Clustering-Based Activity Detection Algorithms for Grant-Free Random Access in Cell-Free Massive MIMO

Unnikrishnan Kunnath Ganesan, *Graduate Student Member, IEEE*, Emil Björnson, *Senior Member, IEEE*, and Erik G. Larsson, *Fellow, IEEE*

***Abstract*— Future wireless networks need to support massive machine type communication (mMTC) where a massive number of devices accesses the network and massive MIMO is a promising enabling technology. Massive access schemes have been studied for co-located massive MIMO arrays. In this paper, we investigate the activity detection in grant-free random access for mMTC in cell-free massive MIMO networks using distributed arrays. Each active device transmits a non-orthogonal pilot sequence to the access points (APs) and the APs send the received signals to a central processing unit (CPU) for joint activity detection. The maximum likelihood device activity detection problem is formulated and algorithms for activity detection in cell-free massive MIMO are provided to solve it. The simulation results show that the macro diversity gain provided by the cell-free architecture improves the activity detection performance compared to co-located architecture when the coverage area is large.**

***Index Terms*— Activity detection, grant-free random access, cell-free massive MIMO, massive machine-type communications (mMTC), Internet-of-Things (IoT).**

I. Introduction

THE data traffic in wireless networks has grown tremendously in the last decade. There is a growing consensus that the future wireless networks should support three generic services namely enhanced mobile broadband (eMBB), massive machine type communications (mMTC) and ultra-reliable low latency communications (URLLC). URLLC and mMTC [2], [3] are two key features of Internet-of-Things (IoT) services envisioned in 5G and beyond 5G wireless networks [3], [4]. URLLC is a critical IoT service for time-critical communications which enables ultra-high reliability and/or ultralow latency at a variety of data rates. MMTC in IoT is meant to support a massive number of low-cost devices which carry very small data packets and require extreme coverage. Massive MIMO for IoT connectivity is still a developing topic and in this paper, we study the cell-free massive MIMO networks for mMTC services. One of the main challenges of mMTC is that the network should be able to support a large number of devices over the same time and frequency resources while keeping the battery lives of the devices as long as possible. Massive multiple input multiple output (MIMO) was shown to be a promising technology to support massive access in [4], [5] by exploiting the spatial degrees of freedom available in the network to let many users transmit simultaneously.

Conventional grant-based massive random access schemes are studied in [6]–[8]. In the grant-based approach, each active device randomly picks a pilot sequence from a shared pool of orthogonal sequences, and uses the selected sequence to inform the base station that it has data to transmit. The base station needs to resolve the collisions that occur and a grant of resources will be provided to selected devices based on collision resolution. In wireless systems, the channel coherence interval is limited and hence, the set of orthogonal pilot sequences is finite. Grant-based protocols permit simple signal processing at the base station. A key feature of mMTC is that the traffic is sporadic with high risk of collision of potentially active users and with very small payloads. Thus, in the massive random access scenario, the probability of multiple active devices selecting the same sequence is quite high. Thus, the grant-based random access protocols suffer from access failure due to collisions, and hence, increase the average latency. Moreover, to resolve the collisions, the signaling overhead is quite large compared to the short payload each device has to send in mMTC applications. Authors in [9] propose a neighbor-aware multiple access protocol that improves system throughput by exploiting the broadcasted acknowledgment signals in the network. However, considering the average latency incurred and short payloads, it is inefficient to use conventional grant-based access methods for mMTC.

To overcome the limitations of grant-based random access schemes, various grant-free protocols [10] have been proposed for the active devices to access the wireless network without a grant. The grant-free approach reduces the access latency

Manuscript received January 13, 2021; revised May 28, 2021; accepted July 25, 2021. Date of publication August 5, 2021; date of current version November 18, 2021. Unnikrishnan Kunnath Ganesan and Erik G. Larsson were supported in part by ELLIIT and in part by Swedish Research Council (VR). Emil Björnson was supported by the Grant 2019-05068 from the Swedish Research Council. This article was presented at 21st IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2020) [1]. The associate editor coordinating the review of this article and approving it for publication was M. Bhatnagar. *(Corresponding author: Unnikrishnan Kunnath Ganesan.)*

Unnikrishnan Kunnath Ganesan and Erik G. Larsson are with the Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden (e-mail: unnikrishnan.kunnath.ganesan@liu.se; erik.g.larsson@liu.se).

Emil Björnson is with the Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden, and is also with the KTH Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: emil.bjornson@liu.se).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3102635.

Digital Object Identifier 10.1109/TCOMM.2021.3102635

and signaling overhead compared to grant-based approaches at the expense of sophisticated signal processing at the base station. In the grant-free random access scheme, each device is assigned a unique pilot sequence and the active devices access the network using this sequence. To enable data transmission for mMTC, channel estimates are required for which identification of active users is crucial. Thus, in the paper, we focus on the activity detection problem in a grant-free random access scheme. To support a massive number of devices in mMTC and due to the limited channel coherence interval length, assigning orthogonal pilot sequences to each device is not feasible. Hence, the users are assigned with unique non-orthogonal pilot sequences and hence the received signal at the base station suffer from severe co-channel interference. Thus, the activity detection is a challenging problem in the grant-free massive random access schemes. The sparse nature of the device activity pattern helps to formulate the activity detection problem as a compressive sensing (CS) problem [11], [12] and greedy pursuit algorithms were leveraged to solve it [13]–[15], with the assumption that the number of active users is known a priori. Bayesian inference based algorithms like approximate message passing (AMP) are utilized in [16]–[20] which are computationally efficient for activity detection. However, the performance of CS based algorithms degrade severely when the number of active devices is larger than the pilot sequence length [18], [21]. Random and structured sparsity learning based multi-user detection was studied in [22]. Deep learning based activity detection approaches are considered in [23], [24].

A covariance-based approach is proposed in [25] for device activity detection which performs better than the existing CS based schemes. The covariance based approach overcomes the limitation in traditional CS based techniques where the number of active devices are required to be smaller than the pilot sequence length. Joint activity and data detection using a covariance-based approach is proposed in [26] and the performance analysis of the covariance-based approach is provided showing the superiority over the AMP based approach by exploiting the asymptotic properties of maximum likelihood estimator. The covariance based method makes better use of the multiple antennas to improve detection accuracy compared to AMP based CS approaches. The activity detection in unsourced random access [27] where all devices use the same codebook is studied in [28] and is shown to have high spectral efficiencies when a covariance based recovery algorithm is employed at the receiver.

To provide a uniform and high per-user data rates in future wireless networks, densification of the network infrastructure by increasing the number of antennas per cell and deploying many access points (APs), is considered. However, network densification increases the inter-cell interference. Towards this, cell-free massive MIMO was proposed wherein the concept of cell is removed and all device communicates with all the APs. In the cell-free architecture, all the APs are connected to a central processing unit (CPU) through a fronthaul as shown in Fig. 1. The CPU is responsible for jointly processing data and serving all the users in the network. Cell-free massive MIMO was shown to be a promising approach to overcome inter-cell interference limitation posed by network densification [29], [30]. In cell-free architecture, the path loss from a device to different APs differ by many orders of magnitude. In [31], grant-free access in cell-free architecture is studied with the assumption of ideal favorable propagation in massive MIMO systems [32], [33]. Favorable propagation in cell-free networks is more likely to occur when more APs are in the systems and the users are spatially well-separated, which cannot be guaranteed in general, and ideal favorable propagation is unlikely.

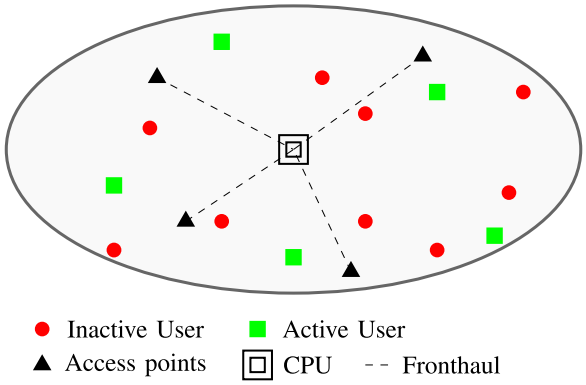

Fig. 1. Cell-Free Network Model for mMTC.

A. Contributions

In this paper, we investigate and develop algorithms for activity detection for grant-free random access in cell-free massive MIMO networks without any assumptions on the network geometry leveraging the covariance-based approach. Most of the existing literature studies random access in a co-located scenario in single and multi-cell environments [34], [35]. In [36, Sec. 3.5], Fengler proposes an activity detection algorithm for cell-free setup. In Fengler's algorithm, each AP decodes the activity pattern of all users and shares their estimates with each other in every iteration. After each iteration, thresholding is done considering the estimates from all APs and the devices are marked as inactive which did not meet threshold criteria by at least one AP. With subsequent iterations, the false alarm rate improves but the missed detection rate does not improve much. In our work, we consider a joint activity detection at the CPU, which improves the activity detection performance in terms of false alarm and missed detection rates. Cell-free MIMO networks can provide better coverage for mMTC due to the shorter propagation distances and are more robust to the shadow fading effects compared to co-located MIMO at the expense of the increased cost of the fronthaul infrastructure. We show that with sophisticated signal processing at the CPU, cell-free networks can obtain better activity detection performance compared to co-located MIMO networks. This reduces the energy consumption of mMTC devices as the fraction of time the device needs to be active becomes shorter. Thereby, keeping the battery life of mMTC devices as long as possible.

