Access Point Placement for Hybrid UAV-Terrestrial Small-Cell Networks

GOVIND R. GOPAL[1] **(Graduate Student Member, IEEE), BHASKAR D. RAO**[1] **(Fellow, IEEE), AND GABRIEL PORTO VILLARDI**[2] **(Senior Member, IEEE)**

[1]Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093, USA

[2]Wireless Network Research Center, National Institute of Information and Communications Technology, Yokosuka 239-0847, Japan

CORRESPONDING AUTHOR: G. R. GOPAL (e-mail: ggopal@ucsd.edu)

ABSTRACT We address the problem of access point (AP) placement in small-cell networks with partial infrastructure flexibility, i.e., a novel class of problem in Beyond 5G, resultant from the utilization of unmanned aerial vehicles (UAVs) with AP functionalities (UAV-APs), to aid fixed wireless networks in coping with momentary peak-capacity requirements. We use the signal-to-generated-interference-plus-noise ratio (SGINR) metric as an alternative to the traditional signal-to-interference-plus-noise ratio (SINR) to quantify the effects of inter-cell interference (ICI) on the per-user capacity. From average SGINR, we derive the ICI-aware distortion measure leading to the Inter-AP Lloyd algorithm to obtain throughput-optimal AP placement for a fully flexible infrastructure. We then impose a hybridity constraint to the AP placement problem which turns a fraction of the network into a fixed infrastructure composed of terrestrial APs (T-APs) while the remainder is constituted by UAV-APs with flexibility in position. This newly formulated AP placement problem is solved by the proposed Lloyd-type algorithm called Hybrid AP Placement Algorithm (HAPPA). Furthermore, we present an initialization method for the Lloyd and Lloyd-type algorithms for Gaussian mixture models (GMMs) that offers an AP allocation leading to a higher rate compared to the k-means++ initialization. Finally, computer simulations show that the Inter-AP Lloyd algorithm can improve the performance of the worst users by up to 42.75% in achievable rate, assuming a fully flexible network. By using HAPPA on hybrid networks, we achieve improvements of up to 71.92% in sum rate over the fixed network and close the performance gap with fully flexible networks down to 2.02%, when an equal number of UAV-APs and T-APs is used. Further, our proposed initialization scheme always results in a balanced AP allocation, which means a more even distribution of users per AP, whereas the k-means++ scheme results in unbalanced allocations at least 30% of the time, resulting in a worse minimum rate.

INDEX TERMS Beyond 5G, hybrid network, inter-cell interference (ICI), Lloyd algorithm, unmanned aerial vehicle (UAV).

I. INTRODUCTION

MASSIVE multiple-input-multiple-output (MIMO) systems have emerged as a prevalent technology for the fifth generation (5G) of wireless communication systems [1]–[5]. Using a large number of antennas, such systems favor channel hardening and reduce interference with increased diversity. For higher rates, distributed antenna systems (DASs) in the form of distributed MIMO are used in lieu of co-located antenna systems [6]–[14]. While cooperative distributed massive MIMO systems, such as cell-free systems [15], [16], can result in enhanced interference mitigation and spectral efficiency, the considerable backhaul necessary for information exchange poses limitations to their practicality [8], [17], [18]. As such, small-cell systems (with limited or no cooperation) are still being preferred over their cooperative counterparts.

For small-cell systems, access point (AP) placement is an integral part of system design and has been studied in the past [7], [10], [13], [19], [20]. None of these works, however, consider the deleterious effects of ICI when calculating AP positions. Recently, unmanned aerial vehicles (UAVs) have been conceived as a means of providing wireless connectivity for the Beyond 5G paradigm. Most of the prior works focus on UAV-terrestrial propagation channel models, routing and energy efficiency, and applications of UAVs to ad hoc networks, civil applications, and Internet-of-Things. Cellular-connected UAVs have seen an increased interest [21]–[23]. Additionally, the mobility inherent to UAVs has made the topic of small-cell AP placement gain even more traction recently. For instance, [24] provides an in-depth tutorial on UAVs, from both the perspective of end-users and APs (UAV-APs). References [25]–[40] (and references therein) provide some examples of UAV placement techniques. The authors in [25] propose the deployment of UAVs at demand hotspots along with an incumbent macro-cell network, by intelligently off-loading sets of users obtained using the K-means algorithm. In [26], mixed integer non-linear optimization with a quadratic constraint is employed to determine the 3-D positions of a single UAV. Placement, however does not aim at throughput maximization and instead attempts to increase coverage to users not previously covered by a terrestrial AP (T-AP). The work in [27] considers the placement of UAV-APs to maximize ground user coverage and alters their altitude to minimize interference to users, however, without considering an existing ground network. Authors in [28] expand the integer non-linear optimization to include a minimum transmit power constraint and separate horizontal and vertical placements. Heuristic successive placement of UAVs with a fixed ground coverage radius in a spiral fashion is performed in [29] to ensure coverage for all users. A cloud radio access network is considered in [30] consisting of terrestrial remote radio heads (RRHs) and UAVs. The positions and contents of the cache-enabled UAVs are optimized, but interference between RRH-user and UAV-user links is not modeled. The work in [31] considers a hybrid architecture consisting of a single ground base station and UAVs flying cyclically around the cell edge to offload users and improve minimum user throughput. By considering a one-hop or two-hop downlink communication scheme between T-APs, UAV-APs, and users, [32] investigates UAV placement coupled with bandwidth and power allocation in the backhaul and access links, however, leaving out ICI considerations. A spectrum sharing scenario is considered in [33] in which UAVs communicate to secondary receivers while minimizing interference to the primary terrestrial transmitter-receiver pair. Two scenarios are considered to maximize the rate of the secondary receiver - one, where the 3-D placements of static UAVs are optimized, and the other, where the trajectories of mobile UAVs are optimized, both with power control. Authors in [34] also address the ground coverage problem (as in [26]). Unlike the previous works, [35] utilizes a user density-driven 3-D UAV placement to attempt maximum coverage of users with minimum data rate requirements. Also, maximizing throughput (utilizing the SNR alone) for mobile users along with random network coding packet scheduling is jointly considered for UAV-AP placement in [36]. In [37], authors take into account the time varying nature of the users' positions to find the optimal placement and coverage radius of the UAV-APs by using cooperative stochastic approximation for Wi-Fi (IEEE 802.11) networks. [38] explores the 3-D deployment of multiple UAVs having directional antennas to minimize the average transmit power of the users. The work in [39] develops user access-based trajectory design for UAV-APs using a value decomposition reinforcement learning algorithm, trained using meta training. Finally, [40] considers UAV-APs to offload traffic from terrestrial networks. Using a weighted expectation maximization algorithm, the user distribution and traffic demand are estimated so that the UAV-APs can be placed by maximizing the utility of the overloaded AP. However, interference between the terrestrial and aerial communication networks is neglected.

