Deliverable 4.2: Semi-blind Algorithms and Methods

Christian Forsch, Zilu Zhao, Laura Cottatellucci, Dirk Slock

January 30, 2026

Abstract

This deliverable presents results related to Tasks 4.2 and 4.3 of the CellFree6G project and investigates distributed semi-blind algorithms for joint activity detection, channel estimation, and data detection (JACD) in grant-free cell-free massive multiple-input multiple-output (GF-CF-MaMIMO) systems. These systems are especially relevant for massive Internet-of-Things (IoT) scenarios, where a large number of sporadically active devices prevents the use of orthogonal pilot sequences and necessitates decentralized signal processing. These systems are fundamentally challenged by (i) pilot contamination due to non-orthogonal pilot reuse, (ii) stringent scalability constraints arising from large numbers of sporadically active devices, and (iii) distributed network architectures.

To address these challenges, the deliverable presents novel Bayesian approximated inference algorithms based on expectation propagation (EP) that enable efficient and scalable JACD without relying on centralized processing of received signals. The proposed framework exploits a distributed factor-graph representation of the joint posterior distribution, allowing channel estimation and refinement to be performed locally at the access points, while user activity detection and data detection rely on lightweight message exchanges and simple fusion operations. This architectural separation significantly reduces fronthaul requirements and supports decentralized implementations with cooperation among processing units.

Two distributed EP-based algorithms are proposed. The first employs Gaussian approximations for channels, while the second introduces a Bernoulli-Gaussian (BG) channel model that explicitly captures channel sparsity induced by sporadic user activity. A key methodological contribution is the derivation of an exponential-family representation of the BG distribution, which enables its seamless integration into the EP framework and leads to closed-form message updates with polynomial complexity. The proposed algorithms are inherently suitable for GF-CF-MaMIMO systems equipped with an arbitrary number of antennas at the access points (APs), as requested in Task 4.3.

The deliverable provides a comprehensive performance evaluation based on numerical simulations, alongside an analysis of computational complexity and information-exchange overhead. Simulation results demonstrate that the proposed distributed algorithms significantly outperform centralized linear receivers and state-of-the-art baselines, particularly under severe pilot contamination, achieving near-optimal performance. Overall, the results demonstrate that the proposed semi-blind distributed approaches effectively mitigate pilot contamination, achieve near-optimal performance, and offer a favorable trade-off between performance, scalability, and implementation complexity, thereby supporting practical deployments of large-scale GF-CF-MaMIMO systems.

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

The explosive growth of Internet-of-Things (IoT) devices and the evolving requirements of massive and critical machine-type communications (MTC) for ultra-low latency, high reliability, and energy efficiency call for new communication system designs [3]. Traditional grant-based multiple access schemes fail to meet these high demands due to signaling overhead and latency, especially under dense device deployments and sporadic burst traffic. Hence, grant-free random access has emerged as a compelling alternative, facilitating efficient resource utilization without the need for an explicit scheduling grant [4–6]. Simultaneously, the cell-free massive multiple-input multiple-output (CF-MaMIMO) network architecture, in which a large number of geographically distributed access points (APs) jointly serve a potentially large number of user equipments (UEs), enables ubiquitous coverage and energy-efficient communication [7–9]. Grant-free CF-MaMIMO (GF-CF-MaMIMO) systems combine the advantages of grant-free random access and CF-MaMIMO and are promising candidates for next-generation IoT networks [10, 11]. At the same time, CF-MaMIMO architectures implemented with centralized processing face fundamental scalability limitations [12] which are further exacerbated in GF-CF-MaMIMO systems, where a large number of sporadically active UEs must be jointly processed, motivating the design of distributed algorithms.

In addition to the classical communication tasks of channel estimation and data detection, grant-free random access requires identifying the set of active UEs transmitting data. To accomplish this, UEs transmit pilot sequences which can be used to jointly estimate both device activities and channels, a process known as joint activity detection and channel estimation (JAC), e.g., [13, 14]. However, the large number of connected devices in massive MTC prevents the use of orthogonal pilot sequences, leading to *pilot contamination (PC)* which degrades overall system performance and further complicates activity and data detection. PC occurs in any communication system where UEs transmit non-orthogonal pilot sequences. In centralized MaMIMO systems, *channel hardening* and *favorable propagation* have been used to alleviate the PC problem for grant-based multiple access in [15–19]. In contrast, such properties typically do not hold in CF-MaMIMO [20–23], rendering such decontamination methods ineffective. The use of centralized minimum mean squared error (MMSE) processing in grant-based centralized MaMIMO was shown to mitigate PC under practical channel statistics in the asymptotic regime of infinitely many antennas [24, 25]. Similar results were obtained for distributed processing in CF-MaMIMO [26]. However, PC still remains a major practical challenge [12], especially in scalable CF-MaMIMO systems, which inherently rely on distributed signal processing, with a limited number of APs and antennas per AP. This problem becomes even more severe in GF-CF-MaMIMO where the additional need for UE activity detection arises and a much larger number of UEs are simultaneously processed, thereby intensifying PC. These challenges motivate the development of distributed pilot decontamination schemes that can effectively mitigate PC while jointly supporting scalability and low fronthaul traffic in GF-CF-MaMIMO architectures.

One promising approach to mitigate PC is to exploit not only the pilot symbols for channel estimation and UE activity detection but also the received data symbols. In grant-based multiple access, where the UE activities are known beforehand and only channels and data symbols need to be estimated, this approach is referred to as joint channel estimation and data detection (JCD). Several studies have explored JCD in CF-MaMIMO systems. The authors in [27] investigated semi-blind methods for JCD in CF-MaMIMO and derived conditions for semi-blind identifiability. In [28], a JCD scheme based on forward-backward splitting was developed, exploiting the sparsity of CF-MaMIMO channels and employing non-orthogonal pilot sequences. A distributed expectation propagation (EP)-based semi-blind JCD algorithm for CF-MaMIMO systems was presented in [29] and was further refined and analyzed under PC in [30]. The authors in [31] combined variational Bayes and EP to develop a semi-blind JCD algorithm based on Bethe free energy optimization.

In grant-free random access, the received data symbols can also be used to additionally enhance the UE activity detection, a task referred to as *joint activity detection, channel estimation, and data detection (JACD)*. In the literature, different JACD schemes have been proposed. One class of algorithms is based on sequences spreading data whereby the data symbols of each UE are

multiplied by a unique spreading signature, e.g., [32–34]. These unique, generally non-orthogonal signatures spread the transmitted symbols in the time/frequency domain, enabling multiple UEs to share the same resources. However, spreading data reduces the spectral efficiency and limits the achievable data rate per UE. Hence, in the following, we focus on JACD schemes that do not rely on spreading data sequences. For centralized MaMIMO systems, several non-spread JACD schemes based on bilinear generalized approximate message passing (BiG-AMP) have been proposed. These include approaches that combine BiG-AMP with loopy belief propagation (BP) [35], introduce vector nodes with correlated channels [36], or iteratively exchange information with the channel decoder [37]. In [38], bilinear Gaussian BP (BiGaBP) was applied to the JACD problem in GF-CF-MaMIMO networks, employing low-coherence pilot sequences. Here, the beliefs of the Bernoulli-Gaussian (BG)-distributed channels and the categorically distributed data symbols are approximated by Gaussian distributions, and a corresponding message-passing algorithm was derived. The authors in [39] developed two JACD algorithms based on forward-backward splitting and deep unfolding for hyperparameter optimization. The proposed centralized algorithms employ Laplace distributions to model the sparsity of channels and data.

In this deliverable, we present two novel distributed JACD algorithms for GF-CF-MaMIMO networks which have been published or submitted for publication in [1, 2]. We formulate the JACD task as a maximum a posteriori (MAP) estimation and detection problem and, then, solve it approximately using EP, a Bayesian learning technique that iteratively computes tractable approximations of factorized probability distributions employing exponential family distributions [40, 41]. To this end, we factorize the a posteriori probability (APP) distribution of user activities, channels, and data symbols to enable a tractable joint inference on a factor graph based on EP. This factor graph approach enables an inherent distributed implementation of the proposed algorithms which is suitable for *decentralized* and *scalable* CF-MaMIMO systems with baseband-processing capabilities at the APs.

Parts of this deliverable have been published in [2]. The remaining material, submitted for publication in [1], introduces as its main novel contribution compared to [2] the adoption of BG distributions within the EP framework to more accurately capture the sparsity of the effective UE channels. This modeling refinement leads to a significant improvement of the JACD performance.

The approach presented in [2] relies on Gaussian approximations for the effective UE channels, which do not explicitly capture channel sparsity and yields the JACD-EP algorithm. To enable EP inference with BG random variables, in [1] we first derive the exponential family representation of the BG distribution and then apply EP message passing to develop the JACD-EP-BG algorithm. Finally, we conduct an extensive performance analysis via Monte Carlo simulations and compare the proposed algorithms with optimal linear and state-of-the-art EP-based algorithms. The contributions of this deliverable are summarized as follows:

- We formulate the JACD problem as a MAP estimation and detection problem and propose a factor graph that enables EP-based inference of the approximate APP distributions of UE activities, channels, and data symbols. The proposed formulation yields a natural task partition between APs and central processing unit (CPU), leading to a distributed algorithm that keeps the computational load at the CPU linear in the number of data signals, supports scalability, and facilitates a structured and efficient design of the information exchanged over the fronthaul.

- We present the exponential family representation of the BG distribution, enabling simple and closed-form multiplication and division of BG distributions and their seamless integration into the EP framework. The BG model provides a more effective representation of sparse random variables such as channels between UEs with unknown activity state and APs compared to Gaussian approximations.

- We develop the two distributed JACD algorithms by applying EP on factor graphs with Gaussian and BG channel models, yielding the proposed JACD-EP and JACD-EP-BG algorithm, respectively. Both algorithms exhibit a polynomial computational complexity and enable scalable signal processing in GF-CF-MaMIMO systems.

- We present an extended numerical analysis of the proposed algorithms' performance for different pilot sequence lengths, modeling different levels of PC. As benchmark, we adopt centralized linear MMSE processing, including genie-aided variants, and show the superior performance of the proposed distributed algorithms, particularly under severe PC. Since centralized MMSE processing provides an upper bound on the performance of distributed linear MMSE schemes, no distributed linear MMSE benchmark is included. The results show that linear MMSE processing is not sufficient to effectively combat PC, highlighting the need for nonlinear estimation and detection schemes.

The remainder of this deliverable is organized as follows. In Section 2 and 3, we introduce the GF-CF-MaMIMO system model and formulate the inference problem, respectively. In Section 4, we derive the exponential family representation of the BG distribution. Then, in Section 5, we propose the JACD-EP and JACD-EP-BG algorithms. In Section 6, the performance of the proposed algorithms is evaluated via Monte Carlo simulations. Finally, conclusions are drawn in Section 8. The detailed analytical derivation of each algorithmic step is provided in the appendices.

For the sake of clarity, each section header specifies whether the corresponding contributions originate from [1, 2] or are common to both manuscripts.

Notation: Lower case, bold lower case, and bold upper case letters, e.g., $x, \mathbf{x}, \mathbf{X}$, represent scalars, vectors, and matrices, respectively. $\mathbf{I}_N$ is the N-dimensional identity matrix. $\mathrm{diag}(\cdot)$ denotes a diagonal matrix whose main diagonal entries are given by the elements inside the brackets. $\delta(\cdot)$ represents the Dirac delta function. The indicator function $\mathbb{1}_{(\cdot)}$ equals one if the condition in the subscript is satisfied and zero otherwise. $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ denote the transpose and conjugate transpose (Hermitian) operation, respectively. The trace of a matrix $\mathbf{X}$ is written as $\mathrm{tr}\{\mathbf{X}\}$. $|\mathcal{S}|$ stands for the cardinality of the set $\mathcal{S}$. $E\{\cdot\}$ denotes the expectation operator. $\mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) = |\pi\mathbf{C}|^{-1}\mathrm{e}^{-(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{H}}\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$ represents the probability density function (PDF) of a proper complex-valued Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. $\pi(x)$ denotes the probability mass function (PMF) of a categorical random variable x. The notation $x \sim p$ indicates that the random variable x follows the distribution p. In a factor graph, the message sent from the factor node Ψ_α to the variable node $\mathbf{x}_\beta$ is denoted as $m_{\Psi_\alpha;\mathbf{x}_\beta}$ and consists of parameters of the distribution $p_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$ in the exponential family. Note that we adopt the same subscript also for the parameters, e.g., $p_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta) = \mathcal{CN}\big(\mathbf{x}_\beta|\boldsymbol{\mu}_{\Psi_\alpha;\mathbf{x}_\beta}, \mathbf{C}_{\Psi_\alpha;\mathbf{x}_\beta}\big)$ with mean $\boldsymbol{\mu}_{\Psi_\alpha;\mathbf{x}_\beta}$ and covariance matrix $\mathbf{C}_{\Psi_\alpha;\mathbf{x}_\beta}$ or $p_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta) = \pi_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$ with probability values $\pi_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$ for a Gaussian or categorical random variable $\mathbf{x}_\beta$, respectively. Analogous notation holds for variable-to-factor messages $m_{\mathbf{x}_\beta;\Psi_\alpha}$.

2 System Model

We consider the uplink of a GF-CF-MaMIMO network comprising L geographically distributed APs, each equipped with N antennas, serving K synchronized single-antenna UEs as illustrated in Fig. 1. All APs are connected to a CPU via fronthaul links. Due to the sporadic traffic characteristics of the network, only a subset of the K UEs is active and transmits data simultaneously. Let $\mathbf{Y}_l = [\mathbf{y}_{l,1} \cdots \mathbf{y}_{l,T}] \in \mathbb{C}^{N\times T}$ denote the signal received at AP l over a coherence block of T channel uses. It is given by

$$\mathbf{Y}_l = \mathbf{H}_l\mathbf{U}\mathbf{X} + \mathbf{N}_l = \sum_{k=1}^{K} \mathbf{h}_{l,k}u_k\mathbf{x}_k^{\mathsf{T}} + \mathbf{N}_l, \tag{1}$$

where $\mathbf{H}_l = [\mathbf{h}_{l,1} \cdots \mathbf{h}_{l,K}] \in \mathbb{C}^{N\times K}$ denotes the channel matrix of AP l with $\mathbf{h}_{l,k} \in \mathbb{C}^{N\times 1}$ being the channel between AP l and UE k. The diagonal activity matrix $\mathbf{U} = \mathrm{diag}(u_1, \ldots, u_K) \in \{0, 1\}^{K\times K}$ contains the binary activity indicators where $u_k = 1$ if UE k is active and $u_k = 0$ otherwise. $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_K]^{\mathsf{T}} \in \mathbb{C}^{K\times T}$ is the transmit symbol matrix with $\mathbf{x}_k \in \mathbb{C}^{T\times 1}$ representing the transmit sequence of UE k and $\mathbf{N}_l \in \mathbb{C}^{N\times T}$ is the matrix of independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) elements $n \sim \mathcal{CN}\big(n|0, \sigma_n^2\big)$. The channels and the user activity indicators are constant during the channel coherence interval. We assume block Rayleigh

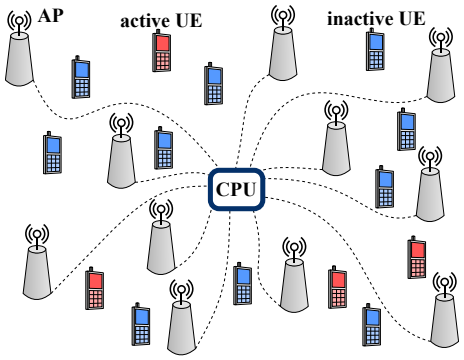


Figure 1: GF-CF-MaMIMO network with geographically distributed APs and UEs exhibiting different activity states.

fading channels, i.e., $\mathbf{h}_{l,k} \sim p_{h_{l,k}}(\mathbf{h}_{l,k}) = \mathcal{CN}(\mathbf{h}_{l,k}|\mathbf{0}_N, \boldsymbol{\Xi}_{l,k})$, where $\boldsymbol{\Xi}_{l,k} \in \mathbb{C}^{N\times N}$ is the spatial correlation matrix and $\xi_{l,k} = \frac{1}{N}\mathrm{tr}\{\boldsymbol{\Xi}_{l,k}\}$ is the associated large-scale fading coefficient. The activity indicator u_k is drawn from a Bernoulli distribution, $u_k \sim p_u(u_k) = (1-\lambda)\mathbb{1}_{u_k=0} + \lambda\mathbb{1}_{u_k=1}$ where λ denotes the user activity probability. The transmit symbol matrix consists of a pilot part $\mathbf{X}^p \in \mathbb{C}^{K\times T_p}$ with known pilot symbols x^p_{kt} and a data part $\mathbf{X}^d \in \mathcal{X}^{K\times T_d}$ with data symbols x^d_{kt} such that $\mathbf{X} = [\mathbf{X}^p\ \mathbf{X}^d]$ and $T_p + T_d = T$. The transmit symbols belong to the constellation $\mathcal{X}$ of cardinality $M = |\mathcal{X}|$ which does not contain the zero-symbol, i.e., $0 \notin \mathcal{X}$. All UEs transmit with the same average transmit power $\sigma_x^2 = E\{|x_{kt}|^2\}$. A similar decomposition holds for the receive matrix, i.e., $\mathbf{Y}_l = [\mathbf{Y}_l^p\ \mathbf{Y}_l^d]$ with received pilots $\mathbf{Y}_l^p \in \mathbb{C}^{N\times T_p}$ and received data $\mathbf{Y}_l^d \in \mathbb{C}^{N\times T_d}$. Finally, we assume $T_p < K$ since the number of UEs can be very large and assigning orthogonal pilot sequences to the UEs is generally impractical. This inevitably leads to the so-called PC effect.

3 Problem Formulation

The non-orthogonality of the pilot sequences degrades both user activity detection and channel estimation performance. The detrimental effect of PC can be mitigated by exploiting the detected data symbols to improve activity detection and channel estimation, which in turn enables a refinement of the detected data symbols and yields an iterative JACD approach. At first, we summarize the received signals from all APs in the global model

$$\mathbf{Y} = \mathbf{HUX} + \mathbf{N}, \tag{2}$$

where, for $\mathbf{A} \equiv \{\mathbf{Y}, \mathbf{H}, \mathbf{N}\}$, we define $\mathbf{A} = [\mathbf{A}_1^\mathsf{T} \cdots \mathbf{A}_L^\mathsf{T}]^\mathsf{T}$. For JACD, the receiver jointly estimates the user-activity, channel, and data matrices $\mathbf{U}$, $\mathbf{H}$, and $\mathbf{X}^d$, respectively. The MAP estimator is given by

$$(\hat{\mathbf{U}}, \hat{\mathbf{H}}, \hat{\mathbf{X}}^d) = \underset{\mathbf{U},\mathbf{H},\mathbf{X}^d}{\arg\max}\ p(\mathbf{U}, \mathbf{H}, \mathbf{X}^d|\mathbf{Y}, \mathbf{X}^p), \tag{3}$$

where the APP distribution $p(\mathbf{U}, \mathbf{H}, \mathbf{X}^d|\mathbf{Y}, \mathbf{X}^p)$ can be factorized using Bayes' rule as

$$p(\mathbf{U}, \mathbf{H}, \mathbf{X}^d|\mathbf{Y}, \mathbf{X}^p) \propto p(\mathbf{Y}|\mathbf{U}, \mathbf{H}, \mathbf{X}) \cdot p(\mathbf{U}) \cdot p(\mathbf{H}) \cdot p(\mathbf{X}). \tag{4}$$

Direct MAP inference in (3) is computationally intractable due to the high dimensionality of the involved variables. Hence, in Section 5, we propose two low-complexity approximate Bayesian learning methods for JACD. Since these methods rely on EP and BG distributions, we review the key properties of the exponential family and provide an exponential family representation of the BG distribution in the following section.

4 Bernoulli-Gaussian Distribution in Exponential Family Form [1]

In this section, we derive the exponential family representation of the BG distribution. To this end, we first recall the general form of an exponential family distribution and, then, show that the BG distribution can be expressed in exponential family form.

4.1 Exponential Family

A probability distribution belongs to the exponential family if it can be expressed as [42]

$$p(\mathbf{x}) = \mathrm{e}^{\boldsymbol{\eta}^{\mathsf{H}}\mathbf{u}(\mathbf{x})-A(\boldsymbol{\eta})}, \tag{5}$$

where $\boldsymbol{\eta}$ is the vector of natural parameters, $\mathbf{u}(\mathbf{x})$ is the vector of sufficient statistics, and $A(\boldsymbol{\eta})$ is the log-partition function. Exponential family distributions enjoy convenient properties, e.g., they allow simple multiplications and divisions of probability distributions, which makes them attractive in the EP framework. Two prominent members of the exponential family are the Bernoulli and Gaussian distributions reviewed below.

Let $x \in \{0, 1\}$ be a Bernoulli-distributed random variable that equals one with probability λ and zero with probability $1 - \lambda$. The corresponding PMF is given by

$$p_{\mathrm{B}}(x) = (1-\lambda)^{\mathbb{1}_{x=0}} \cdot \lambda^{1-\mathbb{1}_{x=0}} = \mathrm{e}^{\eta_{\mathrm{B}} u_{\mathrm{B}}(x) - A_{\mathrm{B}}(\eta_{\mathrm{B}})}, \tag{6}$$

with natural parameter $\eta_{\mathrm{B}} = \log \frac{1-\lambda}{\lambda}$, sufficient statistic $u_{\mathrm{B}}(x) = \mathbb{1}_{x=0}$, and log-partition function $A_{\mathrm{B}}(\eta_{\mathrm{B}}) = -\log\lambda = \log(1 + \mathrm{e}^{\eta_{\mathrm{B}}})$.

