{1,lkt};x_{kt}}$
Message $m_{\Psi_{1,lkt};x_{kt}}$ is obtained by applying (8) and (9). This computation requires the knowledge of messages from variable nodes to factor nodes, namely, $m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}$, $m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}$, and $m_{x_{kt};\Psi_{1,lkt}}$ as we can evince from the factor graph in Fig. 2. In the following, we derive these messages by applying (7). Thus, we obtain $m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}(\boldsymbol{h}_{lk}) \propto m_{\Psi_{2,lk};\boldsymbol{h}_{lk}} \prod_{\tilde{t}\neq t} m_{\Psi_{1,lk\tilde{t}};\boldsymbol{h}_{lk}}$ which is a Gaussian distribution with covariance matrix and mean given by (15) and (16), respectively. As product of Gaussian distributions, the mean and covariance matrix above are obtained by applying the Gaussian multiplication lemma, see, e.g., [25]. Similarly, we derive the other variable-to-factor messages as $m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} = m_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}$ and

$$m_{x_{kt};\Psi_{1,lkt}} \propto m_{\Psi_{3,kt};x_{kt}} \prod_{\tilde{l}\neq l} m_{\Psi_{1,\tilde{l}kt};x_{kt}} \tag{49}$$

which is a categorical distribution. We normalize it as follows

$$m_{x_{kt};\Psi_{1,lkt}}(x_{kt}) = \frac{m_{\Psi_{3,kt};x_{kt}}(x_{kt}) \prod_{\tilde{l}\neq l} m_{\Psi_{1,\tilde{l}kt};x_{kt}}(x_{kt})}{\sum_{x_{kt}\in\mathcal{S}} m_{\Psi_{3,kt};x_{kt}}(x_{kt}) \prod_{\tilde{l}\neq l} m_{\Psi_{1,\tilde{l}kt};x_{kt}}(x_{kt})}. \tag{50}$$

By utilizing the variable-to-factor messages impinging on the factor node $\Psi_{1,lkt}$ computed above, we apply (9) to obtain the distribution before projection as follows

$$\begin{aligned} q_{\Psi_{1,lkt};x_{kt}} &\propto \int \delta(\boldsymbol{z}_{lkt} - x_{kt}\boldsymbol{h}_{lk}) \\ &\quad \times m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} d\boldsymbol{z}_{lkt} d\boldsymbol{h}_{lk} m_{x_{kt};\Psi_{1,lkt}}. \\ &\propto \int \frac{1}{|x_{kt}|^{2N}} m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \left(\frac{\boldsymbol{z}_{lkt}}{x_{kt}}\right) \\ &\quad \times m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(\boldsymbol{z}_{lkt}) d\boldsymbol{z}_{lkt} m_{x_{kt};\Psi_{1,lkt}}, \end{aligned} \tag{51}$$

where the last expression is derived by applying the sifting property of the Dirac delta function, see, e.g., [40]. Then, the factor $m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}\left(\frac{\boldsymbol{z}_{lkt}}{x_{kt}}\right)$ in (51) can be rewritten as shown in (36) at the top of the page. Therefore, using the Gaussian multiplication lemma we obtain

$$\begin{aligned} q_{\Psi_{1,lkt};x_{kt}} &\propto \int \mathcal{N}(\boldsymbol{z}_{lkt}; x_{kt}\boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}, |x_{kt}|^2 \boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}) \\ &\quad \times m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(\boldsymbol{z}_{lkt}) d\boldsymbol{z}_{lkt} \cdot m_{x_{kt};\Psi_{1,lkt}} \\ &\quad \propto \mathcal{N}(\boldsymbol{0}; \boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} - x_{kt}\boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}, \boldsymbol{C}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} \\ &\quad + |x_{kt}|^2 \boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}) \cdot m_{x_{kt};\Psi_{1,lkt}}. \end{aligned}$$

Next, we observe that $q_{\Psi_{1,lkt};x_{kt}}$ is already a categorical distribution in the exponential family. Then, the projection operator in (8) leaves its argument unchanged, i.e., $\text{proj}\{q_{\Psi_{1,lkt};x_{kt}}\} = q_{\Psi_{1,lkt};x_{kt}}$. Finally, by applying (8), we obtain the message $m_{\Psi_{1,lkt};x_{kt}}$ which is given in (30).

