Machine Learning Enabled Preamble Collision Resolution in Distributed Massive MIMO

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

29-12-2020 / 31-12-2020

CITATION

Ding, Jie; Qu, Daiming; Liu, Pei; Choi, Jinho (2020): Machine Learning Enabled Preamble Collision Resolution in Distributed Massive MIMO. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.13498338.v1>

DOI

[10.36227/techrxiv.13498338.v1](https://doi.org/10.36227/techrxiv.13498338.v1)

Machine Learning Enabled Preamble Collision Resolution in Distributed Massive MIMO

Jie Ding, Daiming Qu, Pei Liu, and Jinho Choi, *Senior Member, IEEE*

Abstract

Preamble collision is a bottleneck that impairs the performance of random access (RA) user equipment (UE) in grant-free RA (GFRA). In this paper, by leveraging distributed massive multiple input multiple output (mMIMO) together with machine learning, a novel machine learning based framework solution is proposed to address the preamble collision problem in GFRA. The key idea is to identify and employ the neighboring access points (APs) of a collided RA UE for its data decoding rather than all the APs, so that the mutual interference among collided RA UEs can be effectively mitigated. To this end, we first design a tailored deep neural network (DNN) to enable the preamble multiplicity estimation in GFRA, where an energy detection (ED) method is also proposed for performance comparison. With the estimated preamble multiplicity, we then propose a K -means AP clustering algorithm to cluster the neighboring APs of collided RA UEs and organize each AP cluster to decode the received data individually. Simulation results show that a decent performance of preamble multiplicity estimation in terms of accuracy and reliability can be achieved by the proposed DNN, and confirm that the proposed schemes are effective in preamble collision resolution in GFRA, which are able to achieve a near-optimal performance in terms of uplink achievable rate per collided RA UE, and offer significant performance improvement over traditional schemes.

Index Terms

Preamble collision resolution, grant-free random access, deep learning, distributed massive MIMO, clustering.

Jie Ding and Jinho Choi are with the School of Information Technology, Deakin University, Geelong, VIC 3220, Australia.

Daiming Qu is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, 430074, China.

Pei Liu is with the School of Information Engineering, Wuhan University of Technology, Wuhan, 430070, China.

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 61701186.

I. INTRODUCTION

The Fifth Generation (5G) and future wireless communication focus on three major communication categories: enhanced mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and massive machine-type communication (MTC) (mMTC) [1]. Among them, mMTC has been regarded as an essential communication paradigm for a wide range of applications including healthcare, smart home, smart agriculture, and logistics and tracking [2]. Since mMTC usually features with massive uplink access and limited packet size in nature, it imposes new requirements and challenges in terms of random access (RA) design [3].

In Long Term Evolution (LTE) systems, a typical grant-based RA procedure is used to provide reliable access for human-type communication (HTC) [4]. Since the grant-based RA requires handshaking to issue exclusive channel reservation for each RA user equipment (UE), it is unable to support massive access due to limited channel-resource utilization efficiency and also results in high signalling overhead to mMTC RA UEs. In the light of this, grant-free RA (GFRA) procedure has been recently actively studied in MTC for low signalling overhead and latency [5]–[8]. In GFRA, the request-grant handshaking steps in grant-based RA for channel reservations are skipped, which allows RA UEs to access the network without grant acquisition once they have data to send [5]. As a result, the signalling overhead is reduced [6]. In [7], [8], GFRA schemes by utilizing spreading techniques are proposed to support MTC. Nevertheless, spreading results in a bandwidth expansion, which is inefficient in terms of channel resource utilization. Besides, RA UEs have to contend for RA channel resources in an uncoordinated manner due to no channel reservation. Therefore, making efficient use of channel resources to simultaneously support a large number of RA UEs is essential.

Recently, massive multiple input multiple output (mMIMO) has been a key technology in 5G and future wireless communication to mitigate wireless resource scarcity and increase channel-resource utilization efficiency [9]. As a large number of either co-located antennas (co-located mMIMO) or distributed antennas (distributed mMIMO) are employed at the base station (BS) in mMIMO, mutual channel orthogonality among RA UEs (also known as favorable propagation) can be asymptotically achieved as the number of antennas increases [10], [11]. By taking advantage of this property, RA UEs can share the same channel resource simultaneously without the need of channel reservation, while beamforming techniques can be used to spatially separate them in an effective manner. Thus, mMIMO has been considered a prominent enabler for GFRA.

A number of research works have been undertaken to study the performance of GFRA with co-located mMIMO. As pointed out by [12], preamble collision (i.e., multiple RA UEs choose the same preamble) is the main bottleneck in GFRA with co-located mMIMO that curbs the performance of RA UEs. In fact, since the traffic of mMTC RA UEs is usually random and sporadic, the BS has neither prior information of RA UEs' activity nor their channel state information (CSI) in each GFRA slot. Thus, each RA UE needs to send a preamble prior to data for channel estimation. However, since the number of orthogonal preambles is limited and RA UEs choose preambles in a random and uncoordinated manner, there could be multiple RA UEs that select the same preamble. As a result, the estimated CSI for these collided RA UEs becomes inaccurate and data from them would be incorrectly decoded. Various approaches have been developed to address the preamble collision issue in GFRA with co-located mMIMO. For instance, non-orthogonal preambles are considered in [13] to expand preamble space without constraint on the preamble length and sporadic traffic pattern of RA UEs is exploited to detect and identify active RA UEs. In [14], a super-preamble consisting of multiple short preambles is adopted and features of favorable propagation and channel hardening in mMIMO are exploited to identify the super-preamble of each RA UE. In [15], an ensemble independent component analysis (EICA) based pilot random access is proposed to enable joint active UEs detection and uplink data decoding. These approaches are effective in reducing the preamble collision. In [16], a successive interference cancellations (SIC) based scheme is proposed for preamble collision resolution. However, this type of scheme requires that RA UEs use multiple RA slots to transmit multiple preambles and same data to make SIC possible. Therefore, how to resolve the preamble collision when it occurs on the basis of a single RA slot is still an open issue in GFRA. In fact, since all the signals of collided RA UEs are multiplexed and assembled at the centralized BS in co-located mMIMO, the BS can only deem that the received signals come from a single RA UE, making it practically difficult to find preamble multiplicity (i.e., the number of the RA UEs that select the same preamble) and resolve the preamble collision. Note that conventional preamble decontamination solutions, such as [17], [18], cannot be used to solve the considered problem in GFRA since they require that the information of the number of active UEs and all UEs' partial CSI (e.g., large scale fading coefficients) is available at the BSs as a prior knowledge. In [19], an effective preamble collision resolution scheme is proposed in co-located mMIMO. Nevertheless, it only works in grant-based RA as a feedback after preamble detection from the BS to RA UEs is required.

Different from co-located mMIMO in terms of antenna topology, distributed mMIMO employs a large number of geographically distributed access points (APs) to serve UEs, and each AP is equipped with a single or a few antennas. Compared to co-located mMIMO, distributed mMIMO provides macro-diversity and has enhanced network coverage and capacity [20]–[25]. Nevertheless, the existing works that address the preamble collision issue in distributed mMIMO rely on the condition that the full or partial CSI of UEs is known at the BS [26]–[28], which is not the case in the context of GFRA. Thus, GFRA with distributed mMIMO has not been well investigated yet. On the other hand, since APs are spatially distributed in distributed mMIMO and signals to different APs undergo different levels of large-scale fading, only neighboring APs within a communication range of an UE have non-negligible channel gains [23], [24], which implies signal spatial sparsity in distributed mMIMO [29]. This feature opens up a possibility for preamble collision resolution in GFRA. Specifically, due to the sporadic traffic pattern of RA UEs, collided RA UEs could be separate in space and surrounded by different groups of APs. If the BS is able to identify neighboring APs of a collided RA UE in GFRA and only employs the neighboring APs rather than all the APs to serve the collided RA UE, the interference from other collided RA UEs in the preamble domain could be largely mitigated and its performance is expected to be improved as a result.

Motivated by this, a novel machine learning enabled AP clustering scheme is proposed to resolve the preamble collision in GFRA by leveraging distributed mMIMO. To facilitate preamble collision resolution, collided preamble multiplicity needs to be estimated by the BS. To this end, we first design a tailored deep neural network (DNN) to enable the preamble multiplicity estimation in GFRA, where a data-driven energy detection (ED) method is also proposed for performance comparison. With the estimated preamble multiplicity, we propose a K -means AP clustering algorithm to cluster the neighboring APs of collided RA UEs, and then each AP cluster is employed to decode the received data individually. Under practical wireless environments and different deployments of distributed mMIMO, we investigate and analyze the performance of the proposed DNN and show that decent estimation accuracy and reliability can be achieved. Simulation results further confirm that the proposed machine learning enabled AP clustering schemes are able to achieve a near-optimal performance in terms of preamble collision resolution, and provide significant performance enhancement over the traditional schemes.

The novelty and contribution of this paper are summarized as follows.

