Scalable Cell-Free Massive MIMO with Fully Distributed Large-Scale Fading Decoding

Leonard Paul Schulz, Christian Schappmann, Gerhard Bauch
Institute of Communications
Hamburg University of Technology (TUHH)
Hamburg, Germany
{leonard.schulz, christian.schappmann, bauch}@tuhh.de

***Abstract*—Cell-free massive multiple-input multiple-output (MIMO) is a promising technology for future wireless communication systems, where a large number of access points (APs) serve the user equipments (UEs) in the network coherently. Different ways of operating the system are possible, ranging from a fully centralized approach, where all signal processing is performed at a central processing unit (CPU), to a fully distributed approach, where most of the signal processing is performed at the APs. This paper proposes new heuristics to select the large-scale fading decoding (LSFD) weights in a fully distributed cell-free massive MIMO system. Due to the analytical intractability of the system, we rely on extensive Monte-Carlo simulations to evaluate their performance compared to the other operation levels. Importantly, we consider a *scalable* setup, where the APs are grouped into user-centric clusters. Our results show that a fully distributed system, which locally computes the LSFD weights in each AP using our heuristics, is able to approach the performance of a partially distributed system, where the LSFD weights are selected optimally in a centralized manner, especially for deployments with multiple antennas per AP. Moreover, when considering scalability, the proposed improved fully distributed system achieves substantial gains in terms of spectral efficiency over small-cell systems. Hence, even though fully centralized cell-free systems still achieve the best performance, distributed systems are a viable alternative that can be more practical to implement.**

***Index Terms*—cell-free massive MIMO, uplink, large-scale fading decoding, fully distributed operation, AP cooperation**

I. Introduction

Cell-free massive multiple-input multiple-output (MIMO) is a potential future wireless communication technology that improves spectral efficiency and coverage. Unlike traditional cellular massive MIMO systems that concentrate antennas in a few base stations (BSs), cell-free massive MIMO distributes antennas among numerous access points (APs). These APs, connected to a central processing unit (CPU) via fronthaul links, serve user equipments (UEs) jointly. The geometric distribution of antennas over the coverage area provides higher spatial diversity, enabling aggressive spatial multiplexing. Without cells, there are no cell-edges, eliminating bad service due to interference from neighboring cells. Thus, cell-free massive MIMO enhances network area throughput and service reliability [1].

The obvious drawback of the cell-free massive MIMO architecture is that it is more difficult to practically implement than traditional cellular massive MIMO systems. For this, scalability is a crucial aspect [2]. By forming user-centric clusters, each UE is served by a subset of APs, simplifying signal processing to focus on the most relevant UEs. Different operation modes, also called levels, of cell-free massive MIMO systems have been defined in [3]. The fully centralized (Level 4) operation requires a lot of fronthaul capacity and computational resources at the CPU, but also achieves the best performance. An alternative in the uplink is to distribute some of the signal processing to the APs (Levels 3 and 2) and employ a 2-layer decoding scheme known as large-scale fading decoding (LSFD). The LSFD weights only depend on the *statistical* channel state information (CSI) and can be seen in analogy to the power allocation coefficients in the downlink. The downlink power allocation problem has been tackled both centrally according to different optimization criteria [4] and in a distributed manner by employing a heuristic in each AP [5], [6]. In the uplink, a minimum mean square error (MMSE) optimizer can be employed to compute the LSFD weights centrally (Level 3) [7]. However, computing the LSFD weights in a distributed manner (Level 2) remains relatively unexplored. In this work, we derive different heuristics for this problem and show that they can achieve close to optimal performance. Importantly, we take scalability into account by using a modified (Level 2) version of the pilot assignment and clustering algorithm from [2]. We provide a detailed analysis of the effect of scalability on the different operation modes. Our results reveal that scalability significantly degrades small-cell (Level 1) system performance for setups with multiple antennas per AP, while other modes are only slightly affected. Hence, in scalable systems, substantial improvements over small-cell systems can already be achieved with a fully distributed operation. This is in contrast to unscalable systems, where only more complex implementations of cell-free massive MIMO are competitive with small-cell systems [3].

The paper proceeds as follows: Operation modes of cell-free massive MIMO are detailed in Sec. II, a scalable system is described in Sec. III, and LSFD weight computation is covered in Sec. IV. Detailed numerical results from Monte Carlo simulations are presented in Sec. V.

II. Operation Modes of Cell-Free Massive MIMO

Next, we describe the four different operation modes of cell-free massive MIMO networks as originally defined in [3].

