HKUST SPD - INSTITUTIONAL REPOSITORY

Title	Distributed Expectation Propagation Detection for Cell-Free Massive MIMO
Authors	He, Hengtao; Wang, Hanqing; Yu, Xianghao; Zhang, Jun; Song, Shenghui; Ben Letaief, Khaled
Source	2021 IEEE Global Communications Conference, GLOBECOM 2021 - Proceedings, / IEEE. New York, NY, USA : IEEE, 2021
Version	Accepted Version
DOI	10.1109/GLOBECOM46510.2021.9685334
Publisher	IEEE
Copyright	© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Distributed Expectation Propagation Detection for Cell-Free Massive MIMO

Hengtao He*, Hanqing Wang†, Xianghao Yu*, Jun Zhang‡, S.H. Song*, and Khaled B. Letaief*

*Department of ECE, The Hong Kong University of Science and Technology,
Kowloon, Hong Kong, E-mail: {eehthe, eeyu, eeshsong, eekhaled}@ust.hk
†Huawei Technologies Company Ltd., Shanghai 201206, China, E-mail: wanghanqing3@Huawei.com
‡Department of EIE, The Hong Kong Polytechnic University, Hong Kong, E-mail: jun-eie.zhang@polyu.edu.hk

Abstract—In cell-free massive MIMO networks, an efficient distributed detection algorithm is of significant importance. In this paper, we propose a distributed expectation propagation (EP) detector for cell-free massive MIMO. The detector is composed of two modules, a nonlinear module at the central processing unit (CPU) and a linear module at the access point (AP). The turbo principle in iterative decoding is utilized to compute and pass the extrinsic information between modules. An analytical framework is then provided to characterize the asymptotic performance of the proposed EP detector with a large number of antennas. Simulation results will show that the proposed method outperforms the distributed detectors in terms of bit-error rate.

I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems have been regarded as a key enabling technology for 5G because of its high spectral efficiency (SE), energy efficiency and link reliability [1]. However, the SE gain becomes marginal for the cell-edge user equipments (UEs). To address this problem, a novel network architecture, nameely cell-free massive MIMO, was proposed [2], [3]. This is a disruptive emerging technology which has been recognized as a crucial and core technology for the upcoming beyond 5G and 6G networks [4]. Such technology is expected to bring important benefits, including huge data throughput, ultra-low latency, ultra-high reliability, a huge increase in mobile energy efficiency, and ubiquitous uniform coverage. The fundamental idea is to deploy a large number of distributed access points (APs) connected to a central processing unit (CPU) to serve all the distributed UEs in a wide area. In particular, each AP serves all UEs in the same time-frequency resource block via a time-division duplex (TDD) mode. Compared to conventional colocated massive MIMO, cell-free networks offer more uniform connectivity for all UEs thanks to the macro-diversity gain obtained from the distributed antennas. However, the assumption that each AP serves all UEs renders the system not scalable and incurs huge power and computational resource consumption for decoding, especially at UEs with low signal-to-interference-noise-ratios. To tackle the scalability issue, a *user-centric* dynamic cooperation clustering (DCC) scheme [5] was introduced [6], where each user is only served by a subset of APs.

An efficient data detection algorithm is highly desired in large-scale and complex networks, such as cell-free massive MIMO. In this aspect, some early attempts were made on centralized algorithms where the detection is totally implemented at the CPU with the received pilots and data signals reported from all APs [2], [7]. However, the computational overhead of such a centralized detection scheme is prohibitively high when the network size becomes large. To address this challenge, distributed detectors have been recently investigated. In [8], one centralized and three distributed receivers with different levels of cooperation among APs were compared in terms of SE. Unfortunately, the distributed receivers investigated in [8] are linear receivers and therefore highly suboptimal in terms of the bit-error rate (BER) performance. Therefore, it is of great importance to develop a distributed and *non-linear* receiver to achieve a better BER performance.

In this paper, we propose a non-linear detector for cell-free massive MIMO networks, which is derived based upon the EP principle [9] with a distributed approach [10], [11]. Specifically, by adopting the linear minimum mean-square error (MMSE) estimator, the APs first detect the symbols with the local channel state information and transfer the posterior mean and variance estimates to the CPU. Then, the extrinsic information for each AP is computed and integrated at the CPU by utilizing maximum-ratio combining (MRC). Subsequently, the CPU uses the posterior mean estimator to refine the detection and the extrinsic information is transferred to each AP from the CPU via the fronthaul. Compared to other distributed linear detectors, the proposed distributed EP detector improves the detection performance by introducing more computation overhead mainly at the computationally powerful CPU. Simulation results will demonstrate that the proposed method outperforms existing distributed detectors and even the centralized MMSE detector in terms of BER.