For activity detection in a cell-free network, the optimal method would be, if all APs are contributing to the activity detection for all users. But this is unnecessarily computationally complex. Since only a few APs are close to each user, we consider a cluster of APs with good channels to the user. In [1], we considered a special case where the cluster size was 1. In this paper, we develop algorithms to improve the activity detection performance by considering the information from a

cluster of dominant APs. Thus, we propose novel clustering-based activity detection algorithms utilizing information from a cluster of dominant APs for activity detection, and simulations show that the performance improves with the clustering-approach.

The main contributions of this paper can be summarized as follows:

- We investigate the grant-free massive random access in cell-free massive MIMO networks and formulate the maximum likelihood activity detection problem leveraging the covariance-based approach.
- We study the SNR achieved at the APs for cell-free massive MIMO networks in different deployment scenarios considering shadow fading, dense deployment of APs, and cell area for power limited mMTC devices. We show that the outage probability is less in cell-free networks compared to co-located networks. Hence, cell-free networks provide broad coverage and enable IoT services for many devices.
- We propose an algorithm for device activity detection based on the single dominant AP for each device for cell-free massive MIMO networks. This algorithm was presented in the conference version [1].
- We propose a novel clustering-based activity detection algorithm for activity detection which improves the activity detection performance with the cost of sophisticated signal processing at the CPU.
- We quantify the complexity involved in the detection process and by introducing clustering, we can identify a suitable trade-off between the performance and complexity.
- We propose a novel algorithm which reduces the computational time required for the activity detection, based on clustering and parallelism by exploiting the sparse nature of activity pattern.
- Our simulations show that the cell-free massive MIMO networks can provide better activity detection performance in an mMTC scenario compared to a co-located network. They also show that activity detection performance can be further improved by clustering-approach.
- We study the impact of a capacity-limited fronthaul on the activity detection performance and show that we can achieve a performance similar to lossless fronthaul with contemporary fronthaul technology.

The organization of the rest of the paper is as follows. The signal model and the activity detection problem formulation is explained in Sec. II. The activity detection algorithms are provided in Sec. III. Sec. IV provides the simulation results which shows the performance of the proposed algorithms and Sec. V provides the concluding remarks.

Reproducible research: All the simulation results can be reproduced using the Matlab code available at: https://github.com/emilbjornson/grant-free.

Notations: Bold, lowercase letters are used to denote vectors and bold, uppercase letters are used to denote matrices. $\Re(\cdot)$ and $\Im(\cdot)$ denotes the real and imaginary parts, respectively. $\mathbb{R}$ and $\mathbb{C}$ denote the set of real and complex numbers respectively. The operations $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transpose and Hermitian transpose, respectively. $\mathcal{CN}(0,\sigma^2)$ denotes a circularly symmetric complex Gaussian random variable with zero mean and variance equal to σ^2. $\mathbf{I}_N$ and $\mathbf{0}_N$ represent the $N \times N$ identity matrix and $N \times 1$ zero vector, respectively. The operation $|\cdot|$ denotes the determinant. The operation $|\cdot|_{\mathrm{c}}$ denotes the cardinality of set. $\mathrm{diag}(\mathbf{a})$ represents a diagonal matrix with diagonal entries are elements from $\mathbf{a}$.

II. Signal Model and Problem Formulation

Consider a cell-free massive MIMO wireless network as illustrated in Fig. 1 with M arbitrarily distributed APs each equipped with N antennas and serving K arbitrarily distributed single antenna users. All the M APs are assumed to be connected to a CPU through a lossless infinite capacity fronthaul. Due to the sporadic nature of the traffic in the massive access scenario of mMTC, only a small fraction of the K users are active at any given time instant. In this paper, we assume that each device transmits independently with an activation probability $\epsilon \ll 1$. Let $a_k \in \{0,1\}$ where $a_k = 1$ denotes that the k^{th} device is active and $a_k = 0$ that it is inactive and $\Pr(a_k = 1) = \epsilon$ and $\Pr(a_k = 0) = 1 - \epsilon$. Let $\mathbf{a} = (a_1, a_2, \cdots, a_K)$ denote the activity of K users at any time instant. Due to the sporadic nature of mMTC traffic, the vector $\mathbf{a}$ will be sparse. The set of active users is denoted by $\mathcal{A}$ i.e., $\mathcal{A} = \{k : a_k = 1\}$.

We consider that all the APs and the devices are geographically separated from each other and hence the channels between the devices and the APs can be considered independent. Moreover, we consider that the N antennas at each AP are sufficiently separated to have independent fading between them. Hence, the channel gain between the n^{th} antenna in the m^{th} AP to device k is given by

$$g_{mnk} = \beta_{mk}^{\frac{1}{2}} h_{mnk} \tag{1}$$

where β_{mk} is the large-scale fading coefficient between the m^{th} AP and the user k and $h_{mnk} \sim \mathcal{CN}(0,1)$ is the small-scale fading coefficient. We assume that the large-scale fading coefficient parameters $\{\beta_{mk}\}$ are known at the CPU [37], [38]. Throughout this paper, we consider a block fading scenario where each channel remains constant during a coherence interval [32, Ch.2] and all the channels are independently distributed. Let τ_c be the number of channel uses per coherence interval. Due to the large number of users, typically $K \gg \tau_c$, assigning orthogonal pilot sequences to each user is not feasible. Instead we assign non-orthogonal unique signature sequence, $\mathbf{s}_k \in \mathbb{C}^{L\times 1}$ to each user k, where $L \leq \tau_c$ is the signature sequence length. We assume that the signature sequences of all the users are known at the CPU. Moreover, we assume that all the devices are synchronized during the transmission, which means in an orthogonal frequency division multiplexing system, the time delays from the different devices are well within the cyclic prefix.

The signal $\mathbf{y}_{mn} \in \mathbb{C}^{L\times 1}$ received at the n^{th} antenna of the m^{th} AP is given by

$$\begin{aligned}\mathbf{y}_{mn} &= \sum_{k=1}^{K} a_k \rho_k^{\frac{1}{2}} g_{mnk} \mathbf{s}_k + \mathbf{w}_{mn} \\ &= \mathbf{S}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}}\mathbf{g}_{mn} + \mathbf{w}_{mn},\end{aligned} \tag{2}$$

where $\mathbf{S} = [\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_K] \in \mathbb{C}^{L\times K}$ is the collection of all signature sequences, ρ_k is the power transmitted by user

k, $\mathbf{D_a} = \text{diag}(\mathbf{a})$, $\mathbf{D}_{\boldsymbol{\rho}} = \text{diag}(\rho_1, \rho_2, \ldots, \rho_K)$, $\mathbf{g}_{mn} = [g_{mn1}\ g_{mn2}\ \ldots\ g_{mnK}]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ is the channel vector from all K users to the n^{th} antenna of the m^{th} AP and $\mathbf{w}_{mn} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_L)$ is the independent additive white Gaussian noise vector.

Thus, the signal $\mathbf{Y}_m \in \mathbb{C}^{L\times N}$ received at the m^{th} AP can be expressed as

$$\mathbf{Y}_m = \mathbf{S}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}}\mathbf{G}_m + \mathbf{W}_m, \tag{3}$$

where $\mathbf{G}_m = [\mathbf{g}_{m1}\ \mathbf{g}_{m2}\ \ldots\ \mathbf{g}_{mN}] \in \mathbb{C}^{K\times N}$ is the channel matrix between the K users and the m^{th} AP and $\mathbf{W}_m = [\mathbf{w}_{m1}\ \mathbf{w}_{m2}\ \ldots\ \mathbf{w}_{mN}] \in \mathbb{C}^{L\times N}$ is the noise matrix.