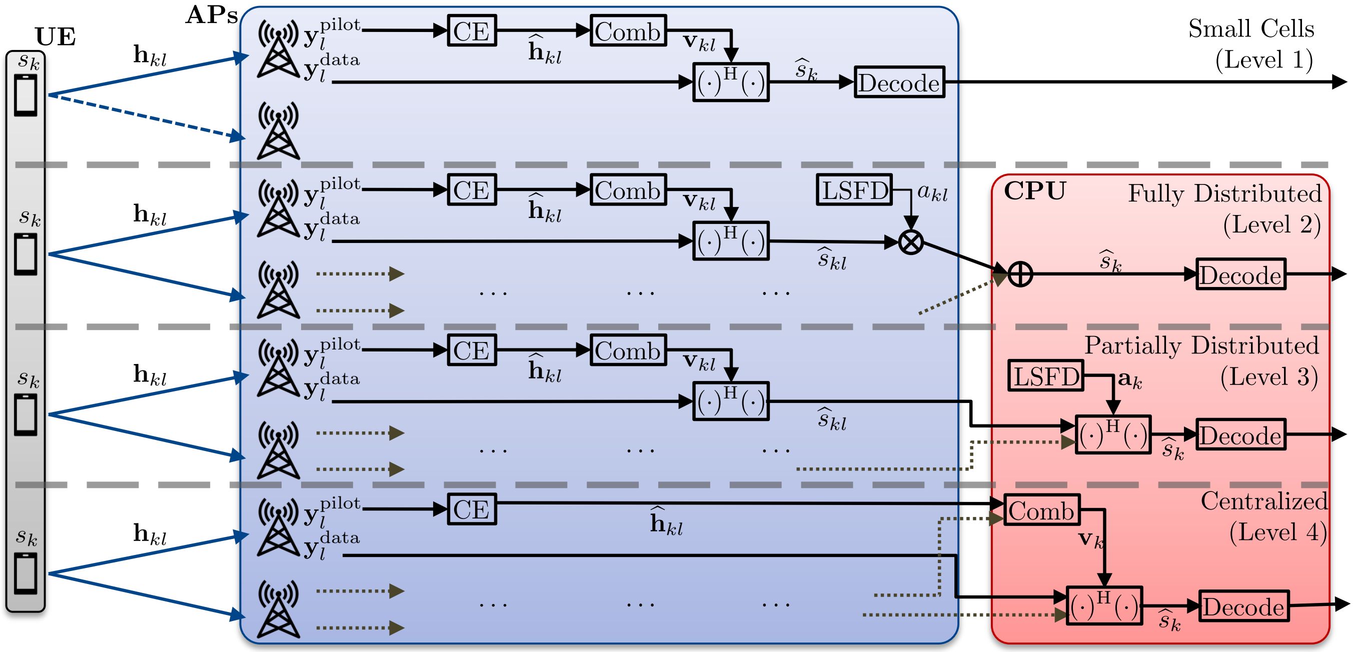


Fig. 1: The four levels of network cooperation as introduced in [3]. (CE = channel estimation, Comb = receive combining, LSFD = large-scale fading decoding weight computation).

Each of the levels of cooperation is visualized in Fig. 1. Note, that for all levels, the channel estimation (CE) is performed locally at the APs, as no gain can be achieved by centralized channel estimation, assuming the channels of different APs are uncorrelated.

Level 4 represents the fully centralized operation, which is the one with the highest amount of cooperation between APs. Here the combining vector $\mathbf{v}_k$ for UE k is selected in a globally optimal way by the CPU. For this, each AP l shares both its received data $\mathbf{y}_l^{\text{data}}$ and its channel estimate $\widehat{\mathbf{h}}_{kl}$ for the channel from UE k via the fronthaul links. At the CPU, the scalar product of $\mathbf{v}_k$ and the concatenation $\mathbf{y}^{\text{data}}$ of all receive vectors $\mathbf{y}_l^{\text{data}}$ yields the signal estimate $\widehat{s}_k$ which is then directly fed into the decoder.

Level 3, which we also call partially distributed operation, assumes only *statistical* CSI is available at the CPU, meaning the combining vector selection has to be done locally at the APs. Based on the locally available instantaneous CSI, a local combining vector $\mathbf{v}_{kl}$ is selected, producing a local signal estimate $\widehat{s}_{kl}$ at each AP. The decoding is hence implemented in a 2-layer scheme, commonly referred to as LSFD: The CPU combines the local estimates from all APs to form the final estimate in a weighted sum. The LSFD weights $\mathbf{a}_k$ are computed based on the statistical CSI and give a higher priority to the APs with better channel conditions as more reliable estimates should have a higher impact on the final estimate.