None of the works described above, especially the ones that study multiple UAVs, consider ICI during the placement process. Hence, with the Lloyd algorithm as the basis, [41] optimizes rate using the signal-to-interference-plus-noise ratio (SINR) and proposes an alteration to the mean squared error (MSE) distortion function which accounts for ICI and succeeds in improving the minimum rate of the small-cell system over the traditional Lloyd algorithm. Following this direction, in [42], we derive a Lloyd-type algorithm (called the Inter-AP Lloyd algorithm) with a simpler, yet effective distortion measure to deal with ICI when compared to the distortion function of the algorithm presented in [41]. Note that all the above papers consider networks with fully flexible infrastructure, in which all APs are position-flexible and can have their locations updated according to the user distribution at a certain time.

In this work, we investigate the AP placement problem, with the objective of maximizing throughput in a small-cell system, with prevalent ICI and varying user distributions. More specifically, we investigate the placement of APs susceptible to ICI in networks with partial infrastructure flexibility, interchangeably referred to as hybrid networks, which is composed of fixed-position T-APs and flexible-position UAV-APs. The degree of infrastructure flexibility varies between the two extremes determined by fully flexible networks and fixed networks, consisting of either only UAV-APs or only T-APs, respectively. In practice, hybrid networks can be deployed in a fast and dynamical manner to tackle the random nature of capacity due to user mobility, to compensate an expected increase in throughput requirements of users attending an event, or even, a sudden reduction in network capacity due to wireless infrastructure being knocked over as a result of natural disasters. From a wireless operator point-of-view, building a fixed infrastructure with high T-AP density, for instance, in a football stadium, will render the network underutilized most of the

time. On the other hand, offloading T-APs by deploying UAV-APs on-demand is more economically appealing. Also, wildfires, earthquakes, and the ensuing tsunamis have, in the recent past, disrupted communication services, leaving populations isolated [43], [44]. Complementing the part of a network that has withstood a natural disaster with UAV-APs is speedy and desirable; however, it requires careful placement considerations since ICI has been shown to significantly hinder the coverage area of emergency network services [45].

CONTRIBUTIONS
To the best of our knowledge, solutions to the AP placement problem for throughput maximization for hybrid UAV-terrestrial small-cell networks and incorporating ICI into the placement process have not been discussed in literature. Hence, in this work, our main contributions are as follows.

- To include ICI in the AP placement process and with the Lloyd algorithm as the basis, we employ an alternate metric to SINR, namely, the signal-to-generated-interference-plus-noise ratio (SGINR). Optimizing the average SGINR, we derive the distortion measure that accounts for ICI through inter-AP distances, establishing the fact that the optimization of SGINR, like SINR, can lead to the Inter-AP Lloyd algorithm, which is known to perform better than the Lloyd algorithm in terms of the rates of the users most affected by ICI.
- Based on the Inter-AP Lloyd algorithm, we formulate and develop an algorithm for hybrid networks that places UAV-APs in an area occupied by T-APs currently providing sub-optimal throughput performances due to a change in user configuration. This is the Hybrid AP Placement Algorithm (HAPPA) that outperforms the T-APs (fixed network) alone. It also provides a performance close to the ideal one of fully flexible networks, despite the fact that only a fraction of the APs can have their positions adjusted to maximize throughput (minimize ICI).
- For unbiased comparison of the presented algorithms, we develop an initialization scheme for the Gaussian mixture model (GMM) user distribution, based on the bit allocation procedure in vector quantization (VQ). The proposed scheme always offers a balanced allocation of APs that results in each AP serving a relatively similar number of users compared to the k-means++ scheme, which can result in unbalanced allocations. The unbalanced allocations are shown to yield a worse minimum rate over the balanced one.

The remainder of this paper is organized as follows. Section II outlines the small-cell system model used throughout the paper. Section III provides a brief description of the VQ framework used throughout this work. The ICI-aware AP placement problem for fully flexible networks, based on SGINR, is mathematically formulated and solved in Section IV. Then, in Section V, introducing hybridity

into the problem, we formulate the UAV-terrestrial AP placement problem and solve using the proposed HAPPA. The simulation methodology which includes the proposed initialization algorithm and the numerical results are presented in Section VI. Finally, we conclude in Section VII.