The exponential family representation of the proper multivariate complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is given by the PDF

$$p_{\mathrm{G}}(\mathbf{x}) = \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) = \mathrm{e}^{\boldsymbol{\eta}_{\mathrm{G}}^{\mathsf{H}}\mathbf{u}_{\mathrm{G}}(\mathbf{x}) - A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})}, \tag{7}$$

with natural parameters $\boldsymbol{\Lambda} = \mathbf{C}^{-1}$, $\boldsymbol{\gamma} = \mathbf{C}^{-1}\boldsymbol{\mu}$, and corresponding vector of natural parameters $\boldsymbol{\eta}_{\mathrm{G}} = \left[\boldsymbol{\gamma}^{\mathsf{T}}, \boldsymbol{\gamma}^{\mathsf{H}}, -\mathrm{vec}\{\boldsymbol{\Lambda}\}^{\mathsf{T}}\right]^{\mathsf{T}}$, sufficient statistics $\mathbf{u}_{\mathrm{G}}(\mathbf{x}) = \left[\mathbf{x}^{\mathsf{T}}, \mathbf{x}^{\mathsf{H}}, \mathrm{vec}\{\mathbf{x}\mathbf{x}^{\mathsf{H}}\}^{\mathsf{T}}\right]^{\mathsf{T}}$, and log-partition function $A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}}) = \boldsymbol{\gamma}^{\mathsf{H}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\gamma} - \log|\pi^{-1}\boldsymbol{\Lambda}|$.

4.2 Bernoulli-Gaussian Distribution

A BG random variable describes two mutually exclusive events. The first event occurs with probability λ and yields a Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. In the complementary event, which occurs with probability $1 - \lambda$, the random vector is zero. We denote the probability λ as activity probability and the probability mass at zero $1 - \lambda$ as inactivity probability. By introducing the Bernoulli indicator as for the Bernoulli distribution in (6), the BG model can be expressed in exponential family form as shown by the following proposition.

Proposition 1. *The exponential family representation of the BG distribution with activity probability λ and the proper complex Gaussian event characterized by mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is given by*

$$p_{\mathrm{BG}}(\mathbf{x}) = \mathcal{BG}(\mathbf{x}|\lambda, \boldsymbol{\mu}, \mathbf{C}) = e^{\boldsymbol{\eta}_{\mathrm{BG}}^{\mathsf{H}}\mathbf{u}_{\mathrm{BG}}(\mathbf{x}) - A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG}})}, \tag{8}$$

with vector of natural parameters $\boldsymbol{\eta}_{\mathrm{BG}} = \left[\kappa, \boldsymbol{\eta}_{\mathrm{G}}^{\mathsf{T}}\right]^{\mathsf{T}}$, sufficient statistics $\mathbf{u}_{\mathrm{BG}}(\mathbf{x}) = \left[\mathbb{1}_{\mathbf{x}=\mathbf{0}}, \mathbf{u}_{\mathrm{G}}(\mathbf{x})^{\mathsf{T}}\right]^{\mathsf{T}}$, and log-partition function $A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG}}) = \log\left(e^{A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})} + e^{\kappa}\right)$, where $\kappa \coloneqq \log\frac{1-\lambda}{\lambda} + A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})$.

Proof. The BG mixture model can be expressed in terms of the Bernoulli indicator as

$$p_{\mathrm{BG}}(\mathbf{x}) = (1-\lambda)^{\mathbb{1}_{\mathbf{x}=\mathbf{0}}} \cdot \left(\lambda \cdot \mathrm{e}^{\boldsymbol{\eta}_{\mathrm{G}}^{\mathsf{H}}\mathbf{u}_{\mathrm{G}}(\mathbf{x}) - A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})}\right)^{1-\mathbb{1}_{\mathbf{x}=\mathbf{0}}}.$$

Collecting all terms into natural parameter and sufficient statistic vectors yields

$$\begin{aligned} p_{\mathrm{BG}}(\mathbf{x}) &= \left(\frac{1-\lambda}{\lambda}\right)^{\mathbb{1}_{\mathbf{x}=\mathbf{0}}} \cdot \mathrm{e}^{A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})\mathbb{1}_{\mathbf{x}=\mathbf{0}}} \cdot \mathrm{e}^{\log\lambda} \cdot \mathrm{e}^{\boldsymbol{\eta}_{\mathrm{G}}^{\mathsf{H}}\mathbf{u}_{\mathrm{G}}(\mathbf{x}) - A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})} \\ &= \mathrm{e}^{\kappa\cdot\mathbb{1}_{\mathbf{x}=\mathbf{0}} + \boldsymbol{\eta}_{\mathrm{G}}^{\mathsf{H}}\mathbf{u}_{\mathrm{G}}(\mathbf{x}) + \log\lambda - A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})}, \end{aligned}$$

which proves the claim with $\boldsymbol{\eta}_{\mathrm{BG}}$, $\mathbf{u}_{\mathrm{BG}}(\mathbf{x})$, and $A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG}})$ defined above. □

We observe that the exponential family representation of the BG model introduces an additional natural parameter and sufficient statistic beyond those of the Gaussian distribution which are related to the inactivity event. The probability distribution $\mathcal{BG}(\mathbf{x}|\lambda,\boldsymbol{\mu},\mathbf{C})$ returns the inactivity probability $1-\lambda$ when $\mathbf{x}=\mathbf{0}$, and for $\mathbf{x}\neq\mathbf{0}$ the corresponding PDF value of the Gaussian event scaled by the activity probability λ.

The key advantage of the exponential family representation is the straightforward computation of normalized products or quotients of the respective distributions by simple addition or subtraction of the natural parameters, respectively. For example, the product of two BG models characterized by the natural parameters $(\kappa_1,\boldsymbol{\gamma}_1,\boldsymbol{\Lambda}_1)$ and $(\kappa_2,\boldsymbol{\gamma}_2,\boldsymbol{\Lambda}_2)$, respectively, yields another BG model with natural parameters $(\kappa_1+\kappa_2,\boldsymbol{\gamma}_1+\boldsymbol{\gamma}_2,\boldsymbol{\Lambda}_1+\boldsymbol{\Lambda}_2)$. To offer further insights, we present in the following the results in the original parameter space, i.e., activity probability λ, mean $\boldsymbol{\mu}$, and covariance matrix $\mathbf{C}$, and explicitly state the normalization constant. Similar results can be derived for quotients of BG models but are omitted for brevity.

Corollary 1 (Bernoulli-Gaussian Product Lemma)**.** *The product of two BG models yields a new unnormalized BG model,*

$$\begin{aligned} &\mathcal{BG}(\mathbf{x}|\lambda_1,\boldsymbol{\mu}_1,\mathbf{C}_1)\cdot\mathcal{BG}(\mathbf{x}|\lambda_2,\boldsymbol{\mu}_2,\mathbf{C}_2) \\ &\quad= \mathcal{BG}(\mathbf{x}|\lambda,\boldsymbol{\mu},\mathbf{C})\cdot\big[\lambda_1\lambda_2\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2,\mathbf{C}_1+\mathbf{C}_2) \\ &\qquad\qquad + (1-\lambda_1)(1-\lambda_2)\big] \end{aligned} \tag{9}$$

with

$$\lambda = \frac{\lambda_1\lambda_2\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2,\mathbf{C}_1+\mathbf{C}_2)}{\lambda_1\lambda_2\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2,\mathbf{C}_1+\mathbf{C}_2)+(1-\lambda_1)(1-\lambda_2)}, \tag{10}$$

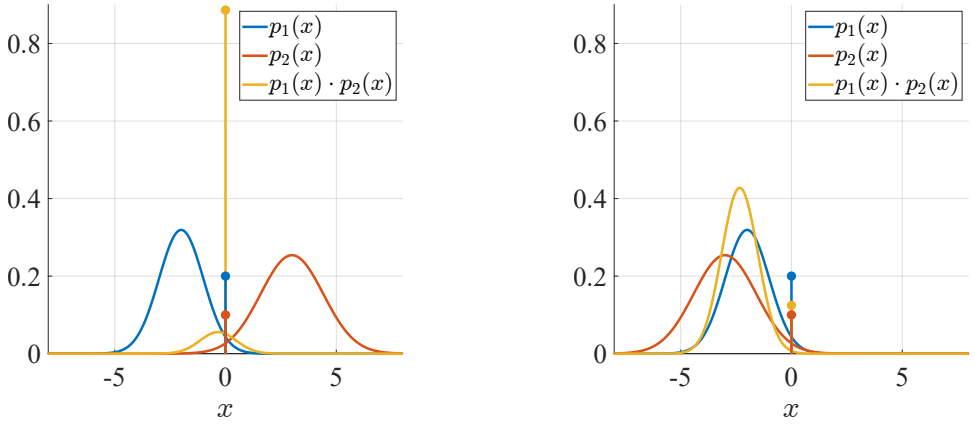
$$\mathbf{C} = \left(\mathbf{C}_1^{-1}+\mathbf{C}_2^{-1}\right)^{-1}, \tag{11}$$

$$\boldsymbol{\mu} = \mathbf{C}\left(\mathbf{C}_1^{-1}\boldsymbol{\mu}_1+\mathbf{C}_2^{-1}\boldsymbol{\mu}_2\right). \tag{12}$$

Proof. The proof is shown in Appendix A. □

The mean (12) and covariance matrix (11) of the Gaussian part are identical to those in the Gaussian product lemma, cf. Lemma 1 in Appendix C. The activity probability (10) provides insights when we explicitly consider the two events a BG random variable characterizes. The event of inactivity occurs with a probability that is proportional to the product of the two factor inactivity probabilities, i.e., $(1-\lambda_1)(1-\lambda_2)$. The activity probability (10) is proportional to the product of the two factor activity probabilities, i.e., $\lambda_1\lambda_2$, and a correction factor which measures how well the two Gaussian densities match, i.e., $\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1-\boldsymbol{\mu}_2,\mathbf{C}_1+\mathbf{C}_2)$. The correction factor is the normalization constant that appears in the Gaussian product lemma. Note that for $\lambda_1=\lambda_2=1$, the BG product lemma reduces to the Gaussian product lemma.

For illustration purposes, two examples of a BG product are shown in Fig. 2. Here, the random variable is modeled as real-valued for simplicity. In Fig. 2a, it can be observed that the area of the Gaussian component corresponding to the activity probability for the resulting product is fairly small, even though the two original BG distributions have an activity probability of $\lambda_1=0.8$ and $\lambda_2=0.9$, respectively, because the respective Gaussian distributions do not significantly overlap. The opposite behavior can be observed in Fig. 2b.



(a) $\mu_1 = -2$, $\mu_2 = 3$.

(b) $\mu_1 = -2$, $\mu_2 = -3$.

Figure 2: Example BG product with $\lambda_1 = 0.8$ and $\lambda_2 = 0.9$.

Besides, it is worth mentioning that computing the product of a BG and a Gaussian distribution is also straightforward with the proposed exponential family representation and reduces to simple addition of the natural parameters corresponding to the Gaussian component. Assuming a BG distribution with natural parameters $(\kappa_1, \boldsymbol{\gamma}_1, \boldsymbol{\Lambda}_1)$ and a Gaussian distribution with $(\boldsymbol{\gamma}_2, \boldsymbol{\Lambda}_2)$, the product is a BG distribution with natural parameters $(\kappa_1, \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2, \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2)$. The same result is obtained by modeling the BG distribution as the sum of a weighted Dirac delta at zero and a weighted Gaussian distribution, i.e., $(1 - \lambda_1)\delta(\mathbf{x}) + \lambda_1 \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{C}_1)$, and then multiplying the sum by the Gaussian distribution $\mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_2, \mathbf{C}_2)$ as proposed in [43]. The equivalence of the two approaches is omitted for brevity.

Finally, we note that the proposed approach for the BG exponential family representation can be readily extended to a broader class of probability distributions with an arbitrary number of discrete point masses at arbitrary positions. Furthermore, the active part that is modeled by a continuous Gaussian distribution in our work can be replaced by another continuous distribution with exponential family representation. The kind of distributions that can be modeled by our approach are related to the so-called *spike and slab priors* [44] where the discrete probability mass characterizes the spike and the continuous probability distribution models the slab. This generality makes the presented approach applicable to a wide range of problems in signal processing.

5 EP-based JACD Algorithms [1, 2]

In this section, we propose two novel EP-based JACD algorithms. The algorithms are derived by first introducing a convenient factorization of $p(\mathbf{U}, \mathbf{H}, \mathbf{X}^d | \mathbf{Y}, \mathbf{X}^p)$ which induces a factor graph as shown in Section 5.1. Then, we assign parametric exponential family representations to the factors to approximate the APP distribution as discussed in Section 5.2. Different choices of these approximate exponential family models yield the two proposed algorithms. Finally, EP message-passing rules are applied to the factor graph as detailed in Section 5.4.

5.1 Factor Graph Representation

Similar to [29], we introduce the auxiliary variables $\mathbf{g}_{l,k} \coloneqq \mathbf{h}_{l,k} u_k$, and $\mathbf{z}_{l,kt} \coloneqq \mathbf{g}_{l,k} x_{kt}$ to decouple activities, channels, and data across UEs. We collect all auxiliary variables in the matrices $\mathbf{G}$ and $\mathbf{Z}$, respectively. The joint MAP estimator of the user activity, channel, data symbols, and auxiliary

variables maximizes the APP distribution given by

$$\begin{aligned} &p(\mathbf{U},\mathbf{H},\mathbf{X}^d,\mathbf{G},\mathbf{Z}|\mathbf{Y},\mathbf{X}^p) \\ &\propto \prod_{l=1}^{L}\prod_{k=1}^{K}\prod_{t=1}^{T}\Big[p(\mathbf{y}_{l,t}|\mathbf{z}_{l,1t},...,\mathbf{z}_{l,Kt})\cdot p(\mathbf{z}_{l,kt}|\mathbf{g}_{l,k},x_{kt}) \\ &\quad\cdot p(\mathbf{g}_{l,k}|\mathbf{h}_{l,k},u_k)\cdot \tilde{p}_{u_k}(u_k)\cdot \tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k})\cdot p_x(x_{kt})\Big], \end{aligned} \tag{13}$$

where the factorization follows from the independence of channel vectors across APs and UEs, the independence of user activities across UEs, and the independence of data symbols across UEs and time indices. The terms $\tilde{p}_{u_k}(u_k)$ and $\tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k})$ denote refined prior information on the user activity u_k and the channel $\mathbf{h}_{l,k}$, respectively, which can be acquired by a pilot-based initialization algorithm. One possible initialization algorithm is the joint activity detection and channel estimation via expectation propagation (JAC-EP) algorithm presented in [2] an described in Appendix B. We denote the resulting prior information for u_k and $\mathbf{h}_{l,k}$ by the two probabilities $\tilde{p}_{u_k}(0)$ and $\tilde{p}_{u_k}(1)$, and by the channel mean vector $\tilde{\boldsymbol{\mu}}_{h_{l,k}}$ and covariance matrix $\tilde{\mathbf{C}}_{h_{l,k}}$, respectively. The probability distributions in (13) are represented by factor nodes (rectangles) in the factor graph illustrated in Fig. 3 and are given by

$$\Psi_{y_{l,t}} \coloneqq p(\mathbf{y}_{l,t}|\mathbf{z}_{l,1t},...,\mathbf{z}_{l,Kt}) = \mathcal{CN}\Big(\mathbf{y}_{l,t}\Big|\sum_{k=1}^{K}\mathbf{z}_{l,kt},\sigma_n^2\mathbf{I}_N\Big), \tag{14}$$

$$\Psi_{z_{l,kt}} \coloneqq p(\mathbf{z}_{l,kt}|\mathbf{g}_{l,k},x_{kt}) = \delta(\mathbf{z}_{l,kt}-\mathbf{g}_{l,k}x_{kt}), \tag{15}$$

$$\Psi_{g_{l,k}} \coloneqq p(\mathbf{g}_{l,k}|\mathbf{h}_{l,k},u_k) = \delta(\mathbf{g}_{l,k}-\mathbf{h}_{l,k}u_k), \tag{16}$$

$$\Psi_{u_k} \coloneqq \tilde{p}_{u_k}(u_k) = \tilde{p}_{u_k}(0)\mathbb{1}_{u_k=0}+\tilde{p}_{u_k}(1)\mathbb{1}_{u_k=1}, \tag{17}$$

$$\Psi_{h_{l,k}} \coloneqq \tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k}) = \mathcal{CN}\big(\mathbf{h}_{l,k}\big|\tilde{\boldsymbol{\mu}}_{h_{l,k}},\tilde{\mathbf{C}}_{h_{l,k}}\big), \tag{18}$$

$$\Psi_{x_{kt}} \coloneqq p_x(x_{kt}) = \begin{cases}\mathbb{1}_{x_{kt}=x_{kt}^p} & \text{for } t\le T_p\\ \frac{1}{M}\mathbb{1}_{x_{kt}\in\mathcal{X}} & \text{for } t>T_p\end{cases}. \tag{19}$$

The variables in the above equations correspond to variable nodes (circles) in the factor graph. The factor and variable nodes are organized to reflect their implementation at the CPU and the APs. As shown in the factor graph, the CPU combines the information from the APs to estimate user activities and data. In contrast, the channels are estimated locally in each AP and do not need to be forwarded to the CPU.

5.2 EP Approximations and Fronthaul Load

To apply the EP message-passing rules to the factor graph in Fig. 3, we assign a parametric distribution representation from the exponential family to each variable node to approximate the corresponding APP distribution. Categorical distributions are chosen for the variables x_{kt} and u_k whereas the variables $\mathbf{z}_{l,kt}$ and $\mathbf{h}_{l,k}$ are modeled by multivariate complex Gaussian distributions. The variable $\mathbf{g}_{l,k}$ is treated differently for the two proposed algorithms. In the JACD-EP algorithm, $\mathbf{g}_{l,k}$ follows a multivariate complex Gaussian distribution, whereas in the JACD-EP-BG algorithm, $\mathbf{g}_{l,k}$ is modeled by a BG vector as presented in Section 4. The parameters characterizing these approximate posterior distributions constitute the messages propagated along the graph.

The fronthaul load is determined solely by the messages associated to x_{kt} and u_k. For $t>T_p$, messages from and towards node x_{kt} consist of $M-1$ probabilities while no messages are exchanged for $t\le T_p$ since the pilot symbols are known a priori. For u_k, a single real-valued parameter suffices to describe its distribution. Hence, the total fronthaul load per iteration amounts to $2LK(T_d(M-1)+1)$ real-valued numbers where the factor two accounts for the bidirectional exchange between CPU and APs.

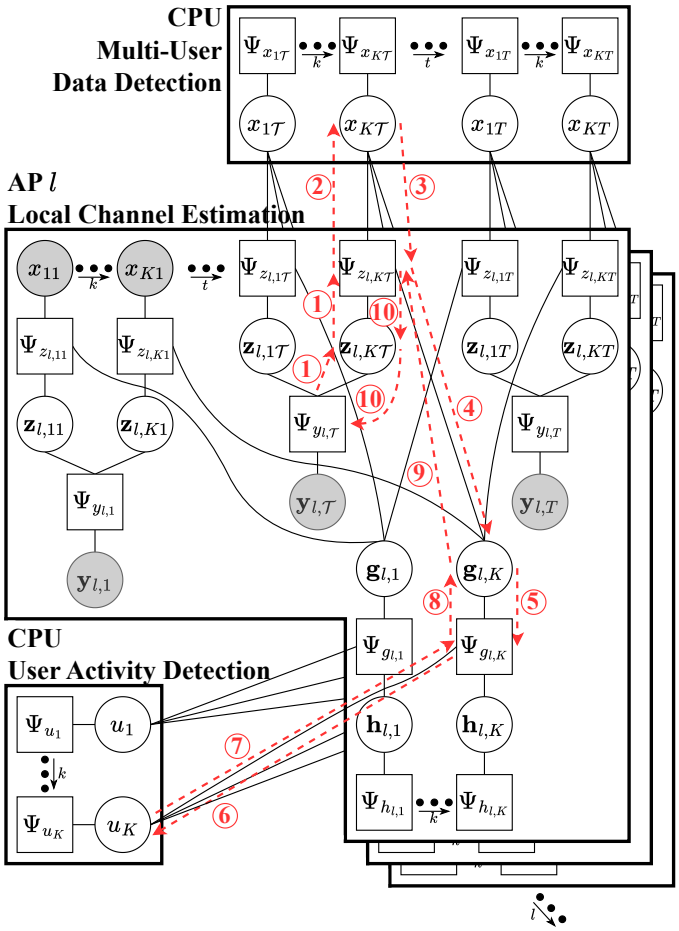


Figure 3: Factor graph for the EP-based JACD algorithms with $\mathcal{T} \coloneqq T_p + 1$. The numbered red dashed arrows show the flow of information according to the scheduling presented in Algorithm 1. Each number corresponds to one message update in Algorithm 1.