Level 2 does not require any CSI at the CPU and therefore represents the fully distributed operation. It was originally defined as a direct simplification of Level 3, where no LSFD weighting vector is used anymore, but rather the final estimate is the average over all local estimates [3]. However, the defining feature of Level 2 is that no statistical CSI is transferred to the CPU, so any LSFD scheme computing the weights locally at the APs can be considered for Level 2. The local weights cannot be optimal, since an AP is not aware of the reliability of signals at the other APs, so it cannot exactly know its relative contribution to the final estimate. However, in Sec. IV-B, we propose different heuristics for finding the LSFD weights locally at the APs, coming close to the performance of the optimal Level 3 scheme. Note, that the local weights in Level 2 operation do not need to be forwarded to the CPU as the local estimates can simply be multiplied by the local weights before being sent to the CPU.

Level 1 is the lowest form of cooperation and corresponds to a cellular system with small cells. No coherent cooperation between APs is performed, rather each UE is served by only one AP. This AP performs channel estimation, combining and makes the decision on the signal estimate locally. No CPU and no fronthaul link is required for Level 1 operation, as the local estimate is the final estimate which can be decoded and sent to the core network. Since there is no AP cooperation in Level 1, it does not qualify as cell-free. Therefore, it is used as a baseline of what performance is achievable with the same deployment, but using conventional cellular techniques.

We refer to [2], [3], [7] for a detailed analysis on the computational complexity and fronthaul requirements of different operation levels. Note, that the computation of the LSFD weights a_{kl} only has to be done when the statistical CSI changes while the combining vectors $\mathbf{v}_{kl}$ have to be updated for each coherence block. All of our in Sec. IV-B proposed heuristics consist of simple computations at each AP, therefore their computational complexity is negligible compared to the channel estimation and receive combining.

III. System Model

We consider a cell-free massive MIMO system with L APs, each equipped with N antennas, and K single-antenna UEs. Hence, the total number of antennas in the system is $M = L \cdot N$. We use the standard block fading channel model, where the channel is constant for one coherence block. The system is operated in a time-division duplexing (TDD) fashion, each coherence block consists of $\tau_c = \tau_p + \tau_u + \tau_d$ symbols, where τ_p symbols are used for pilot transmission, τ_u symbols for uplink data transmission, and τ_d symbols for downlink data transmission. The channel between AP l and UE k is denoted as $\mathbf{h}_{kl} \in \mathbb{C}^N$ and is modeled as spatially correlated Rayleigh fading, meaning the realizations are sampled independently between coherence blocks from a correlated complex Gaussian distribution, i.e., $\mathbf{h}_{kl} \sim \mathcal{CN}(0, \mathbf{R}_{kl})$ as described in [8, Sec. 2.2]. We assume that the *local* statistical CSI of all the links connected to AP l is perfectly known at that AP, as it can be measured with standard methods [9]. Further, we assume all APs and UEs are synchronized in time and frequency.

A. Pilot Assignment and User-Centric Clustering

A cell-free massive MIMO system is called scalable if the computational complexity per AP as well as the fronthaul signaling requirements remain bounded as $K \to \infty$, i.e., as the number of UEs goes to infinity [2]. In order to achieve this, each UE can only be served by a subset of the APs in the network. In the spirit of a cell-free system, each UE is served by a dynamic cooperation cluster (DCC) of APs around it. Further, there are only τ_p orthogonal pilot sequences available in the network, so for $K > \tau_p$, multiple UEs have to share the same pilot sequence. In order to keep pilot contamination low, UEs that are served by a similar set of APs, should be assigned different pilot sequences, making it natural to solve pilot assignment and user-centric clustering jointly. Development of algorithms for this problem is an ongoing research topic (see, e.g., [5], [10]–[12]), we use a modified version of the algorithm from [2] in this work. The pilot assignment step in the original algorithm is not strictly *Level 2*, as it requires some signaling between the APs. In contrast, we let each UE pick its pilot sequence randomly which can be done in a fully distributed manner. Second, each AP can greedily choose to serve the UE with the best channel for each pilot sequence.

For the next sections, we introduce the following notations: $\mathcal{M}_k$ denotes the set of APs that serve UE k. $\mathcal{D}_l$ denotes the set of UEs served by AP l. $\mathcal{S}_k = \{i \in \{1,\dots,K\} \mid \mathcal{M}_i \cap \mathcal{M}_k \neq \emptyset\}$ denotes the set of UEs that are served by partially the same APs as UE k. The matrix $\mathbf{D}_k$ is a block diagonal matrix where each block $\mathbf{D}_{kl}$ is a $N \times N$ identity matrix if AP l serves UE k, and a $N \times N$ zero matrix otherwise. This is useful for expressing the selection of channels, i.e., $\mathbf{D}_k \mathbf{h}_k \in \mathbb{C}^{LN}$ is the channel vector containing zeros in places corresponding to unused channels. The vectors $\mathbf{g}_{ki} = \left[\mathbf{v}_{k1}^{\mathrm{H}} \mathbf{D}_{k1} \mathbf{h}_{i1}, \dots, \mathbf{v}_{kL}^{\mathrm{H}} \mathbf{D}_{kL} \mathbf{h}_{iL}\right]^{\mathrm{T}} \in \mathbb{C}^L$ are the products of the channels from UE i to all APs and the combining vectors of UE k.

