Vector Quantization Methods for Access Point Placement in Cell-Free Massive MIMO Systems

Govind R. Gopal, *Graduate Student Member, IEEE* and Bhaskar D. Rao, *Life Fellow, IEEE*

***Abstract*—We examine the problem of uplink cell-free access point (AP) placement in the context of optimal throughput. In this regard, we formulate two main placement problems, namely the sum rate and minimum rate maximization problems, and discuss the challenges associated with solving the underlying optimization problems with the help of some simple scenarios. As a practical solution to the AP placement problem, we suggest a vector quantization (VQ) approach. The suitability of the VQ approach to cell-free AP placement is investigated by examining three VQ-based solutions. First, the standard VQ approach, that is the Lloyd algorithm (using the squared error distortion function) is described. Second, the tree-structured VQ (TSVQ), which performs successive partitioning of the distribution space is applied. Third, a probability density function optimized VQ (PDFVQ) procedure is outlined, enabling efficient, low complexity, and scalable placement, and is aimed at a massive distributed multiple-input-multiple-output scenario. While the VQ-based solutions do not explicitly solve the cell-free AP placement problems, numerical experiments show that their sum and minimum rate performances are good enough, and offer a good starting point for gradient-based optimization methods. Among the VQ solutions, PDFVQ, with its distinct advantages, offers a good trade-off between sum and minimum rates.**

***Index Terms*—Base station placement, Beyond 5G, Lloyd algorithm, scalability, throughput optimization.**

I. Introduction

THE concept of massive multiple-input-multiple-output (MIMO) [1] has emerged in recent decades as a strong solution for 5G and Beyond wireless communication systems [2]–[4]. By having a large number of antennas, such systems enable higher spectral and energy efficiencies, and reduced interference due to increased diversity [5], [6]. Distributed MIMO, comprising distributed antenna systems (DASs), offer higher average rates over their colocated counterparts [7]–[10]. Distributed massive MIMO can be predominantly split into non-cooperative and cooperative systems. Small-cell systems comprise non-cooperative systems while cell-free systems constitute the latter.

Cell-free massive MIMO, where a very large number of antennas or access points (APs) serve users that are not divided into cells, have become popular of late as they help in mitigating interference among users and in enhanced spectral efficiency over small-cell systems [11]–[14]. The concept of "cell-free" itself arises from network MIMO [15]. However, it is to be noted that these benefits come at the cost of large backhaul requirements necessary for the information exchange between APs [16], [17] and the network controller (NC) or the central processing unit, where all or most processing is performed. The computational and processing requirements of a cell-free system are thus high, and are expected to grow with the anticipated network densification in Beyond 5G/6G wireless systems [18]. Hence, the conventional small-cell architecture is still very much in use in current and near future 5G deployments [19], [20]. Nevertheless, a multitude of problems has been explored based on the cell-free network. Prior works investigate topics such as power optimization and energy efficiency [21]–[24], rate maximization [25], [26], clustering (user- and cell-centric) [27]–[31], limited fronthaul [32], pilot assignment [33], reconfigurable intelligent surfaces [34], and federated learning [35].

In this work, our focus is on AP placement, with the prevailing question: *How do we place APs optimally for a given user distribution?* Here, the optimality is in terms of throughput in cell-free systems. AP placement finds application in scenarios such as stadiums, where the distribution of crowds changes according to the ongoing game, or in offices where the density of employees changes by the hour. Further, deployment of APs using unmanned aerial vehicles (UAVs) can be advantageous in areas affected by natural (or man-made) calamities where existing infrastructure has been destroyed [36]. Past literature has expressed interest in AP placement, however, most of the focus has been on single-cell or small-cell systems (e.g., [37]–[40]). For instance, [37] maximizes cell averaged ergodic capacity of a DAS by using vector quantization (VQ) based codebook design (the Lloyd algorithm) to place antennas. In [38], antennas are placed circularly to optimize the average per-user rate of uniformly distributed users. In [39], a 10-fold improvement in capacity is shown in a distributed system over a colocated in a simulated indoor environment. Further, aligning AP positions with the user density generated around a 40% increase in capacity over uniformly distributed APs. The authors of [40] and their related works study the deployment of heterogeneous wireless sensor networks (placement of both APs and fusion centers) as a source coding problem for optimal power control and with limited communication range. Most optimization problems have mainly considered maximizing the signal-to-noise ratio (SNR) alone, but certain works have incorporated various forms of interference (e.g., two-cell

Govind R. Gopal is with Nokia Standards, Nokia of America Corporation, Naperville, IL 60563, USA (e-mail: govind.gopal@nokia.com). The work of Govind R. Gopal was carried out while he was a PhD student at the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA. The work of Govind R. Gopal was supported in part by the Center for Wireless Communications (CWC), University of California San Diego and in part by Qualcomm Inc. through the Faculty-Mentor-Advisor program.

Bhaskar D. Rao is with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA (e-mail: brao@ucsd.edu). The work of Bhaskar D. Rao was supported by the National Science Foundation (NSF) under Grant CCF-2124929 and Grant CCF-2225617.

leakage interference in [41] and signal-to-interference-plus-noise ratio (SINR) in [42]). Placement of UAVs as APs (UAV-APs) has also been studied (e.g., [43]–[48]). [43] calculates the positions and altitudes of UAV-APs to maximize ground user coverage as well as minimize interference to users. The work of [44] on the other hand places UAV-APs heuristically in a spiral fashion to guarantee user coverage. The authors of [45] find the optimal placement and coverage radius of UAV-APs that serve users with time-varying positions in Wi-Fi (IEEE 802.11) networks. In [46], the 3-D placement of UAV-APs with directional antennas to minimize average user transmit powers is investigated. A hybrid UAV-terrestrial AP network is considered in [47] where position-flexible UAV-APs are deployed in an area occupied by fixed terrestrial APs to maximize throughput while minimizing inter-cell interference (ICI) by using a Lloyd-type algorithm. Finally, [48] provides an extensive tutorial on UAVs as APs and as users (cellular connected UAVs).

The problem of AP placement in cell-free systems is fairly novel and much prior research has not been conducted. As an example, [49] investigates the deployment of UAVs in a cell-free network to maximize minimum SINR using a gradient approach while considering pilot contamination and that not every AP communicates with all users. The authors of [50] consider the placement of APs in a distributed massive MIMO system as a combinatorial problem to minimize transmit powers while considering antenna radiation patterns and different channel models. While a cell-free system is not explicitly defined, the system model considered mimics such a system. A graph-based approach is found to yield significant power savings while ensuring placement with good coverage. In [51], 3-D placement of UAV-APs is considered to maximize the downlink sum rate and uses an alternating optimization method. A prior work by our group [52] examines both sum rate and minimum rate maximization problems. This work solves the two problems using compressed sensing techniques by dividing the geographical area into regular grids. However, the approximations used do not solve the two placement problems optimally.

With the expected densification in future networks, it is necessary to design AP placement schemes that are not only practical, but can scale easily with the number of APs which is characteristic in a massive MIMO scenario. Moreover, as alluded to above in the examples of AP placement scenarios, such methods should be able to easily adapt to changing user environments. While the VQ approach (the Lloyd algorithm) have been utilized in prior works (e.g., [37]), none of the abovementioned works have investigated the application of the Lloyd algorithm and other VQ approaches to solve the cell-free AP placement problem. In the VQ framework as applied to AP placement, the user positions are clustered and the cluster centers are the AP locations. In our prior work [53], we preliminarily investigated the utilization of the standard VQ approach, namely the Lloyd algorithm with the squared Euclidean distance as the distortion function, to place APs in a cell-free network. The VQ technique considers a single user that communicates to its nearest AP with the objective function that utilizes a distortion function averaged over the random position of this user. It does not match the cell-free model where all users communicate to all APs. In spite of these limitations, VQ-based solutions have some features that make them worthy of consideration. By design, VQ offers not only a distributed solution (where APs are placed at different locations as opposed to a colocated solution) but can encourage cooperation (as is expected in the cell-free model) by placing APs closer to one another in areas of higher user density. Additionally, VQ solutions provide good initial points for gradient and learning-based methods to solve specific throughput problems. Hence, in this work, we explore and compare multiple VQ techniques, each with its own benefits, that can solve for AP locations in a cell-free network, with throughput as the performance measure.

Contributions

To the best of our knowledge, analysis of the cell-free AP placement problem in the context of throughput optimality along with the suitability of the VQ framework and the application of VQ-based methods to solve the same have not been addressed in literature. Hence, in this work, our contributions are as follows.

- We formulate the two main throughput optimal cell-free AP placement problems, namely the sum rate and minimum rate maximization problems. While these problems have been previously studied in small-cell works, a detailed discussion in the context of cell-free networks has not been presented. Starting from the simpler sum SNR problem, analysis of the sum rate problem is conducted and simple examples are shown to describe the possible solutions and to highlight the challenges associated with the general AP placement problem. The minimum rate problem is then discussed, also with some examples.
- Three VQ-based techniques to place APs are proposed and explored, namely standard VQ, which is the Lloyd algorithm, tree-structured VQ (TSVQ), and probability density function optimized VQ (PDFVQ). While the Lloyd algorithm provides a well-established method to place APs, there are the disadvantages of complexity and scalability. TSVQ, through successive partitioning of the user area, places APs in such a way so as to foster cooperation. PDFVQ on the other hand, allows an efficient, less computationally intensive, and easily adaptable AP placement solution by using bit allocation, transform coding, and scalar quantization, and is especially suitable for a scaled network with a large number of APs.

The remainder of this manuscript is organized as follows. Section II outlines the cell-free system model used, followed by Section III which explores the two principal throughput optimization problems for AP placement, namely the sum rate and minimum rate maximizations. VQ-based techniques are described in Section IV. In Section V, we state the simulation methodology and results. Finally, concluding remarks are provided in Section VI. Throughout this paper, we will use bold symbols to denote vectors, $\mathbb{E}\{\cdot\}$ is the expectation operator, $||\cdot||$ represents the ℓ_2-norm of a vector, and all logarithms are to the base 2.

II. System Model

The system model outlined in [52], [53], and [54] is used. K single-antenna users are distributed over a geographical area with a probability density function $f_{\mathbf{P}}(\mathbf{p})$, with $\mathbf{p} \in \mathbb{R}^2$ as the random vector denoting the user position. M single-antenna APs serve these users, where $\mathbf{q} \in \mathbb{R}^2$ is the AP location. With $m = 1, 2, \ldots, M$ and $k = 1, 2, \ldots, K$, a narrowband fading channel is considered with

$$g_{mk} = \sqrt{\beta_{mk}} h_{mk}, \tag{1}$$

where β_{mk} and $h_{mk} \sim \mathcal{CN}(0, 1)$ are the large- and small-scale fading coefficients, respectively, independent of each other and over coherent intervals. A general expression for β_{mk} is

$$\beta_{mk} = \frac{c z_{mk}}{||\mathbf{p} - \mathbf{q}_m||^{\gamma}}, \tag{2}$$

where c is a constant, z_{mk} is the shadow fading coefficient, and γ is the pathloss exponent. All APs cooperate with each other and are connected via error-free backhaul links to the NC. We consider the traditional cell-free uplink regime where all APs serve a smaller number of users in the same time-frequency resource. The uplink received signal at AP m is

$$y_m = \sum_{k=1}^{K} \sqrt{\rho_r} g_{mk} s_k + w_m, \tag{3}$$

where for user k, ρ_r is the transmit power, s_k is the data symbol with $\mathbb{E}\{|s_k|^2\} = 1$, and $w_m \sim \mathcal{CN}(0, 1)$ is the additive noise. The received signal vector at the NC from all M APs using (3) can be written as

