

# Repulsive Clustering Based Pilot Assignment for Cell-Free Massive MIMO Systems

Salman Mohebi, Andrea Zanella, Michele Zorzi

Department of Information Engineering, University of Padova, Padova, Italy

E-mails: {surname}@dei.unipd.it

**Abstract**—Thanks to its capability to provide a uniform service rate for the User Equipments (UEs), Cell-free (CF) massive Multiple-Input, Multiple-Output (mMIMO), has recently attracted considerable attention, both in academia and in industry, and so is considered as one of the potential technologies for beyond-5G and 6G. However, the reuse of the same pilot signals by multiple users can create the so-called pilot contamination problem, which can hinder the CF mMIMO from unlocking its full performance. In this paper, we address the challenge by formulating the pilot assignment as a maximally diverse clustering problem and propose an efficient yet straightforward repulsive clustering-based pilot assignment scheme to mitigate the effects of pilot contamination on CF mMIMO. The numerical results show the superiority of the proposed technique compared to some other methods with respect to the achieved uplink per-user rate.

**Index Terms**—cell-free massive MIMO, pilot assignment, pilot contamination, repulsive clustering, maximally diverse clustering

## I. INTRODUCTION

Evolving from the first to the fifth generation, many technologies have been proposed to support growing traffic and service demands in mobile networks. Network densification is a common technique to increase the network coverage and rate for the User Equipments (UEs). Densification can happen both by increasing the number of the Base Stations (BSs), a.k.a. ultra dense networks, or the number of the antennas at the BS, a.k.a. massive Multiple-Input, Multiple-Output (mMIMO). Each of these approaches suffers from some shortages: deploying a large number of BSs increases the inter-cell interference and hence reduces the service quality for the UEs, while in the mMIMO, UEs located at the edge of the cell suffer from high propagation loss because of the long distance from the BS. Cell-free (CF) mMIMO [1] has recently been introduced as an answer to the shortage of the technologies mentioned above by adopting the best of both. The CF mMIMO systems are composed of a large number of distributed Access Points (APs) that jointly serve relatively fewer number of UEs. The operation, unlike the traditional cellular network, takes place in a user-centric fashion, where each UE is surrounded and served by multiple APs. The APs are connected to a central processing units (CPU) through high-capacity error-free channels, where the network synchro-

nization, data detection/precoding/decoding, and some other network management operations take place.

CF mMIMO adopts the block fading models, where time-frequency channels are divided into coherence blocks of  $\tau_c$  channel uses. Each coherent block is further divided into three sub-intervals such that:  $\tau_c = \tau_p + \tau_u + \tau_d$ , where  $\tau_p$  is used for uplink pilot training, and  $\tau_u$  and  $\tau_d$  are used for uplink and downlink data transmission, respectively. Due to the limited number of channel uses in each coherence block, we can only have a limited number of orthogonal pilots, which is typically smaller than the number of UEs. This forces us to reuse the same pilots for different UEs, which introduces some undesirable effects, known as pilot contamination: the fading channel can not be accurately estimated at the APs due to the co-pilot interference among UEs.

The *random* pilot assignment presented in [1] is not efficient, and a proper pilot assignment policy can significantly reduce the effects of the so-called pilot contamination problem. A *greedy* pilot assignment is proposed in [1], which iteratively updates the pilot sequence for the UE with minimum rate. A *structured* pilot assignment scheme is proposed in [2] that maximizes the minimum distance between the co-pilot UEs. A *location-based greedy* pilot assignment is proposed in [3], that utilized the location information of the UEs to improve the initial pilot assignment. The authors in [4] considered the pilot assignment as a topological interference management problem with multiple groupcasting messages. They then formulated two *topological* pilot assignments for known and unknown UE/AP connectivity patterns. Graph theory has also been used for modeling the pilot assignment, where by creating interference graph among the UEs, *graph coloring*, [5] and *weight graphic* [6] is used to assign pilots for different UEs. *Tabu search* is another approach that has already been considered to the pilot assignment problem [7]. Buzzi *et. al.* [8] formulated pilot assignment as a graph matching problem and proposed a *Hungarian* algorithm to solve it. A weighted count-based pilot assignment is presented in [9], which considers the user's prior geographic information and pilot power to maximize the pilot reuse weighted distance. Another scalable pilot assignment algorithm based on *deep learning* is presented in [10] that maps between user locations and pilot assignment schemes. The co-pilot interference, in principle, is because of the pilot reuse in UEs that are close to each other. So, a valid pilot assignment scheme could only rely on the UEs geographical locations instead of adopting a costly channel estimation procedure to form the interference