B. Channel Estimation

In a coherence block, each UE k broadcasts its assigned pilot sequence $\mathbf{\Phi}_t \in \mathbb{C}^{\tau_p}$ with transmit power $\sqrt{p_k}$. To estimate the channel from UE k at AP l, the received superposition of all pilot sequences is correlated with $\mathbf{\Phi}_t$. Assuming the pilot sequences are orthogonal, this cancels out the interference from other UEs except the ones using the same pilot sequence $\mathbf{\Phi}_t$ as UE k. The well-known MMSE estimator can then be applied [13, Sec. 4.2], yielding the channel estimate $\widehat{\mathbf{h}}_{kl} = \widetilde{\mathbf{h}}_{kl} + \mathbf{h}_{kl}$. The channel estimation error $\widetilde{\mathbf{h}}_{kl}$ arises due to noise and pilot contamination, we denote its covariance matrix as $\mathbb{E}\{\widetilde{\mathbf{h}}_{kl}\widetilde{\mathbf{h}}_{kl}^{\mathrm{H}}\} = \mathbf{C}_{kl}$. In a scalable system, each AP l should only need to estimate the channels that are used for transmission, i.e., channels of UEs from the set $\mathcal{D}_l$.

C. Simplified Partial Receive Combining

For detection, the received signal is multiplied with a combining vector which is selected based on the channel estimates $\widehat{\mathbf{h}}_{kl}$. The optimal combining scheme finds the balance between maximizing the signal power and minimizing the interference power, therefore it would have to take the channels from all UEs into account. In a scalable system, only a subset of the channel estimates is used for the selection of the combining vector, leading to so-called *partial* receive combining schemes [14]. Naturally, partial combining schemes yield small losses in performance if the channels that are ignored are weak.

1) Centralized Simplified Partial MMSE: In the centralized operation (Level 4), an MMSE estimator taking channels from different APs into account can be used. To make the scheme scalable, it was proposed in [2] to only take the channels of UEs from the set $\mathcal{S}_k$ into account. However, this approach still requires the knowledge of some channel estimates of channels that have not been selected for transmission by the clustering scheme. Therefore, we further simplify the scheme by ignoring these channels as well (i.e., including the selection matrix $\mathbf{D}_i$ in the equation), leading to the simplified partial (SP)-MMSE scheme with combining vectors

$$\mathbf{v}_k^{\mathrm{SP-MMSE}} = p_k \left(\sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k (\mathbf{D}_i \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}_i^{\mathrm{H}} \mathbf{D}_i + \mathbf{C}_i) \mathbf{D}_k + \sigma_{\mathrm{ul}}^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \widehat{\mathbf{h}}_k. \tag{1}$$

2) Distributed Partial MMSE: In the distributed operation (Level 3 and Level 2), each AP selects its combining vector locally. This cannot be done in an optimal way, since the AP is only aware of its local CSI. However, inspired by the centralized MMSE solution, the local partial (LP)-MMSE combining vector [2], [3] is given by

$$\mathbf{v}_{kl}^{\mathrm{LP-MMSE}} = p_k \left(\sum_{i \in \mathcal{D}_l} p_i (\widehat{\mathbf{h}}_{il} \widehat{\mathbf{h}}_{il}^{\mathrm{H}} + \mathbf{C}_{il}) + \sigma_{\mathrm{ul}}^2 \mathbf{I}_N \right)^{-1} \mathbf{D}_{kl} \widehat{\mathbf{h}}_{kl}. \tag{2}$$