Let the collection of signals be

$$\begin{aligned}\mathbf{Y} &= \begin{bmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\\ \vdots\\ \mathbf{Y}_M\end{bmatrix} = \begin{bmatrix}\mathbf{S}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}}\mathbf{G}_1\\ \mathbf{S}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}}\mathbf{G}_2\\ \vdots\\ \mathbf{S}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}}\mathbf{G}_M\end{bmatrix} + \mathbf{W}\\ &= \begin{bmatrix}\mathbf{S} & \mathbf{0} & \ldots & \mathbf{0}\\ \mathbf{0} & \mathbf{S} & \ldots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{S}\end{bmatrix}\begin{bmatrix}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}} & \mathbf{0} & \ldots & \mathbf{0}\\ \mathbf{0} & \mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}} & \ldots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\rho}}^{\frac{1}{2}}\end{bmatrix}\begin{bmatrix}\mathbf{G}_1\\ \mathbf{G}_2\\ \vdots\\ \mathbf{G}_M\end{bmatrix} + \mathbf{W},\end{aligned} \tag{4}$$

where $\mathbf{W} = [\mathbf{W}_1^{\mathrm{T}}\ \mathbf{W}_2^{\mathrm{T}}\ \ldots\ \mathbf{W}_M^{\mathrm{T}}]^{\mathrm{T}}$. From (4), it can be seen that the columns of $\mathbf{Y}$ are independent and each column is distributed as $\mathbf{Y}(:,i) \sim \mathcal{CN}(\mathbf{0}_{LM}, \mathbf{Q})$, $\forall i = 1, 2, \ldots, N$, where $\mathbf{Q}$ is the covariance matrix given by

$$\mathbf{Q} = \begin{bmatrix}\mathbf{S}\mathbf{D}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\beta}_1}\mathbf{S}^{\mathrm{H}} & \mathbf{0}_L & \ldots & \mathbf{0}_L\\ \mathbf{0}_L & \mathbf{S}\mathbf{D}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\beta}_2}\mathbf{S}^{\mathrm{H}} & \ldots & \mathbf{0}_L\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0}_L & \mathbf{0}_L & \ldots & \mathbf{S}\mathbf{D}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\beta}_M}\mathbf{S}^{\mathrm{H}}\end{bmatrix} + \sigma^2\mathbf{I}_{LM}, \tag{5}$$

where $\mathbf{D}_{\boldsymbol{\beta}_m}$ is a diagonal matrix with diagonal elements corresponding to the large-scale fading coefficient from K users to m^{th} AP, i.e., $\mathbf{D}_{\boldsymbol{\beta}_m} = \text{diag}(\boldsymbol{\beta}_m)$ where $\boldsymbol{\beta}_m = (\beta_{m1}, \beta_{m2}, \ldots, \beta_{mK})$ and $\mathbf{D}_{\boldsymbol{\gamma}} = \text{diag}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = (a_1\rho_1, a_2\rho_2, \ldots, a_K\rho_K)$.

By utilizing the block-diagonal structure of the covariance matrix $\mathbf{Q}$, the likelihood of $\mathbf{Y}$ given $\boldsymbol{\gamma}$ is

$$\begin{aligned}p(\mathbf{Y}|\boldsymbol{\gamma}) &= \prod_{m=1}^{M}\prod_{n=1}^{N}\frac{1}{|\pi\mathbf{Q}_m|}\exp\left(-\mathbf{y}_{mn}^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{y}_{mn}\right)\\ &= \prod_{m=1}^{M}\frac{1}{|\pi\mathbf{Q}_m|^N}\exp(-\text{Tr}(\mathbf{Q}_m^{-1}\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}})),\end{aligned} \tag{6}$$

where $\mathbf{Q}_m = \mathbf{S}\mathbf{D}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\beta}_m}\mathbf{S}^{\mathrm{H}} + \sigma^2\mathbf{I}_L$. The maximum likelihood estimate of $\boldsymbol{\gamma}$ can be found by maximizing $p(\mathbf{Y}|\boldsymbol{\gamma})$ or equivalently minimizing $-\log(p(\mathbf{Y}|\boldsymbol{\gamma}))$ which is given by

$$\begin{aligned}\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}} &\sum_{m=1}^{M}\log|\mathbf{Q}_m| + \text{Tr}\left(\mathbf{Q}_m^{-1}\frac{\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}}{N}\right)\\ &\text{subject to } \boldsymbol{\gamma} \geq \mathbf{0}_K.\end{aligned} \tag{7}$$

To perform the activity detection, all the received signals at the APs need to be passed to the CPU for $L \geq N$. When $L < N$, AP m sends the sample covariance $\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}$ to the CPU to reduce the fronthaul usage. The CPU needs to solve the optimization problem in (7). For $M = 1$, the co-located architecture case, a covariance-based coordinate descent algorithm is proposed in [25] for device activity detection. However, for a cell-free architecture, due to the presence of $M > 1$ summation terms in (7), the brute force approach to solve (7) requires huge complexity and the complexity increases exponentially with M. In Sec. III we develop algorithms for the device activity detection that has affordable complexity, while making use of information obtained at multiple APs.

III. Device Activity Detection

In this section, we study the cost function (7) and exploit the features of cell-free architecture to develop algorithms for activity detection in grant-free random access schemes.

A. Coordinate Descent Cost Function

Let

$$f(\boldsymbol{\gamma}) = \sum_{m=1}^{M}\log|\mathbf{Q}_m| + \text{Tr}\left(\mathbf{Q}_m^{-1}\frac{\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}}{N}\right) \tag{8}$$

be the cost function which needs to be minimized in (7). Define

$$f^m(\boldsymbol{\gamma}) = \log|\mathbf{Q}_m| + \text{Tr}\left(\mathbf{Q}_m^{-1}\frac{\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}}{N}\right) \tag{9}$$

be the cost function associated with the m^{th} block in (8). Then we can write $f(\boldsymbol{\gamma}) = \sum_{m=1}^{M} f^m(\boldsymbol{\gamma})$. Setting $\mathbf{Q}_m$ as a function of $\boldsymbol{\gamma}$, i.e.,

$$\mathbf{Q}_m(\boldsymbol{\gamma}) = \mathbf{S}\mathbf{D}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\beta}_m}\mathbf{S}^{\mathrm{H}} + \sigma^2\mathbf{I}_L \tag{10}$$

$$= \sum_{k=1}^{K}\gamma_k\beta_{mk}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}} + \sigma^2\mathbf{I}_L, \tag{11}$$

we can see $\mathbf{Q}_m$ as a sum of K rank-one updates to $\sigma^2\mathbf{I}_L$. Thus, we can optimize $f(\boldsymbol{\gamma})$ with respect to one argument $\gamma_k, k \in \{1, 2, \ldots, K\}$ in one step and we iterate several times over the whole set of variables until the cost function cannot be further reduced. A random ordering is considered while optimizing to avoid dependency during detection if any. For $k \in \{1, 2, \ldots, K\}$, let us define $f_k^m(d) = f^m(\boldsymbol{\gamma} + d\mathbf{e}_k)$, where $\mathbf{e}_k$ is the k^{th} canonical basis with a single-1 at the k^{th} coordinate. By applying the Sherman-Morrison rank-1 update identity [39] on $\mathbf{Q}_m$, we obtain

$$\left(\mathbf{Q}_m + d\beta_{mk}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\right)^{-1} = \mathbf{Q}_m^{-1} - d\beta_{mk}\frac{\mathbf{Q}_m^{-1}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}}{1 + d\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k}. \tag{12}$$

By applying the determinant identity [40], we can obtain

$$|\mathbf{Q}_m + d\beta_{mk}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}| = (1 + d\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k)|\mathbf{Q}_m|. \tag{13}$$

Now we can write the overall maximum likelihood (ML) cost function in (8) for each coordinate k as $f_k(d) = \sum_{m=1}^{M} f_k^m(d)$, given by

$$f_k(d) = c + \sum_{m=1}^{M} \Bigg(\log(1 + d\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k) - d\beta_{mk}\frac{\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{Q}_{\mathbf{Y}_m}\mathbf{Q}_m^{-1}\mathbf{s}_k}{1 + d\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k} \Bigg), \quad (14)$$

where $c = \sum_{m=1}^{M} \left(\log|\mathbf{Q}_m| + \mathrm{Tr}(\mathbf{Q}_m^{-1}\mathbf{Q}_{\mathbf{Y}_m}) \right)$ is a constant and $\mathbf{Q}_{\mathbf{Y}_m} = \frac{\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}}{N}$. Taking the derivative of $f_k(d)$ with respect to d and equating to zero gives

$$f_k'(d) = \sum_{m=1}^{M} \Bigg(\frac{\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k}{(1 + d\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k)} - d\beta_{mk}\frac{\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{Q}_{\mathbf{Y}_m}\mathbf{Q}_m^{-1}\mathbf{s}_k}{(1 + d\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k)^2} \Bigg) = 0, \quad (15)$$

which is a polynomial of degree $2M - 1$. Hence, finding the value of d which minimizes (14) requires a complexity of $\mathcal{O}(M^4L^2)$ and involves solving higher degree polynomial equations. This huge complexity calls for a low complexity design to ensure scalability of the device activity detection in cell-free massive MIMO networks.

In the cell-free network, where the APs and devices are distributed over the cell area, the large-scale fading coefficients of the device varies significantly in magnitude between the different APs, unless the APs are equidistant from the device. This variation can be up to the order of 50 dB in the presence of shadow fading. Exploiting these variations in channel gain in the cell-free architecture, we propose algorithms for activity detection in the next subsections.