5.3 Initialization and Scheduling

The proposed algorithms are initialized using the activity and channel estimates obtained from the JAC-EP algorithm [2]. This pilot-based initialization algorithm provides the soft user-activity estimates $\tilde{p}_{u_k}(u_k)$ and channel distributions $\tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k})$, equivalently characterized by the expectation of their sufficient statistics $(\tilde{\boldsymbol{\mu}}_{h_{l,k}}, \tilde{\mathbf{C}}_{h_{l,k}})$ or their natural parameters $(\tilde{\boldsymbol{\gamma}}_{h_{l,k}}, \tilde{\boldsymbol{\Lambda}}_{h_{l,k}})$, which serve as priors for the JACD-EP and the JACD-EP-BG algorithm. The exchanged messages are initialized as follows. The initial mean vector and covariance matrix for the message $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ $\forall k, l, t$ are determined from the prior information on $\mathbf{z}_{l,kt}$ induced by $\tilde{p}_{u_k}(u_k)$, $\tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k})$, and $p_x(x_{kt})$,[1]

$$\boldsymbol{\mu}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \begin{cases} \tilde{p}_{u_k}(1) \cdot \tilde{\boldsymbol{\mu}}_{h_{l,k}} \cdot x_{kt}^p & \text{for } t \leq T_p \\ \mathbf{0} & \text{for } t > T_p \end{cases}, \tag{20}$$

$$\mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \begin{cases} \tilde{p}_{u_k}(1)\big(\tilde{\mathbf{C}}_{h_{l,k}} + \tilde{\boldsymbol{\mu}}_{h_{l,k}}\tilde{\boldsymbol{\mu}}_{h_{l,k}}^{\mathsf{H}} \\ \qquad \cdot \tilde{p}_{u_k}(0)\big)\,|x_{kt}^p|^2 & \text{for } t \leq T_p \\ \tilde{p}_{u_k}(1)\big(\tilde{\mathbf{C}}_{h_{l,k}} + \tilde{\boldsymbol{\mu}}_{h_{l,k}}\tilde{\boldsymbol{\mu}}_{h_{l,k}}^{\mathsf{H}}\big)\sigma_x^2 & \text{for } t > T_p \end{cases}. \tag{21}$$

[1] We compute the expectation of the sufficient statistics of $\mathbf{z}_{l,kt}$ by exploiting the independence of u_k, $\mathbf{h}_{l,k}$, and x_{kt}, e.g., the mean $E\{\mathbf{z}_{l,kt}\} = E\{u_k\}\,E\{\mathbf{h}_{l,k}\}\,E\{x_{kt}\}$. The same approach is applied to $\mathbf{g}_{l,k}$.

Algorithm 1 JACD-EP* and JACD-EP-BG† Algorithm

Input: Pilot matrix $\mathbf{X}^p$, transmit power σ_x^2, received signal $\mathbf{Y}$, noise variance σ_n^2, prior distributions on user activities $\tilde{p}_{u_k}(u_k)$ and channels $\tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k})$.

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

$\forall k,l,t$: Initialize $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ (20) (21).
$\forall k,l,t$: Initialize $m_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$ {(22) (23)}* {(24) - (26)}†.
$\forall k,l$: Initialize $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ {(22) (23)}* {(24) - (26)}†.
for $i = 1$ to $i_{\max}$ **do**
  $\forall k,l,t$: Update $m_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}$ (27) (28).
  $\forall k,l,t > T_p$: Update $m_{\Psi_{z_{l,kt}};x_{kt}}$ (29).
  $\forall k,l,t > T_p$: Update $m_{x_{kt};\Psi_{z_{l,kt}}}$ (31).
  $\forall k,l,t$: Update $m_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}$ {(33) (34) via (35) (36)}* {(32) - (34) via (41) - (43)}†.
  $\forall k,l$: Update $m_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}$ {(46) (47)}* {(45) - (47)}†.
  $\forall k,l$: Update $m_{\Psi_{g_{l,k}};u_k}$ (48)* (50)†.
  $\forall k,l$: Update $m_{u_k;\Psi_{g_{l,k}}}$ (51).
  $\forall k,l$: Update $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ {(52) (53)}* {(59) - (61)}†.
  $\forall k,l,t$: Update $m_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$ {(63) (64)}* {(62) - (64)}†.
  $\forall k,l,t$: Update $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ (65) (66).
end for
return $\hat{u}_k$ (69) $\forall k$.
return $\hat{\mathbf{h}}_{l,k}$ (72) $\forall k,l$.
return $\hat{x}_{kt}$ (71) $\forall k, t > T_p$.

Similarly, the initialization of the messages $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ and $m_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$ $\forall k,l,t$ are determined from the prior information on $\mathbf{g}_{l,k}$ and depends on the different EP approximations for $\mathbf{g}_{l,k}$. The Gaussian variables $\mathbf{g}_{l,k}$ in the JACD-EP algorithm are initialized as

$$\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \tilde{p}_{u_k}(1) \cdot \tilde{\boldsymbol{\mu}}_{h_{l,k}}, \tag{22}$$

$$\mathbf{C}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \tilde{p}_{u_k}(1)\big(\tilde{\mathbf{C}}_{h_{l,k}} + \tilde{\boldsymbol{\mu}}_{h_{l,k}} \tilde{\boldsymbol{\mu}}_{h_{l,k}}^{\mathsf{H}} \cdot \tilde{p}_{u_k}(0)\big). \tag{23}$$

In the JACD-EP-BG algorithm, the initialization of the BG variables $\mathbf{g}_{l,k}$ is given by

$$\lambda_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \tilde{p}_{u_k}(1), \tag{24}$$

$$\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \tilde{\boldsymbol{\mu}}_{h_{l,k}}, \tag{25}$$

$$\mathbf{C}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \tilde{\mathbf{C}}_{h_{l,k}}. \tag{26}$$

All other messages are initialized using uninformative priors, namely, uniform distributions for messages involving categorical variables, and zero-mean, zero-precision distributions for messages involving Gaussian variables where the precision matrix is the inverse of the covariance matrix. The uninformative BG distribution exhibits an activity probability equal to 0.5 and a zero-mean, zero-precision Gaussian component.

After initialization, messages are updated according to the scheduling defined in Algorithm 1 and illustrated in Fig. 3 by the red dashed arrows. Algorithm 1 jointly describes both proposed methods which follow the same scheduling and differ only in the message updates. We group update rules for the two algorithms by brackets with different superscripts, i.e., $\{\cdot\}^*$ for JACD-EP and $\{\cdot\}^\dagger$ for JACD-EP-BG. Message updates without brackets are common to both algorithms.

5.4 Message-Passing Update Rules

In this section, we present the EP message-passing rules used in the proposed algorithms. Detailed derivations for the JACD-EP and the JACD-EP-BG algorithm can be found in the extended version of [2][2] and Appendix D, respectively. For a BG random variable, the activity probability λ, the mean vector $\boldsymbol{\mu}$, and the covariance matrix $\mathbf{C}$ can be readily expressed by the natural parameters $\kappa = \log \frac{1-\lambda}{\lambda} + \boldsymbol{\mu}^{\mathsf{H}}\mathbf{C}^{-1}\boldsymbol{\mu} + \log|\pi\mathbf{C}|$, $\boldsymbol{\gamma} = \mathbf{C}^{-1}\boldsymbol{\mu}$, and $\boldsymbol{\Lambda} = \mathbf{C}^{-1}$. The same relations hold for the parameters $\boldsymbol{\mu}$ and $\mathbf{C}$ and natural parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\Lambda}$ of a Gaussian random variable. In the following, we freely switch between these two representations without explicitly mentioning the transformation; whenever $(\lambda_{\Psi_\alpha;\mathbf{x}_\beta}, \boldsymbol{\mu}_{\Psi_\alpha;\mathbf{x}_\beta}, \mathbf{C}_{\Psi_\alpha;\mathbf{x}_\beta})$ are computed, the corresponding $(\kappa_{\Psi_\alpha;\mathbf{x}_\beta}, \boldsymbol{\gamma}_{\Psi_\alpha;\mathbf{x}_\beta}, \boldsymbol{\Lambda}_{\Psi_\alpha;\mathbf{x}_\beta})$ follow automatically and vice versa.

In the first step, the messages $m_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}$ $\forall k, l, t$ are updated. These updates compute the Gaussian beliefs of the auxiliary variables $\mathbf{z}_{l,kt}$ based on the observation $\mathbf{y}_{l,t}$. In practice, the updating rules perform soft interference cancellation using the current estimated interference $\boldsymbol{\mu}_{\Psi_{z_{l,k't}};\mathbf{z}_{l,k't}}$ of all other users $k' \neq k$ and modifying the covariance matrix accordingly. The message mean and covariance are

$$\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} = \mathbf{y}_{l,t} - \sum_{k' \neq k} \boldsymbol{\mu}_{\Psi_{z_{l,k't}};\mathbf{z}_{l,k't}}, \tag{27}$$

$$\mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} = \sigma_n^2 \mathbf{I}_N + \sum_{k' \neq k} \mathbf{C}_{\Psi_{z_{l,k't}};\mathbf{z}_{l,k't}}. \tag{28}$$

These updated beliefs are used to generate the local categorical data symbol beliefs $m_{\Psi_{z_{l,kt}};x_{kt}}$ $\forall k, l, t > T_p$ at each AP. The message update is derived by evaluating how well the product $\mathbf{g}_{l,k}x_{kt}$ matches the relation $\mathbf{z}_{l,kt} = \mathbf{g}_{l,k}x_{kt}$ which is obtained by sampling a Gaussian likelihood $\theta(x_{kt})$,[3]

$$\pi_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) \propto \theta(x_{kt}), \tag{29}$$

with

$$\begin{aligned}\theta(x_{kt}) = \mathcal{CN}\big(\mathbf{0} | \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} - \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} x_{kt}, \\ \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} + \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} |x_{kt}|^2\big).\end{aligned} \tag{30}$$

Then, all local data symbol beliefs are combined at the CPU. The message update $m_{x_{kt};\Psi_{z_{l,kt}}}$ $\forall k, l,\ t > T_p$ combines the categorical beliefs of the data symbol x_{kt} from all APs $l' \neq l$ and forwards them to AP l,

$$\pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \propto \prod_{l' \neq l} \pi_{\Psi_{z_{l',kt}};x_{kt}}(x_{kt}). \tag{31}$$

The updated beliefs of the variables $\mathbf{z}_{l,kt}$ and x_{kt} are used to update the messages $m_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}$ $\forall k, l, t$, yielding the beliefs of the auxiliary variable $\mathbf{g}_{l,k}$. The EP approach first generates the local estimate of $\mathbf{g}_{l,k}$ at the factor node $\Psi_{z_{l,kt}}$ and then removes the knowledge induced by the message $m_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$,

$$\kappa_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \hat{\kappa}_{\mathbf{g}_{l,kt}} - \kappa_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \tag{32}$$

$$\boldsymbol{\gamma}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \hat{\boldsymbol{\gamma}}_{\mathbf{g}_{l,kt}} - \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \tag{33}$$

$$\boldsymbol{\Lambda}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \hat{\boldsymbol{\Lambda}}_{\mathbf{g}_{l,kt}} - \boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}. \tag{34}$$

Note that (32) is computed only for the JACD-EP-BG algorithm, whereas (33) and (34) are evaluated for both algorithms. The computation of the parameters describing the local estimate

[2] https://arxiv.org/abs/2405.09914

[3] Note that $\theta(x_{kt})$ provides unnormalized probability values and $\pi_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt})$ is obtained by normalization.

of $\mathbf{g}_{l,k}$ at $\Psi_{z_{l,kt}}$, i.e., $\hat{\kappa}_{\mathbf{g}_{l,kt}}$, $\hat{\boldsymbol{\gamma}}_{\mathbf{g}_{l,kt}}$, and $\hat{\boldsymbol{\Lambda}}_{\mathbf{g}_{l,kt}}$, differs for the JACD-EP and the JACD-EP-BG algorithm due to the different modeling of $\mathbf{g}_{l,k}$. For the JACD-EP algorithm, we compute the Gaussian parameters

$$\hat{\boldsymbol{\mu}}_{\mathbf{g}_{l,kt}} = \frac{1}{Z_{\Psi_{z_{l,kt}}}} \sum_{x_{kt}\in\tilde{\mathcal{X}}} \frac{\phi(x_{kt})}{x_{kt}} \cdot \check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt}), \tag{35}$$

$$\begin{aligned}\hat{\mathbf{C}}_{\mathbf{g}_{l,kt}} = \frac{1}{Z_{\Psi_{z_{l,kt}}}} \sum_{x_{kt}\in\tilde{\mathcal{X}}} \frac{\phi(x_{kt})}{|x_{kt}|^2} \cdot \big(&\check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x_{kt}) \\ &+ \check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt}) \cdot \check{\boldsymbol{\mu}}^{\mathsf{H}}_{\mathbf{z}_{l,kt}}(x_{kt})\big) - \hat{\boldsymbol{\mu}}_{\mathbf{g}_{l,kt}} \hat{\boldsymbol{\mu}}^{\mathsf{H}}_{\mathbf{g}_{l,kt}},\end{aligned} \tag{36}$$

with $\tilde{\mathcal{X}} = \{x^p_{kt}\}$ for $t \le T_p$, $\tilde{\mathcal{X}} = \mathcal{X}$ for $t > T_p$, and

$$\phi(x_{kt}) = \pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}), \tag{37}$$

$$\check{\boldsymbol{\gamma}}_{\mathbf{z}_{l,kt}}(x_{kt}) = \boldsymbol{\gamma}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} + \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \frac{x_{kt}}{|x_{kt}|^2}, \tag{38}$$

$$\check{\boldsymbol{\Lambda}}_{\mathbf{z}_{l,kt}}(x_{kt}) = \boldsymbol{\Lambda}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} + \boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} |x_{kt}|^{-2}, \tag{39}$$

$$Z_{\Psi_{z_{l,kt}}} = \sum_{x_{kt}\in\tilde{\mathcal{X}}} \phi(x_{kt}). \tag{40}$$

Note that $\pi_{x_{kt};\Psi_{z_{l,kt}}}(x^p_{kt}) = 1$ for $t \le T_p$ since the pilot symbol is known a priori. Hence, the updates for the pilot part simplify to $\hat{\boldsymbol{\mu}}_{\mathbf{g}_{l,kt}} = \check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x^p_{kt})/x^p_{kt}$ and $\hat{\mathbf{C}}_{\mathbf{g}_{l,kt}} = \check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x^p_{kt})/|x^p_{kt}|^2$. For the JACD-EP-BG algorithm, we compute the parameters of the local BG estimate of $\mathbf{g}_{l,k}$ as

$$\hat{\lambda}_{\mathbf{g}_{l,kt}} = \frac{\lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \phi(x^*)}{\lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \phi(x^*) + (1-\lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}) \cdot \theta(0)}, \tag{41}$$

$$\hat{\boldsymbol{\gamma}}_{\mathbf{g}_{l,kt}} = \boldsymbol{\gamma}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} \frac{|x^*|^2}{x^*} + \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \tag{42}$$

$$\hat{\boldsymbol{\Lambda}}_{\mathbf{g}_{l,kt}} = \boldsymbol{\Lambda}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x^*|^2 + \boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \tag{43}$$

with $x^* = x^p_{kt}$ for $t \le T_p$ and

$$x^* = \arg\max_{x_{kt}\in\mathcal{X}} \phi(x_{kt}), \tag{44}$$

for $t > T_p$. Note that using (42) and (43) in (33) and (34), the message updates simplify to $\boldsymbol{\gamma}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \boldsymbol{\gamma}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x^*|^2/x^*$ and $\boldsymbol{\Lambda}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \boldsymbol{\Lambda}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x^*|^2$, respectively. Furthermore, note that we deviate here from the classical EP message-passing rule by computing the local estimate of $\mathbf{g}_{l,k}$ using the most likely symbol x^* instead of averaging across all $x_{kt} \in \mathcal{X}$. This is necessary in order to prevent the Gaussian part of the resulting BG belief to be close to zero which can lead to a high number of false alarms.

Next, the messages $m_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}$ $\forall k, l$ are updated by combining the beliefs of $\mathbf{g}_{l,k}$ from all time slots $t \in \{1, \ldots, T\}$ to generate the updated belief parameters

$$\kappa_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} = \sum_{t=1}^{T} \kappa_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}, \tag{45}$$

$$\boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} = \sum_{t=1}^{T} \boldsymbol{\gamma}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}, \tag{46}$$

$$\boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} = \sum_{t=1}^{T} \boldsymbol{\Lambda}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}. \tag{47}$$

Similar as for the message $m_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}$, (45) is computed only for the JACD-EP-BG algorithm, whereas (46) and (47) are evaluated for both algorithms.

The updated beliefs of $\mathbf{g}_{l,k}$ are then used to compute the local activity probabilities $m_{\Psi_{g_{l,k}};u_k}$ $\forall k,l$ at each AP. Here, the update differs for the two proposed algorithms. The categorical belief for the JACD-EP algorithm is computed by evaluating how well the Gaussian belief of $\mathbf{g}_{l,k}$ matches with the prior of the channel $\mathbf{h}_{l,k}$,

$$\pi_{\Psi_{g_{l,k}};u_k}(u_k) \propto \vartheta(u_k), \tag{48}$$

with

$$\vartheta(u_k) = \mathcal{CN}\big(\mathbf{0}|\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} - \tilde{\boldsymbol{\mu}}_{h_{l,k}}u_k, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \tilde{\mathbf{C}}_{h_{l,k}}u_k\big). \tag{49}$$

For the JACD-EP-BG algorithm, the activity belief incorporates both the Bernoulli and the Gaussian component of the BG variable $\mathbf{g}_{l,k}$,

$$\pi_{\Psi_{g_{l,k}};u_k}(u_k) \propto \begin{cases} 1-\lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} & \text{for } u_k = 0 \\ \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} \cdot \vartheta(1) & \text{for } u_k = 1 \end{cases}. \tag{50}$$

Then, for both the proposed algorithms, the CPU computes the message updates $m_{u_k;\Psi_{g_{l,k}}}$ $\forall k,l$ by combining the categorical beliefs of u_k from all APs $l' \neq l$ with the prior information and forwards them to AP l,

$$\pi_{u_k;\Psi_{g_{l,k}}}(u_k) \propto \tilde{p}_{u_k}(u_k) \cdot \prod_{l' \neq l} \pi_{\Psi_{g_{l',k}};u_k}(u_k). \tag{51}$$

The belief of $\mathbf{g}_{l,k}$ is then updated in the message $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ $\forall k,l$ by combining the information from the Bernoulli variable u_k and the Gaussian variable $\mathbf{h}_{l,k}$. For the JACD-EP algorithm, this is achieved by computing the local Gaussian estimate of $\mathbf{g}_{l,k}$ at the factor node $\Psi_{g_{l,k}}$ and then removing the contribution of $m_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}$,

$$\boldsymbol{\gamma}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \hat{\boldsymbol{\gamma}}_{\mathbf{g}_{l,k}} - \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \tag{52}$$

$$\boldsymbol{\Lambda}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \hat{\boldsymbol{\Lambda}}_{\mathbf{g}_{l,k}} - \boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \tag{53}$$

with

$$\hat{\boldsymbol{\mu}}_{\mathbf{g}_{l,k}} = \frac{1}{Z_{\Psi_{g_{l,k}}}} \cdot \pi_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \cdot \check{\boldsymbol{\mu}}_{\mathbf{g}_{l,k}}, \tag{54}$$

$$\begin{aligned} \hat{\mathbf{C}}_{\mathbf{g}_{l,k}} = \frac{1}{Z_{\Psi_{g_{l,k}}}} \cdot \pi_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \cdot (\check{\mathbf{C}}_{\mathbf{g}_{l,k}} + \check{\boldsymbol{\mu}}_{\mathbf{g}_{l,k}} \check{\boldsymbol{\mu}}^{\mathsf{H}}_{\mathbf{g}_{l,k}}) \\ - \hat{\boldsymbol{\mu}}_{\mathbf{g}_{l,k}} \hat{\boldsymbol{\mu}}^{\mathsf{H}}_{\mathbf{g}_{l,k}}, \end{aligned} \tag{55}$$

and

$$\check{\boldsymbol{\gamma}}_{\mathbf{g}_{l,k}} = \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \tilde{\boldsymbol{\gamma}}_{h_{l,k}}, \tag{56}$$

$$\check{\boldsymbol{\Lambda}}_{\mathbf{g}_{l,k}} = \boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \tilde{\boldsymbol{\Lambda}}_{h_{l,k}}, \tag{57}$$

$$Z_{\Psi_{g_{l,k}}} = \pi_{u_k;\Psi_{g_{l,k}}}(0) \cdot \vartheta(0) + \pi_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1). \tag{58}$$

For the JACD-EP-BG algorithm, the BG belief of $\mathbf{g}_{l,k}$ is given by

$$\lambda_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \pi_{u_k;\Psi_{g_{l,k}}}(1), \tag{59}$$

$$\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \tilde{\boldsymbol{\mu}}_{h_{l,k}}, \tag{60}$$

$$\mathbf{C}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \tilde{\mathbf{C}}_{h_{l,k}}. \tag{61}$$

Next, the messages $m_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$ $\forall k,l,t$ are updated by combining the previously updated belief of $\mathbf{g}_{l,k}$ with the contributions from all time slots $t' \neq t$,

$$\kappa_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \kappa_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \sum_{t'\neq t} \kappa_{\Psi_{z_{l,kt'}};\mathbf{g}_{l,k}}, \tag{62}$$

$$\boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \boldsymbol{\gamma}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \sum_{t'\neq t} \boldsymbol{\gamma}_{\Psi_{z_{l,kt'}};\mathbf{g}_{l,k}}, \tag{63}$$

$$\boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \boldsymbol{\Lambda}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \sum_{t'\neq t} \boldsymbol{\Lambda}_{\Psi_{z_{l,kt'}};\mathbf{g}_{l,k}}. \tag{64}$$

The parameter in (62) is computed only for the JACD-EP-BG algorithm, whereas (63) and (64) are common to both algorithms.