Salman Mohebi has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant agreement No. 813999.

graph. Motivated by the above considerations, in this paper, we consider the pilot assignment in CF mMIMO as a maximally diverse clustering problem, where the UEs are divided into clusters that maximize inter-cluster heterogeneity and intra-cluster homogeneity. We then present a repulsive clustering method to solve it. The remainder of this paper is summarized as follows. Sec. II provides the system model for the CF mMIMO, then in Sec. III, we formulate the pilot assignment problem and propose the repulsive clustering based pilot assignment. The numerical results are presented in Sec. IV, and we conclude the paper in Sec. V.

## II. SYSTEM MODEL

We consider a typical CF mMIMO system, where  $M$  geographically distributed APs equipped with single antenna, coherently serve  $K$  single-antenna UE ( $K \ll M$ ), as exemplified in Fig. 1. All APs are connected to a CPU by an unlimited error-free fronthaul channel. The channel coefficient  $g_{mk}$  between the  $m$ -th AP and the  $k$ -th UE is given as follows:

$$g_{mk} = \beta_{mk}^{1/2} h_{mk}, \quad (1)$$

where  $\{\beta_{mk}\}$  indicate the large-scale fading (LSF) coefficients (i.e., pathloss and shadowing), and  $\{h_{mk}\}$  represent the small-scale fading coefficient which are assumed to be independent identically distributed (i.i.d.) normal random variables  $\mathcal{CN}(0, 1)$ .

### A. Uplink Pilot Training

We assume that there are only  $\tau_p$  mutually orthogonal pilot sequences with length  $\tau_p$  each represented as a column  $\phi \in C^{\tau_p \times 1}$  of a matrix  $\Phi$ , for which we have  $\|\phi_{p_k}^H \phi_{p_{k'}}\| = 1$  if  $p_k = p_{k'}$ , and  $\|\phi_{p_k}^H \phi_{p_{k'}}\| = 0$ , otherwise. The number of available pilots is independent of  $K$  and is limited due to the natural channel variation in the time and frequency domains [11].

In the uplink pilot training phase, all UEs simultaneously transmit their pilots. The  $m$ -th AP receives

$$y_m^p = \sqrt{\tau_p \rho_p} \sum_{k=1}^K g_{mk} \phi_{p_k}^H + n_m^p, \quad (2)$$

where  $\rho_p$  is the normalized Signal-to-Noise Ratio (SNR) of a pilot sequence with respect to noise power, and  $n_m^p \sim \mathcal{CN}(0, 1)$  represents the additive thermal noise.

As shown in [1], the effective channel coefficients between UE  $k$  and AP  $m$  can be estimated employing Minimum Mean Square Error (MMSE) estimator as follows:

$$\hat{g}_{mk} = c_{mk} \left( \sqrt{\tau_p \rho_p} g_{mk} + \sqrt{\tau_p \rho_p} \sum_{k' \neq k} g_{mk'} \phi_{p_k}^H \phi_{p_{k'}} + \phi_{p_k} n_m^p \right), \quad (3)$$

where

$$c_{mk} \triangleq \frac{\sqrt{\tau_p \rho_p} \beta_{mk}}{\tau_p \rho_p \sum_{k'=1}^K \beta_{mk'} |\phi_{p_k}^H \phi_{p_{k'}}|^2 + 1}. \quad (4)$$

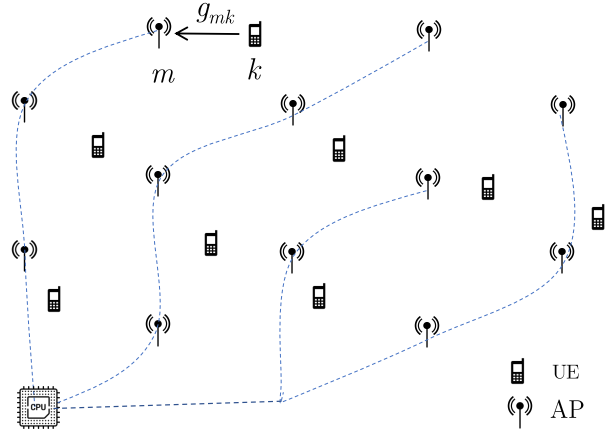


Fig. 1: A CF mMIMO system, where  $M$  distributed APs jointly serve  $K$  UEs ( $K \ll M$ ).