D. Spectral Efficiency

Different lower bounds on the achievable spectral efficiency (SE) of a cell-free massive MIMO system exist [13]. If the decoder is assumed to not be aware of the instantaneous CSI, as is the case with Level 2 and 3, a bound commonly referred to as the Use-and-then-Forget (UatF) bound can be applied and is given by [13, Th. 5.2]

$$\mathrm{SE}_k^{(\mathrm{ul,UatF})} = \frac{\tau_u}{\tau_c}\log_2\left(1+\mathrm{SINR}_k^{(\mathrm{ul,UatF})}\right) \tag{3}$$

with the signal-to-interference-plus-noise ratio (SINR) given by $\mathrm{SINR}_k^{(\mathrm{ul,UatF})} =$

$$\frac{p_k\left|\mathbf{a}_k^{\mathrm{H}}\mathbb{E}\{\mathbf{g}_{kk}\}\right|^2}{\mathbf{a}_k^{\mathrm{H}}\left(\sum_{i=1}^{K}p_i\mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^{\mathrm{H}}\right\}-p_k\mathbb{E}\{\mathbf{g}_{kk}\}\mathbb{E}\{\mathbf{g}_{kk}^{\mathrm{H}}\}+\mathbf{F}_k\right)\mathbf{a}_k} \tag{4}$$

and

$$\mathbf{F}_k = \sigma_{\mathrm{ul}}^2\mathrm{diag}\left(\mathbb{E}\left\{\|\mathbf{D}_{k1}\mathbf{v}_{k1}\|^2\right\},\ldots,\mathbb{E}\left\{\|\mathbf{D}_{kL}\mathbf{v}_{kL}\|^2\right\}\right). \tag{5}$$

The UatF bound is also valid for Level 4 and 1, but may underestimate the performance, since the decoder is aware of the instantaneous CSI in these cases. This is especially the case for Level 1, where there is a low degree of channel hardening. Hence, for Level 4 and 1, we use the more optimistic SE expression given by [13, Th. 5.1]

$$\mathrm{SE}_k^{(\mathrm{ul,tight})} = \frac{\tau_u}{\tau_c}\mathbb{E}\left\{\log_2\left(1+\mathrm{SINR}_k^{(\mathrm{ul,tight})}\right)\right\} \tag{6}$$

$$\mathrm{SINR}_k^{(\mathrm{ul,tight})} = \frac{p_k\left|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_k\right|^2}{\sum_{\substack{i=1\\i\neq k}}^{K}p_i\left|\mathbf{v}_k^{\mathrm{H}}\mathbf{D}_k\widehat{\mathbf{h}}_i\right|^2+\mathbf{v}_k^{\mathrm{H}}\mathbf{Z}_k\mathbf{v}_k+\sigma_{\mathrm{ul}}^2\|\mathbf{D}_k\mathbf{v}_k\|^2}. \tag{7}$$

The expectation in Eq. 6 is with respect to the channel realizations. The matrix $\mathbf{C}_i$ denotes the correlation matrix of the channel estimation errors and $\mathbf{Z}_k = \sum_{i=1}^{K}p_i\mathbf{D}_k\mathbf{C}_i\mathbf{D}_k$.

IV. Large-Scale Fading Decoding

In the distributed operation, every AP computes its local signal estimates for all UEs it is serving and sends those estimates to the CPU for detection. The local estimates for each UE are combined by summing them with the LSFD weights given by the vector $\mathbf{a}_k$. In this section, we will introduce different heuristics (Eqs. (9) to (11)) to compute the LSFD weights in a distributed manner, which is the main contribution of our work. We first describe how to find the optimal LSFD weights in a centralized manner (Level 3), then we propose different heuristics to compute them locally at the APs (Level 2).

A. (Nearly) Optimal Centralized Solution

To achieve optimum performance in Level 3 operation, the LSFD vector can be computed by an MMSE estimator that considers the effective channel statistics of all UEs [13, Sec. 5.2.1]. A scalable ("nearly optimal") variant of this scheme uses only the UEs from the set $\mathcal{S}_k$, yielding the LSFD weights [7]

$$\mathbf{a}_k^{\mathrm{n-opt}} = p_k\left(\sum_{i\in\mathcal{S}_k}p_i\mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^{\mathrm{H}}\right\}+\mathbf{F}_k+\tilde{\mathbf{D}}_k\right)^{-1}\mathbb{E}\{\mathbf{g}_{kk}\}, \tag{8}$$

where $\mathbf{F}_k$ is defined as in Eq. 5 and $\tilde{\mathbf{D}}_k$ is a diagonal matrix with the (l,l)th element being one if $l \notin \mathcal{M}_k$ and zero otherwise.

B. Heuristic Distributed Solutions

As described in Sec. II, any scheme where the LSFD weights are computed locally at the APs is considered Level 2 operation. As the originally proposed trivial weighting vector $\mathbf{a}_k = [1,\ldots,1]^{\mathrm{T}}$ was shown to perform poorly [3], we will now derive different heuristic schemes to compute good LSFD weights a_{kl} locally at each AP l.