II. SYSTEM MODEL
Our system model is based on the small-cell model described in [41], [42], [46], and [47, Ch. 4]. Throughout this manuscript, vectors are denoted by bold letters, $\mathbb{E}\{\cdot\}$ is the expectation operator, $||\cdot||$ represents the ℓ_2-norm of a vector, $|\cdot|$ denotes the cardinality of a set, and all logarithms are to the base 2. We consider a geographical area with K single-antenna users distributed according to a probability density function (pdf) $f_{\mathbf{P}}(\mathbf{p})$, where $\mathbf{p} \in \mathbb{R}^2$ is the random vector denoting the position of a user. These users are served by M single-antenna APs, designated by the set $\mathcal{M}$ such that $|\mathcal{M}| = M$. The location of an AP is denoted by $\mathbf{q} \in \mathbb{R}^2$. Following the theme of the paper, we divide the set of APs $\mathcal{M}$ into T-APs and UAV-APs, denoted by the sets $\mathcal{M}_f$ and $\mathcal{M}_u$, respectively, such that $|\mathcal{M}_f| = M_f$, $|\mathcal{M}_u| = M_u$, $\mathcal{M}_f \cup \mathcal{M}_u = \mathcal{M}$, and $M_f + M_u = M$. For simplicity, the position-flexible UAV-APs are assumed to have the same transmission parameters as the fixed T-APs. Although 3-D channel models with height as a parameter exist for modeling the channel between users and UAV-APs (e.g., [48]), references [49] and [50] use a fixed-height log-distance pathloss model between UAVs and users. We follow this approach and use the same pathloss model between the users and the UAV-APs as well as between the users and T-APs. Moreover, a narrowband flat-fading channel is considered and the channel coefficient between AP m and user k is given by

$$g_{mk} = \sqrt{\beta_{mk}}h_{mk}, \qquad (1)$$

where $m = 1, 2, \ldots, M$ and $k = 1, 2, \ldots, K$. Here, β_{mk} and h_{mk} are the large-scale and small-scale fading coefficients, respectively. $h_{mk} \sim \mathcal{CN}(0, 1)$ is assumed to remain constant in a coherent interval and change independently in the next. Further, h_{mk} is independent of other small-scale coefficients as well as of β_{mk}. The large-scale coefficients are modeled as

$$\beta_{mk} = \begin{cases} c_0, & ||\mathbf{p}_k - \mathbf{q}_m|| \leq r_0, \\ \frac{c_1 z_{mk}}{||\mathbf{p}_k - \mathbf{q}_m||^{\gamma}}, & ||\mathbf{p}_k - \mathbf{q}_m|| > r_0, \end{cases} \qquad (2)$$

where $\mathbf{p}_k$ and $\mathbf{q}_m$ represent the locations of user k and AP m, respectively. Here, γ is the pathloss exponent, z_{mk} is the log-normal shadow fading coefficient, and c_0, c_1, and r_0 are constants. Note that these coefficients can also be estimated by either ray-tracing [51] or data-driven [52] approaches.

We consider an uplink transmission model in which one user in each cell communicates with its AP (serving that same cell), such that there are only M users transmitting at a given time. All APs are in turn connected to a network controller (NC) via high capacity backhaul links and the NC is assumed to have knowledge of the positions of all APs and their respective users. Thus, it is in the NC where the

AP placement algorithms will be loaded and executed. In the small-cell set-up, each of the M cells corresponds to each of the M APs, and pursuant with the uplink model, users in their cells communicating with their serving APs cause interference to all other APs. Now, letting k_m denote a user in the cell associated with AP m, the received signal y_m at this AP is

$$y_m = \sum_{m'=1}^{M} \sqrt{\rho_r} g_{mk_{m'}} s_{k_{m'}} + w_m, \tag{3}$$

where ρ_r is the uplink transmit power, s_{k_m} is the data symbol with $\mathbb{E}\{|s_{k_m}|^2\} = 1$ (unit power), and $w_m \sim \mathcal{CN}(0, 1)$ is the additive noise. A matched filter (MF) employed at the AP m estimates the data symbol s_{k_m} of user k_m as

$$\hat{s}_{k_m} = \frac{g_{mk_m}^*}{|g_{mk_m}|} y_m,$$

$$= \sqrt{\rho_r} |g_{mk_m}| s_{k_m} + \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \sqrt{\rho_r} \frac{g_{mk_m}^*}{|g_{mk_m}|} g_{mk_{m'}} s_{k_{m'}} + v_m, \tag{4}$$

where $v_m \sim \mathcal{CN}(0, 1)$. From this expression, the SINR achieved by user k_m at AP m can be determined as

$$\phi_{k_m} = \frac{\rho_r \beta_{mk_m} |h_{mk_m}|^2}{1 + \rho_r \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \beta_{mk_{m'}} |h_{mk_{m'}}|^2}. \tag{5}$$

III. VECTOR QUANTIZATION FRAMEWORK

Before delving into the details of hybrid AP placement, we provide a brief overview of VQ as applied to AP placement and the basic Lloyd algorithm that can be used to solve for AP locations.

The VQ framework, described in [53], can be applied to the AP placement problem by assuming that the position $\mathbf{p}$ of a single user is the input random vector to be quantized, the Voronoi regions are the cells $\mathcal{C}_m$, and the AP locations $\mathbf{q}_m$ are the output codepoints, for $m = 1, 2, \ldots, M$. Accordingly, the optimization problem to be solved is

$$\underset{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M}{\arg \min} \quad \mathbb{E}_{\mathbf{p}} \{d(\mathbf{p}, \mathbf{q}_{\mathcal{E}(\mathbf{p})})\}, \tag{6}$$

where d is the distortion function that measures the quantization error and $\mathcal{E}$ refers to the encoder of the quantizer. It is to be noted that although d is defined for a user at $\mathbf{p}$, the objective function averages over the entire user distribution. Also, $\mathbf{q}_{\mathcal{E}(\mathbf{p})}$ is the position of the nearest AP, in terms of d, to the user at $\mathbf{p}$ and this is formally expressed as

$$\mathbf{q}_{\mathcal{E}(\mathbf{p})} = \underset{\mathbf{q}_m}{\arg \min} \, d(\mathbf{p}, \mathbf{q}_m). \tag{7}$$