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\rho_r} \mathbf{g}_k s_k + \mathbf{w}, \tag{4}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_M]^T$, $\mathbf{g}_k = [g_{1k}, g_{2k}, \ldots, g_{Mk}]^T$, and $\mathbf{w} = [w_1, w_2, \ldots, w_M]^T$. When a combiner $\mathbf{v}_k$ is used to estimate data symbols of user k as $\hat{s}_k = \mathbf{v}_k^H \mathbf{y}$, the per-user achievable rate is $R_k = \mathbb{E}\{\log(1 + \phi_k^{\mathbf{v}_k})\}$, where the expectation is over all the small-scale and shadow fading coefficients, and the SINR [55] is

$$\phi_k^{\mathbf{v}_k} = \frac{\rho_r \mathbf{v}_k^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}_k}{\mathbf{v}_k^H \mathbf{v}_k + \sum\limits_{\substack{k'=1 \\ k' \neq k}}^{K} \rho_r \mathbf{v}_k^H \mathbf{g}_{k'} \mathbf{g}_{k'}^H \mathbf{v}_k}. \tag{5}$$

One such combiner is the zero forcing (ZF) detector, and results in the processed signal at the NC as $\mathbf{r} = \left(\mathbf{G}^H \mathbf{G}\right)^{-1} \mathbf{G}^H \mathbf{y}$, where $\mathbf{G} = [g_{mk}]$ is a $M \times K$ matrix consisting of the channel coefficients [52]. The achievable per-user SNR in this case is

$$\psi_k^{\text{ZF}} = \frac{\rho_r}{[(\mathbf{G}^H \mathbf{G})^{-1}]_{kk}}. \tag{6}$$

Using an asymptotic approximation for the SNR as outlined in [38], [52], the per-user SNR can also be written as

$$\frac{1}{M} \psi_k^{\text{ZF}} \xrightarrow[M \to \infty]{\text{a.s.}} \rho_r \overline{\beta}_k, \tag{7}$$

where

$$\overline{\beta}_k \triangleq \lim_{M \to \infty} \frac{1}{M} \sum_{m} \beta_{mk}. \tag{8}$$

III. Throughput Formulations for the Cell-Free AP Placement Problem

There are two main formulations for cell-free AP placement in terms of throughput optimality:

- *Sum rate maximization*, which involves the sum of the rates of all users, as follows

$$\arg\max_{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M} \sum_{k=1}^{K} \log(1 + \phi_k^{\mathbf{v}_k}). \tag{9}$$

  Note that this problem is identical to the average rate maximization problem by assuming a sample mean and taking the average over the user distribution $f_{\mathbf{P}}(\mathbf{p})$.
- *Minimum rate maximization*, where the minimum of the rates among all of the users is maximized

$$\arg\max_{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M} \min_{k} \log(1 + \phi_k^{\mathbf{v}_k}). \tag{10}$$

  The notion of fairness, which is important in a cell-free system since all users are served by all APs, is enforced by the minimum rate problem in (10) as opposed to the sum rate problem in (9). In practice, the 95%-likely rate, which represents the best rate among the worst 5% of the users, is used as a measure to evaluate the network minimum rate performance. Hence, the max-min rate is adjusted to the 95%-likely rate, which is more robust. This is subsequently addressed in Section IV-D.

It should be noted that finding solutions that address both of the above metrics, although challenging, is ideal. It is desirable to achieve an optimal trade-off between sum rate and minimum rate. In the ensuing sections, we discuss the above formulations. Our analysis of the sum rate maximization problem is preceded by the simpler sum SNR maximization problem (preliminarily discussed in [53]).

A. Sum SNR Maximization

When throughput is measured by utilizing SNR alone, we can obtain a simpler sum throughput problem to (9) by using the ZF SNR ψ_k^{ZF} from (6) and by replacing the summation with an expectation (the factor of $1/K$ has been neglected). The average is taken over the user position and the set $\mathcal{A} = \{z_{mk}, \forall m\}$ consisting of all shadow fading coefficients between each user and AP. The optimization problem is written as

$$\arg\max_{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M} \mathbb{E}_{\mathcal{A}, \mathbf{p}} \left\{ \psi_k^{\text{ZF}} \right\}. \tag{11}$$

We can simplify the above objective function using the approximation in (7) and β_{mk} from (2) in the following manner

$$\begin{aligned} \mathbb{E}_{\mathcal{A}, \mathbf{p}} \left\{ \psi_k^{\text{ZF}} \right\} &= \mathbb{E}_{\mathcal{A}, \mathbf{p}} \left\{ \rho_r \sum_{m=1}^{M} \beta_{mk} \right\}, \\ &= \mathbb{E}_{\mathbf{p}} \left\{ \rho_r \sum_{m=1}^{M} \mathbb{E}_{\mathcal{A}} \left\{ \beta_{mk} \right\} \right\}, \\ &\overset{(a)}{=} \mathbb{E}_{\mathbf{p}} \left\{ \sum_{m=1}^{M} \frac{c'}{||\mathbf{q}_m - \mathbf{p}||^{\gamma}} \right\}, \\ &\overset{(b)}{=} c' \sum_{m=1}^{M} \mathbb{E}_{\mathbf{p}} \left\{ \frac{1}{||\mathbf{q}_m - \mathbf{p}||^{\gamma}} \right\}, \end{aligned} \tag{12}$$

where $c' = c\rho_r \mathbb{E}_{\mathcal{A}}\{z_{mk}\}$, $\mathbf{p}$ is the position of user k in (a), and (b) uses the fact that the expectation is a linear operator. Note that c' can be ignored as it does not effect the optimization problem.

The following observations can be made regarding the solution of the above objective function in (12).

- A colocated solution is evident since there is no dependence between the terms in the summation associated with each AP m and hence, the optimization for each AP can be performed separately. The suggested colocated solution may not be a unique solution and multiple global and/or local maxima may exist for the optimization problem. The complete characterization of the solution, however, depends on two factors, namely the pathloss exponent γ used as well as the shape of the user distribution $f_{\mathbf{P}}(\mathbf{p})$ over which the expectation is taken, e.g., uni-modal versus multi-modal density functions.
- In (12), although the norm $||\mathbf{q}_m - \mathbf{p}||$, i.e., the Euclidean distance is strictly convex [56] in $\mathbf{q}_m$, its inverse is neither concave nor convex and is also undefined at $\mathbf{q}_m = \mathbf{p}$ (this can, however, be avoided by adding a small positive quantity to the denominator, which could also account for the height of the AP). The sum of the expectation of the inverse over the APs thus also is neither convex or concave. Additionally, when a uni-modal distribution is assumed, we can expect a colocated solution alone with a unique maximum. However, when a multi-modal distribution is considered, it is expected that multiple global maxima exist and distributed solutions may be obtained. This is explored in Section III-C.

It is worth noting that the sum SNR problem has been previously addressed in part in [52]. In this work, while dividing the user area into regular grid points, the sum rate maximization problem is upperbounded to a sum SNR problem and using a compressed sensing framework, approximated to a linear program (called the max-sum algorithm). However, as is expected from a sum SNR problem (as discussed above), most APs are concentrated around high user density regions (a near-colocated solution). Although a high sum rate can be achieved with this solution, users far away from the APs are severely affected in terms of throughput, resulting in poor minimum rate performance. Thus, this solution is not suitable for cell-free AP placement when fairness is considered.

B. Sum Rate Maximization

Returning to the sum rate maximization problem, we rewrite (9) by using the simplifications assumed before, as follows

$$\underset{\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_M}{\arg\max}\ \mathbb{E}_{\mathcal{A},\mathbf{p}}\left\{\log\left(1+\psi_k^{\text{ZF}}\right)\right\}, \tag{13}$$

and the objective function, utilizing the approximation in (7), can be rewritten as

$$\mathbb{E}_{\mathcal{A},\mathbf{p}}\left\{\log\left(1+\psi_k^{\text{ZF}}\right)\right\} = \mathbb{E}_{\mathcal{A},\mathbf{p}}\left\{\log\left(1+\rho_r\sum_{m=1}^{M}\beta_{mk}\right)\right\}. \tag{14}$$

Similar to the SNR problem outlined before, this objective function is neither concave nor convex. However, unlike the former, the term associated with each AP m in (14) cannot be decoupled from the terms associated with the rest of the APs. Hence, for both uni-modal and multi-modal distributions, we can expect only distributed solutions that maximize the sum rate.

In summary, for both the sum SNR and sum rate problems there may be multiple local optima suggesting that the optimization problem is complex and challenging.

C. Examples for Sum SNR and Sum Rate Maximizations

Given the abovementioned complexity in solving the sum SNR and sum rate problems, we now attempt to understand the problems and their solutions better. For this purpose, we explore simple examples where the aforementioned two problems are solved, and where the multiple local optima are studied to develop intuition and insight.

1) User distribution considered: For tractability, we consider a simple 1-D scenario where users are distributed along a line, and the placement of four APs for both the sum SNR and sum rate maximization problems. For this purpose, we assume a bi-modal distribution since it is the simplest among multi-modal distributions that can exhibit the multiple maxima as discussed in Section III-A. With the user position denoted by p, the PDF of the bi-modal Gaussian considered here is

$$f_P(p) = \pi_1\mathcal{N}(p|\mu_1,\sigma_1^2) + \pi_2\mathcal{N}(p|\mu_2,\sigma_2^2), \tag{15}$$

where for each Gaussian i, $i = 1, 2$, π_i is the probability such that $\pi_1 + \pi_2 = 1$, μ_i is the mean, and σ_i is the standard deviation. To generate different solution structures for the two optimization problems, we consider two distinct configurations of the user distribution:

Conf. 1: $\pi_1 = \pi_2 = 0.5$, $\mu_1 = -3$, $\mu_2 = 3$, and $\sigma_1 = \sigma_2 = 1$.
Conf. 2: $\pi_1 = 0.35$, $\pi_2 = 0.65$, $\mu_1 = -3$, $\mu_2 = 4$, and $\sigma_1 = \sigma_2 = 1$.

Configuration 1 is symmetric about the origin while configuration 2 is asymmetric.

2) Definitions of SNR and rate: The four APs have locations q_1, q_2, q_3, and q_4, and the ZF SNR[1] for user k can be calculated as

$$\begin{aligned}\psi_k^{\text{ZF}} &= \beta_{1k} + \beta_{2k} + \beta_{3k} + \beta_{4k},\\ &= \frac{1}{(p_k-q_1)^2} + \frac{1}{(p_k-q_2)^2} + \frac{1}{(p_k-q_3)^2} + \frac{1}{(p_k-q_4)^2},\end{aligned} \tag{16}$$

where, for simplicity, the transmit power is set to unity and the definition of the large-scale fading coefficient β_{mk} (from (2)) assumes that shadow fading is absent, the pathloss exponent is two, and the constant is set to one. These assumptions do not change the conclusions that are obtained in this section. The sum SNR and sum rate quantities are then defined as $\sum_{k=1}^{K}\psi_k^{\text{ZF}}$ and $\sum_{k=1}^{K}\log(1+\psi_k^{\text{ZF}})$, respectively. Note that for implementation purposes, a small quantity ϵ is added to the denominator of β_{ik}, $i = 1, 2, 3, 4$, to prevent ψ_k^{ZF} from approaching infinity.

[1]While the asymptotic approximation for the SNR expression used does strictly not hold here, it is good enough to illustrate the complexities of the maximization problems considered.

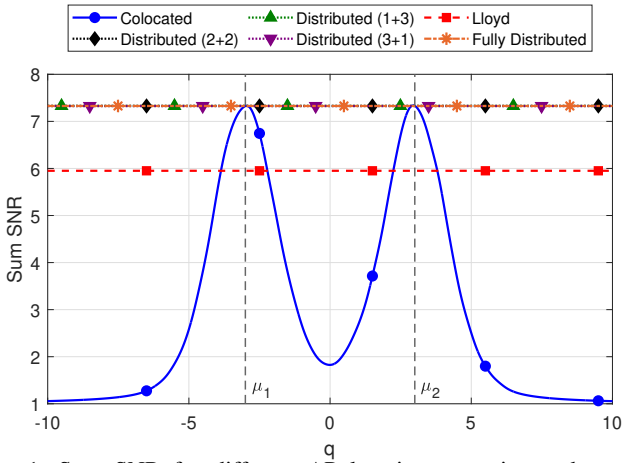


Fig. 1: Sum SNR for different AP location scenarios under user configuration 1.