One intuitive approach to such a scheme might be to start from the (nearly) optimal LSFD solution and simply neglect all interference terms of the other UEs, since they require centralized statistical knowledge. By omitting the matrix inverse in Eq. 8, one arrives at the LSFD weight selection

$$\mathbf{a}_k^{\mathrm{MR}} = \mathbb{E}\{\mathbf{g}_{kk}\}. \tag{9}$$

Since this approach is reminiscent of maximum ratio (MR) combining, where the power of the desired signal is maximized, but all interference is neglected, we call it MR-LSFD in this work.

Another approach would be to directly consider the reliability of the signal estimates, i.e., the SINR, in the LSFD weight computation. The exact computation of the SINR requires the knowledge of the effective channel statistics $\mathbb{E}\{\mathbf{g}_{kk}\}$, however a simpler approximation, that only relies on statistical information about the propagation environment, is desirable. We approximate the power that is received from UE k at AP l by the term $\beta_{kl}p_k$, thus a heuristic selecting the LSFD weights based on the approximate SINR is given by

$$a_{kl}^{\beta-\mathrm{SINR}} = \frac{\beta_{kl}p_k}{\sum_{\substack{i\in\mathcal{D}_l\\i\neq k}}\beta_{il}p_i+\sigma_{\mathrm{ul}}^2}. \tag{10}$$

We denote this heuristic as β-SINR-LSFD. Note, that for scalability reasons, we only sum over the set $\mathcal{D}_l$ of UEs that are served by AP l.

Alternatively, the LSFD weighting heuristic could simply ignore the interference power, since the local combining scheme is already expected to suppress the multi-user interference when computing the local estimates. Then, the sum in the denominator of Eq. 10 can be dropped. The UE power term and the noise power can then also be dropped from the equation, because they are equal for all LSFD weights influencing the same UE. Hence, we compute the LSFD weights as

$$a_{kl}^{\beta^\nu} = \beta_{kl}^{\nu}, \tag{11}$$

where the exponent ν was introduced to be able to adjust the influence of the large-scale fading coefficients: By setting $\nu = 0$, all the weights become equal to 1, i.e., the trivial LSFD weighting vector is recovered and can thus be regarded as a special case of this scheme. With $\nu = 1$, the local estimates are weighted by the expected received power from the UE at the respective APs. Values of $\nu \in (0,1)$ represent a trade-off between both approaches, where the expected received power is considered, but its influence is limited. Finally, setting $\nu > 1$ means that the weighting will be biased towards strong signals even more than in the $\nu = 1$ case. We call this approach β^ν-LSFD in the following. Note, that it has been mentioned in a footnote in [3], but no results were provided.

V. Numerical Evaluation

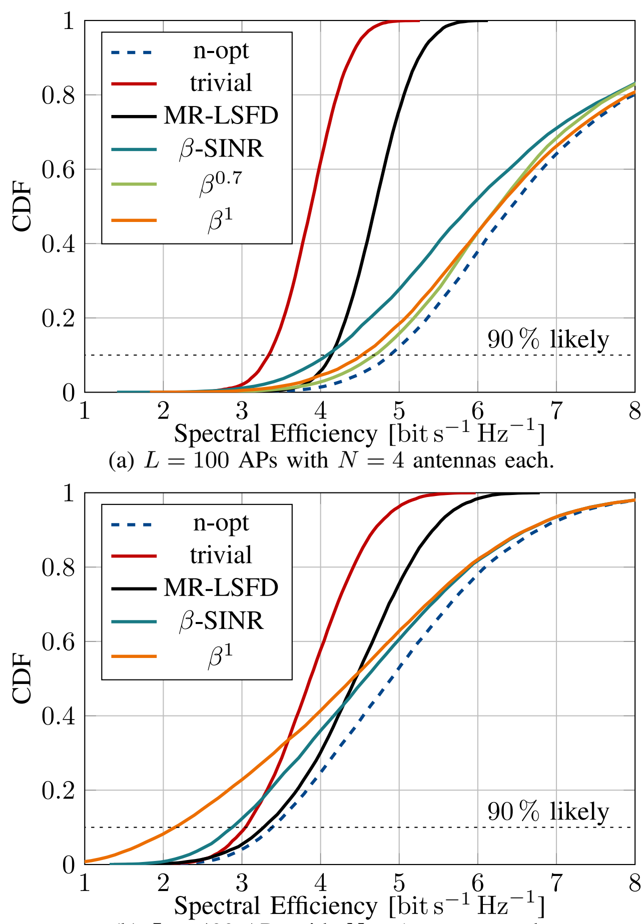
We evaluate the performance of the proposed LSFD heuristics compared to the other levels of operation in a scalable cell-free massive MIMO system using Monte-Carlo simulations. We consider a system with a total of $M = 400$ antennas, that are either deployed on $L = 100$ APs with $N = 4$ antennas each, or on $L = 400$ APs with $N = 1$ antenna each. The APs are arranged on an evenly spaced square grid in an area of $1\,\text{km}^2$. A simulation consists of 300 different drops, where in each drop $K = 40$ UEs are placed uniformly at random in the area. For each drop, the expectations required to compute the SINRs with Eq. 4 are estimated by averaging over 100 channel realizations. All UEs transmit with the same power $p_k = 1\,\text{mW}$, the uplink noise power is set to $\sigma_{\text{ul}}^2 = -96\,\text{dBm}$. The pathloss between UE k and AP l is modeled as