Solution of (6) requires design of both encoder and decoder. It involves finding the best encoder given a fixed decoder using the *Nearest Neighbor Condition* (NNC)

$$\mathcal{C}_m = \{\mathbf{p} : d(\mathbf{p}, \mathbf{q}_m) \leq d(\mathbf{p}, \mathbf{q}_l), \forall l \neq m\}, \tag{8}$$

Algorithm 1 Lloyd Algorithm With Squared Error Distortion

1: **Initialize:**
$\quad \mathbf{q}_m^{(0)}, \forall m \in \mathcal{M}$
$\quad i \leftarrow 0$
2: **while** total distortion less than threshold **do**
3: $\quad$ update $\mathcal{C}_m^{(i+1)}, \forall m \in \mathcal{M}$ using $\qquad\qquad$ ▷ NNC

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

4: $\quad$ update $\mathbf{q}_m^{(i+1)}, \forall m \in \mathcal{M}$ using $\qquad\qquad$ ▷ CC

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

5: $\quad i \leftarrow i + 1$
6: **return** $\mathbf{q}_m, \forall m \in \mathcal{M}$

---

and finding the best decoder given a fixed encoder using the *Centroid Condition* (CC)

$$\mathbf{q}_m = \mathrm{Cent}\{\mathbf{p} | \mathbf{p} \in \mathcal{C}_m\}, \tag{9}$$

where Cent calculates the centroid of the user positions in cell $\mathcal{C}_m$. To solve, the Lloyd algorithm alternates between the NNC and CC steps until convergence to find the optimal AP locations. The most common distortion measure used is the squared Euclidean distance, denoted by

$$d_{\mathrm{SE}}(\mathbf{p}, \mathbf{q}_{\mathcal{E}(\mathbf{p})}) = \left\| \mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})} \right\|^2, \tag{10}$$

and the objective function in (6) then becomes the mean squared error (MSE). We utilize this distortion measure in the Lloyd algorithm and it is provided in Algorithm **1**.

For algorithm implementation, we use the K realization of users at positions $\mathbf{p}_k$, $k = 1, 2, \ldots, K$, as described in Section II. We will use this notation for all the algorithms in this manuscript. Also, in the CC step, as a result of d_{SE}, the centroid calculation to obtain $\mathbf{q}_m$ is replaced by the expectation which is evaluated by using the sample average over the user positions $\mathbf{p}_k$ present in cell $\mathcal{C}_m$. The overall complexity of the Lloyd algorithm is $\mathcal{O}(KMI_L)$, where $\mathcal{O}$ is the 'Big O' notation and I_L is the number of iterations taken for the algorithm to converge. Finally, the term 'Lloyd algorithm' will henceforth be used to denote the basic Lloyd algorithm above using the squared error distance as the distortion measure to solve for the AP locations. For algorithms that use other distortion measures, i.e., the Inter-AP Lloyd algorithm and the Hybrid AP Placement Algorithm (HAPPA) below, we will use the term 'Lloyd-type algorithm'.

Besides the parallels between VQ and the AP placement process discussed at the beginning of this section, we proved mathematically in [42] that the maximization of the average SNR can result in the same optimization problem of the Lloyd algorithm in (6) above, that uses d_{SE}. Further, we found that the Lloyd algorithm could be used as a baseline algorithm and hence, we use it to solve for more complex problems, e.g., accounting for ICI, by modifying its distortion measure.

In addition, it is also important to note here that the VQ framework presented above quantifies the distortion d for a single user at position $\mathbf{p}$ and the average over the distribution of user positions is taken. Finally, VQ does not involve the small-scale and shadow fading components. These quantities, according to the system model, are not dependent on either the user or AP positions and thus do not contribute to the placement process. Consequently, for the subsequent SINR-based and SGINR maximization problems that we will discuss, we will average over the abovementioned quantities.

IV. SGINR-BASED AP PLACEMENT FOR FULLY FLEXIBLE NETWORKS

We have seen in [42] that maximizing the average per-user rate that utilizes the SINR in (5) to determine the throughput-optimal AP locations is solved by using a Lloyd-type algorithm. For this purpose, the large-scale fading coefficients given in (2) for every user-AP pair (shown below for a user at $\mathbf{p}$ and its nearest AP indexed by $\mathcal{E}(\mathbf{p})$), is approximated as

$$\beta_{\mathcal{E}(\mathbf{p})} \approx \frac{c_1 z_{\mathcal{E}(\mathbf{p})}}{||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^\gamma}, \tag{11}$$

since r_0 is much smaller than the area dimensions considered. We also note here that for notational simplicity from Section II, the second subscript has been dropped and we will continue this for the ensuing analyses. Consequently, the SINR-based rate optimization problem for AP placement is

$$\arg\max_{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_M} \mathbb{E}_{\mathcal{A}, \underline{\mathbf{p}}} \left\{ \log\left(1 + \phi_{k_{\mathcal{E}(\mathbf{p})}}\right) \right\}, \tag{12}$$