Lastly, the messages $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$ $\forall k,l,t$ are updated to form the Gaussian belief of $\mathbf{z}_{l,kt}$ which will be used in the next iteration for interference cancellation. The factor node $\Psi_{z_{l,kt}}$ first computes the local estimate of $\mathbf{z}_{l,kt}$ based on the beliefs of $\mathbf{g}_{l,k}$ and x_{kt} and then removes the contribution of $m_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}$,

$$\boldsymbol{\gamma}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \hat{\boldsymbol{\gamma}}_{\mathbf{z}_{l,kt}} - \boldsymbol{\gamma}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}, \tag{65}$$

$$\boldsymbol{\Lambda}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \hat{\boldsymbol{\Lambda}}_{\mathbf{z}_{l,kt}} - \boldsymbol{\Lambda}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}, \tag{66}$$

with the parameters describing the local estimate of $\mathbf{z}_{l,kt}$,

$$\hat{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}} = \frac{1}{Z_{l,kt}} \sum_{x_{kt}\in\tilde{\mathcal{X}}} \phi(x_{kt}) \cdot \check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt}), \tag{67}$$

$$\begin{aligned}\hat{\mathbf{C}}_{\mathbf{z}_{l,kt}} = \frac{1}{Z_{l,kt}} \sum_{x_{kt}\in\tilde{\mathcal{X}}} \phi(x_{kt}) \cdot \big(&\check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x_{kt}) \\ &+ \check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt}) \cdot \check{\boldsymbol{\mu}}^{\mathsf{H}}_{\mathbf{z}_{l,kt}}(x_{kt})\big) - \hat{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}} \hat{\boldsymbol{\mu}}^{\mathsf{H}}_{\mathbf{z}_{l,kt}},\end{aligned} \tag{68}$$

with $\tilde{\mathcal{X}} = \{x^p_{kt}\}$ for $t \leq T_p$, $\tilde{\mathcal{X}} = \mathcal{X}$ for $t > T_p$, and the other parameters defined in (37)-(40). Hence, the updates for $t \leq T_p$ simplify to $\hat{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}} = \check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x^p_{kt})$ and $\hat{\mathbf{C}}_{\mathbf{z}_{l,kt}} = \check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x^p_{kt})$.

Furthermore, to improve numerical stability and convergence, damping is applied in the updating of all factor-to-variable messages using a parameter $\eta \in [0,1]$ as in [29, 45]: the updated parameter is a convex combination of its value from the previous iteration and the newly computed value at the factor node. Finally, messages involving a Gaussian component are updated only if the resulting matrix parameter is positive definite; otherwise, the corresponding parameters from the previous iteration are retained.

5.5 Final Estimation

From the final EP messages, the approximate APP distributions of the user activities, channels, and data are obtained as the product of all the incoming messages at the corresponding variable

nodes. The final estimates are then given by the MAP solution of these APPs,

$$\hat{u}_k = \underset{u_k \in \{0,1\}}{\arg\max}\ \tilde{p}_{u_k}(u_k) \cdot \prod_{l=1}^{L} \pi_{\Psi_{g_{l,k}};u_k}(u_k), \tag{69}$$

$$\hat{\mathbf{h}}_{l,k} = \underset{\mathbf{h}_{l,k} \in \mathbb{C}^N}{\arg\max}\ \tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k}) \cdot p_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}), \tag{70}$$

$$\hat{x}_{kt} = \underset{x_{kt} \in \mathcal{X}}{\arg\max} \prod_{l=1}^{L} \pi_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) \qquad \text{for } t > T_p. \tag{71}$$

Assuming that UE k is active, the solution to (70) is Gaussian, yielding the following closed-form estimate:

$$\hat{\mathbf{h}}_{l,k} = \hat{\mathbf{\Lambda}}^{-1}_{\mathbf{h}_{l,k}} \hat{\gamma}_{\mathbf{h}_{l,k}}, \tag{72}$$

with

$$\hat{\gamma}_{\mathbf{h}_{l,k}} = \tilde{\boldsymbol{\gamma}}_{h_{l,k}} + \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \tag{73}$$

$$\hat{\mathbf{\Lambda}}_{\mathbf{h}_{l,k}} = \tilde{\mathbf{\Lambda}}_{h_{l,k}} + \mathbf{\Lambda}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}. \tag{74}$$

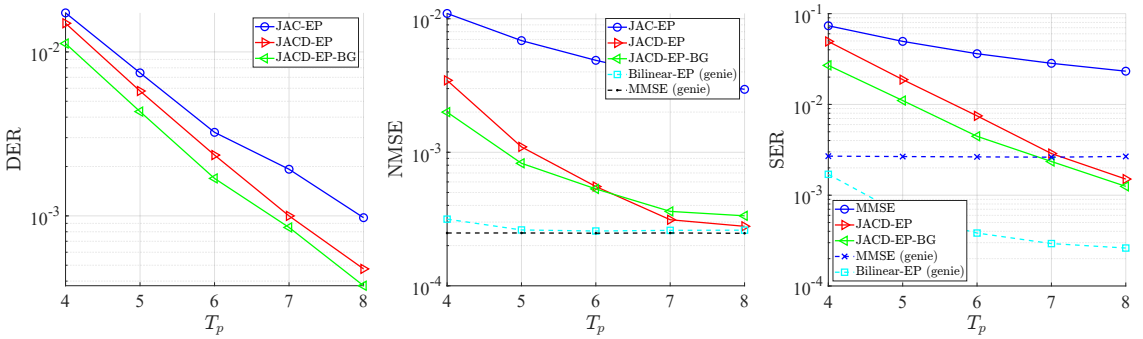
5.6 Computational Complexity and Scalability

In the following, we characterize the computational complexity order of the proposed algorithms neglecting addition and subtraction costs. At the APs, the dominant computational burden for both algorithms arises from the evaluation of the Gaussian likelihood in (30), which requires computing the inverse and determinant of $N-$dimensional covariance matrices. This computation is performed KT_d times for every constellation symbol $x_{kt} \in \mathcal{X}$ and exhibits a complexity of $\mathcal{O}(N^3)$. Consequently, the per-iteration computational complexity is $\mathcal{O}(MN^3KT_d)$ at every AP. From a network scalability perspective, the key observation is that the computational load at the APs scales linearly with the number of UEs K. This dependency can be strongly mitigated by pruning UEs with negligible large-scale fading coefficients, i.e., UEs that are unlikely to contribute meaningfully to the received signal. Such a modification reduces the effective number of UEs processed at each AP without altering the core algorithmic structure.

The complexity at the CPU for both algorithms is mainly determined by the combination of symbol beliefs in (31) which has a complexity order of $\mathcal{O}(LMKT_d)$. Hence, the CPU complexity scales linearly with both the number of UEs and APs, which may limit scalability in very large networks. Nevertheless, the processing of different symbols at the CPU is fully decoupled, enabling straightforward parallel and/or distributed implementations that can effectively remove this scalability bottleneck.

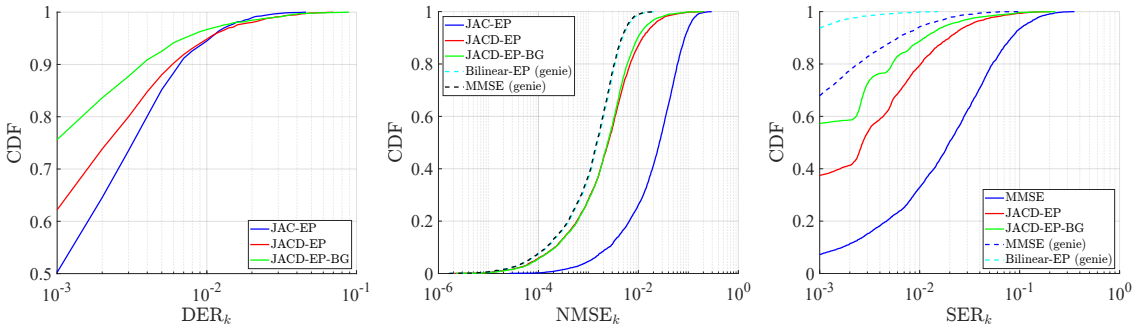
6 Numerical Results

In this section, we evaluate the performance of the proposed EP-based JACD algorithms via extensive Monte Carlo simulations. We consider a square network area of $500 \times 500\,\mathrm{m}^2$ comprising $L = 25$ single-antenna APs, i.e., $N = 1$, placed on a uniform grid at the positions $\{(50 + i \cdot 100, 50 + j \cdot 100)\,\mathrm{m} \,|\, i, j \in \{0, 1, 2, 3, 4\}\}$ and a height of 10 m. The receiver noise power at each AP is set to $\sigma_n^2 = -96\,\mathrm{dBm}$. A total of $K = 40$ UEs are uniformly and independently placed at random ground locations, each with an activity probability $\lambda = 0.3$. An active UE transmits T_p pilot symbols and $T_d = T - T_p$ 4-quadrature amplitude modulation (QAM) data symbols with constant transmit power $\sigma_x^2 = 16\,\mathrm{dBm}$. The T_p columns of the pilot matrix $\mathbf{X}^p$ are chosen from the $K \times K$ discrete Fourier transform (DFT) matrix such that the maximum inner product between any two different pilot sequences is minimized, thereby minimizing the mutual coherence [46]. Due to the limited pilot sequence length and number of UEs in our simulations, the optimal pilot sequences can be found by a full search. For larger systems, optimization-based pilot designs such



(a) DER. (b) NMSE. (c) SER.

Figure 4: Performance metrics versus pilot sequence length for $L = 25$, $N = 1$, $K = 40$, $\lambda = 0.3$, and $T = 60$.



(a) DER. (b) NMSE. (c) SER.

Figure 5: CDFs of performance metrics for $L = 25$, $N = 1$, $K = 40$, $\lambda = 0.3$, $T = 60$, and $T_p = 6$.

as those proposed in [38, 46] would be required. The large-scale fading coefficients follow the 3GPP urban microcell model that incorporates correlated shadow fading [47, Sec. 2.5.2]. Furthermore, we assume a channel coherence time of $T = 60$ which corresponds to 16 ms using a subcarrier spacing of 3.75 kHz as defined in narrowband-IoT over 5G [48]. This choice aligns with the channel coherence time of practical channels, especially low mobility use cases, and, at the same time, limits the number of jointly processed symbols, hence, reducing overall computational complexity and latency. In practice, the number of jointly processed symbols can be chosen arbitrarily according to the complexity and delay requirements, provided that all processed symbols remain within the same channel coherence interval. Note that in the simulation we implement sequentially an algorithm that should be implemented in parallel in a distributed setting. Furthermore, the computations within the CPU and each AP can be parallelized as well. Hence, the algorithm is still scalable even for larger systems which are, however, not considered here for limiting the simulation complexity.

Both the JACD-EP and JACD-EP-BG algorithm use the JAC-EP algorithm from [2] for pilot-based initialization. For benchmarking, we consider the centralized linear MMSE MIMO detector [25] using (1) imperfect activity and channel information obtained from the JAC-EP algorithm and (2) genie-aided perfect activity and channel knowledge. Additionally, we consider the genie-aided MMSE channel estimator with perfectly known data symbols and the bilinear-EP algorithm [30] with perfect UE activity knowledge and pilot-based MMSE channel estimation initialization as performance upper bounds. All the iterative EP-based algorithms employ a damping parameter of $\eta = 0.5$ and perform 20 iterations.

The performance is assessed in terms of the detection error rate (DER) for UE activity detection, the normalized mean squared error (NMSE) for channel estimation, and the symbol error rate

(SER) for data detection. The performance metrics are averaged across all UEs. The DER, which incorporates misdetections and false alarms, is computed as $\text{DER} \coloneqq E\{\sum_k \mathbb{1}_{u_k \neq \hat{u}_k}/K\}$, the NMSE of the channel estimate $\hat{\mathbf{G}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$ is defined as $\text{NMSE} \coloneqq \frac{E\{||\mathbf{G}-\hat{\mathbf{G}}||_F^2\}}{E\{||\mathbf{G}||_F^2\}}$ with $\mathbf{G} = \mathbf{H}\mathbf{U}$, and the SER is given by $\text{SER} \coloneqq E\Big\{\sum_k \sum_t \mathbb{1}_{x_{kt}^d \neq \hat{x}_{kt}^d}/(KT_d) \,|\, u_k = 1\Big\}$. Note that as active UEs are erroneously classified as inactive, the corresponding data symbol estimates are randomly chosen from the symbol constellation $\mathcal{X}$, leading to a higher SER in case of misdetections. The expectation operator in the definitions above is computed with respect to the UE activity, channel, and noise realizations. The results in Fig. 4 are obtained by averaging over 10^3 block transmissions, each with independent UE positions and activities. Fig. 4 illustrates the three performance metrics as a function of the pilot sequence length T_p, which reflects the strength of PC. The proposed algorithms outperform the JAC-EP initialization algorithm in terms of DER and NMSE. The DER performance enhances steadily with increasing T_p, whereas the NMSE performance saturates and approaches the MMSE lower bound for $T_p > 6$ due to the fixed coherence block length $T = 60$. Furthermore, the proposed EP-based algorithms significantly outperform the linear MMSE MIMO detector in terms of the SER. The proposed algorithms reach the same performance and even outperform the genie-aided linear MMSE MIMO detector with perfect activity and channel knowledge for $T_p > 6$. A noticeable performance gap remains with respect to the genie-aided bilinear-EP algorithm due to the challenging UE activity detection under PC. Hence, we can conclude that the overall system performance is limited by the activity detection capability. Furthermore, the JACD-EP-BG algorithm outperforms the JACD-EP algorithm for small T_p which corresponds to strong PC.

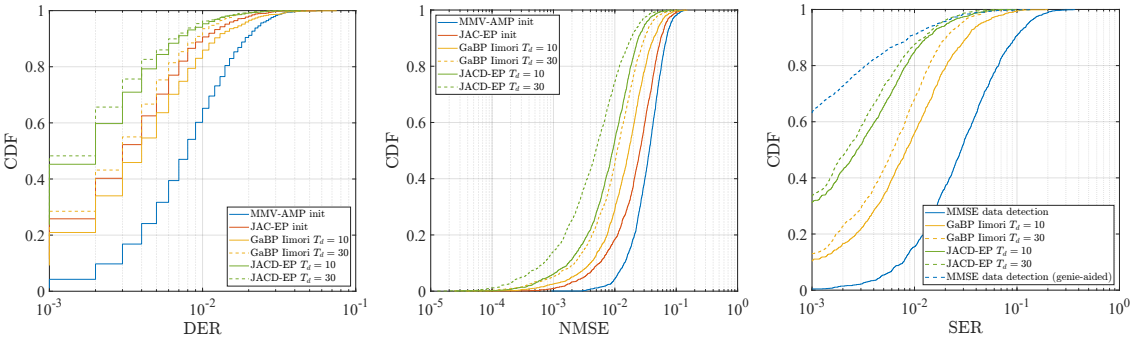
Next, we investigate the quality of service (QoS) distribution within the network by evaluating the empirical cumulative distribution functions (CDFs) of the DER, NMSE, and SER in Fig. 5. The three metrics are computed per UE, i.e., $\text{DER}_k \coloneqq E\{\mathbb{1}_{u_k \neq \hat{u}_k}\}$, $\text{NMSE}_k \coloneqq \frac{E\{||\mathbf{g}_k-\hat{\mathbf{g}}_k||^2\}}{E\{||\mathbf{g}||^2\}}$ with $\hat{\mathbf{g}}_k$ and $\mathbf{g}_k$ being the k^{th} column of $\hat{\mathbf{G}}$ and $\mathbf{G}$, respectively, and $\text{SER}_k \coloneqq E\Big\{\sum_t \mathbb{1}_{x_{kt}^d \neq \hat{x}_{kt}^d}/T_d \,|\, u_k = 1\Big\}$. The CDFs are obtained by considering 100 different realizations of the UE positions, resulting in $100 \cdot K = 4000$ data points. For each realization, the metrics are averaged over 10^3 independent block transmissions, accounting for UE activity, small-scale fading, and noise realizations. Pilot sequences of length $T_p = 6$ are used. Fig. 5 shows again the superior performance of the proposed algorithms. Furthermore, it can be observed that the JACD-EP-BG algorithm outperforms the JACD-EP algorithm, especially for UEs experiencing good propagation conditions.

7 Further Performance Analysis

In this section, we further investigate the performance under PC and how to reduce the fronthaul load for the EP-based JACD algorithms. We focus on the JACD-EP algorithm. Similar results can be obtained for the JACD-EP-BG algorithm. We investigate the performance under stronger PC compared to the simulations presented Section 6 by employing randomly generated pilot sequences instead of mutual coherence-optimized pilot sequences. Furthermore, 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. Additionally, a new scheduling scheme for the message passing is investigated. More details on these approaches are provided in the following.

7.1 Simulations with Randomly Generated Pilot Sequences [2]

We consider a network of size 400×400 m with $L = 16$ APs placed on a regular grid at the following positions $\{(50+i\cdot 100, 50+j\cdot 100)\,\text{m} \,|\, i,j \in \{0,1,2,3\}\}$ and a height of 10 m. On the ground, $K = 16$ UEs are deployed according to a uniform point process (UPP). We consider 100 different outcomes of the UPP. For each UPP outcome, the large-scale fading coefficients are deterministic and computed according to the 3GPP urban microcell model [25] whereas for the random variables, i.e., small-scale fading coefficients, activities, and pilot sequences, we generate 10^3 realizations.



(a) CDF of DER. (b) CDF of NMSE. (c) CDF of SER.

Figure 6: Numerical results for $L = 16$, $N = 1$, $K = 16$, $\lambda = 0.5$, $T_p = 8$, $T_d = \{10, 30\}$, and random BPSK pilots.

This allows us to analyze the QoS distribution in the network. The activity probability of all users is set to $\lambda = 0.5$, and the transmit power is $\sigma_x^2 = 16\,\text{dBm}$. Furthermore, random binary phase-shift-keying (BPSK) pilot sequences of length $T_p = 8$ are utilized and $T_d = \{10, 30\}$ 4-QAM symbols are transmitted during the data transmission phase. The noise power at each AP is set to $\sigma_n^2 = -96\,\text{dBm}$. As benchmark schemes, we consider the centralized linear MMSE data detector and the Gaussian belief propagation (GaBP) algorithm in [38]. Both of these schemes are initialized by the multiple measurement vector AMP (MMV-AMP) algorithm in [14]. Besides, we consider the genie-aided centralized linear MMSE detector with perfect UE activity and channel knowledge. For all iterative algorithms, the damping parameter is $\eta = 0.5$. Furthermore, the maximum number of iterations for the MMV-AMP algorithm is set to 200, whereas the GaBP- and EP-based algorithms perform 20 iterations.

For assessing the user activity detection, channel estimation, and data detection performance, we consider the empirical CDFs of the DER, the NMSE, and the SER, respectively. The empirical CDFs are obtained considering all UEs in each UPP outcome whereas DER, NMSE, and SER are computed by averaging over the 10^3 small-scale fading and UE activity realizations generated for each UPP outcome. The DER is defined as the ratio of the number of erroneously classified UE activities and the total number of UE activity realizations, i.e., $\text{DER} := E\{\mathbb{1}_{\hat{u}_k \neq u_k}\}$, and includes both misdetections and false alarms. The corresponding simulation results are illustrated in Fig. 6a. The NMSE and SER are considered naturally for active users only, i.e., $\text{NMSE} := E\left\{\frac{||\hat{\mathbf{h}}_k - \mathbf{h}_k||}{||\mathbf{h}_k||} \mid u_k = 1\right\}$, $\text{SER} := E\{\mathbb{1}_{\hat{x}_{kt} \neq x_{kt}} \mid u_k = 1, t > T_p\}$. Furthermore, in the computation of the NMSE, we neglect weak channels, i.e., channels whose large-scale fading coefficient multiplied by the transmit power is smaller than the noise power at the APs, to avoid taking into account errors on weak and insignificant channels. The channel estimation and data detection performance is depicted in Figs. 6b and 6c, respectively. The results illustrate the advantages of using jointly pilot and data sequences for both data and activity detection and the superior performance of the JACD-EP algorithm compared to state-of-the-art benchmark schemes. Furthermore, it can be observed that an increase of the data length improves the performance with respect to all considered metrics.

7.2 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. In the following, we investigate how the geographical distance between UEs affect the system

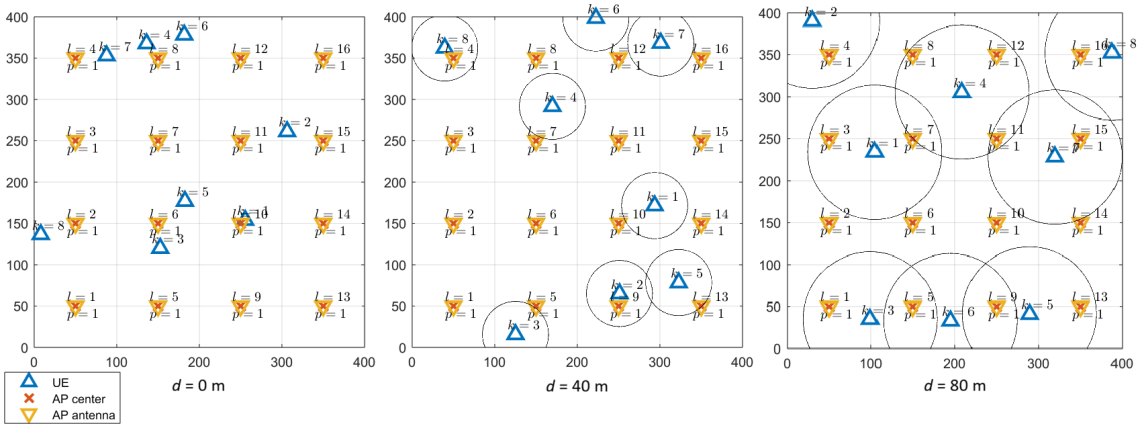


Figure 7: Network realization with different minimum UE distance constraints d.