$$\beta_{kl}[\text{dB}] = -30.5 - 36.7 \log_{10}\left(\frac{d_{kl}}{1\,\text{m}}\right), \tag{12}$$

where d_{kl} is the 3D-distance in m, taking a height difference of $10\,\text{m}$ into account. The channel is modeled as constant over the coherence block length of $\tau_c = 200$ symbols out of which $\tau_p = 10$ are used for pilots. Since we only consider the uplink, we set $\tau_d = 0$ and $\tau_u = 190$.

A. Comparison of LSFD Heuristics

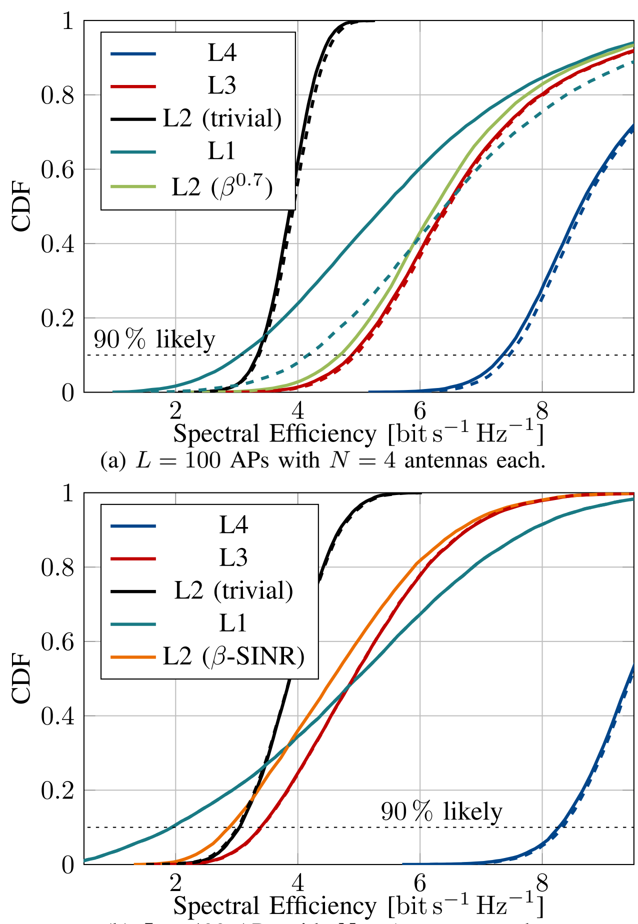
A core goal of cell-free massive MIMO is to improve the service especially for the most unfortunate UEs, which is reflected in the SE achieved by 90 % of the UEs, as marked in the figures. In Fig. 2, we compare the performance of the different LSFD heuristics for the fully distributed (Level 2) operation. The dashed line represents the performance of the nearly-optimal LSFD for Level 3 operation as a benchmark. For the deployment with $N = 4$ antennas per AP (Fig. 2a), clearly all



(a) $L = 100$ APs with $N = 4$ antennas each.

(b) $L = 400$ APs with $N = 1$ antenna each.

Fig. 2: Comparison of the different heuristic schemes for fully distributed (Level 2) LSFD in a scalable system. The dashed line represents partially distributed (Level 3) operation, using the nearly optimal LSFD weights, as a benchmark.



(a) $L = 100$ APs with $N = 4$ antennas each.

(b) $L = 400$ APs with $N = 1$ antenna each.

Fig. 3: Comparison of the performance of the four levels (L1-L4) of operation for scalable (solid lines) and unscalable (dashed lines) implementations. Additionally, an improved Level 2 system with the best-performing heuristic is shown.

of the proposed LSFD heuristics outperform the trivial LSFD scheme. Here, MR-LSFD achieves good performance for the worst UEs, but cannot compete with the other schemes for the best UEs. The best performing scheme is β^ν-LSFD for exponents close to $\nu = 1$, for which the performance is very close to the benchmark. For $\nu = 0.7$, the curve is slightly steeper than for $\nu = 1.0$, meaning that the performance of the worst UEs is improved at the cost of the best UEs. This might be desirable in terms of fairness, but not in terms of the sum-rate in the network. For visibility in Fig. 2, we present results only for $\nu = 0.7$ and $\nu = 1.0$. Nonetheless, we found that for small $\nu < 0.5$, the best UEs lose a substantial amount of performance, while the worst UEs do not gain much. For $\nu > 1$, the performance of the best UEs slightly improve, but at a large cost for the worst UEs. The β-SINR-LSFD shows worse performance in this deployment, but still outperforms the trivial LSFD scheme.