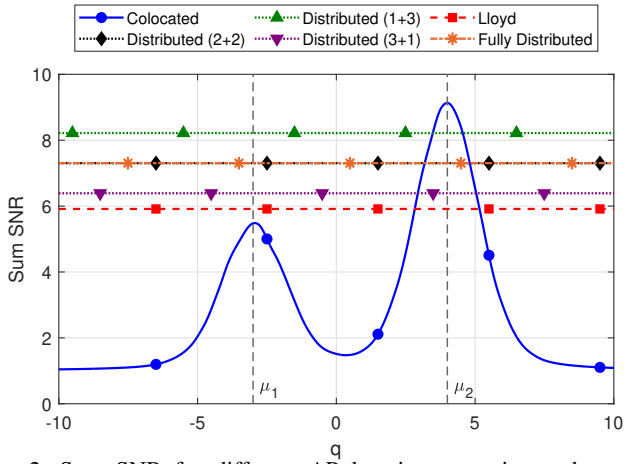


Fig. 2: Sum SNR for different AP location scenarios under user configuration 2.

3) AP location solutions considered: To evaluate and understand the sum SNR and sum rate performances of the system, we study various AP placement scenarios. First, a colocated solution is considered where all four APs are situated at the same location. This location is found by sweeping the AP position denoted by q across the span of the user locations. Second, multiple semi-distributed solutions are selected. Instead of all APs at one location, they can be allocated to each of the Gaussians and placed at their respective means. In this scenario, we consider three situations, namely when two APs each are placed at μ_1 and μ_2, three APs are at μ_1 and one AP is at μ_2, and one AP is at μ_1 and three APs are at μ_2. These three situations are termed 'Distributed (2+2)', 'Distributed (3+1)', and 'Distributed (1+3)', respectively. Third, a fully distributed scenario involves starting from a distributed (2+2) solution and moving the two APs within each Gaussian away from each other until the maximum is achieved. Note that this fully distributed solution represents only a local maximum. Finally, we have the solution obtained by applying the standard Lloyd algorithm to the user distribution. In this iterative solution, the users are clustered using the squared Euclidean distance and the cluster centers (centroids) are determined to be the AP locations. More explicit details are provided in Section IV-A.

4) Results for sum SNR: The results of sum SNR under user configurations 1 and 2 are shown in Fig. 1 and Fig. 2, respectively. Note that for the colocated solution, the sum SNR obtained as a function of location q is plotted. For the other solutions, a line is drawn corresponding to the sum SNRs obtained.

For *configuration 1*, it is observed that the peak for the colocated system occurs at the two means μ_1 and μ_2. The three distributed scenarios as well as the fully distributed scenario offer the same peak sum SNR value as the colocated case. It is noted that the AP locations in the fully distributed case are the same as in distributed (2+2). Further, the Lloyd solution (which is also a fully distributed scenario) offers a lower sum SNR value. The sum SNR maximization for configuration 1 thus has multiple local maxima, including both colocated and semi-distributed solutions.

For *configuration 2*, the Gaussian with mean μ_2 has a higher probability, with the result that a colocated solution with all four APs at mean μ_2 yields the highest sum SNR. Among the distributed solutions, it is observed that a higher allocation of APs at μ_2 favors a higher sum SNR, however, with the colocated solution offering the highest. Again, both the fully distributed and distributed (2+2) solutions have the same AP locations and sum SNR values, with the Lloyd solution performing the worst. In summary, the above results for sum SNR show that the AP locations that maximize the same may be colocated or distributed depending on the user distribution, as discussed in Section III-A.

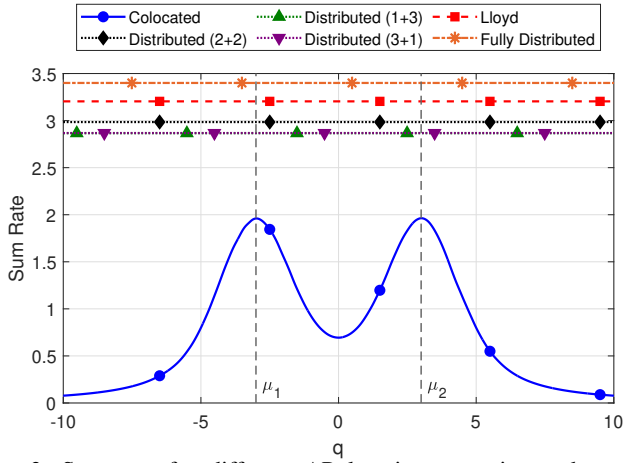


Fig. 3: Sum rate for different AP location scenarios under user configuration 1.

5) Results for sum rate: The results of the sum rate metric for user configurations 1 and 2 are shown in Fig. 3 and Fig. 4, respectively, and differ from the sum SNR plots above. In configuration 1, it is clear that a colocated solution performs poorly and that distributed and fully distributed solutions offer a higher sum rate. Since the distribution is symmetric, both distributed (1+3) and distributed (3+1) solutions have the same sum rate, which is lower than that of distributed (2+2). The Lloyd solution performs better than the distributed solutions as no APs are colocated, and the fully distributed solution performs the best. Note that the fully distributed solution does

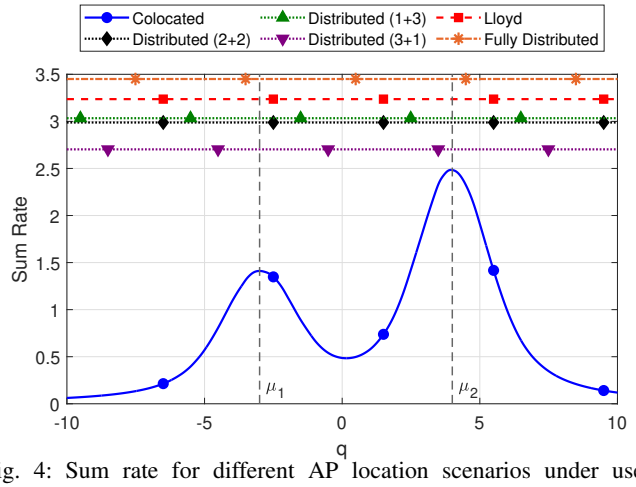


Fig. 4: Sum rate for different AP location scenarios under user configuration 2.

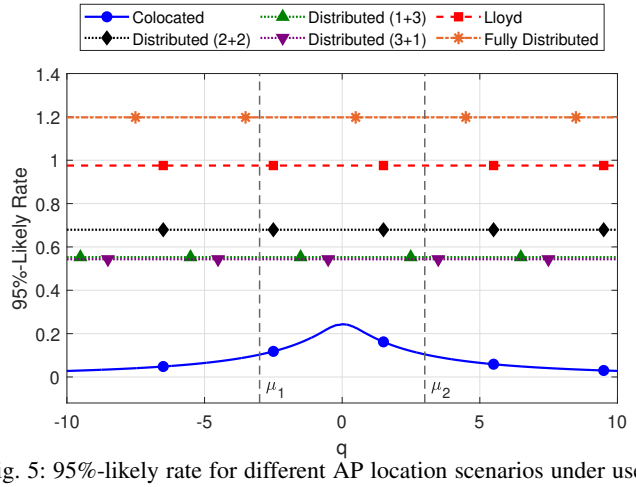


Fig. 5: 95%-likely rate for different AP location scenarios under user configuration 1.

not represent the optimum solution due to the method with which the locations are generated. In configuration 2, the colocated solution still performs the worst. However, since the Gaussian at mean μ_2 carries a higher proportion of users, the distributed (1+3) solution performs better than both distributed (2+2) and distributed (3+1). Finally, as in configuration 1, both the Lloyd and the fully distributed solution perform better than the distributed solutions.

6) Results for minimum rate: We now plot the 95%-likely rates corresponding to the different AP locations for configurations 1 and 2, in Fig. 5 and Fig. 6, respectively. Note that the fully distributed solution in these plots maximizes the 95%-likely rates as opposed to the sum SNR or sum rate. Even though all distributed solutions generate the same peak sum SNR for configuration 1 as shown in Fig. 1, the users will not achieve the same minimum rate performance in all cases. For this user configuration, distributed (1+3) and distributed (3+1) have lower 95%-likely rates than distributed (2+2). Additionally, the Lloyd solution has a higher minimum rate and the fully distributed solution generates the highest 95%-likely rate. Clearly, the colocated solution exhibits the worst performance. The same observations are noted for

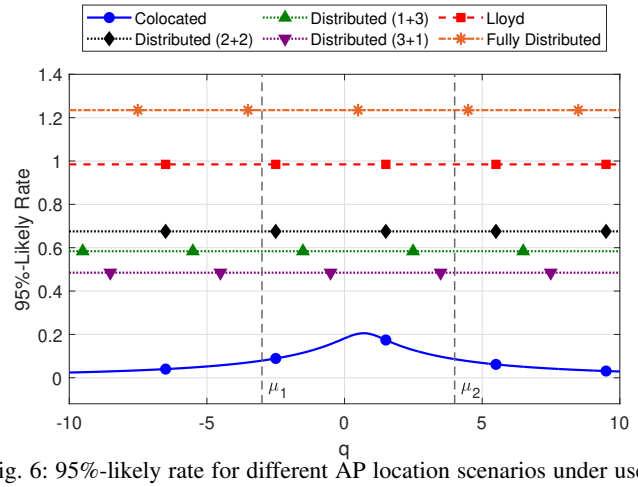


Fig. 6: 95%-likely rate for different AP location scenarios under user configuration 2.

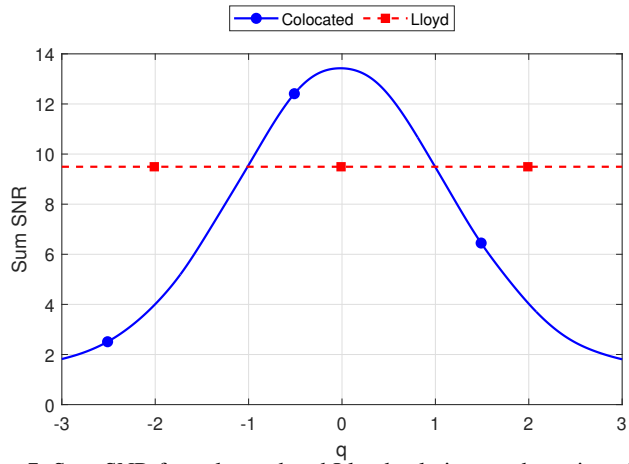


Fig. 7: Sum SNR for colocated and Lloyd solutions under uni-modal distribution.

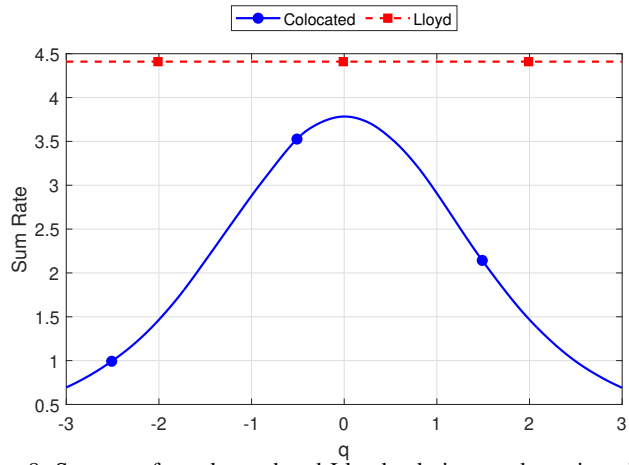


Fig. 8: Sum rate for colocated and Lloyd solutions under uni-modal distribution.

configuration 2, with the difference being that distributed (1+3) generates a higher 95%-likely rate than distributed (3+1) due to the higher user proportion around mean μ_2.