Notations—For any matrix $\mathbf{A}$, $\mathbf{A}^H$ and $\text{tr}(\mathbf{A})$ denote the conjugate transpose and trace of $\mathbf{A}$, respectively. In addition, $\mathbf{I}$ is the identity matrix and $\mathbf{0}$ is the zero matrix. We use $\mathrm{D}z$ to denote the real Gaussian integration measure. That is,

$$\mathrm{D}z = \phi(z)dz, \quad \text{where} \quad \phi(z) \triangleq \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}.$$

A complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance

$\mathbf{\Omega}$ can be described by the probability density function,

$$\mathcal{N}_{\mathbb{C}}(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Omega}) = \frac{1}{\det(\pi\mathbf{\Omega})} e^{-(\mathbf{z}-\boldsymbol{\mu})^H \mathbf{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\mu})}.$$

The remaining part of this paper is organized as follows. Section II formulates the cell-free massive MIMO detection problem. The distributed EP detector is proposed in Section III and an analytical framework is provided in Section IV. Numerical results are then presented in Section V and Section VI concludes the paper.

II. SYSTEM MODEL

In this section, we first present the system model and formulate the cell-free massive MIMO detection problem. Then, four commonly-adopted receivers are briefly introduced.

A. Cell-Free Massive MIMO

As illustrated in Fig. 1, we consider a cell-free massive MIMO network with L distributed APs, each equipped with N antennas to serve K single-antenna UEs. The system can be AP-centric or user-centric. All APs are connected to a CPU that has a high computational capability. Denote $\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{R}_{kl})$ as the channel between the k-th user and the l-th AP, where $\mathbf{R}_{kl} \in \mathbb{C}^{N \times N}$ is the spatial correlation matrix and $\beta_{k,l} = \mathrm{tr}(\mathbf{R}_{kl})/N$ is the large-scale fading coefficient involving the geometric path loss and shadowing. In the uplink data transmission phase, we consider $\mathcal{M}_k \subset \{1, \ldots, L\}$ as the subset of APs that serve the k-th UE and define the DCC matrices $\mathbf{D}_{kl}$ based on $\mathcal{M}_k$ as

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & \text{if } l \in \mathcal{M}_k \\ \mathbf{0}_{N \times N} & \text{if } l \notin \mathcal{M}_k, \end{cases} \quad (1)$$

which is constructed by the DCC strategy. Furthermore, we define $\mathcal{D}_l$ as the set consisting of the UE indices that are served by l-th AP :

$$\mathcal{D}_l = \left\{ k : \mathrm{tr}(\mathbf{D}_{kl}) \geq 1, k \in \{1, \ldots, K\} \right\}, \quad (2)$$

where the cardinality of $\mathcal{D}_l$ is denoted as $|\mathcal{D}_l|$. When each AP serves all users, $\mathcal{D}_l = K$ accordingly. We assume that perfect channel state information (CSI) is available at the local APs. Therefore, the received signal at the l-th AP is given by,

$$\mathbf{y}_l = \sum_{k=1}^{|\mathcal{D}_l|} \sqrt{p_k} \mathbf{h}_{kl} x_k + \mathbf{n}_l, \quad (3)$$

where $x_k \in \mathbb{C}$ is the transmitted symbol drawn from the M-QAM constellation, and $p_k > 0$ is the transmit power at the k-th UE. The additive noise at the l-th AP is denoted as $\mathbf{n}_l \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma^2 \mathbf{I}_N)$. Let $\mathbf{x} = [x_1, \ldots, x_K]$ denote the transmitted vector from all UEs and $\mathbf{h}_l = [\mathbf{h}_{1l}, \ldots, \mathbf{h}_{Kl}]^T \in \mathbb{C}^{N \times K}$ as the channel of the AP l to all UEs. If the k-th UE is not associated with the l-th AP, the channel vector $\mathbf{h}_{kl} = \mathbf{0}$ accordingly. Furthermore, we denote $\mathbf{H} = [\mathbf{h}_k^T, \ldots, \mathbf{h}_K^T]^T \in \mathbb{C}^{LN \times K}$ as the channel matrix between all UEs and APs. The uplink detection problem for cell-free massive MIMO is to detect the transmitted data $\mathbf{x}$ based on the received signals $\mathbf{y}_l$ ($l = 1, 2, \ldots, L$), channel matrix $\mathbf{H}$, and noise power σ^2.
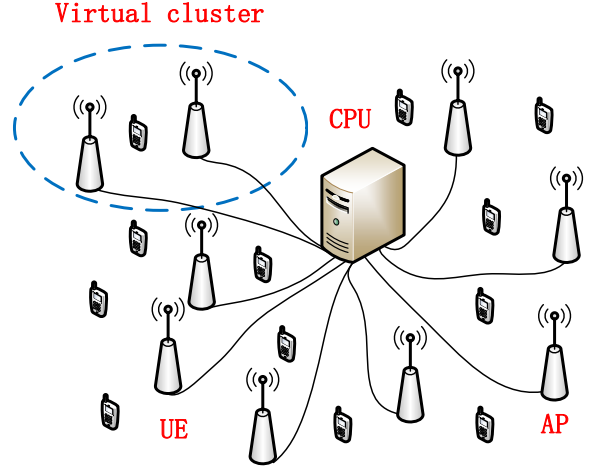
Fig. 1. A typical cell-free massive MIMO system.