Derivation of message $m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$

For the derivation of message $m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$ we consider again factor node $\Psi_{1,lkt}$ in Fig. 2. Thus, the same variable-to-factor messages computed above are necessary to determine $q_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$. The distribution before projection onto the family of exponential functions is

$$\begin{aligned} q_{\Psi_{1,lkt};z_{lkt}} &\propto \int \delta(\boldsymbol{z}_{lkt} - x_{kt}\boldsymbol{h}_{lk}) \\ &\quad \times m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} m_{x_{kt};\Psi_{1,lkt}} dx_{kt} d\boldsymbol{h}_{lk}. \\ &\quad \propto \int \frac{1}{|x_{kt}|^{2N}} m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \left(\frac{\boldsymbol{z}_{lkt}}{x_{kt}}\right) \\ &\quad \times m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(\boldsymbol{z}_{lkt}) m_{x_{kt};\Psi_{1,lkt}} dx_{kt} \\ &\quad \propto \sum_{x'_{kt} \in \mathcal{S}} m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(\boldsymbol{z}_{lkt}) \frac{1}{|x'_{kt}|^{2N}} \\ &\quad \times m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} \left(\frac{\boldsymbol{z}_{lkt}}{x'_{kt}}\right) m_{x_{kt};\Psi_{1,lkt}}(x'_{kt}) \\ &\quad \propto \sum_{x'_{kt} \in \mathcal{S}} \tilde{\omega}(x'_{kt}) \mathcal{N}(\boldsymbol{z}_{lkt}; \tilde{\boldsymbol{\mu}}(x_{kt}), \tilde{\boldsymbol{C}}(x_{kt})) \end{aligned} \tag{30}$$

We observe that $q_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$ is a Gaussian mixture in $|\mathcal{S}|$ components with the parameters $\tilde{\boldsymbol{C}}(x_{kt})$ and $\tilde{\boldsymbol{\mu}}(x_{kt})$ which depend on the specific value x_{kt} and are defined in (22) and (23), respectively, which is obtained by applying again (36) and then the Gaussian multiplication lemma. Additionally, each component of the Gaussian mixture distribution is weighted by the unnormalized factor $\tilde{\omega}(x_{kt})$ given in (37) at the top of the page. Then, as in (9) we have to project the Gaussian mixture distribution onto the family of Gaussian distributions, i.e., we need to determine the Gaussian distribution $\mathcal{N}(\boldsymbol{z}_{lkt}; \hat{\boldsymbol{z}}_{lkt}, \boldsymbol{\Sigma}_{lkt}) := \text{proj}\{q_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}\}$ whose moments are matched to the distribution $q_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$. Denoting with $\omega(x_{kt})$ the normalized weights obtained from $\tilde{\omega}(x_{kt})$ given in (21), the parameters $\hat{\boldsymbol{z}}_{lkt}$ and $\boldsymbol{\Sigma}_{lkt}$ of the moment matched distribution are shown in (19) and

(20), respectively. Finally, we have

$$m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} \propto \frac{\text{proj}\{q_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}\}}{m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}} = \frac{\mathcal{N}(\boldsymbol{z}_{lkt};\hat{\boldsymbol{z}}_{lkt},\boldsymbol{\Sigma}_{lkt})}{m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}}.$$

The parameters $\boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$ and $\boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$ of the updated message are then given by the Gaussian multiplication lemma shown in (17) and (18), respectively.

Derivation of message $m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}$
Lastly, the only message left associated with factor node $\Psi_{1,lkt}$ is $m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}$. Analogously to the derivation discussed before we have for the distribution to be projected

$$\begin{aligned}
q_{\Psi_{1,lkt};h_{lkt=}} &\propto \int \delta(\boldsymbol{z}_{lkt} - x_{kt}\boldsymbol{h}_{lk}) \\
&\quad \times m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}} m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}} m_{x_{kt};\Psi_{1,lkt}} dx_{kt} d\boldsymbol{z}_{lkt}. \\
&\propto \int m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}(\boldsymbol{h}_{lk}) \\
&\quad \times m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(x_{kt}\boldsymbol{h}_{lk}) m_{x_{kt};\Psi_{1,lkt}} dx_{kt} \\
&\propto \sum_{x'_{kt}\in\mathcal{S}} m_{\boldsymbol{h}_{lk};\Psi_{1,lk}}(\boldsymbol{h}_{lk}) \\
&\quad \times m_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}(x'_{kt}\boldsymbol{h}_{lk}) m_{x_{kt};\Psi_{1,lkt}}(x'_{kt})
\end{aligned}$$

$$\begin{aligned}
&\propto \sum_{x'_{kt}\in\mathcal{S}} m_{x_{kt};\Psi_{1,lkt}}(x'_{kt}) \\
&\times \mathcal{N}(\boldsymbol{h}_{lk};\boldsymbol{\mu}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}},\boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}) \\
&\times \frac{1}{|x'_{kt}|^{2N}} \mathcal{N}(\boldsymbol{h}_{lk};\frac{\boldsymbol{\mu}_{\boldsymbol{z}_{lkt};\Psi_{1,lkt}}}{x'_{kt}},\frac{\boldsymbol{C}_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}}{|x'_{kt}|^2}) \\
&\propto \sum_{x'_{kt}\in\mathcal{S}} \tilde{\omega}(x'_{kt}) \mathcal{N}(\boldsymbol{h}_{lk};\bar{\boldsymbol{\mu}}(x_{kt}),\bar{\boldsymbol{C}}(x_{kt})).
\end{aligned}$$