The mean-square of the estimated channel vector  $\hat{g}_{mk}$  is

$$\gamma_{mk} \triangleq \mathbb{E} \left\{ |\hat{g}_{mk}|^2 \right\} = \sqrt{\tau_p \rho_p} \beta_{mk} c_{mk}. \quad (5)$$

### B. Uplink Data Transmission

In CF mMIMO, all APs and UEs use the same time-frequency resources to transmit data. In the uplink, AP  $m$  receives

$$y_m^u = \sqrt{\rho_u} \sum_{k=1}^K g_{mk} \sqrt{\eta_k} q_k + n_m^u, \quad (6)$$

where  $q_k$  is the signal transmitted by UE  $k$  with power  $\mathbb{E} \left\{ |q_k|^2 \right\} = 1$ , while  $\eta \in [0, 1]$  indicates the power control coefficient,  $\rho_u$  denotes the normalized uplink SNR and  $n_m^u \sim \mathcal{CN}(0, 1)$  is the additive noise at receiver.

The Maximum Ratio (MR) combining scheme can be applied to decode the desired signal from a certain UE  $k$ . AP  $m$  sends  $\hat{g}_{mk}^* y_m^u$  to the CPU for data detection. The CPU combines all the received signal for UE  $k$  as:  $r_k^u = \sum_{m=1}^M \hat{g}_{mk}^* y_m^u$ . The signal then can be decomposed at the CPU as follows, as shown in [1]:

$$\begin{aligned} r_k^u = & \underbrace{\sqrt{\rho_u \eta_k} q_k \mathbb{E} \left\{ \sum_{m=1}^M g_{mk} \hat{g}_{mk}^* \right\}}_{\text{DS}_k} \\ & + \underbrace{\sqrt{\rho_u \eta_k} q_k \left( \sum_{m=1}^M g_{mk} \hat{g}_{mk}^* - \mathbb{E} \left\{ \sum_{m=1}^M g_{mk} \hat{g}_{mk}^* \right\} \right)}_{\text{BU}_k} \\ & + \underbrace{\sqrt{\rho_u} \sum_{m=1}^M \sum_{k' \neq k} \sqrt{\eta_{k'}} g_{mk'} \hat{g}_{mk'}^* q_{k'} + \hat{g}_{mk}^* n_k^u}_{\text{CPI}_k}, \end{aligned} \quad (7)$$

where  $\text{DS}_k$ ,  $\text{BU}_k$  and  $\text{CPI}_k$  denoted the desired signal (DS), beamforming uncertainty (BU) and co-pilot interference (CPI), respectively.

The achievable uplink rate for the UE  $k$  can be calculated as (8), shown at the top of the next page.

### III. MAXIMALLY DIVERSE CLUSTERING FOR PILOT ASSIGNMENT

#### A. Problem formulation

An efficient pilot assignment mechanism should maximize the number of effectively estimated channels between the UEs and the APs. Due to the coherent nature of the transmissions in CF mMIMO systems, data can still be potentially detected in the presence of multiple imperfectly estimated channels. So, as in CF mMIMO the ultimate goal is increasing the rate for the UEs, the pilot assignment can be formulated as an uplink rate maximization problem, i.e.,

$$\begin{aligned} \max_{\mathbf{p}} \quad & \sum_{k=1}^K R_k^u \\ \text{s.t.} \quad & \mathbf{p} = \{p_1, \dots, p_K\} \\ & \phi \in \Phi, \quad \forall k \in \{1, \dots, K\}. \end{aligned} \quad (9)$$

#### B. Proposed Scheme

Considering that the distance between UEs has a significant impact on co-pilot interference, to mitigate the effects of pilot contamination an efficient pilot assignment policy should assign the same pilot  $p$  to UEs in a repulsive way, i.e., to the UEs that are geographically far apart or have fewer common serving APs. Hence, we formulate the pilot assignment as a maximally diverse clustering problem, where the data points (UEs) that are assigned to the same cluster have high "dissimilarity", but can be similar to the members from different clusters. To solve the problem, we then proposed a repulsive clustering scheme, that is opposed to typical clustering algorithms which put homogeneous data points in the same clusters. Note that, the inter-cluster similarity is also essential to ensure the fair distribution of data points in clusters.