where the notation $\underline{\mathbf{p}}$ denotes the set containing the served user at $\mathbf{p}$ which is in cell $\mathcal{C}_{\mathcal{E}(\mathbf{p})}$ and the $M - 1$ interfering users from cells $\mathcal{C}_{m'}$, denoted by $\mathbf{p}_{m'}$, where $m' \neq \mathcal{E}(\mathbf{p})$, and $\mathcal{A} = \{h_{\mathcal{E}(\mathbf{p})}, z_{\mathcal{E}(\mathbf{p})}, h_{m'}, z_{m'}\}$. For the user at $\mathbf{p}$, $h_{\mathcal{E}(\mathbf{p})}$ and $z_{\mathcal{E}(\mathbf{p})}$ are the small-scale fading coefficient and shadow fading, respectively, and $h_{m'}$ and $z_{m'}$ correspond to the same quantities, but for the interfering user at $\mathbf{p}_{m'}$. A solution to (12) was the Inter-AP Lloyd algorithm. It is worth noting that implementing power control along with AP placement, i.e., jointly optimizing uplink power with per-user power constraints and AP locations, would increment the rate performance. However, in this work, we wanted to solely investigate the effect of AP placement, with ICI that is prevalent in small-cell systems and that have been ignored by a majority of previous works. For this reason, we will continue to consider that all users transmit at the same power to the APs and leave the power optimization problem to future work.

Accounting for ICI is necessary in a small-cell scenario, and as an alternative to the SINR maximization in (12), we maximize the SGINR. In prior literature, the terms signal-to-leakage-plus-noise ratio (SLNR) and SGINR have been introduced. SLNR is considered in the downlink case [54] and quantifies the ICI by the leaked power by an AP to non-served users. SGINR is considered in the uplink [55] and quantifies the interference generated by the served user to all other APs. The concepts of both SLNR and SGINR are

the same and are fundamentally different to that of SINR. As we model the uplink scenario in our work, we will use the term SGINR. It is worth noting that for SGINR, the knowledge of the selected users from the interfering cells is not necessary, unlike SINR. Additionally, it has been shown that SGINR has computational advantages over SINR in both downlink [56] and uplink [57] beamforming.

Here, we consider the instantaneous SGINR of user k_m at AP m as

$$\psi_{k_m} = \frac{\rho_r \beta_{mk_m} |h_{mk_m}|^2}{1 + \rho_r \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \beta_{m'k_m} |h_{m'k_m}|^2}. \tag{14}$$

Note that while the numerator of the above expression represents the desired signal power, the denominator, unlike SINR, denotes the user-generated interference at other APs. Maximizing the SGINR provides a balance between maximizing user power versus generated ICI. Consequently, we have the following optimization problem

$$\arg\max_{\mathbf{q}_m \in \mathcal{M}} \mathbb{E}_{\mathcal{B}, \mathbf{p}} \left\{ \psi_{k_{\mathcal{E}(\mathbf{p})}} \right\}, \tag{15}$$

where $\mathcal{B} = \{h_{\mathcal{E}(\mathbf{p})}, z_{\mathcal{E}(\mathbf{p})}, h_{m'}, z_{m'}\}$ and $\mathbf{p}$ is the position of served user. Again, we use only a single subscript for simplicity, noting that while $h_{\mathcal{E}(\mathbf{p})}$ and $z_{\mathcal{E}(\mathbf{p})}$ are the small-scale fading coefficient and shadow fading, respectively, for the user at $\mathbf{p}$ to the serving cell, $h_{m'}$ and $z_{m'}$ correspond to the same quantities for the same user, but to the non-serving AP m'. Substituting the expression for SGINR in the objective function above and using Jensen's inequality, we have

$$\mathbb{E}_{\mathcal{B}, \mathbf{p}} \left\{ \psi_{k_{\mathcal{E}(\mathbf{p})}} \right\} \geq \mathbb{E}_{\mathcal{B}} \left\{ \frac{\rho_r c_1 |h_{\mathcal{E}(\mathbf{p})}|^2 z_{\mathcal{E}(\mathbf{p})}}{D} \right\}, \tag{16}$$

where the denominator D (which we now have to minimize) of the above expression is

$$D = \mathbb{E}_{\mathbf{p}} \left\{ ||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^\gamma \left(1 + \rho_r \sum_{\substack{m'=1 \\ m' \neq \mathcal{E}(\mathbf{p})}}^{M} \frac{c_1 |h_{m'}|^2 z_{m'}}{||\mathbf{p} - \mathbf{q}_{m'}||^\gamma} \right) \right\}. \tag{17}$$

The first term $||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^\gamma$ is the distance term of the pathloss and corresponds to the SNR while the other term is the ICI term. D can further be simplified using the independence between the two terms above to

$$D = \mathbb{E}_{\mathbf{p}} \left\{ ||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^\gamma \right\}$$
$$\times \left(1 + \rho_r \sum_{\substack{m'=1 \\ m' \neq \mathcal{E}(\mathbf{p})}}^{M} c_1 |h_{m'}|^2 z_{m'} \mathbb{E}_{\mathbf{p}} \left\{ \frac{1}{||\mathbf{p} - \mathbf{q}_{m'}||^\gamma} \right\} \right). \tag{18}$$

We now simplify the inner term using the following approximation

$$\mathbb{E}_{\mathbf{p}} \left\{ \frac{1}{||\mathbf{p} - \mathbf{q}_{m'}||^\gamma} \right\} \approx \frac{1}{||\mathbf{q}_{m'} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^\gamma}, \tag{19}$$

whose justification is provided in Appendix A, and results in

$$D \approx \mathbb{E}_{\mathbf{p}}\{||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma}\} \left(1 + \rho_r \sum_{\substack{m'=1 \\ m' \neq \mathcal{E}(\mathbf{p})}}^{M} \frac{c_1 |h_{m'}|^2 z_{m'}}{||\mathbf{q}_{m'} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma}} \right).$$

(20)