B. Dominant AP-Based Activity Detection

For a device k, let

$$m' = \underset{m}{\mathrm{argmax}}\{\beta_{mk}\} \quad (16)$$

be the index of the AP with which the device have the dominant large-scale fading coefficient and we call this AP the most dominant AP for the device k. In the proposed dominant AP based activity detection, the updates for any device is given by its corresponding dominant AP. Hence, at the CPU, we minimize the cost function with respect to the dominant AP for device k and the soft information about the device k from this AP is propagated to the other APs. The cost function of device k with respect to the dominant AP m' is given by

$$f_{k,m'}(d) = \log(1 + d\beta_{m'k}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k) - d\beta_{m'k}\frac{\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{Q}_{\mathbf{Y}_{m'}}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k}{1 + d\beta_{m'k}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k}. \quad (17)$$

Taking the derivative of (17) and equating it to zero, we obtain

$$d^* = \frac{\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{Q}_{\mathbf{Y}m'}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k - \mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k}{\beta_{m'k}(\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k)^2}. \quad (18)$$

Note that d^* is the minimizer of $f_{k,m'}(d)$, but need not be the minimizer of $f_k(d)$. To preserve the non-negativity of $\boldsymbol{\gamma}$ in (7), the optimal update step d is given by $\delta = \max\{d^*, -\gamma_k\}$ and the coordinate is updated as $\gamma_k \leftarrow \gamma_k + \delta$. Using (12), the update step d is propagated to all the sub covariance matrices $\mathbf{Q}_m$, $\forall m = 1, 2, \ldots, M$. This procedure will be done over the whole set of random permutation of variables from the set $\{1, 2, \ldots, K\}$ and we iterate the entire procedure until the cost function cannot be further reduced. The proposed algorithm is summarized in Algorithm 1. The complexity of the proposed algorithm based on dominant AP is $\mathcal{O}(IKML^2)$, where I is the maximum number of iterations. The term $\mathcal{O}(L^2)$ considers the matrix-vector multiplications in Algorithm 1.

Algorithm 1 Coordinate Descend Algorithm for Estimating $\boldsymbol{\gamma}$

Input: Observations $\mathbf{Y}_m, \forall m = 1, 2, \ldots M$, $\beta_{mk}, \forall m = 1, 2, \ldots M, k = 1, 2, \ldots K$

Initialize: $\mathbf{Q}_m^{-1} = \sigma^{-2}\mathbf{I}_L, \forall m = 1, 2, \ldots M$, $\hat{\boldsymbol{\gamma}}^0 = \mathbf{0}_K$

Compute $\mathbf{Q}_{\mathbf{Y}m} = \frac{1}{N}\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}, \forall m = 1, 2, \ldots M$
for $i = 1, 2, \ldots, I$ **do**
  Select an index set $\mathcal{K}$ from the random permutation of set $\{1, 2, \ldots, K\}$
  for $k \in \mathcal{K}$ **do**
    Find the strongest link or AP for device k, i.e., $m' = \mathrm{argmax}\{\beta_{mk}\}$
    $\delta = \max\left\{ \frac{\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{Q}_{\mathbf{Y}m'}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k - \mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k}{\beta_{m'k}(\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_{m'}^{-1}\mathbf{s}_k)^2}, -\hat{\gamma}_k \right\}$
    $\hat{\gamma}_k^i = \hat{\gamma}_k^{i-1} + \delta$
    for $m = 1, 2, \ldots, M$ **do**
      $\mathbf{Q}_m^{-1} \leftarrow \mathbf{Q}_m^{-1} - \delta\frac{\beta_{mk}\mathbf{Q}_m^{-1}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}}{1+\delta\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k}$
    end for
  end for
  if $f(\hat{\boldsymbol{\gamma}}^i) \geq f(\hat{\boldsymbol{\gamma}}^{i-1})$ **then**
    $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^{i-1}$
    break
  end if
  $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^i$
end for
return $\hat{\boldsymbol{\gamma}}$

To perform activity detection, the output from Algorithm 1 is compared against a threshold γ_k^{th} for each device k and is given by

$$\hat{a}_k = \begin{cases} 1, & \text{if } \hat{\gamma}_k \geq \gamma_k^{th} \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Let $\hat{\mathcal{A}} = \{k \mid \hat{a}_k = 1, \forall k \in [1, K]\}$ be the estimate of the set of active devices. The probability of miss detection is defined as the average of the ratio of non-detected devices and the number of active devices and the probability of false alarm is defined as the average of inactive devices declared active over inactive devices and are given respectively by

$$P_{md} = 1 - \mathbb{E}\left\{ \frac{|\mathcal{A} \cap \hat{\mathcal{A}}|_{\mathrm{c}}}{|\mathcal{A}|_{\mathrm{c}}} \right\}, P_{fa} = \mathbb{E}\left\{ \frac{|\hat{\mathcal{A}} \setminus \mathcal{A}|_{\mathrm{c}}}{K - |\mathcal{A}|_{\mathrm{c}}} \right\}. \quad (20)$$

The threshold γ_k^{th} is chosen to have a desired probability of miss detection and probability of false alarm performance.

C. Clustering Based Activity Detection

The activity detection in Algorithm 1 uses data from one dominant AP per device and the performance improves when more antennas are used at the AP [25], [26]. However, for activity detection in a cell-free network, the optimal method would be if all APs are contributing to the activity detection for all users, but this is unnecessarily computationally complex as mentioned in Sec. III-A. Since only a few APs are close to each user, we consider a cluster of APs with good channels to the user. In this subsection, we consider the minimization of the cost function in (14), by utilizing the received signals from a cluster of dominant APs for each device. Towards this, we define the function which returns the set of indices of the T maximum values from the set of real numbers $\mathcal{T}$, as

$$\underset{\cdot,T}{\text{indmax}}\{\mathcal{T}\}.$$

Note that the above function reduces to argmax, when $T = 1$.

Algorithm 2 Clustering Based Coordinate Descend Algorithm for Estimating $\boldsymbol{\gamma}$

Input: Observations $\mathbf{Y}_m, \forall m = 1, 2, \ldots M$, $\beta_{mk}, \forall m = 1, 2, \ldots M, k = 1, 2, \ldots, K$

Initialize: $\mathbf{Q}_m^{-1} = \sigma^{-2}\mathbf{I}_L, \forall m = 1, 2, \ldots, M$, $\hat{\boldsymbol{\gamma}}^0 = \mathbf{0}_K$

Compute $\mathbf{Q}_{\mathbf{Y}\,m} = \frac{1}{N}\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}, \forall m = 1, 2, \ldots, M$

Compute $\mathcal{M}_k = \text{indmax}_{m,T}\ \{\beta_{mk}\} \forall k = 1, 2, \ldots, K$

for $i = 1, 2, \ldots, I$ **do**

  Select an index set $\mathcal{K}$ from the random permutation of set $\{1, 2, \ldots, K\}$

  for $k \in \mathcal{K}$ **do**

    for $m \in \mathcal{M}_k$ **do**

      Compute $a_m = \beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k$ and $b_m = \beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{Q}_{\mathbf{Y}_m}\mathbf{Q}_m^{-1}\mathbf{s}_k$

    end for

    Solve the polynomial equation

$$f'_{k,T}(d) = \sum_{m\in\mathcal{M}_k}\left(((a_m + b_m) + a_m^2 d).\prod_{m'\in\mathcal{M}_k\setminus\{m\}}(1 + 2a_{m'}d + a_{m'}^2 d^2)\right) = 0$$

    Compute $\mathcal{D} = \{d : f'_{k,T}(d) = 0,\ \Im(d) = 0,\ \Re(d) \geq -\gamma_k\} \cup \{-\gamma_k\}$

    Let $f_{k,T}(d) = \sum_{m\in\mathcal{M}_k}\left(\log(1 + da_m) - \frac{db_m}{1+da_m}\right)$.

    Compute $\delta = \text{argmin}_{d\in\mathcal{D}}\, f_{k,T}(d)$.