Let us consider  $\mathbf{X}$  as a binary cluster association (pilot assignment) matrix, where  $x_{k,p} = 1$  if UE  $k$  belongs to cluster (pilot)  $p$ , and  $x_{k,p} = 0$  otherwise. So the repulsive clusters can be obtained by solving the following problem:

$$\begin{aligned} \max_{\mathbf{X}} \quad & \sum_{p=1}^{\tau_p} \sum_{k=1}^{K-1} \sum_{k'=k+1}^K x_{k,p} x_{k',p} f_r(k, k') \\ \text{s.t.} \quad & \sum_{p=1}^{\tau_p} x_{k,p} = 1, \quad k \in \{1, \dots, K\} \\ & \left\lfloor \frac{K}{\tau_p} \right\rfloor \leq \sum_{k=1}^K x_{k,p} \leq \left\lceil \frac{K}{\tau_p} \right\rceil + 1, \quad p \in \{1, \dots, \tau_p\} \\ & x_{k,p} \in \{0, 1\}, k \in \{1, \dots, K\}; \end{aligned} \quad (10)$$

where  $f_r(k, k')$  is a customized function that measures the diversity/repulsion score for  $k$  and  $k'$  data points (UEs). The first constraint guarantees that each data point is assigned to one cluster and the second constraint forces the clusters to have similar size. The second constraint is important because

it keeps inter cluster similarity high. This repulsion function can be a predefined static function, i.e., Euclidean distance, or can be parameterized and then learned by, e.g., neural networks. The second approach is favorable as a sophisticated pilot assignment should consider not only the physical location of the UEs but also other parameters like AP locations and their density.

Repulsive clustering has already been considered in the literature under different names: anticlustering [12], [13], and maximally diverse grouping problem [14]. Typically this type of problem is NP-hard, but applying some relaxations can be solved by integer programming [15]. Here, we present Algorithm 1, a simple heuristic yet efficient algorithm to find a feasible (but not necessarily optimal) solution to the repulsive clustering problem. This algorithm first randomly assigns data points to different clusters and then iteratively swaps the UEs among clusters as long as it improves the overall repulsion score.

---

#### Algorithm 1 A Heuristic Algorithm for Repulsive Clustering

---

**Input:** Number of clusters (pilots)  $\tau_p$ , Set of UEs  $\mathcal{K}$

**Output:** Pilot assignment vector  $\mathbf{p}$

---

```

Randomly divide  $\mathcal{K}$  UEs into  $\tau_p$  equal-sized clusters  $\mathcal{C}$ ,
while Performance is improving do
  for  $C1, C2 \in \mathcal{C}$  do
    for  $u \in C1$  and  $w \in C2$  do
      if exchanging clusters of  $u$  and  $w$  increases the
      overall diversity measure as given by (10) then
        Swap the clusters of  $u$  and  $w$ ,
      end if
    end for
  end for
end while
for  $p = 1 : \tau_p$  do
  Assign pilot  $\phi_p$  to UEs in cluster  $\mathbf{p}$ 
end for

```

---

## IV. NUMERICAL RESULTS

#### A. Simulation setup

Let us consider  $M$  APs and  $K$  UEs that are independently and uniformly distributed in a  $1 \times 1$  km<sup>2</sup> square area. We adopt the wrap-around technique to avoid boundary effects at the edge and simulate network behavior in an unlimited area. The 3GPP Urban Microcell model [16] is used to compute the large-scale propagation conditions like path loss and shadow fading. Noise power is calculated by  $P_n = Bk_B T_0 W$ , where  $B = 20$  MHz is the bandwidth,  $k_B = 1.381 \times 10^{-23}$  (Joule per Kelvin) denotes the Boltzmann constant,  $T_0 = 290$  (Kelvin) is the noise temperature and  $W = 9$  represents the noise figure. The transmission powers of the uplink pilot and the uplink data are set to  $\rho_p = 100$  mW and  $\rho_u = 100$  mW, respectively. The channel estimation overhead has been taken into account for defining the per-user uplink throughput as  $T_k^u = B \frac{1-\tau_p/\tau_c}{2} \log 2(1 + \text{SINR}_k^u)$ , where  $\tau_c = 200$  samples. The  $1/2$  in the above equation is due to the co-existence of