- We propose a novel machine learning based framework solution to mitigate the impact

of preamble collision on the performance of collided RA UEs in GFRA with distributed mMIMO, which requires neither prior information of RA UEs' activity nor their CSI. To the best of our knowledge, this is the first work that aims to resolve the preamble collision in GFRA on the basis of a single RA slot by exploiting two-dimensional signal information including signal strength and locations in distributed mMIMO.

- To enable preamble collision resolution in GFRA, the preamble multiplicity is an indispensable parameter that needs to be estimated by the BS. To this end, we for the first time leverage deep learning based classification models to enable the preamble multiplicity estimation in distributed mMIMO, where connections between received preamble signal patterns and preamble multiplicities are exploited.
- With the estimated preamble multiplicity, we further propose a K -means AP clustering algorithm to enable the neighboring AP clustering of collided RA UEs and organize each AP cluster instead of all the APs to decode data of collided RA UEs individually. Thereby, the mutual interference among collided RA UEs in the preamble domain could be effectively mitigated, which results in appreciable performance improvement.

The remainder of this paper is organized as follows. In Section II, the system model of GFRA with distributed mMIMO is introduced and the motivation of this work is explained theoretically by a toy example. In Section III, the proposed DNN for preamble multiplicity estimation is detailed and its estimation performance is investigated. In Section IV, the proposed K -means AP clustering algorithm is presented and its performance in terms of uplink achievable rate per collided RA UE is evaluated. The work is concluded in Section V.

Notation: Boldface lower and upper case symbols represent vectors and matrices, respectively. $\mathbf{I}_n$ is the $n \times n$ identity matrix. The conjugate, transpose, and complex conjugate transpose operators are denoted by $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$. $\|\cdot\|$ denotes the Euclidean norm. $\mathbf{x} \sim \mathcal{CN}(0, \Sigma)$ indicates that $\mathbf{x}$ is a circularly symmetric complex Gaussian (CSCG) random vector with zero-mean and covariance matrix Σ .

II. SYSTEM MODELS AND MOTIVATION

A. System Model

We consider a distributed mMIMO system in a wide area to serve N MTC UEs that are scattered in the area (each UE is equipped with a single antenna). As illustrated in Figure 1, there are M APs uniformly and spatially distributed. We assume that these distributed APs are

connected to a BS central processing unit (CPU) via an error-free fronthaul and each AP is equipped with S antennas.

In a GFRA channel slot¹, suppose that U UEs, indexed as $1, 2, \dots, U$, are active to access the channel for uplink transmission in a grant-free manner, where U follows the binomial distribution $\text{Bino}(N, \rho)$ and ρ ($\rho \ll 1$) is the sporadic activation probability of each UE.

Figure 1: Illustration of distributed mMIMO systems and an example with preamble multiplicity of two.

To enable channel estimation at the APs, each RA UE directly transmits an RA preamble before data, which is randomly selected from an orthogonal preamble pool of size L ($L \ll N$), i.e., $\Omega = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_L\}$, where $\mathbf{p}_l$ denotes the l th orthogonal preamble vector of length L , $\|\mathbf{p}_l\|^2 = L$ and $\mathbf{p}_l^T \mathbf{p}_{l'}^* = 0$, for $l \neq l'$, $l, l' \in \{1, 2, \dots, L\}$. In this paper, as in the RA of LTE, single-root Zadoff-Chu sequence is used to generate the orthogonal preambles thanks to its good auto-correlation properties [4].

Specifically, the received preamble signal, $\mathbf{Y} \in \mathbb{C}^{MS \times L}$, can be given by

$$\mathbf{Y} = \sum_{u=1}^U \sqrt{P_T} \mathbf{g}_u \boldsymbol{\psi}_u^T + \mathbf{N}, \quad (1)$$

where P_T is the transmit power of each RA UE, $\mathbf{g}_u = [\mathbf{g}_{u1}^T, \mathbf{g}_{u2}^T, \dots, \mathbf{g}_{uM}^T]^T \in \mathbb{C}^{MS}$ is the channel response vector between RA UE u and the APs and $\mathbf{g}_{um} = \sqrt{\beta_{um}} \mathbf{h}_{um} \in \mathbb{C}^S$ is the

¹In practice, a GFRA slot consists of multiple channels over frequency and each channel can accommodate a number of RA UEs. Since each channel is independent in frequency, we only focus on a single channel scenario in this work.

channel response vector between RA UE u and AP m , where β_{um} denotes the large-scale fading coefficient and $\mathbf{h}_{um} \sim \mathcal{CN}(0, \mathbf{I}_S)$ stands for the small-scale fading vector, $\psi_u \in \mathbb{C}^L$ is the selected preamble by RA UE u from the preamble pool Ω , and $\mathbf{N}$ is the noise matrix with i.i.d. elements distributed as $\mathcal{CN}(0, \sigma^2)$.

Due to the randomness of preamble selection by each RA UE, a key issue to be addressed is the preamble collision, which constrains the throughput and transmission reliability of collided RA UEs. In the sequel, we detail the performance impairment caused by preamble collision in GFRA and present the intuition and motivation of preamble collision resolution in distributed mMIMO.

B. Performance Impairment due to Preamble Collision

Without loss of generality, we consider the RA UE with index 1 as the RA UE of interest and explain the impact of preamble collision on its performance. In the case of preamble collision, without any prior CSI information, the least-squares (LS) based channel estimation for RA UE 1 can be used and the estimate over all the APs is given by

$$\hat{\mathbf{g}}_1 = \frac{\mathbf{Y}\psi_1^*}{\sqrt{P_T L}} = \mathbf{g}_1 + \sum_{u' \in \Phi_{\psi_1}} \mathbf{g}_{u'} + \frac{1}{\sqrt{\rho_T L}} \mathbf{n}, \quad (2)$$

where Φ_{ψ_1} is the set of indices of RA UEs that select ψ_1 other than RA UE 1, the cardinality of Φ_{ψ_1} is denoted by $|\Phi_{\psi_1}| \geq 1$, $\rho_T = P_T/\sigma^2$ is defined as the uplink transmit signal-to-noise ratio (SNR) corresponding to each RA UE, and $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}_{MS})$. From (2), we see that the estimated channel under preamble collision is distorted by the channels of other RA UEs that select the same preamble.

Following preamble, each RA UE transmits its data. The received data symbol vector $\mathbf{r} \in \mathbb{C}^{MS}$ over all the APs is given by

$$\mathbf{r} = \sum_{u=1}^U \sqrt{P_T} \mathbf{g}_u s_u + \bar{\mathbf{n}}, \quad (3)$$

where $\bar{\mathbf{n}}$ the background noise vector distributed as $\mathcal{CN}(0, \sigma^2 \mathbf{I}_{MS})$ and s_u is a data symbol transmitted by RA UE u and $\mathbb{E}[|s_u|^2] = 1$.

Then, all the APs send their channel estimate and received data signals to the BS CPU. Since the error-free fronthaul links between all the APs and the BS CPU are assumed as in [23]–[29], the information sent by APs can be perfectly gathered at the BS CPU for signal processing.

With (2) and (3), the estimated data symbol of RA UE 1 after conjugate beamforming at the BS CPU is thus given by

$$\hat{s}_1 = \frac{\hat{\mathbf{g}}_1^H \mathbf{r}}{MS\sqrt{P_T}} = \frac{\hat{\mathbf{g}}_1^H \mathbf{g}_1 s_1}{MS} + \frac{\sum_{u=2}^U \hat{\mathbf{g}}_1^H \mathbf{g}_u s_u}{MS} + \frac{\hat{\mathbf{g}}_1^H \bar{\mathbf{n}}}{MS\sqrt{P_T}}. \quad (4)$$

As $M \rightarrow \infty$ (here we fix S), it becomes

$$\begin{aligned} \hat{s}_{1\infty} &= \lim_{M \rightarrow \infty} \left(\frac{\hat{\mathbf{g}}_1^H \mathbf{g}_1 s_1}{MS} + \frac{\sum_{u=2}^U \hat{\mathbf{g}}_1^H \mathbf{g}_u s_u}{MS} + \frac{\hat{\mathbf{g}}_1^H \bar{\mathbf{n}}}{MS\sqrt{P_T}} \right) \\ &\stackrel{(a)}{=} \frac{\mathbf{g}_1^H \mathbf{g}_1}{MS} s_1 + \sum_{u' \in \Phi_{\psi_1}} \frac{\mathbf{g}_{u'}^H \mathbf{g}_{u'}}{MS} s_{u'} \\ &\stackrel{(b)}{=} \lim_{M \rightarrow \infty} \sum_{m=1}^M \frac{\beta_{1m} s_1}{M} + \sum_{u' \in \Phi_{\psi_1}} \lim_{M \rightarrow \infty} \sum_{m=1}^M \frac{\beta_{u'm} s_{u'}}{M} \\ &= \beta_1 s_1 + \sum_{u' \in \Phi_{\psi_1}} \beta_{u'} s_{u'}, \end{aligned} \quad (5)$$

where $\stackrel{(a)}{=}$ and $\stackrel{(b)}{=}$ are obtained based on Chebyshev's Theorem. Specifically, $\stackrel{(a)}{=}$ is obtained by the fact that $\frac{\mathbf{g}_u^H \mathbf{g}_{u'}}{SM} \xrightarrow[M \rightarrow \infty]{P} 0$ when $u \neq u'$ and $\frac{\hat{\mathbf{g}}_1^H \bar{\mathbf{n}}}{SM} \xrightarrow[M \rightarrow \infty]{P} 0$. $\stackrel{(b)}{=}$ is obtained by the fact that, $\frac{\mathbf{g}_u^H \mathbf{g}_u}{S} = \sum_{m=1}^M \frac{\beta_{um} \|\mathbf{h}_{um}\|^2}{S}$. Since $\frac{\|\mathbf{h}_{um}\|^2}{S}$ follows the gamma distribution with shape S and scale $\frac{1}{S}$,

it has mean of 1 and variance of $\frac{1}{S}$. Thus, we have $\frac{\mathbf{g}_u^H \mathbf{g}_u}{MS} \xrightarrow[M \rightarrow \infty]{P} \frac{\sum_{m=1}^M \beta_{um}}{M}$. In addition, $\beta_u \triangleq \lim_{M \rightarrow \infty} \sum_{m=1}^M \frac{\beta_{um}}{M}$, $u = 1, 2, \dots, U$.