We consider two terms in the above expression of D. To minimize D, the first term $\mathbb{E}_{\mathbf{p}}\{||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma}\}$ corresponding to the SNR, which mirrors the objective function of the Lloyd algorithm, is to be minimized. Further, the second ICI term contains $1/||\mathbf{q}_{m'} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma}$, which are the inverses of the inter-AP distances $\forall m' \neq \mathcal{E}(\mathbf{p})$ and need also to be minimized. Since all the abovementioned quantities are positive and have all to be minimized, the distortion function, keeping in line with the VQ optimization in (6), can be written as

$$d_{\mathrm{IA}}\left(\mathbf{p}, \underline{\mathbf{q}}\right) = ||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma} + \sum_{m' \neq \mathcal{E}(\mathbf{p})} \frac{\kappa_{m'}}{||\mathbf{q}_{m'} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma}},$$

(21)

where $\kappa_{m'} \geq 0$, $m' \neq \mathcal{E}(\mathbf{p})$ are the *trade-off factors*. For simplicity, we assume a common $\kappa \geq 0$, and we have

$$d_{\mathrm{IA}}\left(\mathbf{p}, \underline{\mathbf{q}}\right) = ||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma} + \kappa \sum_{m' \neq \mathcal{E}(\mathbf{p})} \frac{1}{||\mathbf{q}_{m'} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^{\gamma}},$$

(22)

The trade-off factor κ is a network design parameter that can be selected to decide the relative importance of the ICI term to the SNR (desired signal) term. This is called the inter-AP distortion measure, denoted by d_{IA}, and coincides with the distortion measure derived from SINR in [42]. This reinforces the fact that the inter-AP distortion function (and correspondingly, the Inter-AP Lloyd algorithm) is helpful in tackling ICI in AP placement. The steps of the Inter-AP Lloyd algorithm were described in [42] for networks with fully flexible infrastructure and hence is omitted here.

V. AP PLACEMENT IN HYBRID UAV-TERRESTRIAL NETWORKS: PROBLEM FORMULATION AND SOLUTION

We now formulate the AP placement problem for hybrid networks. Consider the scenario where M_f T-APs (the fixed network) are optimally placed for the user configuration $f_{\mathbf{P}_1}(\mathbf{p}_1)$. Conceivably, after some time, the user configuration changes, say, to $f_{\mathbf{P}_2}(\mathbf{p}_2)$, and this results in the M_f T-APs being at sub-optimal positions. Thus, the added M_u UAV-APs to the system requires placement such that the performance gap between the scenario where all $M = M_f + M_u$ APs are optimally placed (the fully flexible network) and the scenario where only M_u out of M are movable (the hybrid network) is to be minimized. In other words, the hybrid network consisting of both T-APs and UAV-APs is compared to the ideal case of a fully flexible network that consists of UAV-APs alone. We aim for the performance of the hybrid network to be as close to that of the fully flexible network, which is

the benchmark solution. We quantify the performances using the average SGINRs of both networks.

Let $\psi_{k_{\mathcal{E}(\mathbf{p})}}^{\mathrm{flex}}$ be the SGINR achieved by optimal placement of all M APs in a fully flexible network for the new user distribution $f_{\mathbf{P}}(\mathbf{p})$ (subscript '2' removed for simplicity) and $\psi_{k_{\mathcal{E}(\mathbf{p})}}^{\mathrm{hbd}}$ be the SGINR achieved by using the hybrid UAV-terrestrial system where only M_u out of M APs can be optimally placed. Clearly, the average SGINR of the hybrid network is lower than the average SGINR of the fully flexible network. Hence, the average SGINR of the hybrid network is subtracted from that of the fully flexible network in order to define the performance gap. Formally, the objective function to be minimized (the performance gap) can be written as follows

$$J = \mathbb{E}_{\mathcal{B},\mathbf{p}}\left\{ \psi_{k_{\mathcal{E}(\mathbf{p})}}^{\mathrm{flex}} - \psi_{k_{\mathcal{E}(\mathbf{p})}}^{\mathrm{hbd}} \right\},$$

(23)

where the set $\mathcal{B}$ is as defined above for (15). Alternative to minimizing J, we maximize its negative $J' = -J$ such that

$$J' = \mathbb{E}_{\mathcal{B},\mathbf{p}}\left\{ \psi_{k_{\mathcal{E}(\mathbf{p})}}^{\mathrm{hbd}} \right\} - \mathbb{E}_{\mathcal{B},\mathbf{p}}\left\{ \psi_{k_{\mathcal{E}(\mathbf{p})}}^{\mathrm{flex}} \right\}.$$

(24)

It is assumed that the optimal AP locations for the fully flexible case is known. Note that the second term in (24) is thus independent of the optimal UAV-AP locations of the hybrid network and hence can be neglected. Consequently, maximizing the above lower bound of J' leads to the optimization problem

$$\underset{\mathbf{q}_m \in \mathcal{M}_u}{\arg\max} \quad \mathbb{E}_{\mathcal{B},\mathbf{p}}\left\{ \psi_{k_{\mathcal{E}(\mathbf{p})}}^{\mathrm{hbd}} \right\},$$

(25)

which is performed only over the M_u UAV-APs.