$$R_k^u = \log_2 \left( 1 + \frac{\rho_u \eta_k \left( \sum_{m=1}^M \gamma_{mk} \right)^2}{\rho_u \sum_{k' \neq k}^K \eta_{k'} \left( \sum_{m=1}^M \gamma_{mk} \frac{\beta_{mk'}}{\beta_{mk}} \right)^2 |\phi_{pk}^H \phi_{pk'}| + \rho_u \sum_{k'=1}^K \eta_{k'} \sum_{m=1}^M \gamma_{mk} \beta_{mk'} + \sum_{m=1}^M \gamma_{mk}} \right) \quad (8)$$

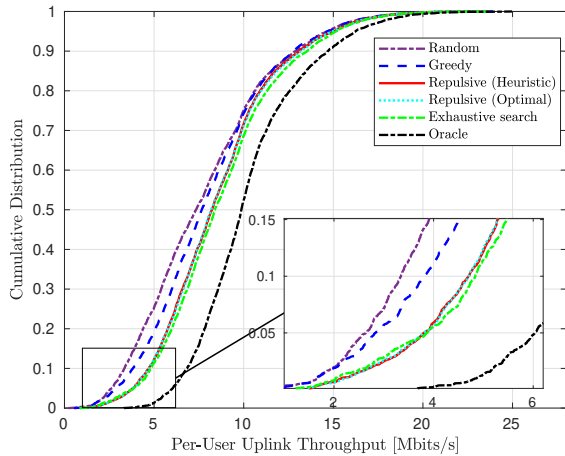


Fig. 2: Cumulative distribution of the per-user uplink throughput for different pilot assignment strategies for a small-scale scenario,  $M = 50$ ,  $K = 12$  and  $\tau_p = 3$ .

the uplink and downlink traffic. We also employed max-min power control [1] to further improve the sum throughput.

In this paper we consider the Euclidean distance for the repulsion function as  $f_r(U_k, U_{k'}) = \sqrt{\sum_{i=1}^{|F|} (U_k[i] - U_{k'}[i])^2}$ , where  $F$  is the feature set (e.g. geographical coordinates) of the UEs. The definition and analysis of more sophisticated repulsive functions are left to future work.

### B. Result and discussion

The CF mMIMO systems aim to provide a uniform service to all the UEs regardless of their physical location. So, the per-user throughput is used to evaluate the performance of the pilot assignment algorithms. The result is compared with the *random* and *greedy* pilot assignment from [1] and *Oracle* pilots assignment, where there is no pilot contamination, i.e.  $\text{CPI}_k = 0$  in (7). The R-package introduced in [12] is used to optimally partition UEs into diverse groups (optimal repulsive clustering).

As the time complexity of exhaustive search and optimal repulsive clustering exponentially grows by the number of UEs, calculating their performance for large  $M$ s is not possible. Fig. 2 shows the Cumulative Distribution Function (CDF) of the per-user uplink throughput for a small-scale scenario, for the sake of comparison. As seen in the figure, the method outperforms the *random* and *greedy* pilot assignments and basically achieves the same (optimal) performance of the exhaustive search, but with far less complexity.

Fig. 3 shows the cumulative distribution of the per-user uplink throughput for different pilot assignment strategies for  $M = \{100, 200, 300\}$ . The superiority of the proposed scheme

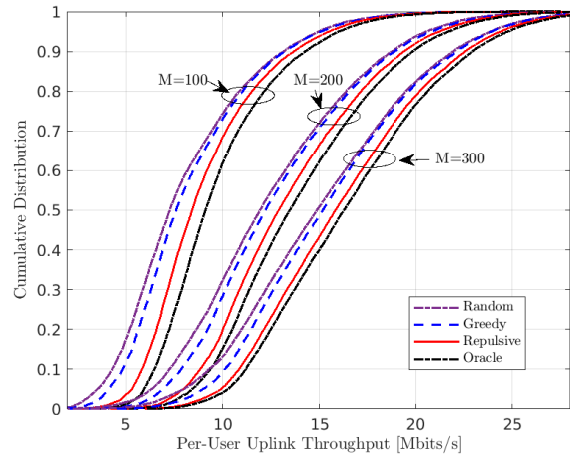


Fig. 3: Cumulative distribution of the per-user uplink throughput for different pilot assignment strategies,  $K = 40$  and  $\tau_p = 10$ .