As a result, the asymptotic signal-to-interference-and-noise ratio (SINR) of RA UE 1 is expressed by

$$\text{SINR}_{1\infty} = \frac{\beta_1^2}{\sum_{u' \in \Phi_{\psi_1}} \beta_{u'}^2}. \quad (6)$$

As we can see, even the number of antennas increases without bound, it does not change the fact that the interference from the collided RA UEs that select the same preamble as RA UE 1 cannot be vanished and could have a significant impact on the performance of RA UE 1.

In co-located mMIMO, since all M APs are geographically centralized, we have $\beta_{u1} = \beta_{u2} = \dots = \beta_{uM} \triangleq \beta_u$ and thus the observation in (6) still holds. Unfortunately, in co-located mMIMO, since all the signals are multiplexed and assembled at the centralized BS, it is difficult to find preamble multiplicity under preamble collision and the performance impairment of collided RA

UEs exists no matter where collided RA UEs are spatially located. However, distributed mMIMO opens up chances for mitigating the impairment thanks to the signal spatial sparsity in distributed mMIMO and random geographic distributions of RA UEs.

C. Preamble Collision Resolution in Distributed mMIMO based GFRA

In distributed mMIMO, considering the distance disparity between an RA UE and different APs, it is demonstrated that only neighboring APs within a communication range of an RA UE have non-negligible channel gains. Since all the collided RA UEs are uniformly and independently distributed in the area, they can be separate in space and surrounded by different groups of APs. If the BS CPU can distinguish the neighboring APs of a collided RA UE in GFRA, it can organize the neighboring APs to serve the collided RA UE, which is expected to improve the performance of collided RA UEs significantly.

Herein, we use a toy example to explain the potential performance gain achieved by such a strategy. In particular, we assume that RA UE 1 is far away from the RA UEs in Φ_{ψ_1} so that the strength of received signals from the other collided RA UEs is negligible at the neighboring APs of RA UE 1 (for example in Figure 1, RA UE 1 and RA UE 2 are the collided RA UEs that select the same preamble but their locations are far away from each other). For simplicity, let $\mathcal{M}_1 = [1, 2, \dots, M_1]$ denote the set of indices of neighboring APs of RA UE 1 and $M_1 = |\mathcal{M}_1| = \omega_1 M$, where ω_1 ($0 < \omega_1 \ll 1$) is a scaling factor that represents the ratio of the sizes of a communication range of RA UE 1 to the considered area.

By only employing the M_1 APs to decode data, similar to (2), the channel estimate of RA UE 1 over the M_1 APs, $\hat{\mathbf{g}}_{1,\mathcal{M}_1} \in \mathbb{C}^{M_1 S}$, can be written by

$$\hat{\mathbf{g}}_{1,\mathcal{M}_1} = \mathbf{g}_{1,\mathcal{M}_1} + \sum_{u' \in \Phi_{\psi_1}} \mathbf{g}_{u',\mathcal{M}_1} + \frac{1}{\sqrt{\rho_T L}} \mathbf{n}_{\mathcal{M}_1}, \quad (7)$$

where $\mathbf{g}_{u,\mathcal{M}_1} = [\mathbf{g}_{u1}^T, \mathbf{g}_{u2}^T, \dots, \mathbf{g}_{uM_1}^T]^T$ and $\mathbf{n}_{\mathcal{M}_1} \sim \mathcal{CN}(0, \mathbf{I}_{M_1 S})$.

Similar to (3), the received data symbol vector over the M_1 APs, $\mathbf{r}_{\mathcal{M}_1} \in \mathbb{C}^{M_1 S}$, is written by

$$\mathbf{r}_{\mathcal{M}_1} = \sum_{u=1}^U \sqrt{P_T} \mathbf{g}_{u,\mathcal{M}_1} s_u + \bar{\mathbf{n}}_{\mathcal{M}_1}, \quad (8)$$

where $\bar{\mathbf{n}}_{\mathcal{M}_1} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{M_1 S})$.

In the considered example, due to significant large-scale fading between RA UE u' in Φ_{ψ_1} and AP m in $\mathcal{M}_1$, $\mathbf{g}_{u',\mathcal{M}_1} \approx \mathbf{0}$, $u' \in \Phi_{\psi_1}$. Thus, we have following approximations:

$$\hat{\mathbf{g}}_{1,\mathcal{M}_1} \approx \mathbf{g}_{1,\mathcal{M}_1} + \frac{1}{\sqrt{\rho_T L}} \mathbf{n}_{\mathcal{M}_1},$$

and

$$\mathbf{r}_{\mathcal{M}_1} \approx \sqrt{P_T} \mathbf{g}_{1,\mathcal{M}_1} s_1 + \sum_{\substack{u \notin \Phi_{\psi_1} \\ u \neq 1}} \sqrt{P_T} \mathbf{g}_{u,\mathcal{M}_1} s_u + \bar{\mathbf{n}}_{\mathcal{M}_1}.$$

Then, the estimated data symbol of RA UE 1 after conjugate beamforming as $M \rightarrow \infty$ (as $M_1 = \omega_1 M$, $M \rightarrow \infty$ leads to $M_1 \rightarrow \infty$) becomes

$$\begin{aligned} \hat{s}_{1\infty} &= \lim_{M_1 \rightarrow \infty} \frac{\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{r}_{\mathcal{M}_1}}{M_1 S \sqrt{P_T}} \\ &\stackrel{(b)}{\approx} \lim_{M_1 \rightarrow \infty} \sum_{m=1}^{M_1} \frac{\beta_{1m} s_1}{M_1} = \beta_{1,\mathcal{M}_1} s_1. \end{aligned} \quad (9)$$

Similarly, $\stackrel{(b)}{\approx}$ is obtained based on Chebyshev's Theorem as $M_1 \rightarrow \infty$ and $\beta_{1,\mathcal{M}_1} \triangleq \lim_{M_1 \rightarrow \infty} \sum_{m=1}^{M_1} \frac{\beta_{1m}}{M_1}$.

From (9), we can see that the received signal of RA UE 1 approximately becomes interference-free from preamble collision under the given scenario as $M \rightarrow \infty$, which indicates the possibility and practicality of preamble collision resolution (mitigating the interference due to preamble collision) in GFRA with distributed mMIMO.

Nevertheless, to achieve the potential preamble collision resolution and improve the performance of collided RA UEs in GFRA by distributed mMIMO, there are two issues remained to be addressed as follows:

- How to detect preamble collision and find preamble multiplicity?
- How to differentiate the neighboring APs of collided RA UEs for performance enhancement?

To address the above issues, it is expected to fully exploit the information obtained from the received preamble signals at APs. To this end, we propose a machine learning based framework solution in this paper.

Specifically, to mitigate the performance impairment of collided RA UEs in GFRA with distributed mMIMO, we first design a simple DNN to enable the preamble multiplicity estimation, where a data-driven ED method is also proposed for performance comparison. With the estimated preamble multiplicity, we then employ the K -means clustering algorithm to separate

the neighboring APs of collided RA UEs and use each associated AP cluster to serve individual collided RA UE.