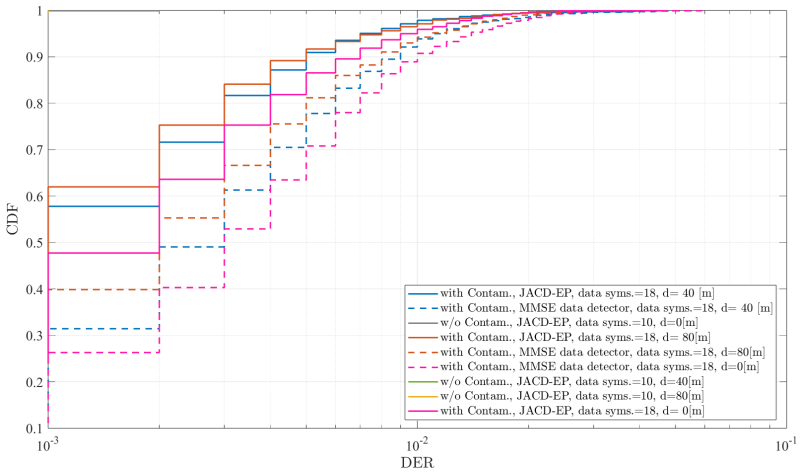


Figure 8: CDF of DER with minimum distance constraint between UEs.

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. 7 for three different values of $d \in \{0, 40, 80\}$ m. Here, the network area of 400×400 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 7.1 except that $T = 26$ and $T_p \in \{8, 16\}$ depending on the scenario with and without PC, respectively. For $T_p = 8$, random BPSK pilot sequences are transmitted by the UEs, whereas for $T_p = 16$ orthogonal pilot sequences are used. The minimum UE distance constraint is set to $d \in \{0, 40, 80\}$ m. Figs. 8, 9, and 10 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 shows the effective exploitation of geographical UE separation in the JACD-EP algorithm for pilot decontamination.

7.3 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

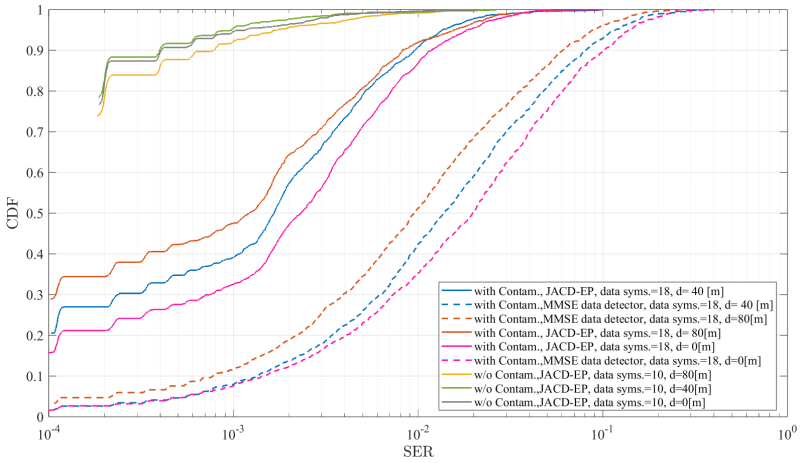


Figure 9: CDF of SER with minimum distance constraint between UEs.

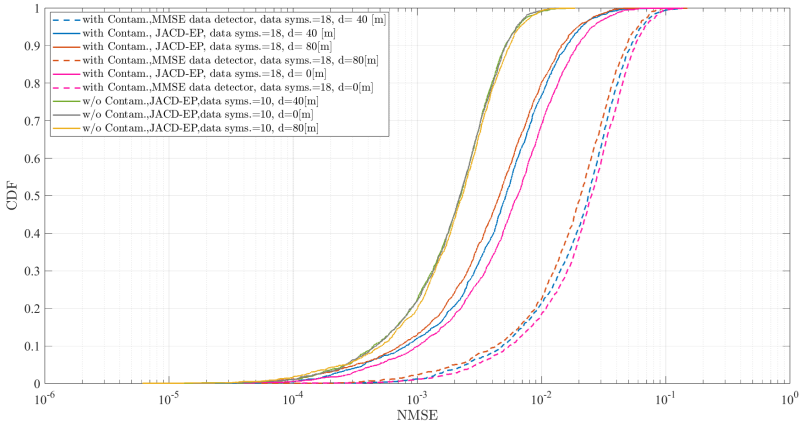


Figure 10: CDF of NMSE with minimum distance constraint between UEs.

APs and the CPU for JACD via the JACD-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 the JACD-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 JACD-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

$$\text{round}\left(i \cdot \frac{i_{\max}}{n_{\text{fh}}}\right), \quad i \in \{1, 2, \ldots, n_{\text{fh}}\}. \tag{75}$$

In this modified framework, the scheduling of messages the $m_{\Psi_{z_{l,kt}};x_{kt}}$ and $m_{\Psi_{g_{l,k}};u_k}$ (from APs to CPU) and $m_{x_{kt};\Psi_{z_{l,kt}}}$ and $m_{u_k;\Psi_{g_{l,k}}}$ (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

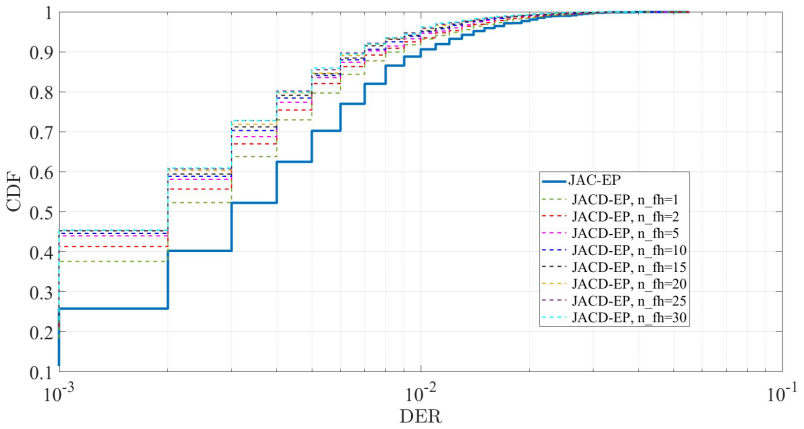


Figure 11: CDF of DER with reduced fronthaul load.

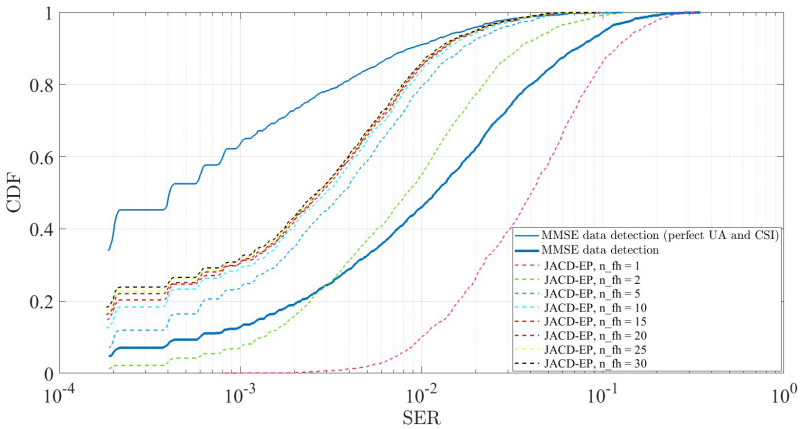


Figure 12: CDF of SER with reduced fronthaul load.

at the CPU and fronthaul usage while maintaining effective iterative refinement of activity, channel, and data estimates, ensuring a balance between system performance and efficiency.

For the following simulations, the same parameters are used as in Section 7.1 with a different number of data symbols $T_d = 14$ 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. 11, 12, and 13. The JACD-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 JACD-EP refines estimates through iterative processing. Furthermore, it can be observed that the JACD-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.

Finally, we consider a new scheduling scheme that addresses the sequential updating of $m_{\Psi_{z_{l,kt}};x_{kt}}$ and $m_{\Psi_{g_{l,k}};u_k}$. So far, first, each AP sends data symbol information to the CPU, then receives updated data symbol information from the CPU, and subsequently uses this updated information to compute new beliefs about user activities, which are then shared with the CPU. This sequential scheduling results in two fronthaul transmissions per iteration, increasing latency and communication overhead. To optimize the process, it is necessary to develop a more efficient scheduling scheme that enables the computation of message updates with only a single fronthaul transmission per iteration, thereby reducing the communication cost and improving system efficiency. To optimize

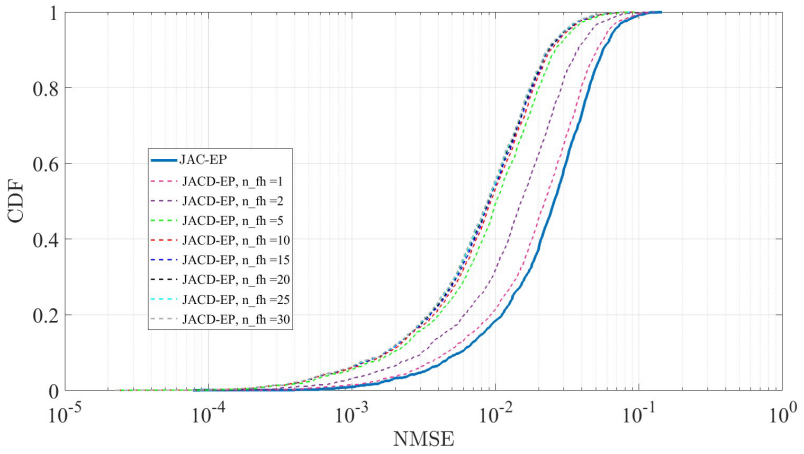


Figure 13: CDF of NMSE with reduced fronthaul load.

this scheduling in the JACD-EP algorithm, the transmission of beliefs on data symbols and user activities is modified to occur simultaneously. The revised scheduling addresses this inefficiency by forwarding the beliefs on data symbols (message update #2) and user activities (message update #6) from the APs to the CPU at the same time. Similarly, the information transfer from the CPU back to the APs (message updates #3 and #7) is performed simultaneously. The new message update scheduling is given by $1 \rightarrow 4 \rightarrow 5 \rightarrow 2 \rightarrow 6 \rightarrow 3 \rightarrow 7 \rightarrow 8 \rightarrow 9 \rightarrow 10$ where the number describes the message update shown in Fig. 3 as red number. The corresponding simulation results did not show significant performance changes compared to the original scheduling, hence, we do not show these plots here. This means that the algorithm with the new scheduling attains the same performance as illustrated in Figs. 11, 12, and 13. Hence, we can conclude that we do not loose in performance by deviating from the sequential updating of $m_{\Psi_{z_{l,kt}};x_{kt}}$ and $m_{\Psi_{g_{l,k}};u_k}$ which results in a lower fronthaul overhead.

8 Conclusion

In this Task 4.2, we considered the uplink of a GF-CF-MaMIMO system and tackled the JACD problem for massive MTC under PC. We developed an EP-based framework for distributed and scalable JACD by appropriately factorizing the APP distribution and applying EP message passing on the associated factor graph. Within this framework, we employed accurate categorical probability distributions for user activities and data, and Gaussian or BG probability distributions for the channels, resulting in the JACD-EP and the JACD-EP-BG algorithms, respectively. To support the latter, we developed an exponential family representation of the BG distribution. The proposed framework is inherently scalable with respect to both the number of APs and UEs. The message-passing structure enables straightforward extensions toward fully distributed implementations, while the computational complexity at each AP scales linearly with the number of UEs in its coverage area. The numerical results showed that the proposed algorithms exhibit strong robustness against PC and outperform state-of-the-art algorithms in terms of DER, NMSE, and SER. Then, we further analyzed the performance of the proposed bilinear-EP algorithm under randomly generated pilot sequences, which results in stronger PC, minimum UE distance constraints, which makes the level of PC in the network controllable to some extent, and under reduced fronthaul capabilities. The results showed that the algorithm performs effective pilot decontamination under different levels of PC and that a reduction of the fronthaul load and overhead is feasible with negligible performance degradation.

A Proof of the Bernoulli-Gaussian Product Lemma

Using the exponential family representation of BG distributions, $\mathcal{BG}(\mathbf{x}|\lambda_i, \boldsymbol{\mu}_i, \mathbf{C}_i) = \mathrm{e}^{\boldsymbol{\eta}_{\mathrm{BG},i}^{\mathsf{H}}\mathbf{u}_{\mathrm{BG}}(\mathbf{x}) - A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},i})}$, $i = \{1, 2\}$, the product of two BG distributions is given by

$$\begin{aligned}
&\mathcal{BG}(\mathbf{x}|\lambda_1, \boldsymbol{\mu}_1, \mathbf{C}_1) \cdot \mathcal{BG}(\mathbf{x}|\lambda_2, \boldsymbol{\mu}_2, \mathbf{C}_2) \\
&\quad = \mathrm{e}^{(\boldsymbol{\eta}_{\mathrm{BG},1}+\boldsymbol{\eta}_{\mathrm{BG},2})^{\mathsf{H}}\mathbf{u}_{\mathrm{BG}}(\mathbf{x}) - A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},1}+\boldsymbol{\eta}_{\mathrm{BG},2})} \\
&\qquad \cdot \mathrm{e}^{A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},1}+\boldsymbol{\eta}_{\mathrm{BG},2}) - A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},1}) - A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},2})} \\
&\quad = \mathcal{BG}(\mathbf{x}|\lambda, \boldsymbol{\mu}, \mathbf{C}) \cdot \mathrm{e}^{A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},1}+\boldsymbol{\eta}_{\mathrm{BG},2}) - A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},1}) - A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG},2})}.
\end{aligned} \tag{76}$$

The first factor in (76) is a normalized BG distribution with vector of natural parameters $\boldsymbol{\eta}_{\mathrm{BG}} = \boldsymbol{\eta}_{\mathrm{BG},1} + \boldsymbol{\eta}_{\mathrm{BG},2}$, or equivalently, $\kappa = \kappa_1 + \kappa_2$, $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2$, and $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2$. From $\boldsymbol{\gamma}$ and $\boldsymbol{\Lambda}$, the covariance matrix and mean of the Gaussian component in (11) and (12), respectively, immediately follows. Recall that $A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}}) = \log \frac{1}{\mathrm{e}^{-A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})}} = \log \frac{1}{\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu},\mathbf{C})}$. Now, using the Gaussian product lemma and denoting $\boldsymbol{\eta}_{\mathrm{G}} = \boldsymbol{\eta}_{\mathrm{G},1} + \boldsymbol{\eta}_{\mathrm{G},2}$, the first natural parameter can be rewritten as

$$\begin{aligned}
\kappa &= \kappa_1 + \kappa_2 \\
&= \log\Big(\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2} \cdot \frac{1}{\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1, \mathbf{C}_1)\,\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_2, \mathbf{C}_2)}\Big) \\
&= \log\Big(\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2} \\
&\qquad \cdot \frac{1}{\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}, \mathbf{C})\,\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \mathbf{C}_1 + \mathbf{C}_2)}\Big) \\
&= \log\Big(\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2} \cdot \frac{1}{\mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \mathbf{C}_1 + \mathbf{C}_2)}\Big) \\
&\quad + A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}}),
\end{aligned} \tag{77}$$

where $\mathbf{C}$ and $\boldsymbol{\mu}$ are given by (11) and (12), respectively. Using the relation $\lambda = \frac{1}{1+\mathrm{e}^{\kappa - A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})}}$ together with (77) yields (10). Noting that $\mathrm{e}^{A_{\mathrm{BG}}(\boldsymbol{\eta}_{\mathrm{BG}})} = \mathrm{e}^{A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}}) - \log\lambda} = \mathrm{e}^{A_{\mathrm{G}}(\boldsymbol{\eta}_{\mathrm{G}})} \cdot \frac{1}{\lambda} = \frac{1}{\lambda \cdot \mathcal{CN}(\mathbf{0}|\boldsymbol{\mu},\mathbf{C})}$ and applying the Gaussian product lemma and (10), it can be shown that the normalization constant for the BG product given in (76) is equal to the normalization constant stated in (9).

B JAC-EP Initialization Algorithm [2]

The JAC-EP algorithm can be obtained by running the JACD-EP algorithm for the pilot part only, i.e., for $t \leq T_p$, with priors $p_{u_k}(u_k)$ and $p_{h_{l,k}}(\mathbf{h}_{l,k})$. However, since we do not have to estimate the symbols x_{kt} during initialization, the factorization of the APP and the corresponding factor graph and algorithm simplify. In the following, we describe the simplified JAC-EP algorithm in more detail. We use the same system model and notation as in the main body of this document except that we only consider signals in the pilot phase, i.e., $1 \leq t \leq T_p$.

B.1 Factor Graph Representation

The task of the JAC-EP algorithm is to find initial estimates on the user activities and channels. Hence, the factorized APP distribution with auxiliary variables $\mathbf{g}_{l,k} \coloneqq \mathbf{h}_{l,k}u_k$ is given by

$$p_{\mathrm{APP}}(\mathbf{U}, \mathbf{H}, \mathbf{G}) \propto \prod_{l=1}^{L}\prod_{k=1}^{K}\prod_{t=1}^{T_p}\Big[p(\mathbf{y}_{l,t}|\mathbf{g}_{l,1},...,\mathbf{g}_{l,K}) \cdot p(\mathbf{g}_{l,k}|\mathbf{h}_{l,k}, u_k) \cdot p_{u_k}(u_k) \cdot p_{h_{l,k}}(\mathbf{h}_{l,k})\Big]. \tag{78}$$

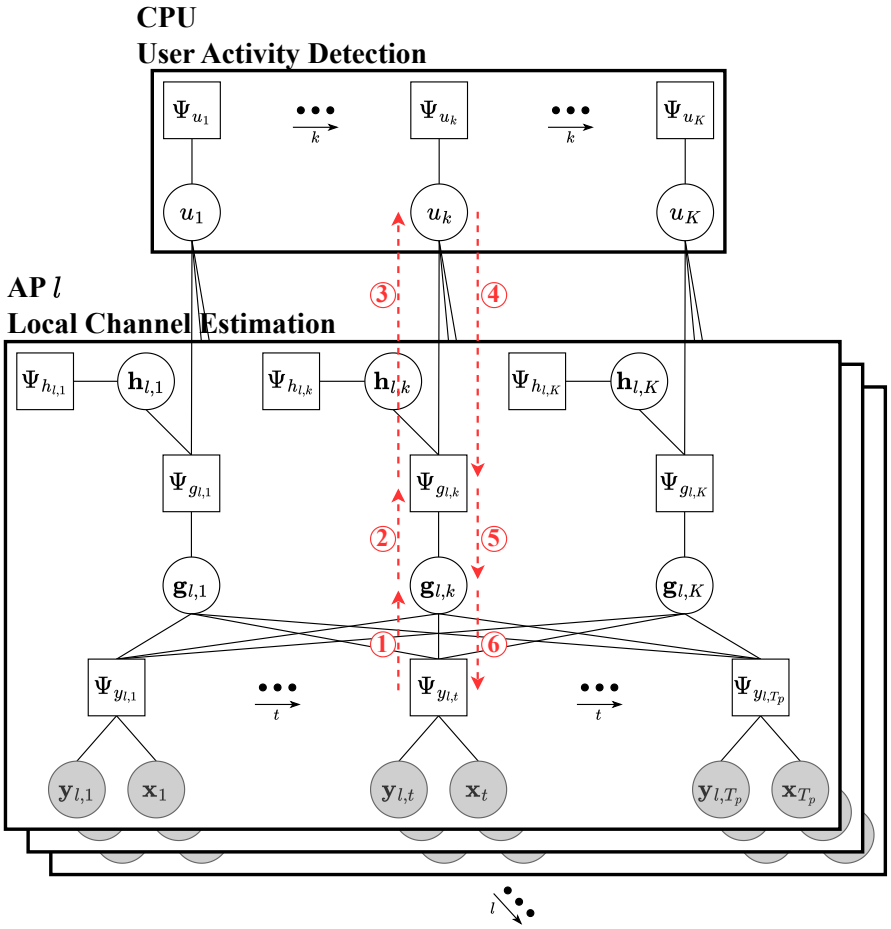


Figure 14: Factor graph for JAC-EP. The numbered red dashed arrows show the flow of information according to the scheduling presented in Algorithm 2. Each number corresponds to one message update in Algorithm 2.

In contrast to (13), the auxiliary variables $\mathbf{z}_{l,kt}$ are irrelevant since the pilot symbols x_{kt} are assumed to be known. The probability distributions in (78) are given by

$$\Psi_{y_{l,t}} \coloneqq p(\mathbf{y}_{l,t}|\mathbf{g}_{l,1},...,\mathbf{g}_{l,K}) = \mathcal{CN}\left(\mathbf{y}_{l,t}\middle|\sum_{k=1}^{K}\mathbf{g}_{l,k}x_{kt},\sigma_n^2\mathbf{I}_N\right), \tag{79}$$

$$\Psi_{u_k} \coloneqq p_{u_k}(u_k) = (1-\lambda)\mathbb{1}_{u_k=0} + \lambda\mathbb{1}_{u_k=1}, \tag{80}$$

$$\Psi_{h_{l,k}} \coloneqq p_{h_{l,k}}(\mathbf{h}_{l,k}) = \mathcal{CN}\big(\mathbf{h}_{l,k}\big|\mathbf{0},\boldsymbol{\Xi}_{l,k}\big), \tag{81}$$

and $\Psi_{g_{l,k}} \coloneqq p(\mathbf{g}_{l,k}|\mathbf{h}_{l,k},u_k)$ is given in (16). The corresponding factor graph is illustrated in Fig. 14.