$M$	Random	Greedy	Repulsive	Oracle
100	3.5	4.1	5.3	5.9
200	6.3	6.9	7.9	8.4
300	7.9	8.9	9.9	10.3

TABLE I: 95th percentile of the per-user uplink throughput [Mbits/s] for different numbers of APs, Here,  $K = 40$  and  $\tau_p = 10$ .

against other approaches by a high margin is evident from the figure. The decreasing gap between the repulsive and *Oracle* pilot assignment by increasing the number of APs shows the robustness of our approach against density.

Tab. I shows the 95th percentile of the per-user throughput extracted from Fig. 3. Our method, for  $M = 100$ , increases the 95th percentile of the per-user throughput by 1.75 Mbps (33%) and 1.18 Mbps (22%) in comparison to *random* and *greedy* assignments, respectively. The improvements for  $M = 300$  are 1.93 Mbps (20%) and 0.92 Mbps (9%). Compared to the situation without pilot contamination, our method successfully reaches 89% to 96% of the 95th percentile of the per-user throughput of *Oracle* pilot assignment, which is a great success.

Fig. 4 illustrates the 95th percentile of the per-user uplink throughput of different pilot assignment schemes against the number of UEs. It can be seen from the figure that by increasing the number of UEs in the network, the throughput for most of the UEs decreases, but the reduction speed varies for different approaches. The increasing gap between the proposed and other approaches shows the superiority of our system. For example for  $K = 60$ , our method improves the

In this paper, we proposed a repulsive clustering based pilot assignment for CF mMIMO systems. We formulated the pilot assignment as a maximally diverse clustering problem and solved it by a repulsive clustering scheme. Numerical results show the effectiveness of the proposed scheme compared to the conventional random and greedy pilot assignment. In future works, we will expand our approach by replacing the Euclidean distance with more sophisticated and parameterized repulsion functions, i.e., Deep Neural Networks (DNNs) that consider different networking factors such as AP locations, and the density of UEs and APs. Another extension will consider pilot assignment jointly with pilot power control, which can further improve the channel estimation performance. The scalability of different pilot assignment strategies is another factor that should be considered in future research.

REFERENCES

- [1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [2] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive MIMO wireless networks," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, May. 2018, pp. 1–6.
- [3] Y. Zhang, H. Cao, P. Zhong, C. Qi, and L. Yang, "Location-based greedy pilot assignment for cell-free massive MIMO systems," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. IEEE, Dec. 2018, pp. 392–396.
- [4] H. Yu, X. Yi, and G. Caire, "Topological pilot assignment in large-scale distributed MIMO networks," *IEEE Transactions on Wireless Communications*, Feb. 2022.
- [5] H. Liu, J. Zhang, S. Jin, and B. Ai, "Graph coloring based pilot assignment for cell-free massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9180–9184, Jun. 2020.
- [6] W. Zeng, Y. He, B. Li, and S. Wang, "Pilot assignment for cell-free massive MIMO systems using a weighted graphic framework," *IEEE Transactions on Vehicular Technology*, Jun. 2021.
- [7] 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 Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2286–2290, Feb. 2019.
- [8] S. Buzzi, C. D'Andrea, M. Fresia, Y.-P. Zhang, and S. Feng, "Pilot assignment in cell-free massive MIMO based on the Hungarian algorithm," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 34–37, Jan. 2020.
- [9] W. Li, X. Sun, and D. Chen, "Pilot Assignment Based on Weighted-Count for Cell-Free Massive MIMO Systems," in *2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*. IEEE, Jan. 2021, pp. 258–261.
- [10] J. Li, Z. Wu, P. Zhu, D. Wang, and X. You, "Scalable Pilot Assignment Scheme for Cell-Free Large-Scale Distributed MIMO With Massive Access," *IEEE Access*, vol. 9, pp. 122 107–122 112, Sep. 2021.
- [11] S. Chen, J. Zhang, J. Zhang, E. Björnson, and B. Ai, "A survey on user-centric cell-free massive MIMO systems," *Digital Communications and Networks*, Dec. 2021.
- [12] M. Papenberg and G. W. Klau, "Using anticlustering to partition data sets into equivalent parts," *Psychological Methods*, vol. 26, no. 2, p. 161, Apr. 2021.
- [13] M. J. Brusco, J. D. Cradit, and D. Steinley, "Combining diversity and dispersion criteria for anticlustering: A bicriterion approach," *British Journal of Mathematical and Statistical Psychology*, vol. 73, no. 3, pp. 375–396, Sep. 2020.
- [14] A. Schulz, "The balanced maximally diverse grouping problem with block constraints," *European Journal of Operational Research*, vol. 294, no. 1, pp. 42–53, Oct. 2021.
- [15] M. Grötschel and Y. Wakabayashi, "A cutting plane algorithm for a clustering problem," *Mathematical Programming*, vol. 45, no. 1, pp. 59–96, Aug. 1989.
- [16] 3GPP, "Further advancements for E-UTRA physical layer aspects," *3rd Generation Partnership Project (3GPP), TR 36.814*, 2010.