    $\hat{\gamma}_k^i = \hat{\gamma}_k^{i-1} + \delta$

    for $m = 1, 2, \ldots, M$ **do**

      $\mathbf{Q}_m^{-1} \leftarrow \mathbf{Q}_m^{-1} - \delta\frac{\beta_{mk}\mathbf{Q}_m^{-1}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}}{1+\delta\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k}$

    end for

  end for

  if $f(\hat{\boldsymbol{\gamma}}^i) \geq f(\hat{\boldsymbol{\gamma}}^{i-1})$ **then**

    $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^{i-1}$

    break

  end if

  $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^i$

end for

return $\hat{\boldsymbol{\gamma}}$

Let

$$\mathcal{M}_k = \underset{m,T}{\text{indmax}}\ \{\beta_{mk}\}, \tag{21}$$

be the cluster of $T < M$ dominant APs of the device k. For $m \in \mathcal{M}_k$, define

$$a_m = \beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k \tag{22}$$

$$b_m = \beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{Q}_{\mathbf{Y}_m}\mathbf{Q}_m^{-1}\mathbf{s}_k. \tag{23}$$

To minimize the cost function (14) by utilizing the signals from the T dominant APs for the user k, we redefine the cost function as

$$f_{k,T}(d) = \sum_{m\in\mathcal{M}_k}\left(\log(1 + da_m) - \frac{db_m}{1 + da_m}\right). \tag{24}$$

Taking the derivative of (24) with respect to d, yields

$$f'_{k,T}(d) = \sum_{m\in\mathcal{M}_k}\frac{a_m}{1 + da_m} + \frac{b_m}{(1 + da_m)^2}. \tag{25}$$

Equating (25) to zero yields

$$\sum_{m\in\mathcal{M}_k}\left(((a_m+b_m)+a_m^2 d)\prod_{m'\in\mathcal{M}_k\setminus\{m\}}(1+2a_{m'}d + a_{m'}^2 d^2)\right) = 0 \tag{26}$$

which is a polynomial equation in d of degree $2T - 1$. Let

$$\mathcal{D} = \{d : f'_{k,T}(d) = 0, \Im(d) = 0, \Re(d) \geq -\gamma_k\} \cup \{-\gamma_k\}, \tag{27}$$

be the set of real roots of (26) and compute

$$\delta = \underset{d\in\mathcal{D}}{\text{argmin}}\, f_{k,T}(d). \tag{28}$$

The value $-\gamma_k$ is added to the set $\mathcal{D}$ to preserve the positivity of $\boldsymbol{\gamma}$ in (7) and the coordinate is updated as $\gamma_k \leftarrow \gamma_k + \delta$. The updating of sub-covariance blocks is carried out as explained in Sec. III-B. The proposed algorithm is outlined in Algorithm 2 and the activity detection can be performed using (19).

When $T = 1$, the clustering based algorithm reduces to Algorithm 1. For $T = 2$, we have degree 3 polynomial equation in (26) and the roots can be solved in closed form [41]. For $T > 2$, we have polynomials of degree 5 and higher and there exists no closed form solutions for the roots [42]. For $T > 2$, the approximate roots of the polynomial in (26) can be found by finding the eigen values of the companion matrix formed using the coefficients [43, Ch. 6] and the computation complexity is $\mathcal{O}(T^3)$ [44]. Thus the overall complexity of the Algorithm 2 is $\mathcal{O}(IK(TL^2 + T^3 + ML^2))$. The term T^3 corresponds to the complexity for finding the coefficients of (26), which can be computed using $2T$ point convolution. The term TL^2 and ML^2 corresponds to the complexities associated with computation of coefficients a_m, b_m and updating of covariance matrices, respectively. By introducing the clustering, we can identify a suitable tradeoff between performance (large T) and complexity (small T).

D. Parallel Architecture of Algorithms

In Algorithms 1 and 2, the update of each user is done sequentially irrespective of whether the user is active or not. In mMTC applications, the probability, ϵ, of device being active is very small, and thus on an average the number of active devices being active at any time is $K\epsilon$, which is much smaller. Hence, the sub-covariance matrices in Algorithm 2 do not change much while updating for an inactive user. Let $\mathcal{K}$ = random permutation of set $\{1, 2, \ldots, K\}$. Divide $\mathcal{K}$ into G random disjoint groups i.e., $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2 \cup \cdots \cup \mathcal{K}_G$. For each group $\mathcal{K}_g$, compute the coefficients a_m, b_m once and find the update δ. Once the updates for each user in group g are obtained, update the covariance matrix and continue to update for next group. This operation can be done in parallel and hence such a parallel architecture can save up to G times the time required for Algorithm 2. The proposed algorithm is summarized in Algorithm 3.

Algorithm 3 Clustering Based Coordinate Descend Parallel Architecture Algorithm for Estimating $\boldsymbol{\gamma}$

Input: Observations $\mathbf{Y}_m, \forall m = 1, 2, \ldots M$, $\beta_{mk}, \forall m = 1, 2, \ldots M, k = 1, 2, \ldots, K$

Initialize: $\mathbf{Q}_m^{-1} = \sigma^{-2}\mathbf{I}_L, \forall m = 1, 2, \ldots, M$, $\hat{\boldsymbol{\gamma}}^0 = \mathbf{0}_K$

Compute $\mathbf{Q}_{\mathbf{Y}_m} = \frac{1}{N}\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}, \forall m = 1, 2, \ldots, M$

Compute $\mathcal{M}_k = \text{indmax}_{m,T}\ \{\beta_{mk}\} \forall k = 1, 2, \ldots, K$

for $i = 1, 2, \ldots, I$ **do**

  Select an index set $\mathcal{K}$ from the random permutation of set $\{1, 2, \ldots, K\}$ and find random disjoint sets such that $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2 \cup \cdots \cup \mathcal{K}_G$

  for $g = 1, 2, \ldots, G$ **do**

    for $m \in \mathcal{M}_k$ and $k \in \mathcal{K}_g$ **do**

      Compute $a_{mk} = \beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k$ and $b_{mk} = \beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{Q}_{\mathbf{Y}_m}\mathbf{Q}_m^{-1}\mathbf{s}_k$

    end for

    for $k \in \mathcal{K}_g$ Parallel processing **do**

      Solve the polynomial equation

$$f'_{k,T}(d) = \sum_{m \in \mathcal{M}_k} \left(((a_m + b_m) + a_m^2 d). \prod_{m' \in \mathcal{M}_k \setminus \{m\}} (1 + 2a_m d + a_m^2 d^2) \right) = 0$$

      Compute $\mathcal{D} = \{d : f'_{k,T}(d) = 0,\ \Im(d) = 0,\ \Re(d) \geq -\gamma_k\} \cup \{-\gamma_k\}$

      Compute $\delta_k = \text{argmin}_{d \in \mathcal{D}} f_{k,T}(d)$, for $f_{k,T}(d) = \sum_{m \in \mathcal{M}_k} \left(\log(1 + da_{mk}) - \frac{db_{mk}}{1 + da_{mk}} \right)$.

      $\hat{\gamma}_k^i = \hat{\gamma}_k^{i-1} + \delta_k$

    end for

    for $m = 1, 2, \ldots, M$ and $k \in \mathcal{K}_g$ **do**

      $\mathbf{Q}_m^{-1} \leftarrow \mathbf{Q}_m^{-1} - \delta_k \frac{\beta_{mk}\mathbf{Q}_m^{-1}\mathbf{s}_k\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}}{1 + \delta_k\beta_{mk}\mathbf{s}_k^{\mathrm{H}}\mathbf{Q}_m^{-1}\mathbf{s}_k}$

    end for

  end for

  if $f(\hat{\boldsymbol{\gamma}}^i) \geq f(\hat{\boldsymbol{\gamma}}^{i-1})$ **then**

    $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^{i-1}$

    break

  end if

  $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}^i$

end for

return $\hat{\boldsymbol{\gamma}}$

E. Convergence of the Algorithms

The cost function $f(\boldsymbol{\gamma})$ in (8) is a sum of concave (log) and convex (trace) functions in $\boldsymbol{\gamma}$. Hence, closed-form expression to find the global minimum of (8) does not exist and in this paper, we use sub-optimal algorithms to find a local minimum of (8). First, we look at each term in $f(\boldsymbol{\gamma})$. $\text{Tr}\left(\mathbf{Q}_m^{-1}\frac{\mathbf{Y}_m\mathbf{Y}_m^{\mathrm{H}}}{N}\right)$, $\forall m = 1, 2, \cdots, M$ are convex and hence, are bounded from below. Also from (11), we can see that $\mathbf{Q}_m(\boldsymbol{\gamma})$ is a sum of positive semi-definite rank-one updates to a positive definite matrix $\sigma^2\mathbf{I}_L$. Hence $|\mathbf{Q}_m(\boldsymbol{\gamma})| > 0$, $\forall m = 1, 2, \cdots, M$. Thus the cost function $f(\boldsymbol{\gamma})$ is bounded from below. With each iteration in the proposed algorithms, the cost function $f(\boldsymbol{\gamma})$ is non-increasing. Hence, the proposed algorithms, where the cost function $f(\boldsymbol{\gamma})$ is monotonically decreasing and being bounded from below, is guaranteed to converge to a local minimum.

IV. Simulation Results

In this section, we characterize the massive connectivity in cell-free massive MIMO architectures and plot the performance of massive activity detection in cell-free massive MIMO with our proposed algorithms. We consider the receiver operating characteristic (ROC) characterized by probability of miss detection and probability of false alarm, as the performance measures for activity detection.