B. Linear Receivers

For the cell-free massive MIMO receiver design, four receivers with different levels of cooperation among APs were introduced and compared in terms of SE [8].

- Level 4 is a fully centralized receiver where the pilot and data signals received at all APs are sent to the CPU for channel estimation and data detection.
- Level 3 involves two stages. First, each AP estimates the channels and uses the linear MMSE detector to detect the received signals. Then, the detected signals are collected at the CPU for joint detection for all UEs by utilizing the large-scale fading decoding (LSFD) method. Compared to Level 4, only the channel statistics are utilized at the CPU but the pilot signals are not required to be sent to the CPU.
- Level 2 is a special case of Level 3. The CPU performs joint detection for all UEs by simply taking the average of the local estimates. Thus, no channel statistics are required to be transmitted to CPU via the fronthaul.
- Level 1 is a fully distributed approach in which the data detection is performed at the APs based on the local channel estimates. No information is required to be transferred to the CPU.

Although the aforementioned four receivers have low complexity, the performance is far from optimal due to the linear structure. This is because all of the four receivers are linear receivers. In contrast, non-linear receivers have shown great advantages in terms of BER while the main limitation is the high computational complexity [13]. Thanks to the relatively high computing abilities at the CPU and the large number of APs in cell-free massive MIMO systems, we can offload parts of the computational-intensive operations to the CPU and distribute the partial computation tasks to APs. Next, we will propose a distributed non-linear detector for cell-free massive MIMO systems, which takes advantage of this idea.

Fig. 2. Block diagram of the proposed distributed EP detector. The "EXT." blocks represents the extrinsic information computation.

III. PROPOSED DISTRIBUTED EP DETECTOR

In this section, we apply the EP principle to develop a powerful distributed MIMO detector for cell-free massive MIMO systems. After introducing the iterative process of the algorithm, we will analyze the computational complexity and fronthaul overhead of the proposed detector.

A. Distributed Bayesian MIMO Detector

We first utilize the Bayesian inference to recover the signals $\mathbf{x}$ from the received signal $\mathbf{y}$ in the data detection stage, which is related to the following linear model $\mathbf{y} = \mathbf{h}\mathbf{x} + \mathbf{n}$. Based on Bayes' theorem, the posterior probability is given by

$$\mathrm{P}(\mathbf{x}|\mathbf{y},\mathbf{H}) = \frac{\mathrm{P}(\mathbf{y}|\mathbf{x},\mathbf{H})\mathrm{P}(\mathbf{x})}{\mathrm{P}(\mathbf{y}|\mathbf{H})} = \frac{\mathrm{P}(\mathbf{y}|\mathbf{x},\mathbf{H})\mathrm{P}(\mathbf{x})}{\int \mathrm{P}(\mathbf{y}|\mathbf{x},\mathbf{H})\mathrm{P}(\mathbf{x})d\mathbf{x}}. \quad (4)$$

Given the posterior probability $\mathrm{P}(\mathbf{x}|\mathbf{y},\mathbf{H})$, the Bayesian MMSE estimate is obtained by

$$\hat{\mathbf{x}} = \int \mathbf{x}\mathrm{P}(\mathbf{x}|\mathbf{y},\mathbf{H})d\mathbf{x}. \quad (5)$$

However, the Bayesian MMSE estimator is not computationally tractable because the marginal posterior probability in (5) involves a high-dimensional integral. The EP algorithm, proposed in [9], provides an iterative method to recover the transmitted $\mathbf{x}$ from the received signal $\mathbf{y}$. It is derived from the factor graph with the messages updated and passed between different pairs of nodes assumed to follow Gaussian distributions. As the Gaussian distribution can be fully characterized by its mean and variance, only mean and variance are required to be calculated and passed. Different from the conventional EP-based [13], the posterior probability in (4) has to be rewritten in a distributed way as follows,

$$\mathrm{P}(\mathbf{x}|\mathbf{y},\mathbf{H}) \propto \mathrm{P}(\mathbf{x}) \prod_{l=1}^{L} \exp(-\|\mathbf{y}_l - \mathbf{h}_l\mathbf{x}\|^2/\sigma^2). \quad (16)$$

Leveraging the computational capability in the AP, we can deploy partial calculation modules of the EP detector at the

---

Algorithm 1: Distributed EP for cell-free massive MIMO detection

Input: Received signal $\mathbf{y}_l$, channel matrix $\mathbf{h}_l$, noise level σ^2.

Output: Recovered signal $\mathbf{x}_{\mathrm{B}}^{\mathrm{post},T}$.