B.2 EP Approximations and Fronthaul Load

We choose the same approximate exponential family distributions for the variable nodes as in the JACD-EP algorithm, i.e., activities u_k are modeled by categorical distributions whereas auxiliary variables $\mathbf{g}_{l,k}$ and channels $\mathbf{h}_{l,k}$ are modeled by multivariate complex Gaussian distributions. Therefore, messages from and towards u_k consist of one probability value which yields a fronthaul load of $2LK$ real-valued numbers. Messages involving $\mathbf{g}_{l,k}$ and $\mathbf{h}_{l,k}$ consist of complex-valued vec-

Algorithm 2 JAC-EP Algorithm

Input: Pilot matrix $\mathbf{X}^p$, received pilot signal $\mathbf{Y}^p$, noise variance σ_n^2, user activity probability λ, channel correlation matrices $\boldsymbol{\Xi}_{l,k}$.

Output: Estimated activities $\hat{u}_k$ and channels $\hat{\mathbf{h}}_{l,k}$.

1: $\forall k, l, t \leq T_p$: Initialize $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ and $m_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}}$ via (82), (83).
2: **for** $i = 1$ to $i_{\max}$ **do**
3: $\forall k, l, t \leq T_p$: Update $m_{\Psi_{y_{l,t}};\mathbf{g}_{l,k}}$ via (90), (91)/(92).
4: $\forall k, l$: Update $m_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}$ via (94), (95).
5: $\forall k, l$: Update $m_{\Psi_{g_{l,k}};u_k}$ via (96).
6: $\forall k, l$: Update $m_{u_k;\Psi_{g_{l,k}}}$ via (93).
7: $\forall k, l$: Update $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ via (97), (98).
8: $\forall k, l, t \leq T_p$: Update $m_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}}$ via (88), (89).
9: **end for**
10: **return** $\hat{u}_k$ calculated via (84) $\forall k$.
11: **return** $\hat{\mathbf{h}}_{l,k}$ calculated via (85) $\forall k, l$.

tors and matrices of dimension N and $N \times N$, respectively, which do not contribute to the fronthaul load since they are only processed within an AP.

B.3 Initialization, Scheduling, and Estimation

As indicated in Appendix B.1, the prior distributions $p_{u_k}(u_k)$ and $p_{h_{l,k}}(\mathbf{h}_{l,k})$ are used for initializing the JAC-EP algorithm. The initial mean vector and covariance matrix of $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ and $m_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}}$ $\forall k, l, t \leq T_p$ are set according to the prior information on $\mathbf{g}_{l,k}$,

$$\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}} = \mathbf{0}, \tag{82}$$

$$\mathbf{C}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}} = \lambda \cdot \boldsymbol{\Xi}_{l,k}. \tag{83}$$

For all other messages, an uninformative initialization is chosen. Then, we update all the messages according to the scheduling given in Algorithm 2 which is illustrated in Fig. 14 as red dashed arrows. Finally, the estimates of the user activities and channels are computed by

$$\hat{u}_k = \underset{u_k \in \{0,1\}}{\arg\max}\ \hat{p}_{u_k}(u_k), \tag{84}$$

$$\hat{\mathbf{h}}_{l,k} = \underset{\mathbf{h}_{l,k} \in \mathbb{C}^N}{\arg\max}\ \hat{p}_{\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) = \frac{1}{Z_{l,k}} \pi_{u_k;\Psi_{g_{l,k}}}(1)\, \vartheta(1)\, \check{\boldsymbol{\mu}}_{l,k}, \tag{85}$$

with the approximations of the posterior distributions

$$\hat{p}_{u_k}(u_k) \propto p_{u_k}(u_k) \cdot \prod_{l=1}^{L} \pi_{\Psi_{g_{l,k}};u_k}(u_k), \tag{86}$$

$$\hat{p}_{\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) \propto p_{h_{l,k}}(\mathbf{h}_{l,k}) \cdot \mathcal{CN}\Big(\mathbf{h}_{l,k} \big| \boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}, \mathbf{C}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\Big). \tag{87}$$

and $\pi_{u_k;\Psi_{g_{l,k}}}(u_k)$, $\vartheta(u_k)$, $\check{\boldsymbol{\mu}}_{l,k}$, $Z_{l,k}$, $\pi_{\Psi_{g_{l,k}};u_k}(u_k)$, $\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}$, and $\mathbf{C}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}$ defined in Appendix B.4.

As in the JACD-EP algorithm, we apply damping to the factor-to-variable messages with damping parameter $\eta \in [0, 1]$ Furthermore, we update the parameters of $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$ in line 7 of Algorithm 2 only if the new covariance matrix obtained by (97) is symmetric positive definite. Otherwise, we keep the parameters from the previous iteration.

B.4 Message-Passing Update Rules

In this section, we present the EP message-passing update rules for the factor graph in Fig. 14. Detailed derivations are omitted since they are very similar to the ones presented in Appendix D. As in the description of the JACD-EP algorithm, we switch between the representation of a Gaussian via the mean vector and covariance matrix and the natural parameters without explicitly mentioning the transformation.

Update of $m_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}}$*:*

$$\mathbf{\Lambda}_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}} = \mathbf{\Lambda}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \sum_{t' \neq t} \mathbf{\Lambda}_{\Psi_{y_{l,t'}};\mathbf{g}_{l,k}}. \tag{88}$$

$$\boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{y_{l,t}}} = \boldsymbol{\gamma}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \sum_{t' \neq t} \boldsymbol{\gamma}_{\Psi_{y_{l,t'}};\mathbf{g}_{l,k}}, \tag{89}$$

Update of $m_{\Psi_{y_{l,t}};\mathbf{g}_{l,k}}$*:*

$$\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{g}_{l,k}} = \left(\mathbf{y}_{l,t} - \sum_{k' \neq k} \boldsymbol{\mu}_{\mathbf{g}_{l,k'};\Psi_{y_{l,t}}} x_{k't} \right) \cdot x_{kt}^{-1}, \tag{90}$$

$$\mathbf{C}_{\Psi_{y_{l,t}};\mathbf{g}_{l,k}} = \left(\sigma_n^2 \mathbf{I}_N + \sum_{k' \neq k} \mathbf{C}_{\mathbf{g}_{l,k'};\Psi_{y_{l,t}}} |x_{k't}|^2 \right) \cdot |x_{kt}|^{-2}. \tag{91}$$

Note that the modification for PC described in Section B.5 can be used to enhance the performance of the JAC-EP algorithm. The update in (91) is then given by

$$\mathbf{C}_{\Psi_{y_{l,t}};\mathbf{g}_{l,k}} = \left(\sigma_n^2 \mathbf{I}_N + \sum_{k' \in \mathcal{P}_k} \mathbf{\Xi}_{l,k} |x_{k't}^p|^2 + \sum_{k' \neq k} \mathbf{C}_{\mathbf{g}_{l,k'};\Psi_{y_{l,t}}} |x_{k't}|^2 \right) \cdot |x_{kt}|^{-2}. \tag{92}$$

Update of $m_{u_k;\Psi_{g_{l,k}}}$*:*

$$\pi_{u_k;\Psi_{g_{l,k}}}(u_k) \propto p_{u_k}(u_k) \cdot \prod_{l' \neq l} \pi_{\Psi_{g_{l',k}};u_k}(u_k), \tag{93}$$

Update of $m_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}$*:*

$$\mathbf{\Lambda}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} = \sum_{t=1}^{T_p} \mathbf{\Lambda}_{\Psi_{y_{l,t}};\mathbf{g}_{l,k}}, \tag{94}$$

$$\boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} = \sum_{t=1}^{T_p} \boldsymbol{\gamma}_{\Psi_{y_{l,t}};\mathbf{g}_{l,k}}. \tag{95}$$

Update of $m_{\Psi_{g_{l,k}};u_k}$*:*

$$\pi_{\Psi_{g_{l,k}};u_k}(u_k) \propto \vartheta(u_k), \tag{96}$$

with

$$\vartheta(u_k) = \mathcal{CN}(\mathbf{0} | \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \mathbf{\Xi}_{l,k} u_k).$$

Update of $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$*:*

$$\mathbf{\Lambda}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \hat{\mathbf{\Lambda}}_{l,k} - \mathbf{\Lambda}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \tag{97}$$

$$\boldsymbol{\gamma}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \hat{\boldsymbol{\gamma}}_{l,k} - \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \tag{98}$$

with

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_{l,k} &= \frac{1}{Z_{l,k}} \cdot \pi_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \cdot \check{\boldsymbol{\mu}}_{l,k}, \\
\hat{\mathbf{C}}_{l,k} &= \frac{1}{Z_{l,k}} \cdot \pi_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \cdot (\check{\mathbf{C}}_{l,k} + \check{\boldsymbol{\mu}}_{l,k}\check{\boldsymbol{\mu}}_{l,k}^H) - \hat{\boldsymbol{\mu}}_{l,k}\hat{\boldsymbol{\mu}}_{l,k}^H. \\
Z_{l,k} &= \pi_{u_k;\Psi_{g_{l,k}}}(0) \cdot \vartheta(0) + \pi_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \\
\check{\boldsymbol{\Lambda}}_{l,k} &= \boldsymbol{\Lambda}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \boldsymbol{\Xi}_{l,k}^{-1}, \\
\check{\boldsymbol{\gamma}}_{l,k} &= \boldsymbol{\gamma}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}.
\end{aligned}$$

B.5 Modification for Pilot Contamination

Since the number of UEs is larger than the pilot sequence length, PC will naturally occur and degrade the performance. In the following, we propose a simple heuristic approach to account for PC. The efficacy of the proposed approach was verified via simulations.

The adverse effect of PC is particularly pronounced for UEs that share fully-correlated pilot sequences, i.e., the same pilot sequence or a rotated version of it, and are close to each other. For these cases, we propose to add a correction term to the covariance matrix in (28) for $t \leq T_p$ when generating the prior information via the pilot-based JAC-EP initialization algorithm. We denote the set of UEs which use fully-correlated pilot sequences with respect to UE k by $\mathcal{P}_k$. By adding the covariance matrix of the received signals from UEs in $\mathcal{P}_k$ which cause PC to UE k at AP l to the covariance matrix in (28), we decrease the reliability of information coming from the factor node $\Psi_{y_{l,t}}$ especially when pilot-sharing users are close to the considered AP. Then, the covariance matrix with the correction term is given by

$$\mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} = \sigma_n^2 \mathbf{I}_N + \sum_{k' \in \mathcal{P}_k} \boldsymbol{\Xi}_{l,k} |x_{k't}^p|^2 + \sum_{k' \neq k} \mathbf{C}_{\Psi_{z_{l,k't}};\mathbf{z}_{l,k't}}. \tag{99}$$

C Properties of Gaussian Distributions

Lemma 1 (Gaussian Product Lemma)**.** *The product of two Gaussian distributions yields a new unnormalized Gaussian distribution,*

$$\mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{C}_1) \cdot \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_2, \mathbf{C}_2) = \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) \cdot \mathcal{CN}(\mathbf{0}|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \mathbf{C}_1 + \mathbf{C}_2) \tag{100}$$

with

$$\mathbf{C} = \left(\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1}\right)^{-1}, \tag{101}$$

$$\boldsymbol{\mu} = \mathbf{C}\left(\mathbf{C}_1^{-1}\boldsymbol{\mu}_1 + \mathbf{C}_2^{-1}\boldsymbol{\mu}_2\right). \tag{102}$$

Proof. See [45, 49]. □

Lemma 2 (Gaussian Quotient Lemma)**.** *The quotient of two Gaussian distributions yields a new unnormalized Gaussian distribution,*

$$\mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{C}_1) / \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_2, \mathbf{C}_2) \propto \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) \tag{103}$$

with

$$\mathbf{C} = \left(\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\right)^{-1}, \tag{104}$$

$$\boldsymbol{\mu} = \mathbf{C}\left(\mathbf{C}_1^{-1}\boldsymbol{\mu}_1 - \mathbf{C}_2^{-1}\boldsymbol{\mu}_2\right). \tag{105}$$

Proof. Follows from Lemma 1 by considering $\mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{C}_1) \propto \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) \cdot \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}_2, \mathbf{C}_2)$. □

Lemma 3 (Gaussian Scaling Lemma)**.** *Let $\mathbf{x} \in \mathbb{C}^N$ be a proper N-dimensional complex Gaussian random vector, $\mathbf{x} \sim \mathcal{CN}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$, and let $\mathbf{y} = c\,\mathbf{x}$ with c being a scalar constant. Then, $\mathbf{y}$ is a proper complex Gaussian random vector with mean $c\,\boldsymbol{\mu}$ and covariance matrix $|c|^2\,\mathbf{C}$, i.e., $\mathbf{y} \sim \mathcal{CN}\big(\mathbf{y} \,|\, c\,\boldsymbol{\mu}, |c|^2\,\mathbf{C}\big)$. Additionally, its PDF can be written in terms of the PDF of $\mathbf{x}$ as*

$$\mathcal{CN}\big(\mathbf{y} \,|\, c\,\boldsymbol{\mu}, |c|^2\,\mathbf{C}\big) = |c|^{-2N} \cdot \mathcal{CN}\big(c^{-1}\,\mathbf{y}|\boldsymbol{\mu}, \mathbf{C}\big)\,. \tag{106}$$

Proof. This lemma can be proven by applying the rules for transformations of random variables [50]. □

D Derivation of Message-Passing Update Rules

In the following, we briefly present the general message-passing update rules for EP on graphs. Next, we derive the message-passing rules for the JACD-EP-BG algorithm presented in Section 5.4. For brevity, we focus on factor-to-variable messages. Variable-to-factor messages can be easily derived by using the corresponding general EP message-passing update rule in (107) and the fact that (107) is applied to exponential family distributions, which makes the computation of products straightforward. Detailed derivations of the message-passing rules utilized in the JACD-EP algorithm can be found in the extended version of [2].

D.1 Expectation Propagation on Graphs

Consider a factor graph with factor nodes Ψ_α and variable nodes $\mathbf{x}_\beta$. Let $\mathbf{x}_\alpha$ be the vector containing all variables connected to Ψ_α. Let N_β denote the set of neighboring factor node indices of $\mathbf{x}_\beta$, i.e., $N_\beta = \{\alpha \,|\, \mathbf{x}_\beta \subseteq \mathbf{x}_\alpha\}$. The variable-to-factor message $m_{\mathbf{x}_\beta;\Psi_\alpha}$ is obtained by computing the parameters of the following distribution,

$$p_{\mathbf{x}_\beta;\Psi_\alpha}(\mathbf{x}_\beta) \propto \prod_{\alpha' \in N_\beta \setminus \alpha} p_{\Psi_{\alpha'};\mathbf{x}_\beta}(\mathbf{x}_\beta). \tag{107}$$

The factor-to-variable message $m_{\Psi_\alpha;\mathbf{x}_\beta}$ is computed by determining the parameters of

$$p_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta) \propto \frac{\mathrm{proj}\left\{q_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)\right\}}{p_{\mathbf{x}_\beta;\Psi_\alpha}(\mathbf{x}_\beta)}, \tag{108}$$

where the distribution $q_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$ is given by

$$q_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta) = \frac{1}{Z_{\Psi_\alpha}} \int \Psi_\alpha(\mathbf{x}_\alpha) \prod_{\beta' \in N_\alpha} p_{\mathbf{x}_{\beta'};\Psi_\alpha}(\mathbf{x}_{\beta'})\, \mathrm{d}\mathbf{x}_\alpha \setminus \mathbf{x}_\beta. \tag{109}$$

The distribution $q_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta)$ is also referred to as the *local belief of $\mathbf{x}_\beta$ at the factor node Ψ_α*. Here, Z_{Ψ_α} is a normalization constant and N_α denotes the set of neighboring variable node indices of Ψ_α, i.e, $N_\alpha = \{\beta \,|\, \mathbf{x}_\beta \subseteq \mathbf{x}_\alpha\}$. Furthermore, proj$\{\cdot\}$ denotes the projection onto the chosen exponential family defined as

$$\mathrm{proj}\{f(\mathbf{x})\} = \underset{g(\mathbf{x}) \in \mathcal{F}}{\arg\min}\, D_{KL}(f(\mathbf{x})||g(\mathbf{x})), \tag{110}$$

where $D_{KL}(f(\mathbf{x})||g(\mathbf{x}))$ is the Kullback-Leibler (KL) divergence between f and g and $\mathcal{F}$ is the exponential family with sufficient statistics $\mathbf{u}(\mathbf{x})$. This minimization is performed by *moment matching*,

$$E_{g(\mathbf{x})}\{\mathbf{u}(\mathbf{x})\} = E_{f(\mathbf{x})}\{\mathbf{u}(\mathbf{x})\}\,. \tag{111}$$

Once message passing converges, the approximate posterior distribution $\hat{p}_{\mathbf{x}_\beta}(\mathbf{x}_\beta)$ of the variable $\mathbf{x}_\beta$ can be computed via

$$\hat{p}_{\mathbf{x}_\beta}(\mathbf{x}_\beta) \propto \prod_{\alpha \in N_\beta} p_{\Psi_\alpha;\mathbf{x}_\beta}(\mathbf{x}_\beta). \tag{112}$$

D.2 JACD-EP Algorithm

Message Updates for Leaf Nodes Ψ_{u_k}, $\Psi_{h_{l,k}}$, and $\Psi_{x_{kt}}$

Since the prior distributions $\tilde{p}_{u_k}(u_k)$, $\tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k})$, and $p_x(x_{kt})$ are in the same exponential family as the approximate distributions of u_k, $\mathbf{h}_{l,k}$, and x_{kt}, respectively, the corresponding message updates simplify significantly. The factor-to-variable messages are constant and consist of the prior information on the variables u_k, $\mathbf{h}_{l,k}$, and x_{kt} which correspond to the distributions

$$p_{\Psi_{u_k};u_k}(u_k) = \tilde{p}_{u_k}(u_k), \tag{113}$$

$$p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) = \tilde{p}_{h_{l,k}}(\mathbf{h}_{l,k}), \tag{114}$$

$$p_{\Psi_{x_{kt}};x_{kt}}(x_{kt}) = p_x(x_{kt}). \tag{115}$$

This result is obtained by applying the message-passing rule in (108) while taking into account the fact that exponential family distributions are closed under multiplication. This makes the projection operation in (108) superfluous. Hence, the distributions corresponding to the updated messages are directly given by the factors Ψ_{u_k}, $\Psi_{h_{l,k}}$, and $\Psi_{x_{kt}}$, respectively, which correspond to the priors according to (17)-(19). The variable-to-factor messages for the leaf nodes are irrelevant in the unfolding of the algorithm and omitted here.