A. Simulation Model

We consider a square area wrapped around the edges to mimic a network with infinite area and to avoid boundary effects. The M APs are arbitrarily and independently distributed. The K users are considered to be uniformly distributed in the network. We will compare grant-free random access in cell-free and co-located architectures and hence, for co-located case, we assume the AP is at the center of the network. We consider such a simulation area with $K = 400$ devices, the activation probability $\epsilon = 0.1$ and the signature sequence length $L = 40$. The following micro-cell propagation model used in [45] is considered for large-scale fading coefficient β_{mk}:

$$\beta_{mk}[\text{dB}] = -30.5 - 36.7 \log_{10}\left(\frac{d_{mk}}{1\text{m}}\right) + F_{mk} \qquad (29)$$

where $F_{mk} \sim \mathcal{N}(0, \sigma_{sh}^2)$ is the shadow fading component with variance σ_{sh}^2, and d_{mk} is the horizontal distance between the k^{th} user and the m^{th} AP in meters ignoring their height differences. We use the same propagation model for the co-located massive MIMO case to ensure that the performance differences are caused by differences in technology characteristics instead of propagation model differences. Note that the shadow fading effects can be larger in co-located architecture. The maximum transmit power for a device is 200 mW and noise power $\sigma^2 = -109$ dBm. We consider a coherence block of 1 ms and 200 kHz, such that $\tau_c = 200$ symbols can be transmitted. However, due to large number of users in the system $K = 400 > \tau_c$, we assign non-orthogonal pilot sequence to each user. We reserve 20% of available symbols for pilot sequences. Hence in this paper, we consider $L = 40$ as the signature sequence length. The signature

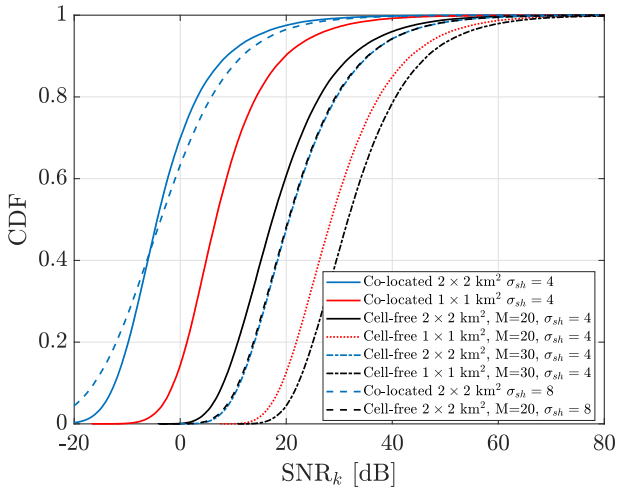

Fig. 2. Active device SNR.

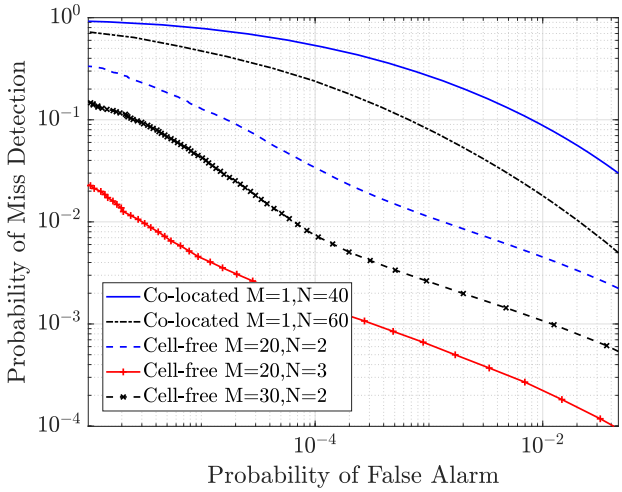

Fig. 3. Performance in 2×2 km^2 cell area. $K = 400$ users, activation probability $\epsilon = 0.1$, sequence length $L = 40$.

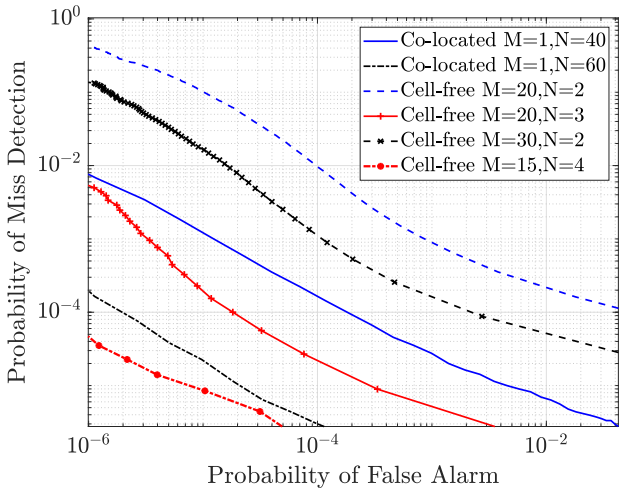

Fig. 4. Same as Fig. 3 but for 1×1 km^2 cell area.

sequence of each device is assumed to be drawn from the Gaussian distribution, $\mathcal{CN}(\mathbf{0}, \mathbf{I}_L)$.

B. Results

First, we look at the SNR achieved at the AP (dominant AP for cell-free) by an active device k transmitting at a power of 200 mW for co-located and cell-free massive MIMO networks of 2×2 km^2 and 1×1 km^2 cell area sizes and is shown in Fig. 2 for different scenarios. The plot shows that there is a significant gap in the SNR achieved for co-located and cell-free massive MIMO and hence the outage probability (probability of not achieving a certain SNR target) is less in the cell-free massive MIMO scenario [46], thereby enabling services to many devices. Due to the distributed topology of APs in the cell-free scenario, a device is highly likely to be close to one of the APs, thus providing a stronger channel gain which improves the SNR. This gain in SNR is referred to as macro-diversity gain of cell-free systems [46, pp. 331]. Also, in cell-free networks when more APs are deployed in the cell area, the SNR achieved at the dominant AP increases. Fig. 2 shows that the co-located MIMO network is sensitive to shadow fading while the cell-free network is robust to shadow fading. Co-located network is likely to have larger shadow fading parameter compared to cell-free network according to standard 3GPP channel models [29], [45].

Next, we look at the performance of the proposed algorithms for activity detection in cell-free massive MIMO network. For simulations, we consider an SNR target at the dominant AP such that 95% of the active devices will be able to achieve the desired SNR and hence access the network. Referring to Fig. 2, target SNR can be computed by finding the SNR where CDF=0.05 for each test cases. The variance of shadow fading parameter is $\sigma_{sh}^2 = 4$. The ROC curve is plotted for different thresholds. We have considered $I = 10$ as the maximum number of iterations and 10^5 Monte-Carlo trials.

The performance of Algorithm 1 is given in Fig. 3 and Fig. 4 for 2×2 km^2 and 1×1 km^2 cell area sizes, respectively. From the plots, it can be seen that, when the number of antennas per AP, N is increased, the activity detection performance improves due to the improvement in the spatial resolution of the devices. For low-power applications like mMTC, co-located MIMO is highly sensitive to the receive SNR and the performance degrades significantly with an increase in cell area. From the plots, it can be seen that the device activity detection performance is better in cell-free massive MIMO networks compared to co-located MIMO networks, as the cell-area increases. Hence, cell-free network can provide a broad coverage in mMTC applications. Note that for co-located architecture, the network will have higher shadow fading effects compared to cell-free architecture which will further reduce the performance owing to the reduction in the SNR. When the AP density in the network defined by $\frac{M}{\text{cell area}}$ increases, the SNR at the dominant AP increases, thereby improving the activity detection performance as shown in Figs. 3 and 4.

The performance of the Algorithm 2 for activity detection in cell-free massive MIMO network is given in Fig. 5 for 2×2 km^2 cell area size. From the plots, we can see that when we consider information from multiple dominant APs ($T > 1$) for activity detection, the performance improves compared to considering only the most dominant AP. Moreover, to study the performance gain from the dominant APs and to have a fair comparison we consider two cases, one with $M = 20$ and other with $M = 25$, such that the AP density is higher

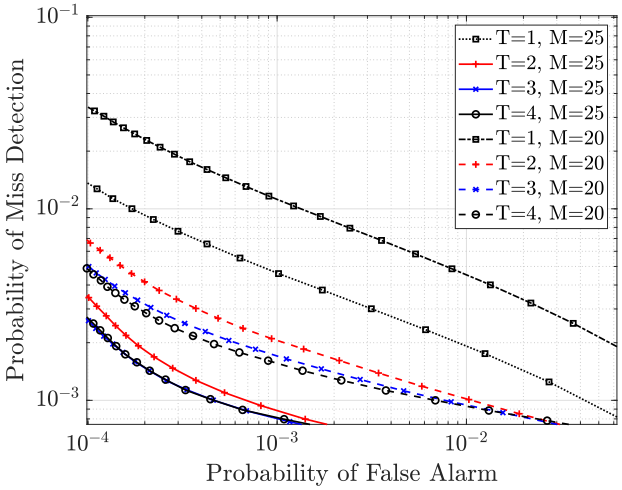

Fig. 5. Activity detection performance with 2×2 km^2 cell area, $N = 2$, $K = 400$, $L = 40$, $\epsilon = 0.1$.