III. DNN BASED PREAMBLE MULTIPLICITY ESTIMATION

For the estimation of preamble multiplicity of an arbitrary preamble, e.g., $\mathbf{p}_l$, $l = 1, 2, \dots, L$, we need to find out the mapping relationship between the preamble multiplicity and the received preamble signal associated with $\mathbf{p}_l$ at APs, i.e.,

$$F: \mathbb{C}^{MS} \rightarrow \mathbb{N}_0$$

$$\mathbf{g}_{\mathcal{B}_l} \mapsto B_l, \quad l = 1, 2, \dots, L \quad (10)$$

where F denotes the mapping function, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathcal{B}_l$ is the set of indices of RA UEs that select $\mathbf{p}_l$ among U RA UEs and $B_l = |\mathcal{B}_l| \in \mathbb{N}_0$ denotes the preamble multiplicity (e.g., $B_l = 0$ indicates that $\mathbf{p}_l$ is not selected by any RA UE), and $\mathbf{g}_{\mathcal{B}_l} \in \mathbb{C}^{MS}$ represents the received preamble signal associated with $\mathbf{p}_l$ at APs, which has a similar expression as in (2) and it is given by

$$\mathbf{g}_{\mathcal{B}_l} = \frac{\mathbf{Y}\mathbf{p}_l^*}{\sqrt{P_T}L} = \sum_{u \in \mathcal{B}_l} \mathbf{g}_u + \frac{1}{\sqrt{\rho_T}L}\mathbf{n}, \quad (11)$$

where $\mathbf{g}_{\mathcal{B}_l} = [\mathbf{g}_{\mathcal{B}_l 1}^T, \mathbf{g}_{\mathcal{B}_l 2}^T, \dots, \mathbf{g}_{\mathcal{B}_l M}^T]^T$ and $\mathbf{g}_{\mathcal{B}_l m}$ is the received preamble signal vector associated with $\mathbf{p}_l$ at AP m .

In the considered problem, obtaining F by traditional programming algorithms is not a trivial task since deriving a general mathematical detection model to recognize subtle patterns associated with different preamble multiplicities and summarize the random patterns of RA UEs' geographic locations and wireless environments in GFRA is too complex and may be infeasible.

A. Proposed DNN Structure

To solve the problem in an effective manner, we design a feed-forward DNN (multi-layer perception) [30] in this section thanks to its powerful approximation and prediction ability. As aforementioned in Section II-C, only neighboring APs within a communication range of an RA UE have non-negligible channel gains in distributed mMIMO. Thus, the neighboring APs of an RA UE usually capture more significant signal energy than the other APs. On the other hand, the separation of RA UEs in space makes the signals of different RA UEs concentrate

Figure 2: A simplified illustration of the proposed DNN diagram for preamble multiplicity estimation, where the circle nodes of different colors represent the neurons of different layers.

in different geographic clusters. By capitalizing these properties, one could envision that there exists connection between the preamble multiplicities and the distribution patterns of received preamble signal energy over M distributed APs. Therefore, the proposed DNN is used to explore the features so that the desired function F can be approximately modelled.

As illustrated in Figure 2, the proposed fully connected DNN consists of $J+2$ layers, including one input layer (layer 0), J hidden layers (layers 1 to J), and one output layer (layer $J+1$). Let N_j denote the number of neurons at layer j , $j = 0, 1, \dots, J+1$.

In the proposed DNN, layer 0 contains $N_0 = M$ neurons, which forwards the instantaneous information of $\mathbf{g}_{B_l}$ to the following layers. Let $\mathbf{E}_{B_l} = [E_{B_l1}, E_{B_l2}, \dots, E_{B_lM}]^T \in \mathbb{R}^M$ denote the received preamble signal energy vector associated with $\mathbf{p}_l$, where

$$E_{B_lm} = \frac{\|\mathbf{g}_{B_lm}\|^2}{S}, m = 1, 2, \dots, M, \quad (12)$$

denotes the received preamble signal energy associated with $\mathbf{p}_l$ at AP m . With $\mathbf{E}_{B_l}$, the input vector of layer 0, denoted by $\mathbf{I}_0 = [I_{01}, \dots, I_{0m}, \dots, I_{0M}]^T \in \mathbb{R}^M$, is given by

$$\mathbf{I}_0 = \text{sort}_D(\mathbf{E}_{B_l}), \quad (13)$$

where $\text{sort}_D(\cdot)$ is a function that sorts the elements in descending order.

In Figure 3, we plot an example to illustrate the pattern features of normalized $\mathbf{I}_0$ corresponding to three different preamble multiplicities with $M = 100$, where we randomly generate three sample sets for each individual preamble multiplicity based on the system setup in Section III-C1 and perform normalization among them. As shown in the example, different preamble

multiplicities lead to different received energy patterns, which are exploited by the proposed DNN to predict preamble multiplicity. Note that we only use nine sample sets in Figure 3 to illustrate the received energy pattern differences of different preamble multiplicities. In practice, the received energy patterns become more complicated as more sample sets are involved for normalization.

Figure 3: An example of pattern differences of normalized $\mathbf{I}_0$ corresponding to different preamble multiplicities with $M = 100$, where the x-label is the index of the elements in $\mathbf{I}_0$.

Building on $\mathbf{I}_0$, the hidden layers of the feed-forward DNN are constructed through the following J iterative processing steps:

$$\mathbf{I}_j = f(\mathbf{W}_{j-1}\mathbf{I}_{j-1} + \mathbf{b}_{j-1}), j = 1, 2, \dots, J, \quad (14)$$

where $\mathbf{I}_j \in \mathbb{R}^{N_j}$ is the output of layer j , $f(\cdot)$ represents a non-linear activation function, and $\mathbf{W}_{j-1} \in \mathbb{R}^{N_j \times N_{j-1}}$ and $\mathbf{b}_{j-1} \in \mathbb{R}^{N_j}$ respectively stand for the weighting matrix and bias vector at layer $j - 1$, which are used to encode the output of layer $j - 1$. In this paper, a sigmoid function defined as $\sigma(x) = \frac{1}{1+e^{-x}}$ is used as the activation function $f(\cdot)$.

As usually done in pattern recognition problems, the predicted output of the proposed DNN in layer $J + 1$ is expressed by

$$\hat{\mathbf{I}}_{J+1} = \text{softmax}(\mathbf{W}_J\mathbf{I}_J + \mathbf{b}_J), \quad (15)$$

where $\hat{\mathbf{I}}_{J+1} = [\hat{I}_{J+1,1}, \hat{I}_{J+1,2}, \dots, \hat{I}_{J+1,N_{J+1}}]^T \in \mathbb{R}^{N_{J+1}}$ and $\text{softmax}(\cdot)$ is the softmax function

[30]. In the proposed DNN, N_{J+1} is set to $T_{\max} + 1$, where $T_{\max}$ is the maximum number of colliding RA UEs that we are interested in detecting. Thus, the output $\hat{\mathbf{I}}_{J+1}$ represents a predicted probability distribution over the $T_{\max} + 1$ different possible preamble multiplicities and the value of $\hat{I}_{J+1,i}$ indicates the predicted probability that the preamble multiplicity equals $i - 1$, $i = 1, 2, \dots, T_{\max} + 1$.

In the light of this, the estimated preamble multiplicity of $\mathbf{p}_l$ according to the output of proposed DNN is given by

$$\hat{B}_l = \arg \max(\hat{\mathbf{I}}_{J+1}) - 1, \quad (16)$$

where $\arg \max(\hat{\mathbf{I}}_{J+1})$ returns the index of the largest entry of $\hat{\mathbf{I}}_{J+1}$.

B. Training Phase

For an accurate approximation of the desired function F , the proposed DNN needs to learn and adjust the parameter sets of $\boldsymbol{\theta}_j = \{\mathbf{W}_j, \mathbf{b}_j\}$, $j = 0, 1, \dots, J$, in the training phase.

Specifically, based on the system configurations in GFRA, we randomly generate Q training sample sets. For sample set q ($q = 1, 2, \dots, Q$), it consists of a pair of the received preamble signal energy $\mathbf{E}_{B^q}$ and the corresponding preamble multiplicity B^q , which is associated with an arbitrary preamble (here we omit the subscript l for notation simplicity). With known training sample set q , we can obtain a mapping pair of the proposed DNN, denoted by $(\mathbf{I}_0^q, \mathbf{I}_{J+1}^q)$, where the input $\mathbf{I}_0^q$ is obtained from $\mathbf{E}_{B^q}$ by (13), and the target output $\mathbf{I}_{J+1}^q$ can be expressed by

$$\mathbf{I}_{J+1}^q = \mathbf{e}_{B^q+1}, \quad (17)$$

where $\mathbf{e}_i \in \mathbb{R}^{T_{\max}+1}$ denotes the standard basis vector that has a single nonzero entry with value 1 at entry i .

By using the Q mapping pairs, the proposed DNN is trained by back-propagation (BP) algorithm to adjust and optimize the parameter sets of all layers, i.e., $\boldsymbol{\theta} = [\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J]$, so that the cross-entropy loss function between the target outputs $\{\mathbf{I}_{J+1}^q\}_{q=1}^Q$ and the predicted outputs $\{\hat{\mathbf{I}}_{J+1}^q\}_{q=1}^Q$, which is given in (18), could be minimized:

$$\mathcal{L}(\boldsymbol{\theta}) = - \sum_{q=1}^Q \sum_{i=1}^{T_{\max}+1} I_{J+1,i}^q \ln \left(\hat{I}_{J+1,i}^q(\mathbf{I}_0^q, \boldsymbol{\theta}) \right). \quad (18)$$

To achieve the minimization of $\mathcal{L}(\theta)$, a number of out-of-the-box gradient methods including the gradient descent and the conjugate gradient can be used. In this paper, we employ the scaled conjugate gradient method [31] to iteratively update parameter sets θ [32].