Initialize: $\boldsymbol{\gamma}_l^{(0)} \leftarrow \mathbf{0}$, $\lambda_l^{(0)} \leftarrow \frac{1}{E_x}$

for $t = 1, \cdots, T$ **do**

    Module A in APs: (1) Compute the posterior mean and variance of $\mathbf{x}_{\mathrm{A},l}$:

$$v_{\mathrm{A},l}^{\mathrm{post}} \leftarrow \boldsymbol{\Sigma}_l^t = (\sigma^{-2}\mathbf{h}_l^H\mathbf{h}_l + \lambda_l^{(t-1)}\mathbf{I}) \quad (6)$$

$$\mathbf{x}_{\mathrm{A},l}^{\mathrm{post}} \leftarrow \boldsymbol{\mu}_l^t = \boldsymbol{\Sigma}_l^t\left(\sigma^{-2}\mathbf{h}_l\mathbf{y}_l + \boldsymbol{\gamma}_l^{(t-1)}\right) \quad (7)$$

    Module B in CPU: (2) Compute the extrinsic mean and variance of $\mathbf{x}_{\mathrm{A},l}$:

$$v_{\mathrm{A},l}^{\mathrm{ext}} \leftarrow \left(\frac{1}{v_{\mathrm{A},l}^{\mathrm{post}}} - \lambda_l^{(t-1)}\right)^{-1} \quad (8)$$

$$\mathbf{x}_{\mathrm{A},l}^{\mathrm{ext}} \leftarrow v_{\mathrm{A},l}^{\mathrm{ext}}\left(\frac{\boldsymbol{\mu}_l^t}{v_{\mathrm{A},l}^{\mathrm{post}}} - \boldsymbol{\gamma}_l^{(t-1)}\right)^{-1} \quad (9)$$

    (3) MRC combining of $\mathbf{x}_{\mathrm{A},l}$:

$$\frac{1}{v_{\mathrm{A}}^{\mathrm{ext}}} = \sum_{l=1}^{L} \frac{1}{v_{\mathrm{A},l}^{\mathrm{ext}}} \quad (10)$$

$$\mathbf{x}_{\mathrm{A}}^{\mathrm{ext}} = v_{\mathrm{A}}^{\mathrm{ext}} \sum_{l=1}^{L} \frac{\mathbf{x}_{\mathrm{A},l}^{\mathrm{ext}}}{v_{\mathrm{A},l}^{\mathrm{ext}}} \quad (11)$$

    (4) Compute the posterior mean and variance of $\mathbf{x}_{\mathrm{B}}$:

$$\mathbf{x}_{\mathrm{B}}^{\mathrm{post}} = \mathrm{E}\{\mathbf{x}|\mathbf{x}_{\mathrm{A}}^{\mathrm{ext}}, v_{\mathrm{A}}^{\mathrm{ext}}\} \quad (12)$$

$$\mathbf{v}_{\mathrm{B}}^{\mathrm{post}} = \mathrm{var}\{\mathbf{x}|\mathbf{x}_{\mathrm{A}}^{\mathrm{ext}}, v_{\mathrm{A}}^{\mathrm{ext}}\} \quad (13)$$

    (5) Compute the extrinsic mean and variance $\mathbf{x}_{\mathrm{B},l}$:

$$\frac{1}{v_{\mathrm{B},l}^{\mathrm{ext}}} \leftarrow \lambda_l^{(t)} = \frac{1}{\mathrm{mean}(\mathbf{v}_{\mathrm{B}}^{\mathrm{post}})} - \frac{1}{v_{\mathrm{A},l}^{\mathrm{ext}}} \quad (14)$$

$$\frac{\mathbf{x}_{\mathrm{B},l}^{\mathrm{ext}}}{v_{\mathrm{B},l}^{\mathrm{ext}}} \leftarrow \boldsymbol{\gamma}_l^{(t)} = \frac{\mathbf{x}_{\mathrm{B}}^{\mathrm{post}}}{\mathrm{mean}(\mathbf{v}_{\mathrm{B}}^{\mathrm{post}})} - \frac{\mathbf{x}_{\mathrm{A},l}^{\mathrm{post}}}{v_{\mathrm{A},l}^{\mathrm{ext}}} \quad (15)$$

AP based on the local information and then send the posterior mean and variance estimates to the CPU for combining. The distributed EP-based detector is illustrated in **Algorithm 1**. The input of the algorithm is the received signal $\mathbf{y}_l$, channel matrix $\mathbf{h}_l$, and noise level σ^2 while the output is the recovered signal $\mathbf{x}_{\mathrm{B}}^{\mathrm{post},T}$ in the T-th iteration. The initial parameters are $\boldsymbol{\gamma}_l^{(0)} = \mathbf{0}$, $\lambda_l^{(0)} = \frac{1}{E_x}$, where

$$E_x = \mathrm{E}\{\|\mathbf{x}\|^2\}/K. \quad (17)$$

$\boldsymbol{\gamma}_l^{(0)}$ and $\lambda_l^{(0)}$ are the initialized extrinsic information and E_x is the power of the transmitted symbol x_k. Furthermore, the

block diagram of the proposed distributed detector is illustrated in Fig. 2, which is composed of module A and module B. Each module uses the turbo principle in iterative decoding. That is, each module passes the extrinsic messages to its next module and this process is executed until convergence.