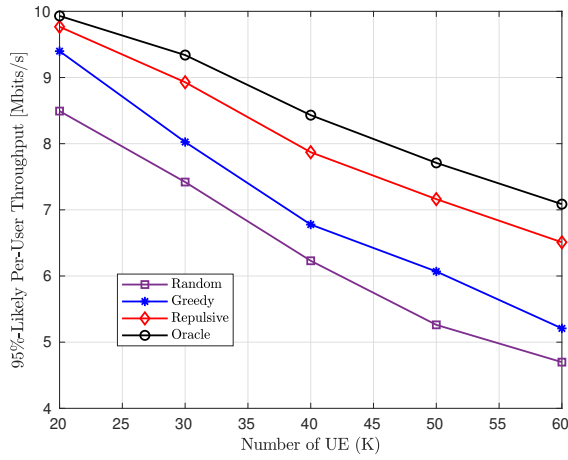


Fig. 4: 95th percentile of the per-user uplink throughput for different numbers of UE,  $M = 100$  and  $\tau_p = 10$ .

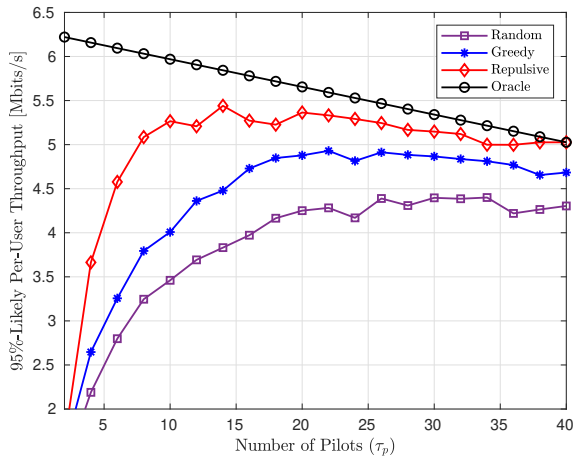


Fig. 5: 95th percentile of the per-user uplink throughput for different numbers of UE,  $M = 100$  and  $K = 40$ .

95th percentile of the per-user throughput by 1.8 Mbps (38%) and 1.3 Mbps (25%), comparing to the *random* and *greedy* approaches, respectively. Also, the gap between the proposed and the *Oracle* pilot assignment scheme grows much slower than for the two other methods, which means that increasing  $K$  does not heavily affect our system.

The 95th percentile of the per-user uplink throughput against the number of pilots ( $\tau_p$ ) for different pilot assignment schemes is presented in Fig. 5. The proposed approach always performs better than the *greedy* and *random* pilot assignments. As can be seen in the figure, increasing the number of pilots will improve the performance for the majority of the UEs only up to a certain point, and reduces afterward. This shows the necessity of finding the optimal number of pilots, a study which is outside the scope of this paper and is left for future research.