To perform the optimization in (25), we derive a Lloyd-type algorithm which maximizes the average per-user SGINR over the UAV-APs, which we call the Hybrid AP Placement Algorithm (HAPPA). Following the steps of the iterative Lloyd-type algorithm, we apply the nearest neighbor condition (NNC) to all M APs as

$$\mathcal{C}_m = \{\mathbf{p} : d_{\mathrm{IA}}(\mathbf{p}, \mathbf{q}_m) \leq d_{\mathrm{IA}}(\mathbf{p}, \mathbf{q}_l), \forall l \neq m\},$$

(26)

where $d_{\mathrm{IA}}(\cdot)$ is the inter-AP distortion function from (22). The centroid condition (CC) step now updates the M_u UAV-AP locations using the steepest descent method whose update equation is given as

$$\mathbf{q}_m^{(j+1)} = \mathbf{q}_m^{(j)} - \delta \frac{\partial}{\partial \mathbf{q}_m^{(j)}} \int_{\mathcal{C}_m} d_{\mathrm{IA}}\left(\mathbf{p}, \mathbf{q}_m^{(j)}\right) f_{\mathbf{P}}(\mathbf{p}) \mathrm{d}\mathbf{p}, \quad (27)$$

for all $m \in \mathcal{M}_u$, where j is the iteration index, δ is the step size, and the gradient expression is given by

$$\frac{\partial}{\partial \mathbf{q}_m} \int_{\mathcal{C}_m} d_{\mathrm{IA}}(\mathbf{p}, \mathbf{q}_m) f_{\mathbf{P}}(\mathbf{p}) \mathrm{d}\mathbf{p}$$
$$= \frac{\gamma}{|\mathcal{C}_m|} \sum_{\mathbf{p}_k \in \mathcal{C}_m} (\mathbf{q}_m - \mathbf{p}_k) ||\mathbf{p}_k - \mathbf{q}_m||^{\gamma - 2}$$
$$+ \kappa \sum_{m' \neq m} \frac{\mathbf{q}_{m'} - \mathbf{q}_m}{||\mathbf{q}_{m'} - \mathbf{q}_m||^{\gamma + 2}}, \qquad (28)$$

Algorithm 2 Hybrid AP Placement Algorithm (HAPPA)

1: **Initialize:**
$\qquad \mathbf{q}_m^{(0)}, \forall m \in \mathcal{M}$
$\qquad i \leftarrow 0$
2: **while** total distortion less than threshold **do**
3: $\quad$ update $\mathcal{C}_m^{(i+1)}, \forall m \in \mathcal{M}$ using $\qquad \triangleright$ NNC

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

4: $\quad$ **if** $m \in \mathcal{M}_f$ **then** $\qquad\qquad \triangleright$ CC for T-APs
5: $\qquad \mathbf{q}_m^{(i+1)} \leftarrow \mathbf{q}_m^{(i)}$
6: $\quad$ **else if** $m \in \mathcal{M}_u$ **then** $\qquad \triangleright$ CC for UAV-APs
7: $\qquad$ update $\mathbf{q}_m^{(i+1)}$ using

$$\mathbf{q}_m^{(j+1)} \leftarrow \mathbf{q}_m^{(j)}$$
$$- \delta \left(\frac{\gamma}{\left|\mathcal{C}_m^{(i+1)}\right|} \sum_{\mathbf{p}_k \in \mathcal{C}_m^{(i+1)}} \left(\mathbf{q}_m^{(j)} - \mathbf{p}_k\right) \left\|\mathbf{p}_k - \mathbf{q}_m^{(j)}\right\|^{\gamma-2} \right.$$
$$\left. + \kappa \sum_{m' \neq m} \frac{\mathbf{q}_{m'}^{(j)} - \mathbf{q}_m^{(j)}}{\left\|\mathbf{q}_{m'}^{(j)} - \mathbf{q}_m^{(j)}\right\|^{\gamma+2}} \right)$$

$\qquad$ until convergence
8: $\qquad \mathbf{q}_m^{(i+1)} \leftarrow \mathbf{q}_m^{(j+1)}$
9: $\quad i \leftarrow i + 1$
10: **return** $\mathbf{q}_m, \forall m \in \mathcal{M}$

The T-APs, however, remain at their original locations as they are fixed in position. The NNC and CC steps are iterated until the overall distortion is less than a threshold or cell assignments no longer change. HAPPA is summarized in Algorithm **2**. HAPPA incurs the same overall complexity as the Inter-AP Lloyd algorithm, i.e., $\mathcal{O}(KMI_SI_L)$, where I_S and I_L are the number of iterations taken for the steepest descent method and the Lloyd-type algorithm to converge, respectively. We do point out that inspite of the same complexity order, the number of operations involved in the CC step of HAPPA is less than that of the Inter-AP Lloyd algorithm and depends on the degree of hybridity, i.e., the proportion of UAV-APs to the total number of APs.

VI. SIMULATION METHODOLOGY AND RESULTS
A. PARAMETERS
The simulation parameters used are listed in Table 1. We consider a geographical area of 2 km × 2 km in which $K = 2000$ users and $M_f = 8$ T-APs are used, unless otherwise stated. For our analysis, by adding $M_u = 2, 4, 6,$ and 8 UAV-APs, we vary the total number of APs up to $M = 16$. For the user distribution, we consider the GMM of the form

$$f_{\mathbf{P}}(\mathbf{p}) = \sum_{l=1}^{L} p_l \mathcal{N}\left(\mathbf{p}|\boldsymbol{\mu}_l, \sigma_l^2 \mathbf{I}\right), \tag{29}$$

where $\mathbf{I}$ is the identity matrix and L is the number of mixture components, called *groups* henceforth. For group l, p_l

TABLE 1. List of simulation parameters.