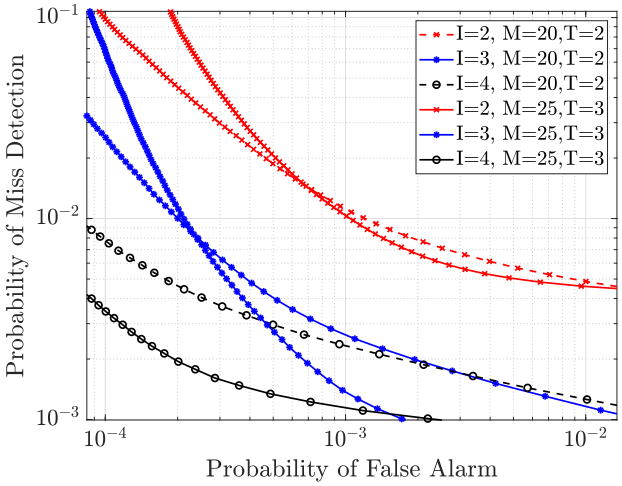

Fig. 6. Convergence of algorithm 2 with 2×2 km^2 cell area, $N = 2$, $K = 400$, $L = 40$.

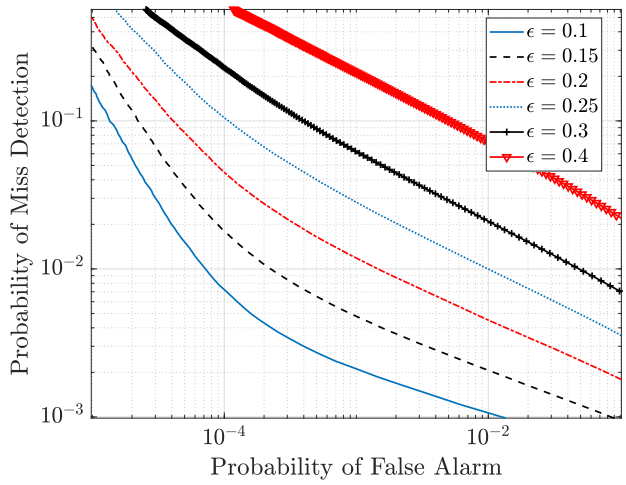

Fig. 7. Activity detection performance with 2×2 km^2 cell area, $M = 20$, $N = 2$, $K = 400$, $L = 40$, $T = 2$ for different device activation probability ϵ.

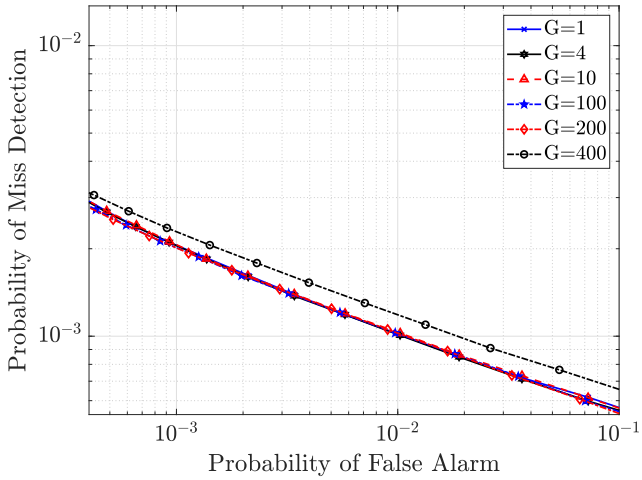

Fig. 8. Activity detection performance with Algorithm 3 for 2×2 km^2 cell area, $M = 20$, $N = 2$, $K = 400$, $L = 40$, $\epsilon = 0.1$, $T = 2$.

in the latter case. From the plot, it can be seen that the performance gain for activity detection improves with the dominance of the APs. However, considering information from three or more dominant APs, the performance is not improving by a significant amount compared to considering information from the two dominant APs. Also, with $T = 2$, the algorithm is computationally efficient as closed form expressions are available for solving a degree 3 polynomial equation.

The convergence of the algorithm with the number of iterations I is plotted in Fig. 6 for different scenarios. Since the cost function (8) is non-increasing and bounded from below, when the proposed algorithms are used, the activity detection algorithms are able to achieve convergence (to a local minimum).

In Fig. 7, we study the impact of different activation probabilities on the activity detection performance of Algorithm 2. From the plot, it can be seen that, we can still achieve a nominal performance for activity detection when $K\epsilon > L$, while using the AMP-based approach, the performance degrades severely in such situations [18], [21]. Hence, the proposed algorithms can support activity detection when total number of active devices are greater than the signature sequence length.

The performance of Algorithm 3 is shown in Fig. 8. The superiority of Algorithm 3 is in the parallelism and hence, the computational time can be cut by G times. From the plot, it can be seen that the performance of Algorithm-3 is similar to Algorithm-2 due to the fact that we have more inactive users compared to active users and hence the cost function will not change much while updating the sub-covariance matrices of inactive users. However, when we apply complete parallelism $G = K$, we slightly loose in performance as shown in the Fig. 8.

C. Capacity Limited Fronthauls

In practice, the fronthauls connecting the APs to the CPU are capacity limited and causes degradation in the performance. The data from APs will be quantized before sending to the CPU. We consider each complex value to be send from each AP is represented using B bits in floating point representation. Out of $\frac{B}{2}$ bits available for a real value, B_M are assigned for mantissa representation and rest for the exponent representation. The performance of a capacity limited fronthaul, for Algorithm 2 is plotted in Fig. 9. The activity detection performance improves as the number of bits

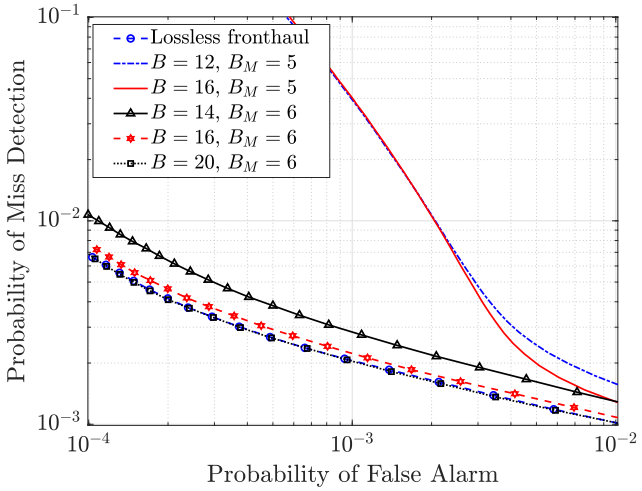

Fig. 9. Activity detection performance with 2×2 km^2 cell area, $M = 20$, $N = 2$, $K = 400$, $L = 40$, $T = 2$, $\epsilon = 0.1$ with a capacity limited fronthaul.

B increases. From the plot, when Algorithm 2 is used with $T = 2$, to achieve the optimal performance of a lossless fronthaul, we require $B = 20$ bits per complex value. The contemporary fronthaul technology has a capacity of 10 Gbps or more, so even if the capacity is not infinite, it won't be the limiting factor for the operations that we are considering.

V. Conclusion

In this paper, we analyzed the grant-free random access scenario in cell-free massive MIMO networks. The paper formulates a novel activity detection problem and proposed algorithms for activity detection based on clustering approach. We show that for low-powered applications like mMTC, co-located massive MIMO is highly sensitive to the large SNR variations. The cell-free massive MIMO network is robust against the shadow fading effects providing macro diversity gains and, hence, can provide better coverage for mMTC applications. Simulation results show that the macro diversity gain offered by cell-free network improves the activity detection performance.

References

[1] U. K. Ganesan, E. Björnson, and E. G. Larsson, "An algorithm for grant-free random access in cell-free massive MIMO," in *Proc. IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, May 2020, pp. 1–5.

[2] E. Dutkiewicz, X. Costa-Perez, I. Z. Kovacs, and M. Mueck, "Massive machine-type communications," *IEEE Network*, vol. 31, no. 6, pp. 6–7, Nov./Dec. 2017.

[3] C. Bockelmann *et al.*, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[4] A.-S. Bana *et al.*, "Massive MIMO for Internet of Things (IoT) connectivity," *Phys. Commun.*, vol. 37, Dec. 2019, Art. no. 100859.

[5] 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.

[6] N. K. Pratas, H. Thomsen, V. C. Stefanović, and P. Popovski, "Code-expanded random access for machine-type communications," in *Proc. IEEE Globecom Workshops*, Dec. 2012, pp. 1681–1686.

[7] J. H. Sørensen, E. De Carvalho, and P. Popovski, "Massive MIMO for crowd scenarios: A solution based on random access," in *Proc. IEEE Globecom Workshops*, Dec. 2014, pp. 352–357.