7) Results for a uni-modal user distribution: As a final note and to complete the discussion on the influence of the

distribution shape on the solutions, we also plot the sum SNR and sum rate performances for a uni-modal Gaussian distribution (with zero mean and unit variance), in Fig. 7 and Fig. 8, respectively. Like the bi-modal examples shown prior, the colocated solution is preferred in the sum SNR case over the distributed Lloyd solution. The opposite is true for the sum rate case. It is also to be noted that unlike the bi-modal scenario, there is only one colocated solution that provides the peak sum SNR.

In conclusion, all observations made in the context of the simple examples support the complex nature of the AP placement for sum SNR and sum rate maximization problems (Section III-A and Section III-B).

D. Minimum Rate Maximization

For the sum SNR maximization case, we showed both colocated and semi-distributed cases were favored. It was also postulated that the minimum rate performance of these solutions would be inferior to that provided by a fully distributed solution (like the Lloyd solution). Even in the sum rate maximization case, we have not addressed the minimum rates of the users explicitly and we can expect that there are users that lie far away from the AP locations, resulting in significantly low throughput to these users, which is not ideal from a fairness point of view. The minimum rate maximization problem is thus of interest and has been defined above in (10). In this paper, we do not discuss this problem as it has been addressed and solved using a grid-based approach in [52], [54] and is called the max-min algorithm. Such a solution is useful since the problem can be converted into a convex problem, which can be solved easily. It is to be noted that the grid structure assumed leads to approximate solutions and a finer grid is necessary to obtain the optimal locations. In [53], we have further analyzed and compared the performance of the max-min algorithm with the Lloyd algorithm.

IV. Vector Quantization Approaches

In this section, as practical solutions to cell-free AP placement, we investigate how VQ techniques can be applied to sum rate and minimum rate maximizations. While acknowledging that all users are served by all APs in the cell-free system under consideration, we assume that for the purposes of AP placement, VQ is implemented by assuming that each user is associated with its geographically nearest AP by adopting the squared Euclidean distance (error) as the distortion measure of interest. We start by presenting the standard VQ approach and motivate why the VQ framework is useful for the placement problem, followed by the other VQ techniques.

A. Standard VQ

The standard and simplest VQ technique is the Lloyd algorithm [57] that utilizes the squared Euclidean distance between APs and users as the distortion measure. This squared error (SE) distortion between a user at $\mathbf{p}$ and an AP at $\mathbf{q}_m$ is denoted as follows

$$d_{\text{SE}}(\mathbf{p}, \mathbf{q}_m) = ||\mathbf{p} - \mathbf{q}_m||^2 . \tag{17}$$

Details of how the Lloyd algorithm can be applied to AP placement have been elucidated in [42]. We provide the algorithm (Algorithm 1) below, where the K user positions denoted by $\mathbf{p}_k$ are generated as realizations from the user distribution $f_{\mathbf{P}}(\mathbf{p})$.

Algorithm 1 Lloyd Algorithm

1: Initialize random AP locations $\mathbf{q}_1^{(0)}, \mathbf{q}_2^{(0)}, \ldots, \mathbf{q}_M^{(0)}$.
2: Use the nearest neighbor condition (NNC) to determine the cells $\mathcal{C}_1^{(i+1)}, \mathcal{C}_2^{(i+1)}, \ldots, \mathcal{C}_M^{(i+1)}$ such that

$$\mathcal{C}_m^{(i+1)} = \left\{ \mathbf{p}_k : d_{\text{SE}}\left(\mathbf{p}_k, \mathbf{q}_m^{(i)}\right) \leq d_{\text{SE}}\left(\mathbf{p}_k, \mathbf{q}_l^{(i)}\right), \forall l \neq m \right\}.$$

3: Use the centroid condition (CC) to determine the AP locations $\mathbf{q}_1^{(i+1)}, \mathbf{q}_2^{(i+1)}, \ldots, \mathbf{q}_M^{(i+1)}$ such that

$$\mathbf{q}_m^{(i+1)} = \frac{1}{\left|\mathcal{C}_m^{(i+1)}\right|} \sum_{\mathbf{p}_k \in \mathcal{C}_m^{(i+1)}} \mathbf{p}_k.$$

4: Repeat from step 2 until convergence (MSE falls below a threshold).

Why Use the VQ Approach? Although we focus on the standard VQ technique as the illustration here, the advantages listed below follow for the subsequent VQ techniques described.

- *Distributedness:* The VQ approaches, through their formulations, are designed to provide a distributed solution. The standard Lloyd algorithm using squared error distortion minimizes the distance of a user to its closest AP, averaged over the entire distribution, obtains a distributed solution by ensuring that at least one AP is close to each user. It is important to note that as a consequence, the Lloyd algorithm solution also addresses the minimum rate of the system in an effective manner, while the minimum rate metric itself is not explicitly contained in its objective function. Additionally, the Lloyd algorithm alone exhibits a space-filling advantage [58] which refers to the generation of the Voronoi cells $(\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_M)$ and therefore the AP locations that span the space of the user distribution efficiently.
- *Cooperation:* While allowing distributedness, the VQ formulations which do not explicitly account for cooperation, also place APs close to one another. This clustering occurs especially in areas of high user density, thus encouraging cooperation that is expected in a cell-free system and addresses the system sum rate. In the specific the case of the Lloyd algorithm, the objective function indirectly fosters cooperation since the AP density is proportional to a power of the user density under a high resolution approximation [57], as follows

$$g_{\mathbf{P}}(\mathbf{p}) = \frac{f_{\mathbf{P}}^{\frac{1}{2}}(\mathbf{p})}{\int f_{\mathbf{P}}^{\frac{1}{2}}(\mathbf{p}')\mathrm{d}\mathbf{p}'}. \tag{18}$$

- *Initialization:* Although it is possible to determine solutions for both problems through gradient methods, such methods usually generate local optima resulting in lower-than-expected performances. The VQ approaches, due

to the distributed nature of their solutions, are able to provide suitable starting points for such gradient-based methods.

B. Tree-Structured VQ

Tree-structured VQ (TSVQ) [57] is an alternate VQ approach where the codebook search time is reduced compared to standard VQ. In this technique, the input training set is partitioned into a hierarchy of Voronoi regions, which allows a tree to be generated for encoding. Thus, TSVQ differs from the standard VQ discussed above in that the final required codebook is generated by the successive splitting of intermediate codebooks, starting with a single centroid. The Lloyd algorithm (Algorithm 1) is applied to each stage of the hierarchy and its application is confined to the partitioned training set of the previous stage.

Why use TSVQ? Designed primarily for relatively fast codebook search properties, TSVQ has the following benefits.

- *Initialization:* An advantage of TSVQ is that initialization of the AP locations is not required, since it starts with a single centroid. Compared to the standard VQ, this avoids the random initializations (considered sub-par) and the calculations needed for advanced initialization methods such as k-means++ [59].
- *Cooperation:* Due to the successive splitting of the intermediate codepoints, we can expect pairs of APs to be closer to each other compared to standard VQ. This enables cooperation among APs, increasing the system sum rate.
- *Flexibility:* When the number of APs is changed for the same user distribution, codepoints can either be merged (when the number of APs is reduced) or split (when increased). The choice of codepoints for merging or splitting will be on the basis of the rate performance associated with the codepoints.

It should be noted that in general, TSVQ does not find the closest AP to each user and there is a small decrease expected in the performance of TSVQ (in terms of mean squared error) when compared to standard VQ.

TSVQ Algorithm. In our implementation of TSVQ for cell-free AP placement, we limit ourselves to balanced binary trees, i.e., at each stage, every intermediate codepoint is split into two codepoints, to favor the lowest complexity. The TSVQ algorithm is outlined in Algorithm 2. The stages are indexed by j (root node is stage 0) and the set of codepoints is represented by $\mathcal{P}_j$ for stage j. The codepoints are indexed by i, and the number of codepoints increases at each stage. The codepoints generated after the algorithm converges are the required AP positions. $\mathcal{R}$ indicates the set of all users while $\mathcal{R}_i$ denotes the set of users associated with codepoint i. The splitting of codepoints is performed by generating two perturbations of the original codepoint.

Algorithm 2 Tree-Structured Vector Quantization Algorithm

1: Initialize $\mathcal{P}_0$ with the codepoint of all users in $\mathcal{R}$.
2: Split the codepoint(s) in $\mathcal{P}_j$ into two.
3: Apply Lloyd algorithm (Algorithm 1) to each split pair for the set of users $\mathcal{R}_i$ at stage j.
4: Partition $\mathcal{R}_i$ at stage j into two sets corresponding to new codepoints.
5: Update $\mathcal{P}_j$ with new codepoints.
6: Repeat from step 2 for next stage $j+1$ until M codepoints are generated.

C. PDF Optimized VQ

So far, we have considered a full-scale version of VQ (the Lloyd algorithm) and TSVQ, which involves hierarchical codebook generation. Both flavors of VQ, however, come with disadvantages such as complexity, scalability, and learnability which will be elaborated shortly. In this section, we outline the probability density function optimized VQ (PDFVQ) procedure, first developed in [60], for use in cell-free AP placement that can address these shortcomings. In PDFVQ, an efficient quantizer is generated by first estimating the distribution PDF (of the users) using the expectation-maximization (EM) algorithm [61] and assuming a Gaussian mixture model (GMM). The GMM considered is of the form

$$f_{\mathbf{P}}(\mathbf{p}) = \sum_{l=1}^{L} p_l \mathcal{N}\left(\mathbf{p}|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l\right), \tag{19}$$

where L is the number of mixture components (clusters), and p_l, $\boldsymbol{\mu}_l$, and $\boldsymbol{\Sigma}_l$ are the probability, mean, and covariance matrix, respectively, of mixture component l. Then, given the total bit budget (which corresponds to the total number of APs), by leveraging both bit allocation and transform coding, closed-form expressions are defined to allocate bits to each cluster and along every dimension (x- and y-coordinates of the 2-D user density) so that scalar quantizers can be used to generate the required codebook (the AP locations).

Why use PDFVQ? Although designed for high-dimensional source inputs, this procedure addresses the shortcomings of the two prior VQ approaches in the following manner.

- *Complexity:* In the two prior VQ approaches, the computational complexity and memory required are high. In TSVQ, the Lloyd algorithm is applied a number of times depending on the codebook splits performed to obtain the required number of APs. The overall complexity of the Lloyd algorithm is $\mathcal{O}(2KMI)$, where the factor '2' is due to the 2-D user density and I denotes the number of iterations required for convergence. It is also observed that the number of executed iterations I increases with the number of APs M. In addition to the simple expressions for AP allocations for each cluster and dimension of the 2-D user density, through the use of scalar quantizers for each dimension, PDFVQ enjoys lower complexity as the quantizers work on both a lower number of users and APs, consequently performing a lower number of iterations. Further, the scalar quantizers can also be implemented by means of look-up tables [62] that determine the codebook. Thus, there exists a simple closed-form mapping from user density to the codepoints

(AP locations). The memory requirements in PDFVQ are also small as a result of transform coding.