C. Performance Analysis of Proposed DNN

1) *Simulation Setup:* In this subsection, we evaluate the performance of the proposed DNN for preamble multiplicity estimation under practical wireless environments, where the large-scale fading coefficient, depending on the RA UE's location and the propagation environment, is modelled as [33],

$$\beta_{um} = \frac{X_{um}}{1 + \text{PL}(d_0) \left(\frac{d_{um}}{d_0} \right)^v}, \quad (19)$$

where X_{um} stands for the the shadow fading that is a log-normal random variable with standard deviation σ_{SF} (dB), $\text{PL}(d_0)$ is the path loss at a reference distance d_0 (m), d_{um} is the distance between UE u and AP m , and v is the path loss exponent. The additive thermal noise is assumed to have a power spectral density of -174 dBm/Hz, while the front-end receiver at the AP is assumed to have a noise figure of 9 dB according to [28]. Thus, the noise power σ^2 is -112 dBm with a narrow bandwidth of $B_w = 200$ KHz.

We consider a square area of 1 km^2 and the distributed APs are deployed on a square grid². Three different deployments are considered: 1) $M = 10 \times 10$ APs with $S = 2$ antenna; 2) $M = 10 \times 10$ APs with $S = 1$ antenna; and 3) $M = 7 \times 7$ APs with $S = 2$ antennas. The rest of system parameters are summarized in Table I.

For the proposed DNN, we simulate $Q = 10^5$ realizations to generate sample sets. To improve generalization and avoid overfitting, the sample sets are randomly divided into training sample sets (80% of total instances), validation sample sets (10% of total instances), and test sample sets (10% of total instances). Moreover, the minimum performance gradient is set to be 10^{-6} . The maximum number of epochs to train is set to be 1000. And the maximum number of validation checks is set to be 8. For each realization, the active RA UEs are randomly distributed

²Note that the real AP deployments could be irregular due to network planning and other topological and demographic factors. However, the APs cannot be randomly distributed in practice because it may lead to APs being placed very close to each other, which generally does not make sense. Thus, APs are typically distributed more regularly than randomly distributed. For this reason, the widely accepted square-grid-based AP deployment model is considered in this paper. Moreover, this paper takes into account the shadow fading. The randomness caused by shadow fading coefficients can be seen as displacing the APs and varying the distances between the APs and the RA UEs [24]. Therefore, the considered model is a sensible model for AP deployments.

Table I: System parameters

Number of UEs N	2000
Activation Probability ρ	0.01
Number of Preambles L	20
Transmit Power P_T	17 dBm
Shadow Fading σ_{SF}	0 or 8 dB
Reference Distance d_0	1 m
Path Loss PL(d_0)	30 dB
Path Loss Exponent v	3.8

in the considered area and their number U is generated following the binomial distribution $\text{Bino}(N, \rho)$. Since each preamble is selected by RA UEs uniformly at random in GFRA, the preamble multiplicity B associated with an arbitrary preamble follows the binomial distribution $\text{Bino}(N, \rho/L)$ as mentioned earlier. With the given system parameters, i.e., $N = 2000$, $\rho = 0.01$, and $L = 20$, over 99% realizations are generated in the way that a preamble is selected by 4 RA UEs at most. As a result, we set $T_{\max} = 4$ as the maximum preamble multiplicity that we are interested in estimating. In the following, we discuss the performance of the proposed DNN in terms of classification accuracy and reliability in different deployments.

2) *Performance Analysis for Different Deployments:* We first consider the performance of deployment 1 with $M = 10 \times 10$ and $S = 2$. In this deployment, the proposed DNN consists of 4 hidden layers, whose numbers of neurons are 128, 128, 64, and 32, respectively.

Table II: Confusion matrix for the proposed DNN in the deployment of $M = 100$ and $S = 2$.

		Predicted $\hat{B}$				
		0	1	2	3	4
Target B	0	1	0	0	0	0
	1	0.002	0.997	0.001	0	0
	2	0	0.012	0.979	0.009	0
	3	0	0	0.067	0.918	0.015
	4	0	0	0.002	0.120	0.878
$\sigma_{SF} =$ 0 dB	0	1	0	0	0	0
	1	0.002	0.991	0.007	0	0
	2	0	0.046	0.923	0.031	0
	3	0	0	0.151	0.838	0.011
	4	0	0	0.002	0.211	0.787
$\sigma_{SF} =$ 8 dB	0	1	0	0	0	0
	1	0.002	0.991	0.007	0	0
	2	0	0.046	0.923	0.031	0
	3	0	0	0.151	0.838	0.011
	4	0	0	0.002	0.211	0.787

To understand the classification accuracy of the proposed DNN for preamble multiplicity, confusion matrices for different σ_{SF} in Table II are included. In the scenario with no shadow fading, i.e., $\sigma_{\text{SF}} = 0$, it is seen that the proposed DNN model is able to predict (estimate) the preamble multiplicity in GFRA with high accuracy over a wide range of multiplicities. In terms of the performance of preamble detection, i.e., determining whether a preamble is selected or not by any RA UE, the proposed DNN model provides almost error-free performance, with negligible false alarm and missed detection errors (0.2% missed detection probability occurs merely when preamble is selected by only one RA UE). When the preamble collision occurs, about 98% and 88% estimation accuracy can be achieved when the preamble multiplicity equals 2 and 4, respectively. The main reason that the estimation accuracy declines as the preamble multiplicity increases is due to the fact that, a larger preamble multiplicity means that more collided RA UEs select the same preamble. As a consequence, there are comparably more chances that some of the collided RA UEs are co-located in vicinity, which could make the proposed DNN mistakenly treat these close-located RA UEs as a single RA UE and results in an incorrect estimated preamble multiplicity that is smaller than the actual one. Therefore, the estimation performance is degraded. Nevertheless, it is noticed that, almost all the incorrect estimated multiplicities are only offset by 1 compared to the actual ones. For example, when the preamble multiplicity equals 2 and 3, the proposed DNN only gets the multiplicity wrong by ± 1 (mostly by -1). When the preamble multiplicity equals 4, the proposed DNN guarantees an estimation result that is either correct or incorrect by ± 1 with a high probability of 99.8% (only gets the multiplicity incorrect by -2 with a as little as 0.2% probability). These observations demonstrate the accuracy as well as the reliability achieved by the proposed DNN for preamble multiplicity estimation.

In addition, we also consider a practical channel scenario with shadow fading $\sigma_{\text{SF}} = 8$. Under such a condition, it is not surprising that, due to the impact of shadow fading variations on channel gains, the classification accuracy of the proposed DNN degrades compared to the case without shadow fading. As we can see, although shadow fading has little impact on the preamble detection performance of the proposed DNN, it incurs certain accuracy degradations for estimating collided preamble multiplicities. For instance, the estimation accuracy for a preamble multiplicity of 4 is decreased from 87.8% to 78.7% and more errors are introduced by incorrect estimation to multiplicity 3, which indicates that the channel randomness induced by shadow fading inherently increases confusion between adjacent multiplicity classes. Nevertheless, an

estimation accuracy of 78.7% for preamble multiplicity 4 is still considered decent, under such an amount of collided RA UEs coexists at the same time. Besides, similar to what we observed in the case with no shadow fading, almost all the incorrect estimated multiplicities differ from the true ones by ± 1 when $\sigma_{\text{SF}} = 8$, which reveals that although the accuracy performance of the proposed DNN is affected by the shadow fading, its estimation reliability remains uninfluenced.

Table III: Confusion matrix for the proposed DNN in the deployment of $M = 100$ and $S = 1$.

		Predicted $\hat{B}$				
		0	1	2	3	4
Target B	0	1	0	0	0	0
	1	0.004	0.992	0.004	0	0
	2	0	0.038	0.943	0.019	0
	3	0	0	0.116	0.864	0.020
	4	0	0	0.002	0.218	0.780
$\sigma_{\text{SF}} = 8 \text{ dB}$	0	1	0	0	0	0
	1	0.006	0.981	0.013	0	0
	2	0	0.051	0.919	0.030	0
	3	0	0	0.174	0.786	0.040
	4	0	0	0.004	0.269	0.727

We also consider other two deployments, i.e., deployment 2 with $M = 10 \times 10$ and $S = 1$ and deployment 3 with $M = 7 \times 7$ and $S = 2$. Their confusion matrices are illustrated in Table III and Table IV, respectively. In deployment 2, the proposed DNN consists of 4 hidden layers, whose numbers of neurons are the same as those in deployment 1. In deployment 3, the proposed DNN consists of 4 hidden layers, whose numbers of neurons are 64, 128, 64, and 32, respectively. As observed in Table II, similar observations and conclusions can be drawn from Tables III and IV. In terms of estimation accuracy of the proposed DNN, we can see that it is slightly degraded in deployment 2 compared to that in deployment 1, which is mainly due to a loss of channel diversity in deployment 2 with $S = 1$. Nevertheless, a 72.7% estimation accuracy for preamble multiplicity 4 is still achievable with $\sigma_{\text{SF}} = 8$ in deployment 2. Moreover, with the roughly same amount of antennas in deployments 2 and 3, a reasonably close multiplicity estimation performance is observed without considering shadow fading. However, results show that the performance in deployment 3 seems more sensitive to the channel randomness resulted from shadow fading. In particular, under $\sigma_{\text{SF}} = 8$, its estimation accuracy for preamble multiplicities 3 and 4 is significantly degraded compared to that under $\sigma_{\text{SF}} = 0$. This could be explained

by the fact that the antenna distribution in deployment 3 is more sparse, which makes that the shadow fading comparably has more significant impact on the channel fluctuations. As a result, the classification confusion between preamble multiplicities 3 and 4 gets more pronounced.