[8] E. Bjornson, E. de Carvalho, E. G. Larsson, and P. Popovski, "Random access protocol for massive MIMO: Strongest-user collision resolution (SUCR)," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[9] Y. Yang, G. Song, W. Zhang, X. Ge, and C. Wang, "Neighbor-aware multiple access protocol for 5G mMTC applications," *China Commun.*, vol. 13, no. 2, pp. 80–88, 2016.

[10] 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, 3rd Quart., 2020.

[11] C. Bockelmann, H. F. Schepker, and A. Dekorsy, "Compressive sensing based multi-user detection for machine-to-machine communication," *Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 4, pp. 389–400, Jun. 2013.

[12] F. Monsees, M. Woltering, C. Bockelmann, and A. Dekorsy, "Compressive sensing multi-user detection for multicarrier systems in sporadic machine type communication," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.

[13] Z. Gao, L. Dai, Z. Wang, S. Chen, and L. Hanzo, "Compressive-sensing-based multiuser detector for the large-scale SM-MIMO uplink," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8725–8730, Oct. 2016.

[14] Y. Du *et al.*, "Efficient multi-user detection for uplink grant-free NOMA: Prior-information aided adaptive compressive sensing perspective," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2812–2828, Dec. 2017.

[15] H. Xiao, B. Ai, and W. Chen, "A grant-free access and data recovery method for massive machine-type communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[16] 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, Jun. 2018.

[17] 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, Sep. 2018.

[18] K. Senel and E. G. Larsson, "Device activity and embedded information bit detection using AMP in massive MIMO," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2017, pp. 1–6.

[19] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, Dec. 2018.

[20] Z. Chen, F. Sohrabi, and W. Yu, "Multi-cell sparse activity detection for massive random access: Massive MIMO versus cooperative MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4060–4074, Aug. 2019.

[21] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, 2009.

[22] T. Ding, X. Yuan, and S. C. Liew, "Sparsity learning-based multiuser detection in grant-free massive-device multiple access," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3569–3582, Jul. 2019.

[23] Y. Bai, B. Ai, and W. Chen, "Deep learning based fast multiuser detection for massive machine-type communication," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5.

[24] Z. Zhang, Y. Li, C. Huang, Q. Guo, C. Yuen, and Y. L. Guan, "DNN-aided block sparse Bayesian learning for user activity detection and channel estimation in grant-free non-orthogonal random access," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12000–12012, Dec. 2019.

[25] S. Haghighatshoar, P. Jung, and G. Caire, "Improved scaling law for activity detection in massive MIMO systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 381–385.

[26] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Covariance based joint activity and data detection for massive random access with massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[27] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2523–2527.

[28] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Grant-free massive random access with a massive MIMO receiver," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 23–30.

[29] 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.

[30] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 197, Dec. 2019.

[31] 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, Mar. 2021.

[32] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[33] 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, Nov. 2018.

[34] X. Xu, X. Rao, and V. K. N. Lau, "Active user detection and channel estimation in uplink CRAN systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2727–2732.

[35] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection in multi-cell massive MIMO exploiting channel large-scale fading," 2021, *arXiv:2103.00782*. [Online]. Available: http://arxiv.org/abs/2103.00782

[36] A. Fengler, "Sparse recovery based grant-free random access for massive machine-type communication," Ph.D. dissertation, Technische Universität Berlin, Berlin, Germany, 2021, doi: 10.14279/depositonce-11526.

[37] C. Wang, O. Y. Bursalioglu, H. Papadopoulos, and G. Caire, "On-the-fly large-scale channel-gain estimation for massive antenna-array base stations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[38] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021, doi: 10.1109/TIT.2021.3065291.

[39] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 124–127, 1950.

[40] J. J. Sylvester, "On the relation between the minor determinants of linearly equivalent quadratic functions," *Philos. Mag.*, vol. 1, no. 4, pp. 295–305, 1851.

[41] G. C. Holmes, "The use of hyperbolic cosines in solving cubic polynomials," *Math. Gazette*, vol. 86, no. 507, pp. 473–477, Nov. 2002.

[42] V. B. Alekseev, *Abel's Theorem in Problems and Solutions: Based on the Lectures of Professor VI Arnold*. Moscow, Russia: Springer, 2004.

[43] J. McNamee, *Numerical Methods for Roots of Polynomials-Part I*. Amsterdam, The Netherlands: Elsevier, 2007.

[44] O. Aberth, "Iteration methods for finding all zeros of a polynomial simultaneously," *Math. Comput.*, vol. 27, no. 122, pp. 339–344, Apr. 1973.

[45] 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.

[46] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends Signal Process.*, vol. 14, nos. 3–4, pp. 162–472, 2021.

Unnikrishnan Kunnath Ganesan (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the University of Calicut, Kerala, India, in 2011, and the M.E. degree in telecommunication engineering from Indian Institute of Science, Bengaluru, India, in 2014. From 2014 to 2017, he worked as a Modem Systems Engineer with Qualcomm India Private Ltd., Bengaluru. From 2017 to 2019, he worked as a Senior Firmware Engineer with Intel. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering (ISY), Linköping University, Sweden. His primary research interests include MIMO wireless communications, space-time coding, network coding, and information theory.

Emil Björnson (Senior Member, IEEE) received the M.S. degree in engineering mathematics from Lund University, Sweden, in 2007, and the Ph.D. degree in telecommunications from the KTH Royal Institute of Technology, Sweden, in 2011.

From 2012 to 2014, he held a joint post-doctoral position at Alcatel-Lucent Chair on Flexible Radio, Supelec, France, and the KTH Royal Institute of Technology. He joined Linköping University, Sweden, in 2014, where he is currently an Associate Professor. In September 2020, he became a part-time Visiting Full Professor at the KTH Royal Institute of Technology. He has authored the textbooks *Optimal Resource Allocation in Coordinated Multi-Cell Systems* in 2013, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* in 2017, and *Foundations of User-Centric Cell-Free Massive MIMO* in 2021. He is dedicated to reproducible research and has made a large amount of simulation code publicly available. He has performed MIMO research for over 15 years, his articles have received more than 14 000 citations, and has filed more than 20 patent applications. He is a host of the podcast wireless future and has a popular YouTube channel. He performs research on MIMO communications, radio resource allocation, machine learning for communications, and energy efficiency.

Dr. Björnson has been a member of Online Editorial Team of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since 2020. He has received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA, the 2015 Ingvar Carlsson Award, the 2016 Best Ph.D. Award from EURASIP, the 2018 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 EURASIP Early Career Award, the 2019 IEEE Communications Society Fred W. Ellersick Prize, the 2019 IEEE Signal Processing Magazine Best Column Award, the 2020 Pierre-Simon Laplace Early Career Technical Achievement Award, the 2020 CTTC Early Achievement Award, and the 2021 IEEE ComSoc RCC Early Achievement Award. He also coauthored articles that received Best Paper Awards at the conferences, including WCSP 2009, IEEE CAMSAP 2011, IEEE SAM 2014, IEEE WCNC 2014, IEEE ICC 2015, and WCSP 2017. He has been on the Editorial Board of IEEE TRANSACTIONS ON COMMUNICATIONS since 2017. He has been an Area Editor in *IEEE Signal Processing Magazine* since 2021. He has also been a guest editor of multiple special issues.

Erik G. Larsson (Fellow, IEEE) received the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 2002. He is currently a Professor of communication systems with Linköping University (LiU), Linköping, Sweden. He was with the KTH Royal Institute of Technology, Stockholm, Sweden; The George Washington University, USA; the University of Florida, USA; and Ericsson Research, Sweden. His main professional interests are within the areas of wireless communications and signal processing. He has coauthored *Space-Time Block Coding for Wireless Communications* (Cambridge University Press, 2003) and *Fundamentals of Massive MIMO* (Cambridge University Press, 2016).

He is currently a member of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS Steering Committee and an Editorial Board Member of *IEEE Signal Processing Magazine*. He served as the Chair for the IEEE Signal Processing Society SPCOM Technical Committee (2015–2016), the Chair for the IEEE WIRELESS COMMUNICATIONS LETTERS Steering Committee (2014–2015), the General and Technical Chair for the Asilomar SSC Conference (2015 and 2012), the Technical Co-Chair for the IEEE Communication Theory Workshop (2019), and a member for the IEEE Signal Processing Society Awards Board (2017–2019). He was an Associate Editor for, among others, IEEE TRANSACTIONS ON COMMUNICATIONS (2010–2014) and IEEE TRANSACTIONS ON SIGNAL PROCESSING (2006–2010).

He received the *IEEE Signal Processing Magazine* Best Column Award twice, in 2012 and 2014, the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, the IEEE ComSoc Best Tutorial Paper Award in 2018, and the IEEE ComSoc Fred W. Ellersick Prize in 2019.