Message Updates for $\Psi_{y_{l,t}}$

Incoming messages to factor node $\Psi_{y_{l,t}}$

$$p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt}) = p_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}). \tag{116}$$

Outgoing messages from factor node $\Psi_{y_{l,t}}$

Message Update for $m_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}$

$$p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \propto \frac{\mathrm{proj}\Big\{q_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})\Big\}}{p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt})}, \tag{117}$$

with

$$
\begin{aligned}
&q_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \\
&\quad\propto \int\cdots\int \mathcal{CN}\left(\mathbf{y}_{l,t}\middle|\sum_{k'=1}^{K}\mathbf{z}_{l,k't},\sigma_n^2\mathbf{I}_N\right)\cdot p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt})\cdot\prod_{k'\neq k}p_{\mathbf{z}_{l,k't};\Psi_{y_{l,t}}}(\mathbf{z}_{l,k't})\,d\mathbf{z}_{l,k't} \\
&\quad\overset{(a)}{=} p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt})\cdot\int\cdots\int\mathcal{CN}\left(\mathbf{z}_{l,k''t}\middle|\mathbf{y}_{l,t}-\sum_{k'\neq k''}\mathbf{z}_{l,k't},\sigma_n^2\mathbf{I}_N\right) \\
&\qquad\cdot\prod_{k'\neq k}\mathcal{CN}\Big(\mathbf{z}_{l,k't}\big|\boldsymbol{\mu}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}},\mathbf{C}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}}\Big)\,d\mathbf{z}_{l,k't} \\
&\quad\overset{(b)}{=} p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt})\cdot\int\cdots\int\mathcal{CN}\big(\mathbf{z}_{l,k''t}|\boldsymbol{\mu}_{\text{tmp}},\mathbf{C}_{\text{tmp}}\big)\,d\mathbf{z}_{l,k''t} \\
&\qquad\cdot\mathcal{CN}\left(\mathbf{0}\middle|\mathbf{y}_{l,t}-\sum_{k'\neq k''}\mathbf{z}_{l,k't}-\boldsymbol{\mu}_{\Psi_{z_{l,k''t};\mathbf{z}_{l,k''t}}},\sigma_n^2\mathbf{I}_N+\mathbf{C}_{\Psi_{z_{l,k''t};\mathbf{z}_{l,k''t}}}\right) \\
&\qquad\cdot\prod_{k'\neq\{k,k''\}}\mathcal{CN}\Big(\mathbf{z}_{l,k't}\big|\boldsymbol{\mu}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}},\mathbf{C}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}}\Big)\,d\mathbf{z}_{l,k't} \\
&\quad\overset{(c)}{=} p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt})\cdot\int\cdots\int\mathcal{CN}\left(\mathbf{y}_{l,t}\middle|\sum_{k'\neq\{k''\}}\mathbf{z}_{l,k't}+\boldsymbol{\mu}_{\Psi_{z_{l,k''t};\mathbf{z}_{l,k''t}}},\sigma_n^2\mathbf{I}_N+\mathbf{C}_{\Psi_{z_{l,k''t};\mathbf{z}_{l,k''t}}}\right) \\
&\qquad\cdot\prod_{k'\neq\{k,k''\}}\mathcal{CN}\Big(\mathbf{z}_{l,k't}\big|\boldsymbol{\mu}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}},\mathbf{C}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}}\Big)\,d\mathbf{z}_{l,k't} \\
&\quad=\ldots \\
&\quad= p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt})\cdot\mathcal{CN}\left(\mathbf{z}_{l,kt}\middle|\mathbf{y}_{l,t}-\sum_{k'\neq k}\boldsymbol{\mu}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}},\sigma_n^2\mathbf{I}_N+\sum_{k'\neq k}\mathbf{C}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}}\right),
\end{aligned}
\tag{118}
$$

where (a) is obtained by a basic transformation of the Gaussian distribution and using (116), (b) is obtained by the Gaussian product rule (100) with $\boldsymbol{\mu}_1=\mathbf{y}_{l,t}-\sum_{k'\neq k''}\mathbf{z}_{l,k't}$, $\boldsymbol{\mu}_2=\boldsymbol{\mu}_{\Psi_{z_{l,k''t};\mathbf{z}_{l,k''t}}}$, $\mathbf{C}_1=\sigma_n^2\mathbf{I}_N$, and $\mathbf{C}_2=\mathbf{C}_{\Psi_{z_{l,k''t};\mathbf{z}_{l,k''t}}}$ which yields $\mathbf{C}_{\text{tmp}}$ and $\boldsymbol{\mu}_{\text{tmp}}$ according to (101) and (102), respectively, and (c) is obtained by integrating over $\mathbf{z}_{l,k''t}$ and applying a basic transformation of the Gaussian distribution. The final result in (118) is obtained by repeatedly applying the above described steps for all $k'\neq k$. Since $p_{\mathbf{z}_{l,kt};\Psi_{y_{l,t}}}(\mathbf{z}_{l,kt})$ is Gaussian distributed, we can conclude by utilizing the Gaussian product lemma that $q_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ is Gaussian distributed as well. Hence, the projection operation in (117) is superfluous since it projects $q_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ onto a Gaussian distribution which is the EP exponential family approximation choice of $\mathbf{z}_{l,kt}$. Therefore, the denominator of (117) cancels with the first term in (118), and the final message update rule is given by the distribution

$$
p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})=\mathcal{CN}\Big(\mathbf{z}_{l,kt}|\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}},\mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\Big),
\tag{119}
$$

with

$$
\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}=\mathbf{y}_{l,t}-\sum_{k'\neq k}\boldsymbol{\mu}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}},
\tag{120}
$$

$$
\mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}=\sigma_n^2\mathbf{I}_N+\sum_{k'\neq k}\mathbf{C}_{\Psi_{z_{l,k't};\mathbf{z}_{l,k't}}}.
\tag{121}
$$

Message Updates for $\Psi_{z_{l,kt}}$

Incoming messages to factor node $\Psi_{z_{l,kt}}$

$$\begin{aligned} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) &\propto p_{\Psi_{x_{kt}};x_{kt}}(x_{kt}) \cdot \prod_{l' \neq l} p_{\Psi_{z_{l',kt}};x_{kt}}(x_{kt}), \\ &\propto \prod_{l' \neq l} p_{\Psi_{z_{l',kt}};x_{kt}}(x_{kt}), \end{aligned} \tag{122}$$

$$p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) = p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}), \tag{123}$$

$$\begin{aligned} p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k}) &\propto p_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \cdot \prod_{t' \neq t} p_{\Psi_{z_{l,kt'}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \\ &\propto \mathcal{CN}\Big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big), \end{aligned} \tag{124}$$

with

$$\mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \left(\mathbf{C}^{-1}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \sum_{t' \neq t} \mathbf{C}^{-1}_{\Psi_{z_{l,kt'}};\mathbf{g}_{l,k}} \right)^{-1}, \tag{125}$$

$$\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} = \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \left(\mathbf{C}^{-1}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} \boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \sum_{t' \neq t} \mathbf{C}^{-1}_{\Psi_{z_{l,kt'}};\mathbf{g}_{l,k}} \boldsymbol{\mu}_{\Psi_{z_{l,kt'}};\mathbf{g}_{l,k}} \right), \tag{126}$$

which is obtained by applying the Gaussian product lemma multiple times. The update in (122) is obtained by noticing that $p_{\Psi_{x_{kt}};x_{kt}}(x_{kt})$ (115) is uninformative according to (19) for $t > T_p$.

Outgoing messages from factor node $\Psi_{z_{l,kt}}$

Message Update for $m_{\Psi_{z_{l,kt}};x_{kt}}$

$$p_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) \propto \frac{\text{proj}\Big\{ q_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) \Big\}}{p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt})}, \tag{127}$$

with

$$\begin{aligned} q_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) &\propto \int\int \delta(\mathbf{z}_{l,kt} - \mathbf{g}_{l,k} x_{kt}) \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) \\ &\quad \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{z}_{l,kt}\, d\mathbf{g}_{l,k} \\ &\overset{(a)}{=} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \int p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k} x_{kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{g}_{l,k} \\ &\overset{(b)}{=} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \int \mathcal{CN}\Big(\mathbf{g}_{l,k} x_{kt} | \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\Big) \\ &\quad \cdot \mathcal{CN}\Big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big)\, d\mathbf{g}_{l,k} \\ &\overset{(c)}{=} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \int |x_{kt}|^{-2N} \cdot \mathcal{CN}\big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\text{tmp}}, \mathbf{C}_{\text{tmp}}\big) \\ &\quad \cdot \mathcal{CN}\Big(\mathbf{0} | \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} x_{kt}^{-1} - \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x_{kt}|^{-2} + \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big)\, d\mathbf{g}_{l,k} \\ &= p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot |x_{kt}|^{-2N} \\ &\quad \cdot \mathcal{CN}\Big(\mathbf{0} | \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} x_{kt}^{-1} - \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x_{kt}|^{-2} + \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big) \\ &= p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}), \end{aligned} \tag{128}$$

where (a) is obtained by the sifting property of the Dirac delta function [51], (b) is obtained by using (123), and (c) is obtained by utilizing the Gaussian scaling lemma on $\mathcal{CN}\Big(\mathbf{g}_{l,k}x_{kt}|\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\Big)$ and then the Gaussian product lemma with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} x_{kt}^{-1}$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$, $\mathbf{C}_1 = \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}|x_{kt}|^{-2}$, and $\mathbf{C}_2 = \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$ which yields $\mathbf{C}_{\text{tmp}}$ and $\boldsymbol{\mu}_{\text{tmp}}$ according to (101) and (102), respectively. The final result in (178) is obtained by applying the Gaussian scaling lemma once again with

$$\theta(x_{kt}) = \mathcal{CN}\Big(\mathbf{0}|\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} - \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} x_{kt}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} + \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}|x_{kt}|^2\Big). \tag{129}$$

Evaluating (178) at $x_{kt} \in \mathcal{X}$ for $t > T_p$ yields a categorical distribution for $q_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt})$. Hence, the projection operation in (127) is superfluous since it projects $q_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt})$ onto a categorical distribution, and the denominator of (127) cancels with the first term in (178). Thus, the final message update rule is given by the distribution

$$p_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) \propto \theta(x_{kt}). \tag{130}$$

Message Update for $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$

$$p_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \propto \frac{\text{proj}\Big\{q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})\Big\}}{p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt})}, \tag{131}$$

with

$$\begin{aligned}
&q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \\
&\quad\propto \sum_{x_{kt}} \int \delta(\mathbf{z}_{l,kt} - \mathbf{g}_{l,k}x_{kt}) \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{g}_{l,k} \\
&\quad\overset{(a)}{=} \sum_{x_{kt}} |x_{kt}|^{-2N} \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\left(\frac{\mathbf{z}_{l,kt}}{x_{kt}}\right) \\
&\quad= \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \mathcal{CN}\Big(\mathbf{z}_{l,kt}|\check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}), \check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt})\Big) \cdot \theta(x_{kt}),
\end{aligned} \tag{132}$$

where (a) is obtained by applying the scaling and sifting property of the Dirac delta function [51] and using (123). The final result in (189) is obtained by the Gaussian scaling and product lemma with $\theta(x_{kt})$ given in (129) and

$$\check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}) = \left(\mathbf{C}^{-1}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} + \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}|x_{kt}|^{-2}\right)^{-1}, \tag{133}$$

$$\check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}) = \check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt})\left(\mathbf{C}^{-1}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} + \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\frac{x_{kt}}{|x_{kt}|^2}\right). \tag{134}$$

The normalization constant for $q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ is given by

$$\begin{aligned}
\tilde{Z}_{\Psi_{z_{l,kt}}} &= \int \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \mathcal{CN}\Big(\mathbf{z}_{l,kt}|\check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}), \check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt})\Big) \cdot \theta(x_{kt})\, d\mathbf{z}_{l,kt} \\
&= \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}).
\end{aligned} \tag{135}$$

According to (189), $q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ is a Gaussian mixture with mean vector and covariance matrix

$$\hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \frac{1}{\tilde{Z}_{\Psi_{z_{l,kt}}}} \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}) \cdot \check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}), \tag{136}$$

$$\begin{aligned}\hat{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \frac{1}{\tilde{Z}_{\Psi_{z_{l,kt}}}} \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}) \cdot \Big(&\check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}) \\ &+ \check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}) \check{\boldsymbol{\mu}}^H_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x_{kt}) \Big) - \hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} \hat{\boldsymbol{\mu}}^H_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}.\end{aligned} \tag{137}$$

Note that in the pilot phase for $t \leq T_p$, the transmitted symbol x^p_{kt} is already known. Hence, $q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ reduces to a Gaussian distribution for $t \leq T_p$ with mean vector $\hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x^p_{kt})$ and covariance matrix $\hat{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(x^p_{kt})$. The final message update can be computed according to (131) using moment matching to find the projected distribution and the Gaussian quotient lemma,

$$p_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) = \mathcal{CN}\Big(\mathbf{z}_{l,kt} | \boldsymbol{\mu}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}, \mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}\Big), \tag{138}$$

with

$$\mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \Big(\hat{\mathbf{C}}^{-1}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} - \mathbf{C}^{-1}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\Big)^{-1}, \tag{139}$$

$$\boldsymbol{\mu}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} = \mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} \Big(\hat{\mathbf{C}}^{-1}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} \hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}} - \mathbf{C}^{-1}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\Big). \tag{140}$$

Message Update $m_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}$

$$p_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \propto \frac{\text{proj}\Big\{q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k})\Big\}}{p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})}, \tag{141}$$

with

$$\begin{aligned}
&q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \\
&\quad\propto \sum_{x_{kt}} \int \delta(\mathbf{z}_{l,kt} - \mathbf{g}_{l,k} x_{kt}) \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{z}_{l,kt} \\
&\quad\overset{(a)}{=} \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{g}_{l,k} x_{kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k}) \\
&\quad= \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \mathcal{CN}\Big(\mathbf{g}_{l,k} | \check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}), \check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt})\Big) \cdot \theta(x_{kt}),
\end{aligned} \tag{142}$$

where (a) is obtained by the sifting property of the Dirac delta function and using (123). The final result in (142) is obtained by the Gaussian scaling and product lemma with θ given in (129) and

$$\check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}) = \Big(\mathbf{C}^{-1}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x_{kt}|^2 + \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big)^{-1}, \tag{143}$$

$$\check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}) = \check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}) \left(\mathbf{C}^{-1}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} \frac{|x_{kt}|^2}{x_{kt}} + \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \right). \tag{144}$$

According to (142), $q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k})$ is a Gaussian mixture with mean vector and covariance matrix

$$\hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \frac{1}{\tilde{Z}_{\Psi_{z_{l,kt}}}} \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}) \cdot \check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}), \tag{145}$$

$$\begin{aligned}\hat{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \frac{1}{\tilde{Z}_{\Psi_{z_{l,kt}}}} \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}) \cdot \Big(&\check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}) \\ &+ \check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}) \check{\boldsymbol{\mu}}^H_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x_{kt}) \Big) - \hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} \hat{\boldsymbol{\mu}}^H_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}.\end{aligned} \tag{146}$$

Note that in the pilot phase for $t \leq T_p$, the transmitted symbol x^p_{kt} is already known. Hence, $q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k})$ reduces to a Gaussian distribution for $t \leq T_p$ with mean vector $\hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \check{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x^p_{kt})$ and covariance matrix $\hat{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \check{\mathbf{C}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(x^p_{kt})$. The final message update can be computed according to (141) using moment matching and the Gaussian quotient lemma,

$$p_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) = \mathcal{CN}\Big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}, \mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}\Big), \tag{147}$$

with

$$\mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \Big(\hat{\mathbf{C}}^{-1}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} - \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big)^{-1}, \tag{148}$$

$$\boldsymbol{\mu}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} = \mathbf{C}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} \Big(\hat{\mathbf{C}}^{-1}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} \hat{\boldsymbol{\mu}}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} - \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big). \tag{149}$$

Message Updates for $\Psi_{g_{l,k}}$

Incoming messages to factor node $\Psi_{g_{l,k}}$

$$p_{u_k;\Psi_{g_{l,k}}}(u_k) \propto p_{\Psi_{u_k};u_k}(u_k) \cdot \prod_{l' \neq l} p_{\Psi_{g_{l',k}};u_k}(u_k), \tag{150}$$

$$p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}) = p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}), \tag{151}$$

$$\begin{aligned}p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) &\propto \prod_{t=1}^{T} p_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \\ &\propto \mathcal{CN}\Big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}\Big),\end{aligned} \tag{152}$$

with

$$\mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} = \left(\sum_{t=1}^{T} \mathbf{C}^{-1}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}\right)^{-1}, \tag{153}$$

$$\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} = \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} \left(\sum_{t=1}^{T} \mathbf{C}^{-1}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}} \boldsymbol{\mu}_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}\right), \tag{154}$$

which is obtained by applying the Gaussian product lemma multiple times.

Outgoing messages from factor node $\Psi_{g_{l,k}}$

Message Update $m_{\Psi_{g_{l,k}};u_k}$

$$p_{\Psi_{g_{l,k}};u_k}(u_k) \propto \frac{\mathrm{proj}\Big\{q_{\Psi_{g_{l,k}};u_k}(u_k)\Big\}}{p_{u_k;\Psi_{g_{l,k}}}(u_k)}, \tag{155}$$

with

$$\begin{aligned} q_{\Psi_{g_{l,k};u_k}}(u_k) &\propto \int\int \delta(\mathbf{g}_{l,k} - \mathbf{h}_{l,k}u_k) \cdot p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) \cdot p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k})\, d\mathbf{g}_{l,k}\, d\mathbf{h}_{l,k} \\ &\overset{(a)}{=} p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot \int p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}u_k) \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})\, d\mathbf{h}_{l,k} \\ &= p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot \vartheta(u_k), \end{aligned} \tag{156}$$

where (a) is obtained by the sifting property of the Dirac delta function and using (151). The final result in (187) is obtained by considering the fact that u_k is a binary random variable with $u_k \in \{0, 1\}$, utilizing (114) and the Gaussian product rule, and, then, integrating over $\mathbf{h}_{l,k}$ with

$$\vartheta(u_k) = \mathcal{CN}\Big(\mathbf{0}|\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} - \tilde{\boldsymbol{\mu}}_{h_{l,k}}u_k, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \tilde{\mathbf{C}}_{h_{l,k}}u_k\Big). \tag{157}$$

The projection operation in (155) is superfluous since $q_{\Psi_{g_{l,k};u_k}}(u_k)$ (187) is already categorically distributed. Hence, the denominator of (155) cancels with the first term in (187). The final message update rule is given by the distribution

$$p_{\Psi_{g_{l,k};u_k}}(u_k) \propto \vartheta(u_k). \tag{158}$$

Message Update $m_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}$

$$p_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) \propto \frac{\text{proj}\Big\{q_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})\Big\}}{p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,kt})}, \tag{159}$$

with

$$\begin{aligned} q_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) &\propto \sum_{u_k} \int \delta(\mathbf{g}_{l,k} - \mathbf{h}_{l,k}u_k) \cdot p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) \cdot p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k})\, d\mathbf{g}_{l,k} \\ &\overset{(a)}{=} p_{u_k;\Psi_{g_{l,k}}}(0) \cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{0}) \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) + p_{u_k;\Psi_{g_{l,k}}}(1) \cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}) \\ &\quad \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) \\ &= p_{u_k;\Psi_{g_{l,k}}}(0) \cdot \vartheta(0) \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) + p_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \\ &\quad \cdot \mathcal{CN}\Big(\mathbf{h}_{l,k}|\check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}, \check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\Big), \end{aligned} \tag{160}$$

where (a) is obtained by making the sum explicit for u_k and utilizing the sifting property of the Dirac delta function and (151). The final result in (160) is obtained by using (114), (157), and the Gaussian multiplication lemma with

$$\check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} = \Big(\mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \tilde{\mathbf{C}}^{-1}_{h_{l,k}}\Big)^{-1}, \tag{161}$$

$$\check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} = \check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\left(\mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} + \tilde{\mathbf{C}}^{-1}_{h_{l,k}}\tilde{\boldsymbol{\mu}}_{h_{l,k}}\right). \tag{162}$$

The normalization constant for $q_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})$ is given by

$$\begin{aligned} \tilde{Z}_{\Psi_{g_{l,k}}} &= \int p_{u_k;\Psi_{g_{l,k}}}(0) \cdot \vartheta(0) \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) \\ &\qquad + p_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \cdot \mathcal{CN}\Big(\mathbf{h}_{l,k}|\check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}, \check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\Big)\, d\mathbf{h}_{l,k} \\ &= p_{u_k;\Psi_{g_{l,k}}}(0) \cdot \vartheta(0) + p_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1). \end{aligned} \tag{163}$$

According to (160), $q_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})$ is a Gaussian mixture in two components with mean vector and covariance matrix

$$\hat{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} = \frac{1}{\tilde{Z}_{\Psi_{g_{l,k}}}}\Big(p_{u_k;\Psi_{g_{l,k}}}(0)\cdot\vartheta(0)\cdot\tilde{\boldsymbol{\mu}}_{h_{l,k}} + p_{u_k;\Psi_{g_{l,k}}}(1)\cdot\vartheta(1)\cdot\check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\Big), \tag{164}$$

$$\begin{aligned}\hat{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} = \frac{1}{\tilde{Z}_{\Psi_{z_{l,kt}}}}\Big(&p_{u_k;\Psi_{g_{l,k}}}(0)\cdot\vartheta(0)\cdot\big(\tilde{\mathbf{C}}_{h_{l,k}} + \tilde{\boldsymbol{\mu}}_{h_{l,k}}\tilde{\boldsymbol{\mu}}^H_{h_{l,k}}\big) + p_{u_k;\Psi_{g_{l,k}}}(1)\cdot\vartheta(1)\\ &\cdot\big(\check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} + \check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\check{\boldsymbol{\mu}}^H_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\big)\Big) - \hat{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}\hat{\boldsymbol{\mu}}^H_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}.\end{aligned} \tag{165}$$

Note that the estimated posterior distribution (112) of $\mathbf{h}_{l,k}$ is proportional to the product of $p_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})$ and $p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,kt})$ which is proportional to $\mathrm{proj}\Big\{q_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})\Big\}$ according to (151) and (159). Hence, the final estimate of $\mathbf{h}_{l,k}$ can be computed via (164) after the last iteration of the EP algorithm.