Table IV: Confusion matrix for the proposed DNN in the deployment of $M = 49$ and $S = 2$.

		Predicted $\hat{B}$				
		0	1	2	3	4
Target B	$\sigma_{\text{SF}} = 0 \text{ dB}$	0	1	0	0	0
		1	0.003	0.994	0.003	0
		2	0	0.034	0.946	0.020
		3	0	0	0.138	0.830
		4	0	0	0.002	0.765
	$\sigma_{\text{SF}} = 8 \text{ dB}$	0	1	0	0	0
		1	0.003	0.971	0.026	0
		2	0	0.061	0.924	0.015
		3	0	0	0.313	0.663
		4	0	0	0.018	0.586

D. Comparison to Threshold-Based ED Method

To further validate the effectiveness of the proposed DNN based method for preamble multiplicity estimation, we consider a threshold-based ED method for performance comparison in this subsection. As indicated in Figure 3, different preamble multiplicities, to some extent, lead to different levels of received energy after normalization among training sample sets. For this reason, a simple threshold-based ED (T-ED) method for preamble multiplicity estimation can be developed by the following steps:

- 1) Perform energy normalization among Q training sample sets between -1 and 1 and store the normalization setting. For sample set q ($q = 1, 2, \dots, Q$), its normalized energy set is denoted by $\tilde{\mathbf{I}}_0^q$.
- 2) Denote $\mathcal{V}_B$ as the set of indices of sample sets associated with preamble multiplicity B ($B = 0, 1, \dots$). Calculate the average normalized energy per AP for preamble multiplicity B as $\zeta_B = \frac{\sum_{q \in \mathcal{V}_B} \|\tilde{\mathbf{I}}_0^q\|_1}{M|\mathcal{V}_B|}$, where $\|\cdot\|_1$ is the Taxicab norm.
- 3) Obtain threshold Th_B for preamble multiplicity estimation, where $\text{Th}_0 = -1$ and $\text{Th}_B = \frac{\zeta_B + \zeta_{B+1}}{2}$ for $B \neq 0$.

- 4) For a new received energy set $\mathbf{I}_0$, apply the normalization setting to it and obtain its average normalized energy per AP as $\frac{\|\tilde{\mathbf{I}}_0\|_1}{M}$. If $\text{Th}_B < \frac{\|\tilde{\mathbf{I}}_0\|_1}{M} \leq \text{Th}_{B+1}$, its preamble multiplicity is estimated as B .

Table V: Confusion matrix of the T-ED method in deployment 1 with $\sigma_{\text{SF}} = 8$.

		Predicted $\hat{B}$				
		0	1	2	3	4
Target B	0	1	0	0	0	0
	1	0.107	0.892	0.001	0	0
	2	0	0.131	0.844	0.025	0
	3	0	0	0.155	0.764	0.081
	4	0	0	0	0.281	0.719

Like the proposed DNN, the developed T-ED method is a data-driven based method. For performance comparison, its confusion matrix in deployment 1 with $\sigma_{\text{SF}} = 8$ is presented in Table V. Comparing the results to the ones in Table II, we see that the estimation performance of the T-ED method is not as good as that of the proposed DNN. This is because that the T-ED method is unable to exploit the received energy patterns over APs of different preamble multiplicities as the proposed DNN does. Since similar conclusions can be drawn as in deployment 1, we omit the confusion matrices of the T-ED method in deployments 2 and 3. In Section V, performance comparison between the DNN based scheme and the T-ED based scheme in terms of uplink achievable rate per collided RA UE will be illustrated in Figure 6.

Discussion: For the proposed DNN, training with the setup of $M = 100$ requires approximately 400 iterations, which in total take around 10 minutes. The training time becomes shorter with a smaller M . We note that the training cost is incurred offline and the cost burden of the proposed DNN is on the BS CPU, whose capabilities are rapidly improving with artificial intelligence (AI) processors and with little computational and energy restrictions. Thus, the incurred cost and computational complexity may not be a limiting factor. Besides, the training frequency depends on the environment stability, e.g., stability of the placement and the number of APs. Since the placement and the number of APs over a covered area should remain unchanged over a long period of time in practice, the training should be performed infrequently.

In the next section, the estimated preamble multiplicity information will be used to cluster neighboring APs of collided RA UEs for their performance enhancement.

IV. PROPOSED AP CLUSTERING ALGORITHM

The estimated preamble multiplicity $\hat{B}$ (associated with an arbitrary preamble) based on the proposed DNN indicates the status of associated preamble in GFRA, i.e., whether or not it is selected by any RA UE, and if selected then how many RA UEs select it. When $\hat{B} \geq 2$, the BS CPU assumes that preamble collision occurs. Under such conditions, as revealed in Section II-C, it is expected that the BS CPU only allocates the neighboring APs of a collided RA UE (rather than all the APs) to decode its data so that the mutual interference among collided RA UEs in the preamble domain can be mitigated. In the light of this, we propose a K -means AP clustering algorithm in this section.

A. K -Means AP Clustering Algorithm

In this paper, we denote M_c as the average number of neighboring APs to decode for each collided RA UE in the case of preamble collision. Ideally, the neighboring APs of a collided RA UE can be the M_c APs with its strongest channel gains [34]. In practice, this scenario is desirable, but unattainable in GFRA since the BS CPU has no prior CSI of RA UEs. As a compromised solution, the K -means AP clustering algorithm is proposed to cluster neighboring APs for collided RA UEs.

On one hand, the neighboring APs in the vicinity of an RA UE usually capture more significant signal energy than other APs. As the collided RA UEs are randomly distributed in space, it can be reasonably envisaged that the APs with $M_c \hat{B}$ strongest received preamble energy are most likely composed by the neighboring APs of $\hat{B}$ collided RA UEs. On the other hand, the K -means clustering algorithm is one of the most popular clustering algorithms, which aims to partition observations into K clusters where each observation belongs to exactly one cluster with the nearest mean cluster centroid [35]. For these reasons, it motivates us to propose the K -means AP clustering algorithm that iteratively partitions the APs corresponding to largest $M_c \hat{B}$ entries of $\mathbf{E}_B$ into $\hat{B}$ clusters based on their coordinates. Note that the deployment of distributed APs along with their coordinates are pre-determined and known at the BS CPU.

Herein, we denote $\mathcal{A}_{\hat{B}}$ as the set of indices of APs corresponding to the largest $M_c \hat{B}$ entries of $\mathbf{E}_B$ and $|\mathcal{A}_{\hat{B}}| = M_c \hat{B}$. Then, we have $\mathcal{C}_{\hat{B}} = \{\mathbf{c}_m \mid m \in \mathcal{A}_{\hat{B}}\}$ as the coordinate set of the APs in $\mathcal{A}_{\hat{B}}$, where $\mathbf{c}_m = [x_m, y_m]^T$ denotes the coordinate of AP m , $m \in \mathcal{A}_{\hat{B}}$, in a 2-dimensional Euclidean space.

With $\mathcal{C}_{\hat{B}}$, the proposed K -means AP clustering algorithm is described in Algorithm 1.

Algorithm 1 Proposed K -means AP clustering algorithm

Input: M_c , $\hat{B}$, and $\mathcal{C}_{\hat{B}}$;
Output: A set of $\hat{B}$ clusters, i.e., $\mathcal{Z}_k = \{m \mid z_m = k, m \in \mathcal{A}_{\hat{B}}\}$, $k = 1, 2, \dots, \hat{B}$;
Initialization: Randomly select $\hat{B}$ coordinates from $\mathcal{C}_{\hat{B}}$ as the initial cluster centroids $\mu_1, \mu_2, \dots, \mu_{\hat{B}}$;
1: **repeat**
2: AP assignment, i.e, assign each AP in $\mathcal{C}_{\hat{B}}$ to its closest cluster centroid with label:
 $z_m = \arg \min_k \|\mathbf{c}_m - \mu_k\|^2$;
3: Update the cluster centroids, i.e., compute the mean coordinates of APs assigned in each
cluster to obtain new cluster centroid: $\mu_k = \frac{\sum_{m=1}^{M_c \hat{B}} \mathbb{1}(z_m=k) \mathbf{c}_m}{\sum_{m=1}^{M_c \hat{B}} \mathbb{1}(z_m=k)}$;
4: **until** {Cluster centroids are stabilized}

With the AP clusters $\{\mathcal{Z}_k\}_{k=1}^{\hat{B}}$, the BS CPU deems that there exists one collided RA UE in the vicinity of each AP cluster, and organizes each cluster to decode the received data individually.