This is a Gaussian mixture in $|S|$ components as well with the parameters $\bar{\boldsymbol{C}}(x_{kt})$ and $\bar{\boldsymbol{\mu}}(x_{kt})$ which depend on the specific value x_{kt} and are defined in (28) and (29), respectively, which is obtained by applying the Gaussian multiplication lemma. Additionally, each component of the Gaussian mixture distribution is weighted by the unnormalized factor $\tilde{\omega}(x_{kt})$ given in (37). We need to project this down onto the Gaussian distribution $\mathcal{N}(\boldsymbol{h}_{lk};\hat{\boldsymbol{h}}_{lk},\hat{\boldsymbol{\Sigma}}_{lkt}) := \text{proj}\{q_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}\}$ whose moments are matched to the distribution $q_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}$. The parameters $\hat{\boldsymbol{h}}_{lk}$ and $\hat{\boldsymbol{\Sigma}}_{lk}$ of the moment matched distribution are shown in (26) and (27), respectively. Finally, we have

$$m_{\Psi_{1,lkt};\boldsymbol{h}_{lk}} \propto \frac{\text{proj}\{q_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}\}}{m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}} = \frac{\mathcal{N}(\boldsymbol{h}_{lk};\hat{\boldsymbol{h}}_{lk},\hat{\boldsymbol{\Sigma}}_{lk})}{m_{\boldsymbol{h}_{lk};\Psi_{1,lkt}}}.$$

The parameters $\boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}$ and $\boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{h}_{lk}}$ of the updated message are then given by the Gaussian multiplication lemma shown in (24) and (25), respectively.

Derivation of message $m_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}$
The product of variable messages entering the node $\Psi_{0,lt}$ is

$$\prod_{k'=1}^{K} m_{\boldsymbol{z}_{lkt};\Psi_{0,lt}} = \prod_{k'=1}^{K} m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}$$

Therefore, we have for the distribution to be projected

$$\begin{aligned} q_{\Psi_{0,lt};\boldsymbol{z}_{lkt}} &= \int \prod_{k'=1}^{K} m_{\Psi_{1,lk't};\boldsymbol{z}_{lk't}} \\ &\quad \times \mathcal{N}(\boldsymbol{y}_{lt}; \sum_{k'=1}^{K} \boldsymbol{z}_{lk't}, \sigma^2 \boldsymbol{I}_N) \prod_{\tilde{k}=1,\tilde{k}\neq k}^{K} d\boldsymbol{z}_{l\tilde{k}t} \\ &= m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} \int \prod_{k'=1,k'\neq k}^{K} m_{\Psi_{1,lk't};\boldsymbol{z}_{lk't}} \\ &\quad \times \mathcal{N}(\boldsymbol{y}_{lt}; \sum_{k'=1}^{K} \boldsymbol{z}_{lk't}, \sigma^2 \boldsymbol{I}_N) \prod_{\tilde{k}=1,\tilde{k}\neq k}^{K} d\boldsymbol{z}_{l\tilde{k}t} \end{aligned}$$

By repeatedly applying the Gaussian multiplication lemma on the functions being integrated, we obtain

$$\begin{aligned} q_{\Psi_{0,lt};\boldsymbol{z}_{lkt}} &\propto m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}} \\ &\quad \times \mathcal{N}(\boldsymbol{z}_{lkt}; \boldsymbol{y}_{lt} - \sum_{\tilde{k}=1,\tilde{k}\neq k}^{K} \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}, \\ &\quad \sigma^2 \boldsymbol{I}_N + \sum_{\tilde{k}=1,\tilde{k}\neq k}^{K} \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}) \end{aligned}$$

Since $q_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}$, as a product of two Gaussians is already Gaussian and thus of the desired distribution, no further projection onto a Gaussian is necessary and we obtain the the updated message by

$$\begin{aligned} m_{\Psi_{0,lt};\boldsymbol{z}_{lkt}} &\propto \frac{q_{\Psi_{0,lt};\boldsymbol{z}_{lkt}}}{m_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}} \\ &= \mathcal{N}(\boldsymbol{z}_{lkt}; \boldsymbol{y}_{lt} - \sum_{\tilde{k}=1,\tilde{k}\neq k}^{K} \boldsymbol{\mu}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}, \\ &\quad \sigma^2 \boldsymbol{I}_N + \sum_{\tilde{k}=1,\tilde{k}\neq k}^{K} \boldsymbol{C}_{\Psi_{1,lkt};\boldsymbol{z}_{lkt}}). \end{aligned}$$

This concludes the update of all desired factor-to-variable messages.