- *Learnability:* It is expected that the user density varies over time, in which case the AP positions must be adapted to account for this change. Both prior VQ schemes are not amenable to a learning environment as they cannot adapt quickly to changes in the user distribution as well as number of APs, and their computationally complex procedures must be repeated to optimize to the new environment. In PDFVQ, when the source distribution changes, both the density estimation and codebook generation steps must be implemented, however, only the parameters of the density have to be learned (the EM algorithm is easily updated using existing values). If the number of APs only is changed, then the closed-form expressions and scalar quantizers alone are needed, resulting in computational savings and faster adaptability.
- *Scalability:* The above discussions on complexity and learnability also point to the scalability aspects of the Lloyd, TSVQ, and PDFVQ algorithms. For the Lloyd algorithm alone, as the number of APs increases, the whole procedure must be repeated with increased complexity. In the case of TSVQ, additional successive splitting steps must be performed, with the number of times that the Lloyd algorithm is implemented doubling at every step. In PDFVQ, the closed-form expressions that perform allocations do not change with the increase in APs. These allocations and scalar quantizers can be easily and quickly implemented, enabling an 'on-the-fly' placement. Thus, PDFVQ facilitates scalability, both with respect to dimensions (of the user distribution) as well as the number of APs.

Due to the abovementioned advantages, PDFVQ is thus suitable for AP placement both in an environment where the user distribution changes over time as well as the large-scale massive MIMO scenario with a very high number of APs. It is also worth noting that the PDFVQ procedure is independent of the chosen distortion measure for the scalar quantizers, but for simplicity, we will consider that the Lloyd algorithm is used.

PDFVQ Algorithm. The PDFVQ procedure from [60] introduced in the beginning of this section applied to AP placement (involving only two dimensions) is outlined in Algorithm 3 below. Note that in this paper, we model the user distributions as GMMs (Section V), which avoids the need for parameter estimation. The total number of bits available b_{tot} is $\log_2 M$. It is important to note that in both step 1 and step 2, the resulting bit allocations b_l and $b_{l,j}$, respectively, are *not* expected to be integers. The number of levels computed in step 2 corresponding to the bit allocation $b_{l,j}$, that is, $V_{l,j} = 2^{b_{l,j}}$, is ultimately rounded off to the nearest integer, and represents the number of APs. As 'bit' generally refers to an integer quantity, to avoid subsequent confusion, we omit the word 'bit' when discussing allocations. The codebooks generated in step 3 for each cluster are the required AP locations.

Algorithm 3 PDF Optimized Vector Quantization Algorithm

1: Determine the allocation b_l to cluster l given the total budget b_{tot} using

$$2^{b_l} = 2^{b_{\text{tot}}} \frac{\sqrt{p_l c_l}}{\sum_{j=1}^{L} \sqrt{p_j c_j}}, \tag{20}$$

where $c_l = \sqrt{\lambda_{l,1}\lambda_{l,2}}$, $\boldsymbol{\lambda}_l = \text{diag}(\lambda_{l,1}, \lambda_{l,2})$, and $\boldsymbol{\Sigma}_l = \mathbf{Q}_l \boldsymbol{\lambda}_l \mathbf{Q}_l^T$ is the eigen value decomposition.

2: With each cluster l, compute the allocation along each dimension $b_{l,j}$, $j = 1, 2$, using

$$b_{l,j} = \frac{b_l}{2} + \frac{1}{2} \log \left[\frac{\lambda_{l,j}}{c_l} \right]. \tag{21}$$

Compute and round off the corresponding level $V_{l,j} = 2^{b_{l,j}}$.

3: Generate the codebook $\mathcal{R}_l$ for each cluster using

$$\mathcal{R}_l = \{\mathbf{q} | \mathbf{q} = \mathbf{Q}_l \mathbf{y} + \boldsymbol{\mu}_l, \mathbf{y} \in \mathcal{T}_l\}, \tag{22}$$

where $\mathcal{T}_l$ is the set of vectors given by the Cartesian product $\mathcal{T}_l = \mathcal{S}_{l,1} \times \mathcal{S}_{l,2}$, with $\mathcal{S}_{l,j}$, $j = 1, 2$ being the optimal $V_{l,j}$-level scalar quantizer of a univariate Gaussian with variance $\lambda_{l,j}$.

D. Gradient Approaches

As alluded to above, the VQ approaches do not explicitly solve the sum rate and minimum rate maximization problems, but provide a practical solution. However, if gradient ascent were to be applied to solve the problems so that either the sum rate or minimum rate can be improved further, the proper initialization of the AP locations is critical to avoid sub-par local optima. Hence, the VQ-based methods could provide reasonably good starting points to apply gradient ascent for both problems. Accordingly, we present both of the gradient calculations below, and are called the *max-sum* and *max-min gradient*, respectively.

1) Max-sum gradient: To maximize sum rate, the gradient update expression with j as the iteration index is

$$\mathbf{q}_m^{(j+1)} = \mathbf{q}_m^{(j)} + \delta \frac{\partial}{\partial \mathbf{q}_m^{(j)}} \left\{ \sum_{k=1}^{K} \log \left(1 + \rho_r \sum_{m=1}^{M} \beta_{mk} \right) \right\}, \ \forall m, \tag{23}$$

where in the objective function from (14), we have neglected the shadow fading z_{mk} and replaced the expectation with the sum over the users. The gradient in (23) is calculated as

$$\frac{\partial}{\partial \mathbf{q}_m^{(j)}} \left\{ \sum_{k=1}^{K} \log \left(1 + \rho_r \sum_{m=1}^{M} \beta_{mk} \right) \right\} = \frac{\gamma \rho_r}{2} \sum_{k=1}^{K} \frac{1}{1 + \psi_k^{\text{ZF}}} \frac{(\mathbf{p}_k - \mathbf{q}_m^{(j)})}{||\mathbf{p}_k - \mathbf{q}_m^{(j)}||^{\gamma+2}}. \tag{24}$$

2) Max-min gradient: For minimum rate, taking the gradient of the rate of the worst user is fragile since the absolute value of the minimum rate can vary significantly across the iterations of gradient ascent causing convergence

issues. Additionally, we are also interested in evaluating the performance of the cell-free system in terms of the 95%-likely rate, which quantifies the best rate of the worst 5% of the users. Accordingly, we consider the sum of rates corresponding to the worst 5% of the users, represented by $\mathcal{K}_{5\%}$, as

$$\mathbf{q}_m^{(j+1)} = \mathbf{q}_m^{(j)} + \delta \frac{\partial}{\partial \mathbf{q}_m^{(j)}} \left\{ \sum_{k \in \mathcal{K}_{5\%}} \log \left(1 + \rho_r \sum_{m=1}^{M} \beta_{mk} \right) \right\}, \ \forall m, \tag{25}$$

which uses the same simplifications as in (23), and the gradient is

$$\frac{\partial}{\partial \mathbf{q}_m^{(j)}} \left\{ \log \left(1 + \rho_r \sum_{m=1}^{M} \beta_{mk} \right) \right\} = \frac{\gamma \rho_r}{2} \sum_{k \in \mathcal{K}_{5\%}} \frac{1}{1 + \psi_k^{\text{ZF}}} \frac{(\mathbf{p}_k - \mathbf{q}_m^{(j)})}{||\mathbf{p}_k - \mathbf{q}_m^{(j)}||^{\gamma+2}}. \tag{26}$$

V. Simulation Methodology and Results

A. Simulation Parameters

In a geographical area of dimensions 2 km $\times$ 2 km, we consider $M = 32$ APs and $K = 4$ users, since $M \gg K$ for a cell-free system. For the purposes of placement, however, we use 2000 users that are distributed according to a GMM of the form in (19), with covariance $\mathbf{\Sigma}_l = \sigma_l^2 \mathbf{I}$, where σ_l is the standard deviation of mixture component l and $\mathbf{I}$ is the identity matrix. We choose the GMM as it can approximate a broad class of distributions [63], [64] and the PDF estimation step of PDFVQ can be avoided for simplicity. The GMM used in our simulations has parameters $L = 3$, $\boldsymbol{\mu}_1 = [0.5, -0.5]^T$, $\boldsymbol{\mu}_2 = [0, 0.5]^T$, $\boldsymbol{\mu}_3 = [-0.5, 0]^T$, $\sigma_1 = \sigma_2 = \sigma_3 = 100$, $p_1 = 0.6$, and $p_2 = p_3 = 0.2$. The pathloss model from [54, (4.34)] is used with shadow fading ignored for simplicity and the transmit power ρ_r of the users is increased from 5 to 30 dB. The max-sum and max-min gradient methods utilize step sizes of $\delta = 10^3$ and $\delta = 3 \times 10^4$, respectively.

B. Performance Measures

The per-user achievable rate is used, and is defined for user k as

$$R_k = \mathbb{E}\left\{ \log_2 \left(1 + \psi_k^{\text{ZF}} \right) \right\}, \tag{27}$$

with ψ_k^{ZF} from (6) and the rate values are generated using Monte Carlo iterations. Further, algorithms are run multiple times and the solution that yields the best result is chosen. The maximum number of iterations for the Lloyd algorithm is set to 50. For comparison among the algorithms, both the sum rate and the 95%-likely rate measures are used. The relative performance between algorithms (say, algorithm 2 over algorithm 1) can be calculated by using the following measure expressed as percentage

$$\text{Improvement Ratio} = \frac{\mathcal{P}^{\text{Algorithm 2}} - \mathcal{P}^{\text{Algorithm 1}}}{\mathcal{P}^{\text{Algorithm 1}}} \times 100, \tag{28}$$

where $\mathcal{P}$ is either the sum rate or 95%-likely rate.

We note here that while the sum rate and minimum rate maximization problems are solved by the max-sum and max-min algorithms, respectively, in [52], we do not include the results here. As described in Section III-A, the max-sum algorithm does not solve the sum rate problem accurately and generates a colocated-like solution with a remarkably poor 95%-likely rate performance (reflected in the numerical simulations of [52]). Further, in [53], we compare the grid-based max-min performance with the Lloyd algorithm and show that their rate performances are comparable. Even though a higher 95%-likely rate is achievable with the max-min algorithm, the simplicity of the Lloyd algorithm outweighs the complexity associated with the increased grid resolution needed.

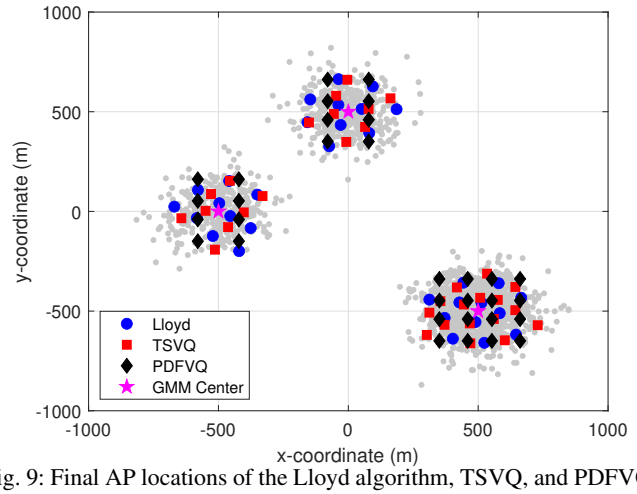


Fig. 9: Final AP locations of the Lloyd algorithm, TSVQ, and PDFVQ at $\rho_r = 30$ dB.

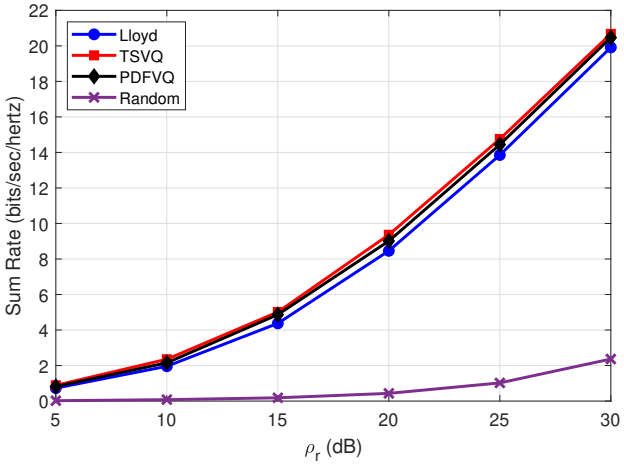


Fig. 10: Sum rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random AP locations

.