To better understand the distributed EP detection algorithm, we elaborate the details for each iteration in **Algorithm 1**. Specifically, module A is the linear MMSE (LMMSE) estimator performed at the APs according to the following linear model

$$\mathbf{y}_l = \mathbf{h}_l \mathbf{x} + \mathbf{n}_l. \tag{18}$$

In the t-iteration of the algorithm, the explicit expression for the posterior covariance matrix $\boldsymbol{\Sigma}_l^t$ and mean vector $\boldsymbol{\mu}_l^t$ are given by (6) and (7), respectively. Note that each AP only uses the local channel $\mathbf{h}_l$ to detect the transmitted signal $\mathbf{x}$. For ease of notation, we omit the iteration index t for all mean and variance estimates. Then, the variance $v_{\mathrm{A},l}^{\mathrm{post}} = \mathrm{tr}(\boldsymbol{\Sigma}_l^t)/|\mathcal{D}_l|$ and mean estimate $\mathbf{x}_{\mathrm{A},l}^{\mathrm{post}} = \boldsymbol{\mu}_l^t$ are transferred to the CPU to compute the extrinsic information $v_{\mathrm{A},l}^{\mathrm{ext}}$ (8) and $\mathbf{x}_{\mathrm{A},l}^{\mathrm{ext}}$ (9), respectively. The extrinsic information $\mathbf{x}_{\mathrm{A},l}^{\mathrm{ext}}$ can be regarded as the AWGN observation given by

$$\mathbf{x}_{\mathrm{A},l}^{\mathrm{ext}} = \mathbf{x} + \mathbf{n}^{\mathrm{eq}}, \tag{19}$$

where $\mathbf{n}_l^{\mathrm{eq}} \sim \mathcal{N}_{\mathbb{C}}(0, v_{\mathrm{A},l}^{\mathrm{ext}}\mathbf{I})$ [11]. Therefore, the linear model in (3) is decoupled into K parallel and independent AWGN channels with equivalent noise $v_{\mathrm{A},l}^{\mathrm{ext}}$. Subsequently, the CPU collects all extrinsic means $\{\mathbf{x}_{\mathrm{A},l}^{\mathrm{ext}}\}_{l=1}^L$ and variances $\{v_{\mathrm{A},l}^{\mathrm{ext}}\}_{l=1}^L$ and performs MRC. The MRC expressions (10) and (11) are obtained by maximizing the post-combination signal-to-noise ratio (SNR) of the final AWGN observation $\mathbf{x}_{\mathrm{A}}^{\mathrm{ext}}$ at the CPU, given by

$$\mathbf{x}_{\mathrm{A}}^{\mathrm{ext}} = \mathbf{x} + \mathbf{n}^{\mathrm{eq}}, \tag{20}$$

where $\mathbf{n}_{\mathrm{A}}^{\mathrm{ext}} \sim \mathcal{N}_{\mathbb{C}}(0, v_{\mathrm{A}}^{\mathrm{ext}}\mathbf{I})$. The CPU uses the posterior mean estimator to detect the signal $\mathbf{x}$ from the equivalent AWGN model (20). Then, the posterior mean and variance are computed by the posterior MMSE estimator for the equivalent AWGN model in (12) and (13). As the transmitted symbol is assumed drawn from the M-QAM set $\mathcal{S} = \{s_1, s_2, \ldots, s_M\}$, the corresponding expressions for each element in (12) and (13) are given by

$$[x_{\mathrm{B}}^{\mathrm{post}}]_k = \frac{\sum_{s_i \in \mathcal{S}} s_i \mathcal{N}_{\mathbb{C}}(s_i; [x_{\mathrm{A}}^{\mathrm{ext}}]_k, v_{\mathrm{A}}^{\mathrm{ext}}) p(s_i)}{\sum_{s_i \in \mathcal{S}} \mathcal{N}_{\mathbb{C}}(s_i; [x_{\mathrm{A}}^{\mathrm{ext}}]_k, v_{\mathrm{A}}^{\mathrm{ext}}) p(s_i)} \tag{21}$$

$$[v_{\mathrm{B}}^{\mathrm{post}}]_k = \frac{\sum_{s_i \in \mathcal{S}} |s_i|^2 \mathcal{N}_{\mathbb{C}}(s_i; [x_{\mathrm{A}}^{\mathrm{ext}}]_k, v_{\mathrm{A}}^{\mathrm{ext}}) p(s_i)}{\sum_{s_i \in \mathcal{S}} \mathcal{N}_{\mathbb{C}}(s_i; [x_{\mathrm{A}}^{\mathrm{ext}}]_k, v_{\mathrm{A}}^{\mathrm{ext}}) p(s_i)} - |[x_{\mathrm{B}}^{\mathrm{post}}]_k|^2. \tag{22}$$