B. Exemplary Outputs of Proposed Algorithm

With a predetermined M_c , the outcome of the proposed AP clustering algorithm relies on the estimated preamble multiplicity of proposed DNN. As observed and discussed in Section III-C, the proposed DNN is able to provide a decent estimation accuracy for each preamble multiplicity. In most error multiplicity estimation, the proposed DNN only gets the multiplicity wrong by -1 . For instance, for preamble multiplicity 3 in deployment 2 with $\sigma_{\text{SF}} = 8$ dB, an estimation accuracy of 78.6% is achieved and an estimation error of 17.4% is caused by mistakenly classifying it as multiplicity 2. Based on these facts, we present two kinds of representative outcomes of the proposed AP clustering algorithm with $M_c = 4$ in Figure 4(a) and Figure 4(b), respectively. Specifically, under deployment 2 with $\sigma_{\text{SF}} = 8$ dB, the outcome in Figure 4(a) represents a typical clustering output for correctly estimated preamble multiplicity 3, while the outcome in Figure 4(b) represents a typical clustering output for incorrectly estimated preamble multiplicity 3. In both subfigures, red empty circles represent the locations of $M = 100$ deployed APs in a 2-dimensional Euclidean space, blue diamonds represent the locations of collided RA UEs (associated with a certain preamble) that are randomly generated, and the APs filled with same color form an AP cluster outputted by the proposed AP clustering algorithm. Note that there exists other RA UEs in space, however, since their signals are orthogonal to those of the displayed collided RA UEs in the preamble domain, we omit them herein.

(a) With correct preamble multiplicity estimation.

(b) With incorrect preamble multiplicity estimation.

Figure 4: Exemplary outputs of Algorithm 1 with $M_c = 4$, in deployment 2 with $\sigma_{SF} = 8$ dB.

As shown in Figure 4(a), RA UEs 1, 5, and 14 select the same preamble over the same GFRA channel. With the correct preamble multiplicity estimation, the proposed AP clustering algorithm divides the 12 APs with the strongest received preamble signal energy into 3 clusters. Since the three collided RA UEs are geographically separate, different collided RA UEs are surrounded by different clusters of APs, each of which represents a group of APs capturing the strongest signals from a specific collided RA UE. Using each individual AP cluster rather

than all the APs to decode the data of the collided RA UE in its vicinity is beneficial since the mutual interference due to preamble collision is significantly discarded. In Figure 4(b), RA UEs 1, 5, and 7 select the same preamble. Since RA UEs 1 and 5 are located in vicinity to each other, they are treated as a single collided RA UE by the proposed DNN, which thus mistakenly estimates the preamble multiplicity as 2. With the incorrect preamble multiplicity estimation, the proposed AP clustering algorithm divides the 8 APs with the strongest received preamble signal energy into 2 clusters. In this example, RA UEs 1 and 5 share the same AP cluster. Although it is unable to mitigate the mutual interference between them, the interference from RA UE 7 is mitigated by only using their neighboring clustered APs rather than all the APs. On the other hand, for RA UE 7, the interference from the other two RA UEs becomes small at the clustered APs surrounding it and using them only to decode its data is thus beneficial.

In fact, the selection of M_c is of particular importance in the proposed K -means AP clustering. On one hand, since the signals of an RA UE are usually concentrated at its few neighboring APs, selecting too large M_c brings little benefit of increasing the amount of desired signals, but introduces comparably large interference from all of the other RA UEs in GFRA. On the other hand, selecting too small M_c could have little benefit of further mitigating interference, but provides comparably small amount of desired signals due to limited array gain. Therefore, a proper value of M_c in the proposed K -means AP clustering needs to be selected. Unfortunately, it is difficult to determine M_c in a closed-form expression in the considered scenario. In this paper, we shed light on its proper value through simulation by evaluating the performance in terms of the 95% likely achievable rate per collided RA UE.

V. SIMULATION RESULTS

In this section, performance evaluation on the uplink achievable rate per collided RA UE is conducted by simulation to validate the effectiveness of the proposed AP clustering schemes in preamble collision resolution in GFRA. The simulation setup and parameters are the same as given in Section III-C1.

Like in Section II-C, we consider RA UE 1 as the collided RA UE of interest under preamble collision ($|\Phi_{\psi_1}| \geq 1$) and $\mathcal{M}_1$ as the set of indices of APs employed for decoding data of RA

UE 1. Based on (7) and (8), the estimated data symbol of RA UE 1 is given by

$$\begin{aligned}\hat{s}_1 &= \frac{\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{r}_{\mathcal{M}_1}}{M_1 S \sqrt{P_T}} \\ &= \frac{\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{g}_{1,\mathcal{M}_1} s_1}{M_1 S} + \frac{\sum_{u' \in \Phi_{\psi_1}} \hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{g}_{u',\mathcal{M}_1} s'_u}{M_1 S} + \frac{\sum_{\substack{u \notin \Phi_{\psi_1} \\ u \neq 1}} \hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{g}_{u,\mathcal{M}_1} s_u}{M_1 S} + \frac{\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \bar{\mathbf{n}}_{\mathcal{M}_1}}{M_1 S \sqrt{P_T}}.\end{aligned}\quad (20)$$

From (20), the uplink achievable rate of RA UE 1 under preamble collision is given by

$$\mathcal{R}_1 = B_w \log(1 + \text{SINR}_1), \quad (21)$$

where SINR_1 is the uplink SINR of RA UE 1, which is given by [34]

$$\text{SINR}_1 = \frac{\rho_T |\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{g}_{1,\mathcal{M}_1}|^2}{\sum_{u' \in \Phi_{\psi_1}} \rho_T |\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{g}_{u',\mathcal{M}_1}|^2 + \sum_{\substack{u \notin \Phi_{\psi_1} \\ u \neq 1}} \rho_T |\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \mathbf{g}_{u,\mathcal{M}_1}|^2 + |\hat{\mathbf{g}}_{1,\mathcal{M}_1}^H \bar{\mathbf{n}}_{\mathcal{M}_1}|^2}. \quad (22)$$

To show the performance superiority of the proposed DNN based K -means AP clustering scheme in terms of $\mathcal{R}_1$ in GFRA, simulation results are presented in the sequel. Throughout the simulations, the three deployments in Section III-C1 are considered only with $\sigma_{\text{SF}} = 8$, and the following five schemes are compared:

- DNN based K -means AP clustering scheme: in this proposed scheme, the AP cluster closest to RA UE 1 is employed as set $\mathcal{M}_1$.
- T-ED based K -means AP clustering scheme: different from the DNN based scheme, the T-ED method developed in Section III-D is used for preamble multiplicity estimation.
- All-AP scheme: in this scheme, without preamble multiplicity estimation, all M APs are employed as set $\mathcal{M}_1$.
- M_c -strongest-AP scheme: in this scheme, without preamble multiplicity estimation, the M_c APs with the strongest received preamble signal energy are simply employed as set $\mathcal{M}_1$.
- Genie-aided scheme: in this genie-aided scheme, we assume that the set of APs that have the M_c largest channel gains of RA UE 1 is perfectly known at the BS CPU and these APs are employed as set $\mathcal{M}_1$. The performance of this scheme can be seen as an upper bound of the proposed scheme, which is desirable but unattainable in practical GFRA.

We first investigate the performance of the proposed DNN based K -means AP clustering scheme in terms of $\mathcal{R}_1$ in GFRA with different M_c in the considered three deployments with $\sigma_{\text{SF}} = 8$. In Figure 5, uplink achievable rate complementary cumulative distribution functions

Figure 5: Uplink achievable rate CCDF per collided RA UE of the proposed scheme, with different M_c and $\sigma_{SF} = 8$ dB.

(CCDF) per collided RA UE with different M_c are presented. The performance of the all-AP scheme is also considered in the figure as a baseline. As observed, compared to the all-AP scheme, the proposed DNN based scheme is able to significantly improve the achievable rate for collided RA UEs with a wide range of M_c , which verify its effectiveness in resolving the preamble collision in GFRA. In addition, different M_c leads to different performance. Evidently, neither a too small nor a too large value of M_c is desirable in the proposed scheme. By comparing the 95%-likely performance, it is clear to see that, throughout the three deployments, the proposed DNN based scheme with $M_c = 4$ provides the highest achievable rate, for example, achieving about 16 dB performance gain to the case of $M_c = 1$ and 4dB to the case of $M_c = 16$ in deployment 1. These observations validate the analysis provided at the end of Section IV. In the following, we further study the performance of the proposed DNN based scheme with only $M_c = 4$ in different deployments.

Figure 6: Comparisons of uplink achievable rate CCDF per collided RA UE among different schemes, with $\sigma_{SF} = 8$ dB and $M_c = 4$.