For the deployment with $N = 1$ antenna per AP (Fig. 3b), the performance of the heuristics changes: Now, β-SINR-LSFD achieves the best performance for the best UEs, but gets outperformed by MR-LSFD for the worst UEs. The β^ν-LSFD scheme shows substantial losses in performance for the worst UEs. None of the heuristics achieve close to optimal performance for all UEs for $N = 1$.

The difference in performance between the two deployments can be explained by considering the differences in spatial directivity that is present locally in each AP. For $N = 4$, even a locally selected combining vector can effectively cancel the interference, while there is no such directivity for $N = 1$. Since the β-SINR-LSFD scheme puts emphasis on the interference for weight selection, it performs better for $N = 1$ in comparison to $N = 4$. Note, that the overall performance of Levels 2 and 3 in the $N = 1$ case is substantially worse than in the $N = 4$ case, while the performance of the Level 4 operation is improved. Splitting up the antennas into more geographically distributed APs increases the spatial diversity, which can be exploited by the optimal centrally selected combining vector in Level 4 operation. In contrast, for Level 2 and 3 operation, the locally selected combining vectors are less effective, the fewer antennas are available per AP.

B. Comparison of Operation Levels

In [3], the four levels of operation were compared without taking scalability into account. We show how scalability affects the performance for each level for both deployments in Fig. 3 by comparing their scalable (with clustering) and unscalable (without clustering, i.e., each UE is served by all APs) implementations. The performance losses due to the clustering are minor for all cases except for Level 1 operation with $L = 100$, $N = 4$. This is because the unscalable implementation of Level 1 operation assumes every AP knows the channels of UEs it does not serve, i.e., UEs in other cells. For $N = 1$ the Level 1 system performs poorly either way, but for $N = 4$ the increased spatial directivity allows each AP to cancel a large amount of the inter-cell interference. This means that for deployments with multiple antennas per AP, our improved cell-free Level 2 system substantially outperforms the Level 1 small-cell setup, while still being low in complexity and fronthaul requirements. This stands in contrast to the previous belief that only more complex implementations of cell-free massive MIMO are competitive with small-cell systems.

VI. Conclusion

In this work we put the scalable fully distributed cell-free massive MIMO operation (Level 2) into a new light. We first improve the performance of such a network by introducing new heuristics to find the LSFD weights. Our results show, that for multi-antenna APs, the best heuristic achieves close to optimal performance. While a fully centralized (Level 4) system can achieve much better performance at the cost of much higher complexity, the fully distributed (Level 2) system still achieves substantial performance gains over a conventional small-cell (Level 1) system when taking scalability into account.

References

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

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

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

[4] S. Chakraborty, Ö. T. Demir, E. Björnson, and P. Giselsson, "Efficient downlink power allocation algorithms for cell-free massive MIMO systems," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 168–186, 2021.

[5] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6.

[6] M. Zaher, Ö. T. Demir, E. Björnson, and M. Petrova, "Learning-based downlink power allocation in cell-free massive MIMO systems," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[7] O. T. Demir, E. Bjoernson, and L. Sanguinetti, "Cell-free massive MIMO with large-scale fading decoding and dynamic cooperation clustering," in *WSA 2021; 25th International ITG Workshop on Smart Antennas*, Nov. 2021, pp. 1–6.

[8] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*. 2017, vol. 11.

[9] L. Sanguinetti, M. Morelli, and L. Marchetti, "A random access algorithm for LTE systems," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 1, pp. 49–58, 2013.

[10] S. Chen, J. Zhang, E. Björnson, J. Zhang, and B. Ai, "Structured massive access for scalable cell-free massive MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1086–1100, Apr. 2021.

[11] M. Mussbah, S. Schwarz, and M. Rupp, "Access point clustering-based pilot assignment for cell-free massive MIMO," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*, Oct. 2022, pp. 722–726.

[12] F. Göttsch, N. Osawa, T. Ohseki, *et al.*, "User-centric clustering under fairness scheduling in cell-free massive MIMO," in *2023 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2023, pp. 292–297.

[13] Ö. T. Demir, E. Björnson, and L. Sanguinetti, *Foundations of user-centric cell-free massive MIMO*, Feb. 22, 2022.

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