where $[x_{\mathrm{B}}^{\mathrm{post}}]_k$, $[v_{\mathrm{B}}^{\mathrm{post}}]_k$, and $[x_{\mathrm{A}}^{\mathrm{ext}}]_k$ are the k-th element in $\mathbf{x}_{\mathrm{B}}^{\mathrm{post}}$, $\mathbf{v}_{\mathrm{B}}^{\mathrm{post}}$, and $[\mathbf{x}_{\mathrm{A}}^{\mathrm{ext}}]_k$, respectively. The posterior mean and variance $\mathbf{x}_{\mathrm{B}}^{\mathrm{post}}$ and $\mathbf{v}_{\mathrm{B}}^{\mathrm{post}}$ are then utilized to compute the extrinsic information $\lambda_l^{(t)}$ and $\gamma_l^{(t)}$ for each AP in (14) and (15), where the function $\mathrm{mean}(\cdot)$ is used to compute the mean. Finally, the extrinsic information $\lambda_l^{(t)}$ and $\gamma_l^{(t)}$ are transferred to each AP in the next iteration. The whole

procedure is executed iteratively until terminated by a certain stopping criterion or a maximum number of iterations.

TABLE I. Complexity of different detectors

Detectors	Distributed EP	Level 4	Level 3&2	Level 1
AP	$O(T|\mathcal{D}_l|N^2)$	0	$O(|\mathcal{D}_l|N^2)$	$O(|\mathcal{D}_l|N^2)$
CPU	$O(T|\mathcal{D}_l|^2)$	$O(LN)^3$	$O(|\mathcal{D}_l|)$	0

B. Computational Complexity and Fronthaul Overhead

In the following, we provide the complexity analysis for different detectors in Tables I and II. For the proposed distributed EP detector, the computational complexity at each AP is dominated by the LMMSE estimator for estimating the signal $\mathbf{x}_{\mathrm{A},l}^{\mathrm{post}}$, which is $O(|\mathcal{D}_l|^3)$ because of the matrix inversion is required in (6) while the computational complexity at the CPU is $O(|\mathcal{D}_l|^2)$ in each iteration. Furthermore, if the number of antennas N is less than $|\mathcal{D}_l|$, we can use the matrix inversion lemma to carry out the matrix inversion in (6) as follows

$$(\sigma^{-2}\mathbf{h}_l^H \mathbf{h}_l + \mathbf{D})^{-1} = \mathbf{D}^{-1} \tag{23}$$
$$- \sigma^{-2}\mathbf{D}^{-1}\mathbf{h}_l^H (\mathbf{I} + \sigma^{-2}\mathbf{h}_l \mathbf{D}^{-1}\mathbf{h}_l^H)^{-1}\mathbf{h}_l \mathbf{D}^{-1},$$

where $\mathbf{D} = \lambda_l^{(t-1)}\mathbf{I}$ and the computational complexity is reduced to $O(|\mathcal{D}_l|N^2)$. Therefore, the overall computational complexity at each AP is $O(T|\mathcal{D}_l|N^2)$ while the overall computational complexity at the CPU is $O(T|\mathcal{D}_l|^2)$ for T iterations. As can be observed in Table I, the distributed EP detector mainly increases the computational complexity at the CPU when compared with other distributed linear receivers (Levels 1-3).

TABLE II. Fronthaul overhead

Detectors	Coherence block	Statistical parameters
Distributed EP	$\sum_{l=1}^{L}(\tau_c - \tau_p)2T(|\mathcal{D}_l| + 1)$	0
Level 4	$\tau_c NL$	$KLN^2/2$
Level 3	$(\tau_c - \tau_p)KL$	$KL + (L^2K^2 + KL)/2$
Level 2	$(\tau_c - \tau_p)KL$	0
Level 1	0	0

We compare the number of complex scalars that need to be transmitted from the APs to the CPU via the fronthauls in Table II. We assume that τ_c and τ_p are the coherence time and pilot length, respectively. The results for the 4 level detectors are from [8]. With the proposed detector, $\sum_{l=1}^{L}(\tau_c - \tau_p)2T(|\mathcal{D}_l| + 1)$ scalars need to be passed from the APs to the CPU and no statistical parameters are required to be passed, where T denotes the total iterations. The fronthaul overhead of the proposed distributed EP detector is similar to Level 4 and the detailed comparison is determined by the value of the system parameters.

IV. State Evolution Analysis

In this section, we provide an analytical framework to predict the asymptotic performance of the distributed EP in the large system limit. We consider $L, K, N \to \infty$ and fix

$$\alpha_l = \frac{K}{N}, \alpha = \frac{K}{LN}. \tag{24}$$

Then, we have following proposition.