TABLE I: Rate Improvement of TSVQ and PDFVQ Relative to the Lloyd Algorithm at $\rho_r = 30$ dB

Algorithm	Sum Rate	95%-Likely Rate
TSVQ	3.82%	−7.98%
PDFVQ	2.78%	−0.82%

C. Numerical Results

Experiment 1. We compare the performances of TSVQ and PDFVQ to the standard Lloyd algorithm. Their final AP

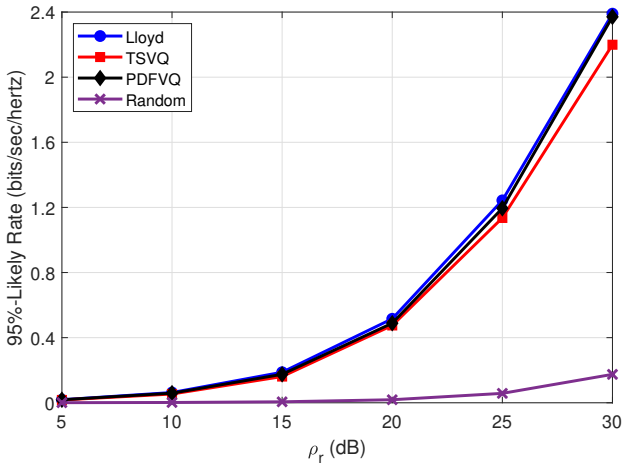


Fig. 11: 95%-likely rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random AP locations

.

locations and the distribution of the 2000 users (as gray circles) are shown in Fig. 9, where the Lloyd algorithm results in APs that are more distributed than TSVQ. This is expected since the TSVQ algorithm successively splits Voronoi regions and after completion does not necessarily associate each user to its closest AP. For PDFVQ, the allocation procedure within each cluster results in 3.85, 2.93, and 2.93 APs along each dimension, respectively, since the GMM clusters are symmetric along both x- and y-coordinates. Although rounding off these values would result in 4, 3, and 3 APs along each coordinate, the total number of APs would then be 34. Thus, to limit to $M = 32$ APs, we select 4, 2, and 2 APs along the x-coordinate and 4, 4, and 4 APs along the y-coordinate. This combination is selected as it provides the best result through repeated trials. The AP locations are observed to be more regular and grid-like due to the scalar quantizers used in the transform coding design, when compared to the standard Lloyd algorithm or TSVQ. The sum and 95%-likely rates of the three VQ-based algorithms are plotted in Fig. 10 and Fig. 11, respectively. Additionally, for comparison, the performances when the APs are randomly placed are plotted (the AP locations themselves are omitted in Fig. 9 to avoid cluttering the figure), showing an expected low performance. The improvements in sum and 95%-likely rates, expressed as percentages, of TSVQ and PDFVQ over the Lloyd algorithm at transmit power $\rho_r = 30$ dB are noted in Table I. It is seen that while the sum rate of TSVQ is greater than that of the Lloyd algorithm (by nearly 4%), the 95%-likely rate is worse (by nearly 8%). For PDFVQ, the sum rate offered is close to that of TSVQ while the 95%-likely rate approaches but does not equal that of the Lloyd algorithm (with the difference being less than 1%), due to the fact that the AP positions are far away from a small group of users in two mixture components. While for the user distribution considered PDFVQ offers a solution providing higher sum rate and similar 95%-likely rate to the Lloyd algorithm, it can be expected that this performance will reduce as the variance of the mixture components is increased.

Experiment 2. In Experiment 1, we compared the VQ approaches for a GMM consisting of Gaussian mixture components with spherical (proportionate to the identity matrix) covariance matrices. In order to show an example of PDFVQ applied to a full (non-diagonal) covariance matrix, we now consider a user distribution where the second component of the GMM considered is modified to have a covariance matrix of $\mathbf{\Sigma}_2 = \begin{bmatrix} \sigma^2 & 2\sigma^2/3 \\ 2\sigma^2/3 & 2\sigma^2 \end{bmatrix}$ with $\sigma = 100$ and the mean of the third component $\boldsymbol{\mu}_3 = [-0.5, -0.5]^T$. When PDFVQ is applied to this distribution, 4, 2, 3 APs along the x-coordinate and 4, 5, and 3 APs along the y-coordinate are allocated, totaling 35 APs. The best allocation for the desired 32 APs is found to be 4, 2, and 4, and 4, 4, and 2 APs along the x- and y-coordinates, respectively. The positions of such APs along with those of the Lloyd algorithm and TSVQ are shown in Fig. 12. The sum rate and 95%-likely rates corresponding to these locations are provided in Fig. 13 and Fig. 14, respectively, and the rate improvements are tabulated in Table II. Again, the performances of the random AP locations are significantly worse than the VQ-based methods. Similar to the GMM with spherical Gaussians in Experiment 1, we observe that PDFVQ is able to match the sum rate performance of TSVQ. However, unlike above, the 95%-likely rate, like TSVQ, is lower than the Lloyd algorithm by nearly 13% since the space-filling advantage of the Lloyd algorithm is lost. Based on the above two experiments, it could be concluded that PDFVQ is a reasonable alternative to TSVQ that provides a similar or superior performance. We also note that the conclusions regarding the relative performance shown above between PDFVQ and the Lloyd algorithm remain the same when the number of APs is increased as in a massive MIMO system (the results are not shown here due to space constraints).

TABLE II: Rate Improvement of TSVQ and PDFVQ Relative to the Lloyd Algorithm With a Full Covariance Matrix at $\rho_r = 30$ dB

Algorithm	Sum Rate	95%-Likely Rate
TSVQ	5.59%	-13.67%
PDFVQ	4.98%	-12.85%

Although it is not the case here, we also note that the clusters of the GMM may be placed quite close to one another or may even overlap marginally. When PDFVQ is applied, the APs associated with the overlapping clusters may be closely placed and is inefficient. In such a scenario, starting from the cluster pair with most overlap (the degree of which can be computed by a measure such as overlap rate [65]), the AP allocation of the cluster with the higher 95%-likely rate could be decreased and re-allocated to another cluster with a poor performance. This process can be repeated for all overlapping clusters to improve the 95%-likely rate of the network.

Experiment 3. In this experiment, we use the gradient approaches outlined in Section IV-D to improve the sum rate and 95%-likely rate performances shown in Experiment 2 for the GMM with a full covariance matrix. First, we show that using the max-sum gradient ascent, we can increase the sum rates of each of the Lloyd, TSVQ, and PDFVQ performances, as shown in Fig. 15. Note that in order to view all the curves better, we have included an additional figure that focuses on the power levels $\rho_r = 20$, 25, and 30 dB. For comparison with an existing method, we also plot the performance when

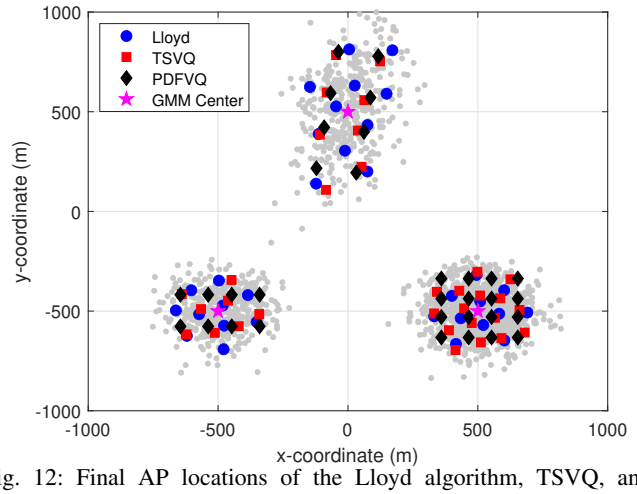


Fig. 12: Final AP locations of the Lloyd algorithm, TSVQ, and PDFVQ with a full covariance matrix at $\rho_r = 30$ dB.

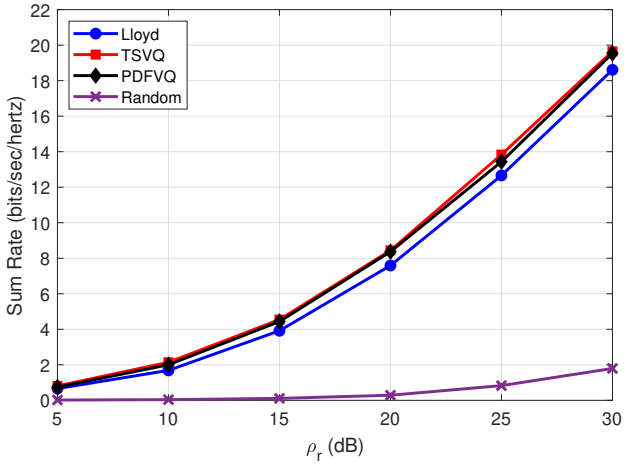


Fig. 13: Sum rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random AP locations with a full covariance matrix.

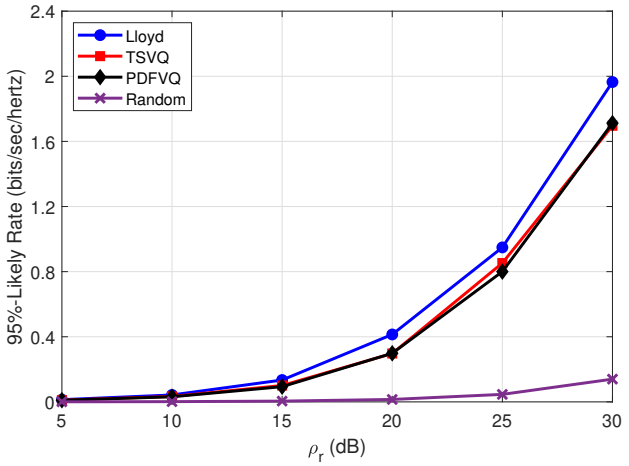


Fig. 14: 95%-likely rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random AP locations with a full covariance matrix.

the max-sum gradient ascent algorithm is applied to randomly initialized AP positions. This initialization considers that the APs are allocated to each GMM cluster according to PDFVQ and then positioned using the i.i.d. Gaussian distribution of the cluster. Table III lists the percentage rate improvements of the max-sum gradient applied to each VQ approach, over the Lloyd algorithm alone. The gradient ascent applied to both the TSVQ and PDFVQ AP locations yields the highest sum rate (nearly 10% over the Lloyd algorithm) while the rate obtained when the ascent operation is applied with the Lloyd algorithm as the starting point is nearly the same as the sum rates of PDFVQ and TSVQ. This occurrence is due to the fact that the gradient ascent iterations with the Lloyd AP solutions as the initial points converge to a local optimum which is different from that obtained when the ascent is applied to PDFVQ or TSVQ. The 95%-likely rate performances shown in Fig. 16, when the max-sum gradient is applied to PDFVQ and TSVQ, do not change significantly (an increase is observed at $\rho_r = 30$ dB) and decrease when the gradient is applied to the Lloyd AP positions. Thus, in terms of sum rate, PDFVQ provides the best solution out of all VQ approaches and a further increase in sum rate (about 4% over PDFVQ) without negatively affecting the minimum rate performance (about 5% increase) is achieved by using the max-sum gradient ascent. For the randomly initialized case, it is observed that for a similar sum rate performance as the other techniques, the 95%-likely rate is remarkably inferior, thus showing the superiority of the VQ-initialized gradient approaches. Next, the max-min gradient ascent is applied and the 95%-likely rate is observed to increase as shown in Fig. 17. The best rate is obtained when the ascent algorithm is applied to the Lloyd solution as opposed to when it is applied to either TSVQ or PDFVQ where it is able to match the performance of the Lloyd algorithm. Table IV informs us that while a 14% improvement in the 95%-likely rate is achieved by applying the gradient to the Lloyd solution, the difference of the gradient applied to TSVQ or PDFVQ from the Lloyd algorithm is only up to 4%. The sum rates corresponding to the 95%-likely rates are plotted in Fig. 18 (which also includes an additional zoomed in figure), where as a result of the 95%-likely rate improvement, the sum rate when the ascent is applied to the Lloyd algorithm is the least value. In contrast, despite the increase in 95%-likely rate, the sum rate performances when ascent is applied to TSVQ and PDFVQ are nearly the same as TSVQ and PDFVQ itself. Hence, with its simpler design, PDFVQ along with max-min gradient offers a good tradeoff between a 95%-likely rate similar to and a sum rate higher (by over 5%) than the Lloyd algorithm. It should, however, be mentioned that if 95%-likely rate (or minimum rate) is the sole performance measure of interest, the Lloyd algorithm alone is a straightforward choice. The gradient approach requires the choice of an appropriate step size for convergence while the Lloyd algorithm is known to converge [66]. Finally, as we observed for the max-sum case, while a similar sum rate is observed when the max-min gradient is applied to randomly initialized AP locations, the 95%-likely rate is significantly inferior to the other situations.