B References

[1] 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 J. Signal Process.*, vol. 5, pp. 284–293, 2024.

[2] C. Forsch, Z. Zhao, D. Slock, and L. Cottatellucci, "Bayesian learning for pilot decontamination in cell-free massive MIMO," in *Proc. 28th Int. Workshop Smart Antennas (WSA)*, pp. 19–25, 2025.

[3] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.

[4] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, 2018.

[5] H. Yang and T. L. Marzetta, "Energy efficiency of massive MIMO: Cell-free vs. cellular," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC-Spring)*, pp. 1–5, 2018.

[6] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive MIMO networks: A survey of opportunities, challenges and solutions," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 611–652, 2022.

[7] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 942–953, 2014.

[8] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, 2018.

[9] R. Gholami, L. Cottatellucci, and D. Slock, "Favorable propagation and linear multiuser detection for distributed antenna systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020.

[10] R. Gholami, L. Cottatellucci, and D. Slock, "Channel models, favorable propagation and MultiStage linear detection in cell-free massive MIMO," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2020.

[11] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[12] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012.

[13] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[14] L. Cottatellucci, R. R. Müller, and M. Vehkapera, "Analysis of pilot decontamination based on power control," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC-Spring)*, 2013.

[15] R. R. Müller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, 2014.

[16] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2990–3003, 2016.

[17] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, 2020.

[18] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends® Signal Process.*, vol. 14, no. 3-4, pp. 162–472, 2021.

[19] H. Wang, A. Kosasih, C.-K. Wen, S. Jin, and W. Hardjawana, "Expectation propagation detector for extra-large scale massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2036–2051, 2020.

[20] T. P. Minka, *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Inst. Technol., Cambridge, 2001.

[21] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell. (UAI)*, pp. 362–369, 2001.

[22] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, 2014.

[23] K. Ghavami and M. Naraghi-Pour, "MIMO detection with imperfect channel state information using expectation propagation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8129–8138, 2017.

[24] 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, 2018.

[25] 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 Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6145–6161, 2020.

[26] Y. Zhang, Z. Yuan, Q. Guo, Z. Wang, J. Xi, and Y. Li, "Bayesian receiver design for grant-free NOMA with message passing based structured signal estimation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8643–8656, 2020.

[27] Y. Dong, H. Li, C. Gong, X. Wang, and X. Dai, "An enhanced fully decentralized detector for the uplink M-MIMO system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13030–13042, 2022.

[28] H. Li, Y. Dong, C. Gong, X. Wang, and X. Dai, "Decentralized groupwise expectation propagation detector for uplink massive MU-MIMO systems," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 5393–5405, 2023.

[29] A. Kosasih, V. Miloslavskaya, W. Hardjawana, V. Andrean, and B. Vucetic, "Improving cell-free massive MIMO detection performance via expectation propagation," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC2021-Fall)*, pp. 1–5, 2021.

[30] H. He, H. Wang, X. Yu, J. Zhang, S. H. Song, and K. B. Letaief, "Distributed expectation propagation detection for cell-free massive MIMO," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021.

[31] H. He, X. Yu, J. Zhang, S. H. Song, and K. B. Letaief, "Cell-free massive MIMO detection: A distributed expectation propagation approach," *arXiv*, 2023.

[32] K. Ghavami and M. Naraghi-Pour, "Noncoherent SIMO detection by expectation propagation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017.

[33] C. Schülke, *Statistical physics of linear and bilinear inference problems*. PhD thesis, Université Paris Diderot, Sapienza Università di Roma, 2016.

[34] T. Minka, "Divergence measures and message passing," Tech. Rep. MSR-TR-2005-173, Microsoft Research Ltd., 2005.

[35] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, "Approximate inference techniques with expectation constraints," *J. Stat. Mech.*, vol. 2005, no. 11, p. P11015, 2005.

[36] H. Iimori, T. Takahashi, K. Ishibashi, G. T. F. de Abreu, and W. Yu, "Grant-free access via bilinear inference for cell-free MIMO with low-coherence pilots," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7694–7710, 2021.

[37] 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 *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021.

[38] Z. Zhao and D. Slock, "Decentralized message-passing for semi-blind channel estimation in cell-free systems based on Bethe free energy optimization," in *Proc. 58th Asilomar Conf. Signals, Syst., and Comput.*, pp. 1443–1447, 2024.

[39] A. Á. Polegre, L. Sanguinetti, and A. G. Armada, "Pilot decontamination processing in cell-free massive MIMO," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3990–3994, 2021.

[40] C. Candan, "Proper definition and handling of Dirac delta functions [lecture notes]," *IEEE Signal Process. Mag.*, vol. 38, no. 3, pp. 186–203, 2021.