Proposition 1. *In the large-system limit, the asymptotic behavior (such as MSE and BER) of Algorithm 1 can be described by the following equations:*

$$v_{\mathrm{A},l}^{\mathrm{ext},t} = \frac{\alpha_l \sigma^2 + (\alpha_l - 1)\lambda_l^{(t-1)}}{2}$$

$$+ \frac{\sqrt{\left(\alpha_l \sigma^2 + (\alpha_l - 1)\lambda_l^{(t-1)}\right)^2 + 4\alpha_l \sigma^2 \lambda_l^{(t-1)}}}{2} \tag{25a}$$

$$v_{\mathrm{A}}^{\mathrm{ext},t} = \left(\sum_{l=1}^{L} \frac{1}{v_{\mathrm{A},l}^{\mathrm{ext},t}} \right)^{-1} \tag{25b}$$

$$\lambda_l^{(t)} = \frac{1}{\mathrm{MSE}(v_{\mathrm{A}}^{\mathrm{ext},t})} - \frac{1}{v_{\mathrm{A},l}^{\mathrm{ext},t}}. \tag{25c}$$

∎

The function $\mathrm{MSE}(\cdot)$ is given by

$$\mathrm{MSE}(v_{\mathrm{A},l}^{\mathrm{ext},t}) \triangleq \mathsf{E}\left\{ |x - \mathsf{E}\{x | x_{\mathrm{A},l}^{\mathrm{ext},t}, v_{\mathrm{A},l}^{\mathrm{ext},t}|^2 \right\} \tag{26}$$

and the expectation is with respect to x. The state equations can be proved using the method in [12], and we give an intuitive explanation here, which is linked to **Algorithm 1**. By substituting (6) into (8), we have following expression as

$$v_{\mathrm{A},l}^{\mathrm{ext},t} = \frac{1}{\frac{1}{K}\mathrm{tr}(\sigma^{-2}\mathbf{h}_l^H \mathbf{h}_l + \lambda_l^{(t-1)}\mathbf{I})} - \lambda_l^{(t-1)}. \tag{27}$$

Notably, $v_{\mathrm{A},l}^{\mathrm{ext}}$ can be characterized by the eigenvalues of $\mathbf{h}_l^H \mathbf{h}_l$, which converge to a deterministic distribution in the large system limit. Because of the limited space, we only show the final result of the **Proposition 1** and omit the derived process. By adopting the result of the $\mathcal{R}$-transform of the average empirical eigenvalue distribution, we have an asymptotic expression for $v_{\mathrm{A},l}^{\mathrm{ext},t}$ in (25). Then, the asymptotic expressions for $v_{\mathrm{A}}^{\mathrm{ext},t}$ and $\lambda_l^{(t)}$ can be derived from (10) and (14). Finally, the asymptotic MSE can be interpreted as the MSE of the decoupled scalar AWGN channels (20) and is related to distribution of transmitted signal $\mathbf{x}$. Next, we will give a specific example for **Proposition 1**.

Example 1: If the data symbol is drawn from a quadrature phase-shift keying (QPSK) constellation, the MSE is expressed by

$$\mathrm{MSE} = 1 - \int \mathrm{D}z \tanh(\frac{1}{v_{\mathrm{A},l}^{\mathrm{ext},t}} + \sqrt{\frac{1}{v_{\mathrm{A},l}^{\mathrm{ext},t}}}z). \tag{28}$$

Furthermore, the final BER w.r.t. $\mathbf{x}$ can also be evaluated through the equivalent AWGN channel (20) with an equivalent

$\mathrm{SNR} = 1/v_{\mathrm{A},l}^{\mathrm{ext},T}$ and is given by [14]

$$\mathrm{BER} = 2Q\left(\sqrt{\frac{1}{v_{\mathrm{A},l}^{\mathrm{ext},T}}}\right) - [Q\left(\sqrt{\frac{1}{v_{\mathrm{A},l}^{\mathrm{ext},T}}}\right)]^2, \tag{29}$$

where $Q(x) = \int_x^\infty \mathrm{D}z$ is the Q-function. In fact, the MSE and BER are determined on the basis of the knowledge of the AWGN channel (20) with $\mathrm{SNR} = 1/v_{\mathrm{A},l}^{\mathrm{ext},T}$, which is known as the decoupling principle. Thus, if the data symbol is drawn from other M-QAM constellations, then the corresponding BER can be easily obtained using the closed-form BER expression in [15]. Note that the state evolution equations are accurate when each AP serves all UEs in the network and the channel matrix $\mathbf{H}$ is a Rayleigh fading channel. This assumption can be achieved by perfect power control to compensate the large-scaled fading effect.

V. Simulation Results

In this section, we provide sample simulation results to demonstrate the performance of the proposed distributed EP detector for cell-free massive MIMO. Consider $L = 16$, $N = 8$, $K = 16$, and the other simulation parameters are the same as those in [8]. The SNR is defined as $\mathrm{SNR} = 1/\sigma^2$.
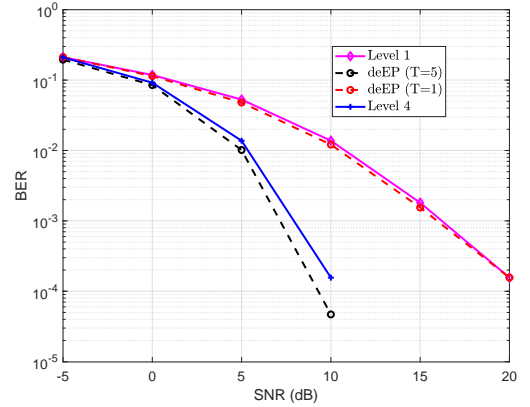

Fig. 3. BER performance comparisons of different detectors in the original cell-free massive MIMO system.