Experiment 4. In this experiment, we quantify the effect of a time-varying user density on the network performance and show the need for an easily adaptable AP placement algorithm. We consider a simple situation where the users are *initially* distributed as a single-cluster GMM ($L = 1$ in (19)) with mean

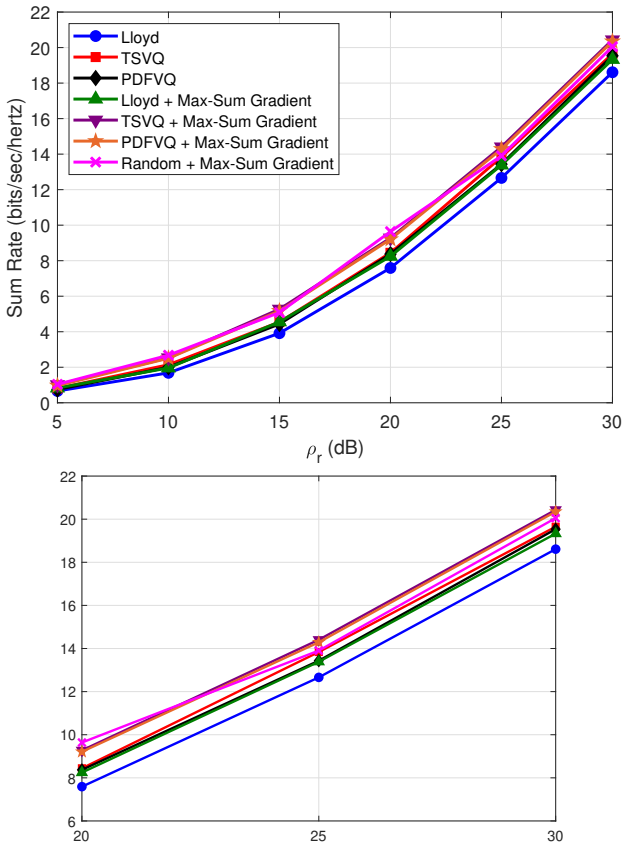


Fig. 15: Sum rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random locations with a full covariance matrix along with max-sum gradient. The figure on the bottom zooms in on the sum rates for $\rho_r = 20$, 25, and 30 dB.

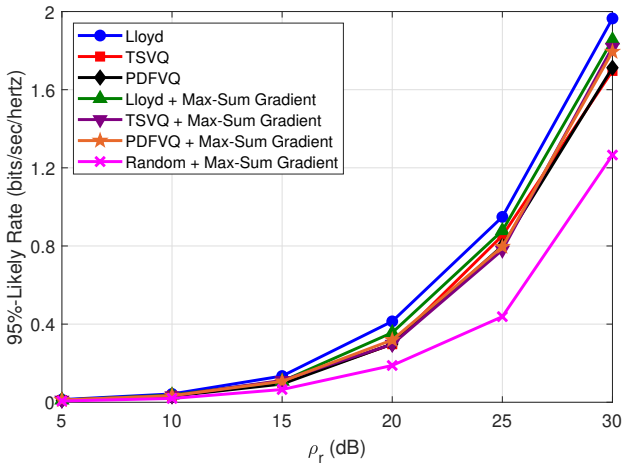


Fig. 16: 95%-likely rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random locations with a full covariance matrix along with max-sum gradient.

TABLE III: Rate Improvement of the VQ Approaches with the Max-Sum Gradient Relative to the Lloyd Algorithm at $\rho_r = 30$ dB

Algorithm	Sum Rate	95%-Likely Rate
Lloyd + Max-Sum Gradient	3.88%	−5.60%
TSVQ + Max-Sum Gradient	9.80%	−7.52%
PDFVQ + Max-Sum Gradient	9.34%	−8.66%

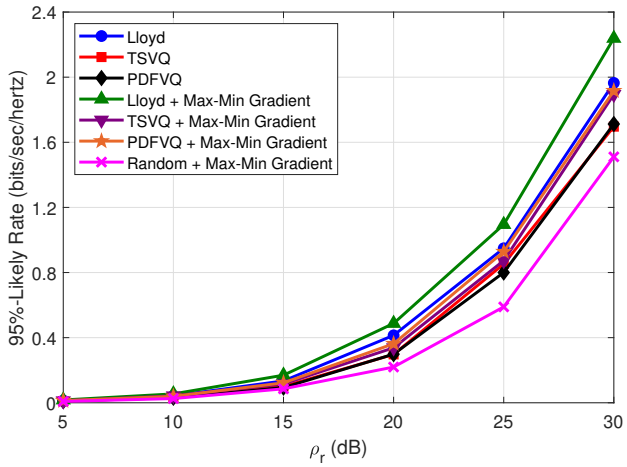


Fig. 17: 95%-likely rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random locations with a full covariance matrix along with max-min gradient.

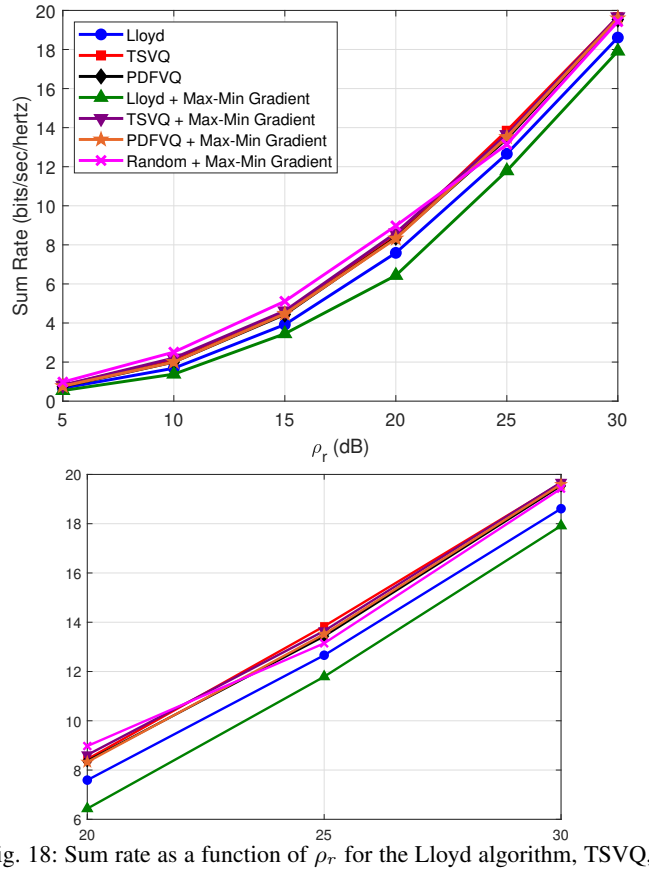
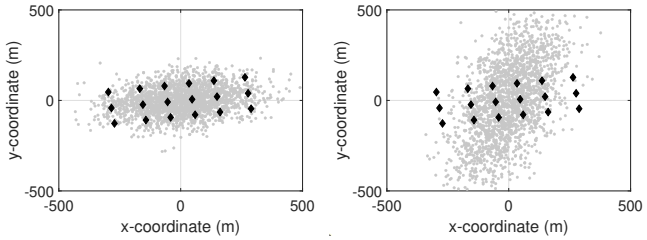


Fig. 18: Sum rate as a function of ρ_r for the Lloyd algorithm, TSVQ, PDFVQ, and random locations with a full covariance matrix along with max-min gradient. The figure on the bottom zooms in on the sum rates for $\rho_r = 20$, 25, and 30 dB.

TABLE IV: Rate Improvement of the VQ Approaches with the Max-Min Gradient Relative to the Lloyd Algorithm at $\rho_r = 30$ dB

Algorithm	95%-Likely Rate	Sum Rate
Lloyd + Max-Min Gradient	13.93%	−3.73%
TSVQ + Max-Min Gradient	−3.60%	5.73%
PDFVQ + Max-Min Gradient	−2.48%	5.30%

$\boldsymbol{\mu} = [0, 0]^T$ and covariance $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma^2/3 \\ \sigma^2/3 & \sigma^2/2 \end{bmatrix}$ where $\sigma = 200$, and call it density A. Assuming a total of $M = 18$ APs, we



(a) Density A. (b) Density B.

Fig. 19: The two user densities A and B considered for Experiment 4 along the PDFVQ AP locations matched to density A.

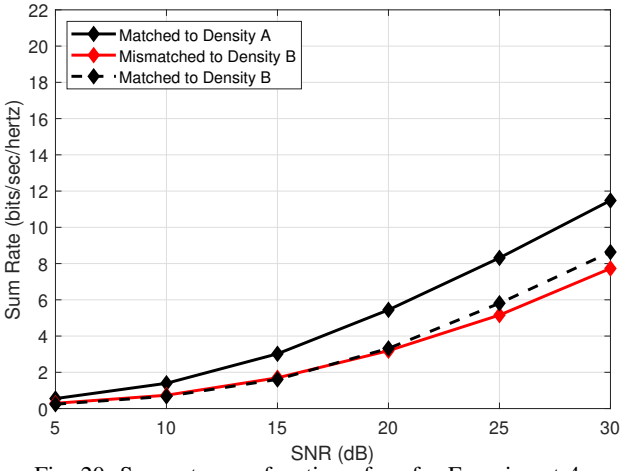


Fig. 20: Sum rate as a function of ρ_r for Experiment 4.

calculate the PDFVQ AP locations for this user density. Over time, we consider that the user distribution changes to density B, where the users are more spread out along another direction to that of density A, with covariance matrix $\mathbf{\Sigma} = \begin{bmatrix} \sigma^2/2 & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 \end{bmatrix}$ and $\sigma = 300$. These two user densities and the PDFVQ AP locations determined for density A are shown in Fig. 19. We evaluate the sum rate and 95%-likely rate performances when the AP locations are matched to density A and consider both user densities A and B, so that the loss due to mismatch of the AP locations to the user density B is also shown. For completeness, we also compute the rate values when PDFVQ AP locations are determined and matched to density B. As expected, in both the sum rate plotted in Fig. 20 and the 95%-likely rate plotted in Fig. 21, there is a significant loss in performance for the PDFVQ AP locations which are matched to density A, when the users re-position to density B. The relative losses in the sum rate and 95%-likely rate at $\rho_r = 30$ dB are 32.67% and 89.54%, respectively. To prevent this performance decrease, there is a need to re-calculate the AP locations for the new user density. Clearly, when the AP locations are matched via PDFVQ to the user density B, the rate performances in Fig. 20 and Fig. 21 are improved over the diminished performance of the mismatched case. With easy adaptability and the other aforementioned advantages, PDFVQ offers the best method for cell-free AP placement among the VQ techniques discussed in this paper. In future work, we will address how the changing user densities can be learned for use in PDFVQ.