In Figure 6, the uplink achievable rate CCDF per collided RA UE in different deployments are

shown. Evidently, the proposed DNN and T-ED based schemes both perform much better than the all-AP scheme, which validates the effectiveness of the proposed framework that incorporates preamble multiplicity estimation and AP clustering to address the preamble collision problem in distributed mMIMO. Since the DNN can provide more accurate preamble multiplicity estimation than the T-ED (as shown in the example given in Section III-D), the DNN based scheme outperforms the T-ED based scheme under different deployments. For instance, around 4 dB performance gain in term of the 95%-likely uplink achievable rate per collided RA UE can be achieved in deployment 1, which is approximately equivalent to 150% performance improvement. Besides, both the all-AP and M_c -strongest-AP schemes provide poor 95%-likely achievable rate performance for collided RA UEs because these two schemes have no capability of dealing with preamble collision and as a result, introduce strong interference originating from the preamble collision. In addition, it is shown that both schemes in deployment 1 only exhibit slightly better performance than that in deployment 2. This is due to the fact that employing $S = 2$ not only brings in an array gain to the desired signals, but also to the interference due to the preamble collision. Therefore, increasing S has trivial benefit of improving the performance of these two schemes. Interestingly, we see that the all-AP scheme is more preferable to the M_c -strongest-AP scheme in terms of the 95%-likely performance. This results from that, without preamble multiplicity information and sensible AP clustering, the M_c -strongest-AP scheme leads to severe clustering errors, i.e., the M_c APs could be the neighboring APs of other collided RA UEs that are far away from the target RA UE, which make these APs contain little desired signals of the target RA UE, but strong interference from other collided RA UEs in most times. This implies the necessity of preamble multiplicity estimation to enable correct AP clustering. Contrastingly, based on the proposed DNN based preamble multiplicity estimation, the proposed AP clustering scheme is significantly superior to the all-AP and M_c -strongest-AP schemes. For instance in deployment 1, compared to the two schemes, the 95%-likely achievable rate of a collided RA UE can be largely enhanced by 26 dB and 34 dB, respectively. This performance superiority is also validated in Figure 7, where the uplink ergodic achievable rates per collided RA UE in GFRA in different schemes and deployments are presented. Compared to the all-AP and M_c -strongest-AP schemes, the proposed DNN based scheme is able to provide an improvement of 187% and 150%, respectively, in terms of the ergodic rate per collided RA UE in deployment 1. Similar improvement can also be seen in other two deployments. Furthermore, it is shown that the CCDF curves of the proposed DNN based scheme are close to those of the genie-aided

scheme. In deployment 1, for example, a 7 dB performance gap is observed in terms of 95%-likely performance. Considering that the performance of the proposed DNN based scheme is achieved without any CSI information of RA UEs, we claim that this performance gap is quite acceptable. This insight is also supported by the results in Figure 7. In particular, only around 7.0% and 8.5% ergodic rate losses in deployment 1 and deployment 3 are observed respectively, by comparing the proposed DNN based scheme to the genie-aided one. These observations indicate that the proposed machine learning based framework solution is able to achieve a near-optimal performance under preamble collision, validating its effectiveness in terms of preamble collision resolution under given deployments of distributed mMIMO.

Figure 7: Comparisons of uplink ergodic achievable rate per collided RA UE among different schemes, with $\sigma_{\text{SF}} = 8$ dB and $M_c = 4$.

VI. CONCLUSION

In this paper, a novel machine learning based framework was proposed to mitigate the impact of preamble collision on the performance of collided RA UEs in GFRA with distributed mMIMO. By taking advantages of signal spatial sparsity in distributed mMIMO and sporadic traffic pattern of mMTC RA UEs, we first developed a tailored DNN to enable the preamble detection and estimate the preamble multiplicity in GFRA, where a T-ED method was also proposed for performance comparison. Under practical wireless environments and different deployments of distributed mMIMO, we analyzed and compared the performance of the proposed DNN and

showed that decent estimation accuracy and reliability can be achieved. With the estimated preamble multiplicity, we then proposed a K -means AP clustering algorithm to cluster the neighboring APs of collided RA UEs rather than all the APs to decode the received data individually. Simulation results verified the effectiveness of the proposed schemes in preamble collision resolution in GFRA. Particularly, as examples shown in the simulation, the proposed DNN based scheme is able to achieve a close performance to the genie-aided scheme in terms of uplink achievable rate per RA UE under preamble collision. In the considered deployments of distributed mMIMO, the proposed DNN based scheme provided its best performance when $M_c = 4$, which exhibits a significant performance gain of up to 26 dB over the all-AP scheme.

The machine learning based framework solution could be served as a groundwork for preamble collision resolution in GFRA with distributed mMIMO. To further improve the performance, other powerful and efficient machine learning enabled solutions can be explored. For example, it might be better to integrate the preamble multiplicity estimation and AP clustering into one machine learning based algorithm, which is left for future research.

REFERENCES

- [1] 3GPP TS 38.913, "Study on scenarios and requirements for next generation access technologies (release 15)," v.15.2.0, June 2018.
- [2] J. Ding, M. Nemati, C. Ranaweera, and J. Choi, "IoT connectivity technologies and applications: A survey," *IEEE Access*, vol. 8, pp. 67646-67673, April 2020.
- [3] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *IEEE Commun. Surv. Tutor.*, pp. 4-16, March 2014.
- [4] S. Sesia, M. Baker, and I. Toufik, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons, 2011.
- [5] N. H. Mahmood, N. Pratas, T. Jacobsen, and P. Mogensen, "On the performance of one stage massive random access protocols in 5G systems," *International Symp. on Turbo Codes & Iterative Information Processing*, pp. 340-344, September 2016.
- [6] 3GPP, "Discussions on 2 steps RACH procedure," TR R1-1700668, January 2017.
- [7] H. F. Schepker, C. Bockelmann, and A. Dekorsy, "Exploiting sparsity in channel and data estimation for sporadic multi-user communication," *IEEE ISWCS*, August 2013, pp. 1-5.
- [8] J. Choi, "Two-stage multiple access for many devices of unique identifications over frequency-selective fading channels," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 162-171, February 2017.
- [9] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, November 2010.
- [10] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in mMIMO," *Proc. EUSIPCO*, pp. 76-80, September 2014.

- [11] G. N. Kamga, M. Xia, and S. Aïssa, "Spectral-efficiency analysis of massive MIMO systems in centralized and distributed schemes," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1930-1941, May 2016.
- [12] J. Ding, D. Qu, H. Jiang, and T. Jiang, "Success probability of grant-free random access with mMIMO," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 506-516, February 2019.
- [13] L. Liu and W. Yu, "Massive connectivity with massive MIMO-part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933-2946, June 2018.
- [14] H. Jiang, D. M. Qu, J. Ding, and T. Jiang, "Multiple preambles for high success rate of grant-free random access with massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4779-4789, October 2019.
- [15] H. Han, Y. Li, W. Zhai and L. Qian, "A grant-free random access scheme for M2M communication in massive MIMO systems," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3602-3613, April 2020.
- [16] J. H. Sørensen, E. De Carvalho, C. Stefanovic, and P. Popovski, "Coded pilot random access for massive MIMO systems," *IEEE Transactions on Wireless Commun.*, vol. 17, pp. 8035-8046, October 2018.
- [17] L. You, X. Gao, X.-G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352-3366, June 2015.
- [18] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Transactions on Wireless Commun.*, vol. 17, no. 1, pp. 574-590, January 2018.
- [19] H. Han, X. Guo, and Y. Li, "A high throughput pilot allocation for M2M communication in crowded massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9572-9576, October 2017.
- [20] X. H. You, D. M. Wang, B. Sheng, X. Q. Gao, X. S. Zhao, and M. Chen, "Cooperative distributed antenna systems for mobile communications [coordinated and distributed MIMO]," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 35-43, June 2010.
- [21] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE Journal of Selected Topics in Signal Process.*, vol. 8, no. 5, pp. 954-965, October 2014.
- [22] A. Yang, Y. Jing, C. Xing, Z. Fei, and J. Kuang, "Performance analysis and location optimization for massive MIMO systems with circularly distributed antennas," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5659-5671, October 2015.
- [23] J. Wang and L. Dai, "Downlink rate analysis for virtual-cell based large-scale distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1998-2011, March 2016.
- [24] 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, November 2018.
- [25] P. Liu, K. Luo, D. Chen, and T. Jiang, "Spectral efficiency analysis of cell-free massive MIMO systems with zero-forcing detector," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 795-807, February 2020.
- [26] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive MIMO wireless networks," *IEEE ICC*, Kansas City, MO, US, May 2018, pp. 1-6.
- [27] H. Liu, J. Zhang, X. Zhang, A. Kurniawan, T. Juhana, and B. Ai, "Tabu-search-based pilot assignment for cell-free massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2286-2290, February 2020.
- [28] 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, March 2017.
- [29] J. Choi, "Compressive random access for MTC in distributed input distributed output systems," *IEEE 85th Vehicular Technology Conference (VTC Spring)*, Sydney, NSW, pp. 1-5, June 2017.
- [30] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, 2006.

- [31] M. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, pp. 525-533, 1993.
- [32] L. E. Scales, *Introduction to Non-Linear Optimization*, Springer-Verlag, New York, 1985.
- [33] P. Liu, S. Jin, T. Jiang, Q. Zhang, and M. Matthaiou, "Pilot power allocation through user grouping in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 65, no. 4, pp. 1561-1574, April 2017.
- [34] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: user-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706-709, December 2017.
- [35] Stuart P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. on Information Theory*, vol. 28, pp. 129-137, 1982.