A. Original Cell-Free Massive MIMO

In this subsection, we consider the BER of the proposed detector in the original cell-free massive MIMO where each AP serves all UEs. We assume that perfect CSI can be obtained at each AP. Fig. 3 compares the achievable BER of the proposed distributed EP detector with other detectors investigated in [8]. The results are obtained by the Monte Carlo simulation with $10,000$ independent channel realizations. We denote "deEP" as the distributed EP detector. It can be observed that the proposed distributed EP detector is comparable with the Level 1 detector with only one EP iteration and outperforms the centralized Level 4 detector with $T = 5$ iterations.

We next the accuracy of the analytical framework in Fig. 4 with different modulation schemes. As shown in the figure, the
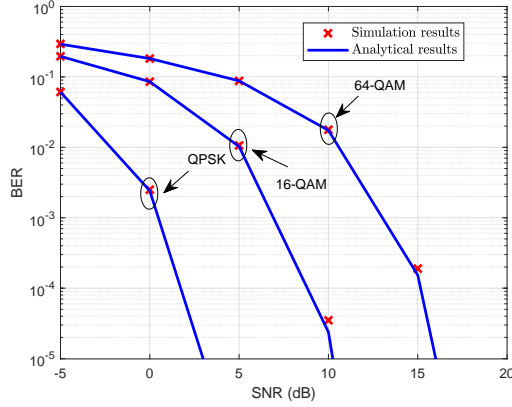
Fig. 4. BER performance comparisons of the analytical and simulation results of the distributed EP detector with different modulation schemes.

Fig. 5. BER performance comparisons of different detectors in cell-free massive MIMO with dynamic cluster cooperation.

BERs of the proposed detector match well with the derived analytical results, which demonstrate the accuracy of the analytical framework. Therefore, instead of performing time-consuming Monte Carlo simulations to obtain the corresponding performance metrics, we can predict the theoretical behavior by state equations. Furthermore, the analytical framework can be further utilized to optimize the system design.

B. Scalable Cell-Free Massive MIMO

As the *user-centric* approach is more attractive for cell-free massive MIMO, we then investigate the distributed EP detector with DCC. The accessing UE first appoints a master AP according to the large-scale fading factor and assigns a pilot to the appointed AP. Then, other neighboring APs determine whether they serve the accessing UE according to the assigned pilot. Finally, the cluster for the k-th UE is constructed. Fig. 5 shows that the distributed EP detector outperforms both the centralized and distributed MMSE detectors. Furthermore, the performance loss is acceptable when compared to original cell-free massive MIMO and the computational complexity is significantly decreased from $O(KN^2)$ to $O(|\mathcal{D}_l|N^2)$.

VI. CONCLUSION

In this paper, we proposed a distributed EP detector for cell-free massive MIMO. It is shown that such detector can achieve better performance than other linear receivers for both original and scalable cell-free massive MIMO networks. Compared to other distributed detectors, it can achieve a better BER performance and mainly increase the computational overhead at the CPU. An analytical framework was also provided to describe the asymptotic performance of the proposed detector in a large system setting. Simulation results have also been provided and demonstrated that the proposed method outperforms the existing distributed detectors for cell-free massive MIMO in terms of BER performance.
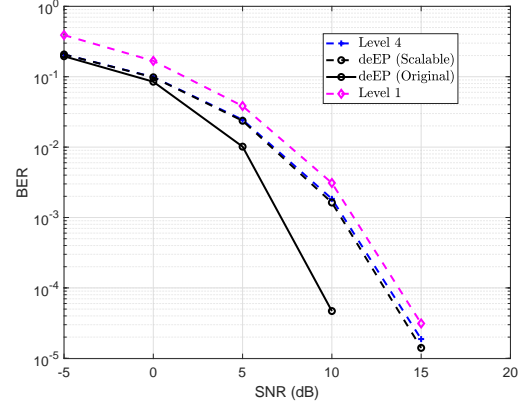
REFERENCES

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

[2] 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, Mar. 2017.

[3] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99878–99888, Sep. 2019.

[4] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J.-A. Zhang, "The roadmap to 6G: AI empowered wireless networks" *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[5] E. Björnson, N. Jaldén, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, Dec. 2011.

[6] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

[7] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free massive MIMO systems with MMSE and LSFD receivers," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 203–207.

[8] 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, Jan. 2020.

[9] T. P. Minka, "A family of algorithms for approximate Bayesian Inference," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., MIT, Cambridge, MA, USA, 2001.

[10] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, "Decentralized equalization with feedforward architectures for massive MU-MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4418–4432, Sep. 2019.

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

[12] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Trans. Inf. Theory.*, vol. 66, no. 1, pp. 368–386, Oct. 2019.

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

[14] H. He, C.-K. Wen, and S. Jin, "Bayesian optimal data detector for hybrid mmWave MIMO-OFDM systems with low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 469–483, Jun. 2018.

[15] J. G. Proakis, *Digital Communications*. Boston, USA: McGraw-Hill Companies, 2007.