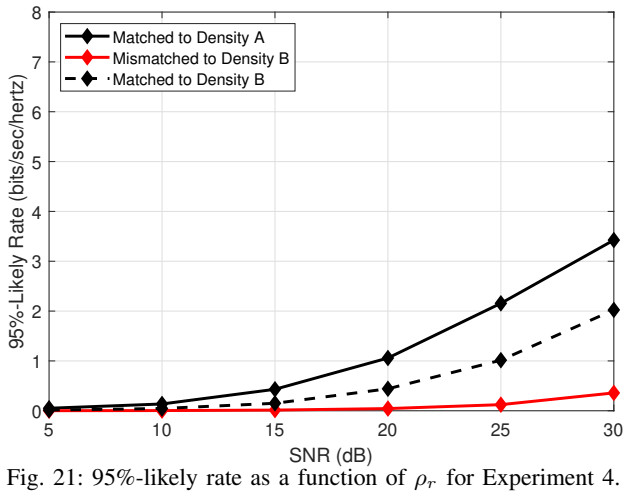


Fig. 21: 95%-likely rate as a function of ρ_r for Experiment 4.

VI. Conclusion

In this paper, we have addressed access point (AP) placement in cell-free massive multiple-input-multiple-output (MIMO) systems under a throughput criteria. We investigated the two main optimization problems in this regard, namely the sum rate and minimum rate maximization problems. To understand their solution frameworks, simple examples were constructed and analyzed exposing the difficulty in solving the problems. Therefore, as a practical approach, the use of vector quantization (VQ) based methods, namely the popular Lloyd algorithm, tree-structured VQ (TSVQ), and probability density function optimized VQ (PDFVQ), to cell-free AP placement, was investigated. Among the three algorithms presented, although the tree-structured VQ (TSVQ) provides better sum rate (as it fosters cooperation among APs by placing them closer) compared to the Lloyd algorithm, it suffers from high complexity, poor scalability, and the inability to easily adapt to new environments. PDFVQ, which overcomes the aforementioned shortcomings, allowed a more efficient generation of the codebook and generated a sum rate similar to and 95%-likely rate higher than TSVQ and close to the Lloyd algorithm. Additionally, for gradient-based maximization methods, PDFVQ is found to provide good initial points. It was observed numerically that, over the Lloyd algorithm, an increase of 9% in sum rate and a difference of just 2.5% in the 95%-likely rate was achieved by applying max-sum and max-min gradient ascent algorithms, respectively, with the PDFVQ AP locations as starting points. Thus, PDFVQ offers a convenient, less computationally intensive, and easily scalable AP placement technique for cell-free networks.

References

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

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[3] R. Chataut and R. Akl, "Massive MIMO systems for 5G and beyond networks — overview, recent trends, challenges, and future research direction," in *Sensors*, vol. 20, no. 10, May 2020, Art. ID 2753.

[4] J. Jeon, G. Lee, A. A. Ibrahim, J. Yuan, G. Xu, J. Cho, E. Onggosanusi, Y. Kim, J. Lee, and J. C. Zhang, "MIMO evolution toward 6G: Modular massive MIMO in low-frequency bands," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 52–58, Nov. 2021.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[6] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[7] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 69–73, Jan. 2007.

[8] K. T. Truong and R. W. Heath, "The viability of distributed antennas for massive MIMO systems," in *Proc. 2013 47th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2013, pp. 1318–1323.

[9] J. Wang, H. Zhu, and N. J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 675–683, May 2012.

[10] E. Koyuncu, "Performance gains of optimal antenna deployment for massive MIMO systems," in *Proc. 2017 IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[11] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-free massive MIMO systems," in *Proc. 2015 49th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2015, pp. 695–699.

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

[13] A. A. I. Ibrahim, A. Ashikhmin, T. L. Marzetta, and D. J. Love, "Cell-free massive MIMO systems utilizing multi-antenna access points," in *Proc. 2017 51th Asilomar Conf. Signals, Syst., Comput.*, Oct. 2017, pp. 1517–1521.

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

[15] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Proc. 2007 41th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2007, pp. 83–87.

[16] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[17] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[18] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/June 2020.

[19] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127 639–127 651, Sept. 2019.

[20] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11ax high efficiency WLANs," *IEEE Commun. Surveys Tut.*, vol. 21, no. 1, pp. 197–216, Firstquarter 2019.

[21] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, July 2017.

[22] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. and Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[23] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 651–663, Sept. 2019.

[24] N. Rajapaksha, K. B. S. Manosha, N. Rajatheva, and M. Latva-Aho, "Deep learning-based power control for cell-free massive MIMO networks," in *Proc. 2021 IEEE Int. Conf. Commun. (ICC)*, June 2021, pp. 1–7.

[25] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink max–min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.

[26] M. Farooq, H. Q. Ngo, E.-K. Hong, and L.-N. Tran, "Utility maximization for large-scale cell-free massive MIMO downlink," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 7050–7062, Oct. 2021.

[27] M. Alonzo and S. Buzzi, "Cell-free and user-centric massive MIMO at millimeter wave frequencies," in *Proc. 2017 IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.

[28] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[29] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *Proc. 2019 IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[30] S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, Feb. 2020.

[31] Y. Al-Eryani, M. Akrout, and E. Hossain, "Multiple access in cell-free networks: Outage performance, dynamic clustering, and deep reinforcement learning-based design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1028–1042, Apr. 2021.

[32] M. Bashar, H. Q. Ngo, K. Cumanan, A. G. Burr, P. Xiao, E. Björnson, and E. G. Larsson, "Uplink spectral and energy efficiency of cell-free massive MIMO with optimal uniform quantization," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 223–245, Jan. 2021.

[33] 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 Commun. Lett.*, vol. 10, no. 1, pp. 34–37, Jan. 2021.

[34] Y. Zhang, B. Di, H. Zhang, J. Lin, C. Xu, D. Zhang, Y. Li, and L. Song, "Beyond cell-free MIMO: Energy efficient reconfigurable intelligent surface aided cell-free MIMO communications," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 412–426, June 2021.

[35] T. T. Vu, H. Quoc Ngo, T. L. Marzetta, and M. Matthaiou, "How does cell-free massive MIMO support multiple federated learning groups?" in *Proc. 2021 IEEE 22nd Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Sept. 2021, pp. 401–405.

[36] J. Košmerl and A. Vilhar, "Base stations placement optimization in wireless networks for emergency communications," in *Proc. 2014 IEEE Int. Conf. Commun. (ICC) Workshops*, June 2014, pp. 200–205.

[37] X. Wang, P. Zhu, and M. Chen, "Antenna location design for generalized distributed antenna systems," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 315–317, May 2009.

[38] 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, Oct. 2015.

[39] S. T. Abraha, D. F. Castellana, X. Liang, A. Ng'oma, and A. Kobyakov, "Experimental study of distributed massive MIMO (DM-MIMO) in in-building fiber-wireless networks," in *Proc. Opt. Fiber Commun. Conf. Expo. (OFC)*, 2018.

[40] S. Karimi-Bidhendi, J. Guo, and H. Jafarkhani, "Energy-efficient node deployment in heterogeneous two-tier wireless sensor networks with limited communication range," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 40–55, Jan. 2021.

[41] E. Park, S. Lee, and I. Lee, "Antenna placement optimization for distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2468–2477, July 2012.

[42] G. R. Gopal, E. Nayebi, G. P. Villardi, and B. D. Rao, "Modified vector quantization for small-cell access point placement with inter-cell interference," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6387–6401, Aug. 2022.

[43] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1647–1650, Aug. 2016.

[44] J. Lyu, Y. Zeng, R. Zhang, and T. J. Lim, "Placement optimization of UAV-mounted mobile base stations," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 604–607, Mar. 2017.

[45] Y. Zhang and L. Dai, "Joint optimization of placement and coverage of access points for IEEE 802.11 networks," in *Proc. 2020 IEEE Int. Conf. Commun. (ICC)*, June 2020, pp. 1–7.

[46] J. Guo, P. Walk, and H. Jafarkhani, "Optimal deployments of UAVs with directional antennas for a power-efficient coverage," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5159–5174, Aug. 2020.

[47] G. R. Gopal, B. D. Rao, and G. P. Villardi, "Access point placement for hybrid UAV-terrestrial small-cell networks," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1826–1841, July 2021.

[48] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. IEEE*, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.

[49] C. Diaz-Vilor, A. Lozano, and H. Jafarkhani, "On the deployment problem in cell-free UAV networks," in *Proc. 2021 IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.
[50] Y.-H. Zhu, G. Callebaut, H. Çalık, L. Van der Perre, and F. Rottenberg, "Energy efficient access point placement for distributed massive MIMO," *Netw.*, vol. 2, no. 2, pp. 288–310, May 2022. [Online]. Available: https://www.mdpi.com/2673-8732/2/2/19
[51] L. Wang and Q. Zhang, "Cell-free massive MIMO with UAV access points: UAV location optimization," in *Proc. 2022 IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2022, pp. 262–267.
[52] E. Nayebi and B. D. Rao, "Access point location design in cell-free massive MIMO systems," in *Proc. 2018 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 985–989.
[53] G. R. Gopal, G. P. Villardi, and B. D. Rao, "Is vector quantization good enough for access point placement?" in *Proc. 2021 55th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2021.
[54] E. Nayebi, "TDD massive MIMO systems: Channel estimation, power optimization, and access point location design," Ph.D. dissertation, University of California San Diego, 2018.
[55] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free massive MIMO systems with MMSE and LSFD receivers," in *Proc. 2016 50th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2016, pp. 203–207.
[56] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[57] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer Academic Publishers, 1991.
[58] T. Lookabaugh and R. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1020–1033, Sept. 1989.
[59] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, ser. SODA '07. USA: Society for Industrial and Applied Mathematics, Jan. 2007, p. 1027–1035.
[60] A. Subramaniam and B. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 2, pp. 130–142, Mar. 2003.
[61] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, Apr. 1984.
[62] A. K. Jain, *Fundamentals of Digital Image Processing*. USA: Prentice-Hall, Inc., 1989.
[63] G. McLachlan and D. Peel, *Finite Mixture Models*. USA: John Wiley & Sons, Inc., 2000.
[64] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
[65] H. Sun and S. Wang, "Measuring the component overlapping in the Gaussian mixture model," *Data Mining Knowl. Discovery*, vol. 23, pp. 479–502, Nov. 2011.
[66] M. Sabin and R. Gray, "Global convergence and empirical consistency of the generalized Lloyd algorithm," *IEEE Trans. Inf. Theory*, vol. 32, no. 2, pp. 148–155, Mar. 1986.

Govind R. Gopal (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the National Institute of Technology Karnataka, Surathkal, India, in 2016, and the M.S. and Ph.D. degrees in electrical and computer engineering with an emphasis on communication theory and systems from the University of California San Diego, La Jolla, CA, USA, in 2018 and 2023, respectively. He is currently a Radio Research Specialist at Nokia Standards, Nokia of America Corporation, Naperville, IL, USA. His previous experience is on routing in device-to-device communications, wireless communication theory, signal processing for wireless applications, multiple-input-multiple-output small-cell and cell-free systems, vector quantization, and access point location design. His current research interests include beyond 5th generation and 6th generation wireless systems, artificial intelligence and machine learning applications to wireless networks, and 3GPP wireless standards.

Bhaskar D. Rao (Life Fellow, IEEE) is currently a Distinguished Professor in the Department of Electrical and Computer Engineering and the holder of the Ericsson Endowed Chair in wireless access networks at the University of California San Diego, La Jolla, CA, USA. He received the 2016 IEEE Signal Processing Society Technical Achievement Award. His research interests include digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and biomedical signal processing.