By applying moment matching and the Gaussian quotient lemma to (159), we obtain for the final message update the following distribution,

$$p_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) = \mathcal{CN}\Big(\mathbf{h}_{l,k}|\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}, \mathbf{C}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\Big), \tag{166}$$

with

$$\mathbf{C}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} = \Big(\hat{\mathbf{C}}^{-1}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} - \tilde{\mathbf{C}}^{-1}_{h_{l,k}}\Big)^{-1}, \tag{167}$$

$$\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} = \mathbf{C}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\Big(\hat{\mathbf{C}}^{-1}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}\hat{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}} - \tilde{\mathbf{C}}^{-1}_{h_{l,k}}\tilde{\boldsymbol{\mu}}_{h_{l,k}}\Big). \tag{168}$$

Message Update $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$

$$p_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \propto \frac{\mathrm{proj}\Big\{q_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k})\Big\}}{p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,kt})}, \tag{169}$$

with

$$\begin{aligned}q_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) &\propto \sum_{u_k}\int \delta(\mathbf{g}_{l,k} - \mathbf{h}_{l,k}u_k)\cdot p_{u_k;\Psi_{g_{l,k}}}(u_k)\cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k})\cdot p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k})\,d\mathbf{h}_{l,k}\\ &\overset{(a)}{=} p_{u_k;\Psi_{g_{l,k}}}(0)\cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k})\cdot\delta(\mathbf{g}_{l,k}) + p_{u_k;\Psi_{g_{l,k}}}(1)\cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k})\\ &\quad\cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{g}_{l,k})\\ &= p_{u_k;\Psi_{g_{l,k}}}(0)\cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k})\cdot\delta(\mathbf{g}_{l,k}) + p_{u_k;\Psi_{g_{l,k}}}(1)\cdot\vartheta(1)\\ &\quad\cdot\mathcal{CN}\Big(\mathbf{g}_{l,k}|\check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}, \check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}\Big),\end{aligned} \tag{170}$$

where (a) is obtained by making the sum explicit for u_k and utilizing the sifting property of the Dirac delta function and (151). The final result in (188) is obtained by using (114), (157), and the Gaussian multiplication lemma with

$$\check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}, \tag{171}$$

$$\check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{h}_{l,k}}. \tag{172}$$

The mean vector and covariance matrix of the Bernoulli-Gaussian distribution $q_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k})$

in (188) is given by

$$\hat{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \frac{1}{\tilde{Z}_{\Psi_{g_{l,k}}}} \cdot p_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \cdot \check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}, \tag{173}$$

$$\begin{aligned}\hat{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \frac{1}{\tilde{Z}_{\Psi_{g_{l,k}}}} \cdot p_{u_k;\Psi_{g_{l,k}}}(1) \cdot \vartheta(1) \cdot \Big(\check{\mathbf{C}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} + \check{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} \check{\boldsymbol{\mu}}^H_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}\Big) \\ - \hat{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} \hat{\boldsymbol{\mu}}^H_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}.\end{aligned} \tag{174}$$

By applying moment matching and the Gaussian quotient lemma to (169), we obtain for the final message update the following distribution,

$$p_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) = \mathcal{CN}\Big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}, \mathbf{C}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}\Big), \tag{175}$$

with

$$\mathbf{C}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \Big(\hat{\mathbf{C}}^{-1}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} - \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}\Big)^{-1}, \tag{176}$$

$$\boldsymbol{\mu}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} = \mathbf{C}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} \Big(\hat{\mathbf{C}}^{-1}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} \hat{\boldsymbol{\mu}}_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}} - \mathbf{C}^{-1}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}\Big). \tag{177}$$

D.3 JACD-EP-BG Algorithm

Message Update for $m_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}$

This message update is identical for the JACD-EP-BG and the JACD-EP algorithm, and, hence, its derivation can be found in Appendix D.2 or the extended version of [2].

Message Update for $m_{\Psi_{z_{l,kt}};x_{kt}}$

The local belief of x_{kt} at the factor node $\Psi_{z_{l,kt}}$ is given by

$$\begin{aligned}
&q_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt}) \\
&\quad\propto \iint \delta(\mathbf{z}_{l,kt} - \mathbf{g}_{l,k} x_{kt}) \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \\
&\qquad \cdot p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{z}_{l,kt}\, d\mathbf{g}_{l,k} \\
&\quad\overset{(a)}{=} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \int p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k} x_{kt}) \\
&\qquad \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{g}_{l,k} \\
&\quad\overset{(b)}{\approx} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \int \mathcal{CN}\big(\mathbf{g}_{l,k} x_{kt} | \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\big) \\
&\qquad \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \mathcal{CN}\big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\big)\, d\mathbf{g}_{l,k} \\
&\quad\overset{(c)}{=} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \int |x_{kt}|^{-2N} \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \\
&\qquad \cdot \mathcal{CN}\big(\mathbf{g}_{l,k} | \boldsymbol{\mu}_{\mathrm{tmp}}, \mathbf{C}_{\mathrm{tmp}}\big) \cdot \mathcal{CN}\big(\mathbf{0} | \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} x_{kt}^{-1} \\
&\qquad - \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x_{kt}|^{-2} + \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\big)\, d\mathbf{g}_{l,k} \\
&\quad= p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot |x_{kt}|^{-2N} \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \\
&\qquad \cdot \mathcal{CN}\big(\mathbf{0} | \boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} x_{kt}^{-1} - \boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}} |x_{kt}|^{-2} + \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\big) \\
&\quad= p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \theta(x_{kt}),
\end{aligned} \tag{178}$$

where (a) is obtained by the sifting property of the Dirac delta function [51], (b) is obtained by using $p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) = p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ and considering only the Gaussian part of the BG distribution $p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$ which models the active component that is necessary to detect the transmitted symbol x_{kt}, and (c) is obtained by utilizing the Gaussian scaling and product lemma which yields $\boldsymbol{\mu}_{\text{tmp}}$ and $\mathbf{C}_{\text{tmp}}$ accordingly. The final result in (178) is obtained by applying the Gaussian scaling lemma once again with $\theta(x_{kt})$ given in (30). For $t > T_p$ and $x_{kt} \in \mathcal{X}$, (178) is a categorical distribution. Hence, the projection operation in (108) is superfluous since it projects $q_{\Psi_{z_{l,kt}};x_{kt}}(x_{kt})$ onto a categorical distribution, and the denominator of (108) cancels with the first term in (178). Thus, the final message update rule reduces to (29).

Message Update for $m_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}$

The local belief of $\mathbf{g}_{l,k}$ at the factor node $\Psi_{z_{l,kt}}$ is given by

$$\begin{aligned}
&q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \\
&\quad\propto \sum_{x_{kt}} \int \delta(\mathbf{z}_{l,kt} - \mathbf{g}_{l,k}x_{kt}) \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \\
&\qquad \cdot p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{z}_{l,kt} \\
&\quad= \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{g}_{l,k}x_{kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k}),
\end{aligned} \tag{179}$$

where (179) is obtained by the sifting property of the Dirac delta function and using $p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) = p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$. The local BG belief is further computed by considering the two parts modeling activity and inactivity of the BG distribution separately. For the BG inactive component, only the Dirac at $\mathbf{g}_{l,k} = \mathbf{0}$ contributes, i.e., $p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k} = \mathbf{0}) = 1 - \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$. Thus,

$$\begin{aligned}
&q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k} = \mathbf{0}) \\
&\quad\propto \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(0) \cdot (1 - \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}) \\
&\quad= \theta(0) \cdot (1 - \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}),
\end{aligned} \tag{180}$$

with $p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{0}) = \theta(0)$ and $\theta(x)$ defined in (30). For the BG active component, only the Gaussian part is considered, i.e., $p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k}) = \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \mathcal{CN}\big(\mathbf{g}_{l,k}|\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\big)$,

$$\begin{aligned}
&q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k} \neq \mathbf{0}) \\
&\quad\propto \sum_{x_{kt}} p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \mathcal{CN}\big(\mathbf{g}_{l,k}x_{kt}|\boldsymbol{\mu}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}, \mathbf{C}_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}\big) \\
&\qquad \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \mathcal{CN}\big(\mathbf{g}_{l,k}|\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\big) \\
&\quad= \sum_{x_{kt}} \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \phi(x_{kt}) \\
&\qquad \cdot \mathcal{CN}\big(\mathbf{g}_{l,k}|\check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt})x_{kt}^{-1}, \check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x_{kt})|x_{kt}|^{-2}\big),
\end{aligned} \tag{181}$$

which is obtained by applying the Gaussian scaling and product lemma with $\phi(x_{kt})$, $\check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt})$, and $\check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x_{kt})$ defined in (37), (38), and (39), respectively. Thus, the active component of $q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k})$ in (181) is a Gaussian mixture whose components correspond to the transmit symbols x_{kt}. The computation of the expectation of the sufficient statistics of this mixture distribution for moment matching may yield a mean close to zero, which would indicate activity with a weak channel. To prevent this issue which determines a high number of false alarms, we select the mixture component with the largest weight, denoted by x^* in (44). Note that in the pilot phase

for $t < T_p$, the transmitted symbol $x_{kt} = x_{kt}^p$ is known a priori and, hence, $x^* = x_{kt}^p$. With this pragmatic approximation, we obtain

$$q_{\Psi_{z_{l,kt}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \approx \mathcal{BG}\big(\mathbf{g}_{l,k}|\hat{\lambda}_{\mathbf{g}_{l,kt}}, \hat{\boldsymbol{\mu}}_{\mathbf{g}_{l,kt}}, \hat{\mathbf{C}}_{\mathbf{g}_{l,kt}}\big) \tag{182}$$

with parameters $\hat{\lambda}_{\mathbf{g}_{l,kt}}$, $\hat{\boldsymbol{\mu}}_{\mathbf{g}_{l,kt}}$, and $\hat{\mathbf{C}}_{\mathbf{g}_{l,kt}}$ given by (41), (42), and (43), respectively. By applying the EP update rule (108), the final message is defined by the parameters in (32), (33), and (34).

Message Update for $m_{\Psi_{g_{l,k}};u_k}$

The local belief of u_k at the factor node $\Psi_{g_{l,k}}$ is given by

$$\begin{aligned}
&q_{\Psi_{g_{l,k}};u_k}(u_k) \\
&\quad\propto \iint \delta(\mathbf{g}_{l,k} - \mathbf{h}_{l,k}u_k) \cdot p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) \\
&\qquad\cdot p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k})\, d\mathbf{g}_{l,k}\, d\mathbf{h}_{l,k} \\
&\quad= \int p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}u_k) \\
&\qquad\cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})\, d\mathbf{h}_{l,k}
\end{aligned} \tag{183}$$

which is obtained by the sifting property of the Dirac delta function and using $p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}) = p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})$. The local categorical belief is further computed by treating separately the two components modeling the Bernoulli distribution. For the inactive case, i.e., $u_k = 0$, only the Dirac component of the BG distribution $p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k})$ contributes, i.e., $p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}u_k = \mathbf{0}) = 1 - \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}$,

$$\begin{aligned}
&q_{\Psi_{g_{l,k}};u_k}(u_k = 0) \\
&\quad\propto \int p_{u_k;\Psi_{g_{l,k}}}(0) \cdot (1 - \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}) \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})\, d\mathbf{h}_{l,k} \\
&\quad= p_{u_k;\Psi_{g_{l,k}}}(0) \cdot (1 - \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}).
\end{aligned} \tag{184}$$

For the active case, i.e., $u_k = 1$, only the Gaussian component of the BG distribution $p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k})$ is relevant, i.e., $p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) = p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}) = \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} \cdot \mathcal{CN}\big(\mathbf{h}_{l,k}|\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}\big)$,

$$\begin{aligned}
&q_{\Psi_{g_{l,k}};u_k}(u_k = 1) \qquad (185)\\
&\quad\propto \int p_{u_k;\Psi_{g_{l,k}}}(1) \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} \\
&\qquad\cdot \mathcal{CN}\big(\mathbf{h}_{l,k}|\boldsymbol{\mu}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}\big) \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})\, d\mathbf{h}_{l,k} \\
&\quad= p_{u_k;\Psi_{g_{l,k}}}(1) \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} \cdot \vartheta(1), \qquad (186)
\end{aligned}$$

which is obtained by applying the Gaussian product lemma with $\vartheta(u_k)$ defined in (49). Then, combining the active and inactive components, we obtain

$$q_{\Psi_{g_{l,k}};u_k}(u_k) \propto p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot \begin{cases} 1 - \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} & \text{for } u_k = 0 \\ \lambda_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}} \cdot \vartheta(1) & \text{for } u_k = 1 \end{cases}. \tag{187}$$

The projection in (108) is superfluous since $q_{\Psi_{g_{l,k}};u_k}(u_k)$ is already a categorical distribution. Hence, the denominator of (108) cancels with the first term in (187). Thus, the final message update rule is given by (50).

Message Update for $m_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}$

The local belief of $\mathbf{g}_{l,k}$ at the factor node $\Psi_{g_{l,k}}$ is given by

$$\begin{aligned}
&q_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k}) \\
&\quad\propto \sum_{u_k} \int \delta(\mathbf{g}_{l,k} - \mathbf{h}_{l,k}u_k) \cdot p_{u_k;\Psi_{g_{l,k}}}(u_k) \cdot p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) \\
&\qquad \cdot p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k})\, d\mathbf{h}_{l,k} \\
&\quad\overset{(a)}{=} p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) \cdot \big(p_{u_k;\Psi_{g_{l,k}}}(0) \cdot \delta(\mathbf{g}_{l,k}) \\
&\qquad + p_{u_k;\Psi_{g_{l,k}}}(1) \cdot p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{g}_{l,k})\big) \\
&\quad= p_{\mathbf{g}_{l,k};\Psi_{g_{l,k}}}(\mathbf{g}_{l,k}) \cdot \mathcal{BG}\big(\mathbf{g}_{l,k}|p_{u_k;\Psi_{g_{l,k}}}(1), \tilde{\boldsymbol{\mu}}_{h_{l,k}}, \tilde{\mathbf{C}}_{h_{l,k}}\big)
\end{aligned} \tag{188}$$

where (a) is obtained by explicitly evaluating the sum over u_k and utilizing the sifting property of the Dirac delta function along with the identity $p_{\mathbf{h}_{l,k};\Psi_{g_{l,k}}}(\mathbf{h}_{l,k}) = p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k})$. The final equation (188) is obtained by using $p_{\Psi_{h_{l,k}};\mathbf{h}_{l,k}}(\mathbf{h}_{l,k}) = \mathcal{CN}\big(\mathbf{h}_{l,k}|\tilde{\boldsymbol{\mu}}_{h_{l,k}}, \tilde{\mathbf{C}}_{h_{l,k}}\big)$ and noting that the second factor in (a) written in brackets is a BG distribution. The projection in (108) is superfluous since $q_{\Psi_{g_{l,k}};\mathbf{g}_{l,k}}(\mathbf{g}_{l,k})$ is the product of two BG distributions which is again a BG distribution. Hence, the denominator of (108) cancels with the first term in (188). Thus, the final message parameters are given by (59), (60), and (61).

Message Update for $m_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}$

The local belief of $\mathbf{z}_{l,kt}$ at the factor node $\Psi_{z_{l,kt}}$ is given by

$$\begin{aligned}
&q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \\
&\quad\propto \sum_{x_{kt}} \int \delta(\mathbf{z}_{l,kt} - \mathbf{g}_{l,k}x_{kt}) \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \\
&\qquad \cdot p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}(\mathbf{g}_{l,k})\, d\mathbf{g}_{l,k} \\
&\quad\overset{(a)}{=} \sum_{x_{kt}} |x_{kt}|^{-2N} \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \\
&\qquad \cdot p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big(\frac{\mathbf{z}_{l,kt}}{x_{kt}}\Big) \\
&\quad\overset{(b)}{\approx} \sum_{x_{kt}} |x_{kt}|^{-2N} \cdot p_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt}) \\
&\qquad \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot \mathcal{CN}\Big(\frac{\mathbf{z}_{l,kt}}{x_{kt}}\Big|\mu_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}, \mathbf{C}_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}\Big) \\
&\quad= \sum_{x_{kt}} \pi_{x_{kt};\Psi_{z_{l,kt}}}(x_{kt}) \cdot \theta(x_{kt}) \cdot \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \\
&\qquad \cdot \mathcal{CN}\big(\mathbf{z}_{l,kt}|\check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt}), \check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x_{kt})\big),
\end{aligned} \tag{189}$$

where (a) is obtained by applying the scaling and sifting property of the Dirac delta function [51] and using $p_{\mathbf{z}_{l,kt};\Psi_{z_{l,kt}}}(\mathbf{z}_{l,kt}) = p_{\Psi_{y_{l,t}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$, and (b) is obtained by approximating the BG distribution $p_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}}$ by its Gaussian part. The final result in (189) is obtained by applying the Gaussian scaling and product lemma with $\theta(x_{kt})$, $\check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt})$, and $\check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x_{kt})$ given by (30), (38), and (39), respectively. The normalization constant for $q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ is given by $\tilde{Z}_{\Psi_{z_{l,kt}}} = \lambda_{\mathbf{g}_{l,k};\Psi_{z_{l,kt}}} \cdot Z_{\Psi_{z_{l,kt}}}$ where $Z_{\Psi_{z_{l,kt}}}$ is defined in (40). According to (189), $q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ is a Gaussian mixture with mean vector and covariance matrix given by (67) and (68), respectively.

Note that in the pilot phase for $t < T_p$, the transmitted symbol $x_{kt} = x_{kt}^p$ is known a priori. Hence, $q_{\Psi_{z_{l,kt}};\mathbf{z}_{l,kt}}(\mathbf{z}_{l,kt})$ reduces to a Gaussian distribution for $t < T_p$ with mean vector $\hat{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}} = \check{\boldsymbol{\mu}}_{\mathbf{z}_{l,kt}}(x_{kt}^p)$ and covariance matrix $\hat{\mathbf{C}}_{\mathbf{z}_{l,kt}} = \check{\mathbf{C}}_{\mathbf{z}_{l,kt}}(x_{kt}^p)$. The final message parameters are given by (65) and (66), which are computed according to (108).

E References

[1] C. Forsch and L. Cottatellucci, "Expectation propagation for distributed inference in grant-free cell-free massive MIMO," *submitted to IEEE Trans. Signal Process.*, 2025, preprint available at: https://arxiv.org/abs/2601.04166.

[2] C. Forsch, A. Karataev, and L. Cottatellucci, "Distributed joint user activity detection, channel estimation, and data detection via expectation propagation in cell-free massive MIMO," in *Proc. IEEE 25th Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, 2024, pp. 531–535, extended version available at: https://arxiv.org/abs/2405.09914.

[3] N. H. Mahmood *et al.*, "Machine type communications: key drivers and enablers towards the 6G era," *EURASIP J. Wireless Commun. Netw.*, vol. 2021, no. 1, p. 134, 2021.

[4] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, 2018.

[5] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 2020.

[6] Z. Gao *et al.*, "Compressive-sensing-based grant-free massive access for 6G massive communication," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 7411–7435, 2024.

[7] 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.

[8] 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.

[9] M. Mohammadi, Z. Mobini, H. Quoc Ngo, and M. Matthaiou, "Next-generation multiple access with cell-free massive MIMO," *Proc. IEEE*, vol. 112, no. 9, pp. 1372–1420, 2024.

[10] H. Wang, J. Wang, and J. Fang, "Grant-free massive connectivity in massive MIMO systems: Collocated versus cell-free," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 634–638, 2021.

[11] U. K. Ganesan, E. Björnson, and E. G. Larsson, "Clustering-based activity detection algorithms for grant-free random access in cell-free massive MIMO," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7520–7530, 2021.

[12] Y. Xu, E. G. Larsson, E. A. Jorswieck, X. Li, S. Jin, and T.-H. Chang, "Distributed signal processing for extremely large-scale antenna array systems: State-of-the-art and future directions," *IEEE J. Sel. Topics Signal Process.*, vol. 19, no. 2, pp. 304–330, 2025.

[13] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, 2018.

[14] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 764–779, 2020.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] ——, "Channel models, favorable propagation and MultiStage linear detection in cell-free massive MIMO," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2020.

[24] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, 2018.

[25] 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.

[26] 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.

[27] 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.

[28] H. Song, T. Goldstein, X. You, C. Zhang, O. Tirkkonen, and C. Studer, "Joint channel estimation and data detection in cell-free massive MU-MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4068–4084, 2022.

[29] 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.

[30] 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)*, 2025, pp. 19–25.

[31] 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.*, 2024, pp. 1443–1447.

[32] S. Jiang, X. Yuan, X. Wang, C. Xu, and W. Yu, "Joint user identification, channel estimation, and signal detection for grant-free NOMA," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6960–6976, 2020.

[33] 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.

[34] C. Zhang, Y. Liu, J. Hu, and K. Yang, "Joint user identification, channel estimation, and data detection for grant-free NOMA in LEO satellite communications," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 1, pp. 107–121, 2025.

[35] Q. Zou, H. Zhang, D. Cai, and H. Yang, "A low-complexity joint user activity, channel and data estimation for grant-free massive MIMO systems," *IEEE Signal Process. Lett.*, vol. 27, pp. 1290–1294, 2020.

[36] S. Zhang, Y. Cui, and W. Chen, "Joint device activity detection, channel estimation and signal detection for massive grant-free access via BiGAMP," *IEEE Trans. Signal Process.*, vol. 71, pp. 1200–1215, 2023.

[37] X. Bian, Y. Mao, and J. Zhang, "Joint activity detection, channel estimation, and data decoding for grant-free massive random access," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14 042–14 057, 2023.

[38] 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.

[39] G. Sun, M. Cao, W. Wang, W. Xu, and C. Studer, "Deep-unfolded massive grant-free transmission in cell-free wireless communication systems," *IEEE Trans. Signal Process.*, vol. 73, pp. 1094–1109, 2025.

[40] T. P. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, 2001.

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

[42] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends® Mach. Learn.*, vol. 1, no. 1–2, pp. 1–305, 2007.

[43] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proc. 45th Asilomar Conf. Signals, Syst., Comput.*, 2011, pp. 799–803.

[44] D. Hernández-Lobato, J. M. Hernández-Lobato, and P. Dupont, "Generalized spike-and-slab priors for Bayesian group feature selection using expectation propagation," *J. Mach. Learn. Res.*, vol. 14, no. 1, p. 1891–1945, 2013.

[45] 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.

[46] C. Rusu, N. González-Prelcic, and R. W. Heath, "Algorithms for the construction of incoherent frames under various design constraints," *Signal Process.*, vol. 152, pp. 363–372, 2018.

[47] 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.

[48] M. Chen, Y. Miao, Y. Hao, and K. Hwang, "Narrow band internet of things," *IEEE Access*, vol. 5, pp. 20 557–20 577, 2017.

[49] P. A. Bromiley, "Products and convolutions of Gaussian distributions," Tech. Rep. Tina Memo No. 2003-003, 2003.

[50] A. Papoulis and S. U. Pillai, *Probability, random variables and stochastic processes*, 4th ed. McGraw-Hill, 2002.

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