Parameter	Value
Number of T-APs M_f	8
Number of UAV-APs M_u	$\{2, 4, 6, 8\}$
Number of users K	2000
GMM-1 $\{L, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3,$	$\{3, [0.5, -0.5]^T, [0, 0.5]^T, [-0.5, 0]^T\}$
$p_1, p_2, p_3, \sigma_1, \sigma_2, \sigma_3\}$	$\{0.6, 0.6, 0.2, 100, 100, 100\}$
GMM-2 $\{L, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3,$	$\{3, [0.52, -0.52]^T, [0, 0.5]^T, [-0.5, 0]^T\}$
$p_1, p_2, p_3, \sigma_1, \sigma_2, \sigma_3\}$	$\{0.6, 0.6, 0.2, 200, 100, 100\}$
GMM-3 $\{L, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3,$	$\{3, [0.27, -0.27]^T, [0, 0.25]^T, [-0.25, 0]^T\}$
$p_1, p_2, p_3, \sigma_1, \sigma_2, \sigma_3\}$	$\{0.6, 0.6, 0.2, 200, 100, 100\}$
Pathloss $\{\gamma, c_0, c_1, r_0\}$	$\{2, 75.86, 7.59 \times 10^{-7}, 0.001 \text{ km}\}$
Power ρ_r	200 mW
Bandwidth	20 MHz

is the mixture component weight, $\boldsymbol{\mu}_l$ is the mean, and σ_l is the variance. In line with our problem formulation, we consider two user configurations, namely, GMM-1 $f_{\mathbf{P}_1}(\mathbf{p}_1)$ and GMM-2 $f_{\mathbf{P}_2}(\mathbf{p}_2)$. For simplicity, we assume that the new user configuration GMM-2 has the same parameters as GMM-1, except that $\boldsymbol{\mu}_1 = [0.52, -0.52]^T$ and $\sigma_1 = 200$, i.e., GMM-2 causes users of group 1 to be more dispersed than in GMM-1. We also model a compact version of GMM-2, called GMM-3. GMM-3 has the same parameters as GMM-2, but with means $\boldsymbol{\mu}_1 = [0.27, -0.27]^T$, $\boldsymbol{\mu}_2 = [0, 0.25]^T$, and $\boldsymbol{\mu}_3 = [-0.25, 0]^T$. Note that the distinct parameters of GMM-2 and GMM-3 aim at modeling sparse and dense AP deployments, respectively. The pathloss model defined in (2) is used with the Cost 231 Hata propagation model described in [47, eqs. (4.36) and (4.37)] and shadow fading is ignored as it is averaged out in (25) during the placement process.

B. PERFORMANCE MEASURES
1) PER-USER ACHIEVABLE RATE
We use the SINR quantity ϕ_{k_m} in (5) to calculate the achievable per-user rate of user k_m as

$$R_{k_m} = \mathbb{E}\left\{\log_2\left(1 + \phi_{k_m}\right)\right\}. \tag{30}$$

Further, [47, Ch. 4] provides an alternative calculation

$$R_{k_m} = \frac{1}{\ln 2} e^{\mu_k} \text{Ei}(\mu_k), \tag{31}$$

where

$$\mu_k = \frac{1 + \rho_r \sum_{\substack{m'=1 \\ m' \neq m}}^{M} \beta_{mk_{m'}}}{\rho_r \beta_{mk_m}}, \tag{32}$$

and $\text{Ei}(x) = \int_x^\infty \frac{e^{-t}}{t} dt$ is the exponential integral.

2) SUM RATE
The sum of all M per-user achievable rates, where each user is selected from a cell, is given by

$$S = \sum_{m=1}^{M} R_{k_m}. \tag{33}$$

The maximum number of iterations for the Lloyd and the proposed algorithms is set to 50. We use Monte Carlo simulations to generate achievable and sum rates, considering 10,000 iterations, where we choose a set of users for transmission each time. For visualization, we use cumulative distribution function (CDF) plots, normalizing the rate values by its maximum quantity for relative comparison of the algorithms. Further, in order to quantify the performance, we utilize the 95%-likely metric that represents the best rate of the worst 5% of the users (users closer to cell borders). We denote this by $\mathcal{P}^{5\%}$, where $\mathcal{P}$ is R_{k_m} or S. The improvement ratio (IR) then quantifies the relative performance, expressed as percentage

$$\text{IR} = \frac{\mathcal{P}^{5\%,\text{Placement B}} - \mathcal{P}^{5\%,\text{Placement A}}}{\mathcal{P}^{5\%,\text{Placement A}}} \times 100. \quad (34)$$

C. INITIALIZATION OF THE ALGORITHMS

It is well known that random initializations result in convergence of the Lloyd or the proposed Lloyd-type algorithms presented in this paper to different local minima. Many researchers try to circumvent this problem by running multiple instances of the entire algorithm and then averaging the results. To have a common starting point and for unbiased comparison of all algorithms, we develop, specifically for the GMM user distribution, a novel initialization scheme based on the bit allocation problem in VQ. The number of APs allocated to each group l of the GMM user distribution is

$$u_l = \bar{u} + \log_2 \frac{h_l}{H} + \log_2 \frac{g_l}{G}, \quad (35)$$

where

$$\bar{u} = \frac{M}{L}, \qquad h_l = 4\sqrt{|\boldsymbol{\Sigma}_l|}, \qquad g_l = K_l,$$

$$H = \left(\prod_{l=1}^{L} h_l\right)^{\frac{1}{L}}, \qquad G = \left(\prod_{l=1}^{L} g_l\right)^{\frac{1}{L}}. \quad (36)$$

Here, L is the total number of groups, K_l is the number of users in group l, and $\boldsymbol{\Sigma}_l$ is the (sample) covariance of the users in group l. After allocating the M APs to the L groups, they are initialized randomly within each group. The proof of (35) in presented in Appendix B. It is to be noted that although the above initialization scheme is designed keeping the GMM in mind, it can be applied to any general distribution as the GMM presents a convenient way to express other distributions.

D. NUMERICAL RESULTS

Experiment 1: To simulate the Inter-AP Lloyd algorithm, we first generate the initial AP locations with the novel initialization procedure discussed above. For $M = 16$, the APs are allocated among the $L = 3$ groups as $u_1 = 6$ and $u_2 = u_3 = 5$. It can be ascertained from (35) that the mixture component weights are primarily responsible for this allocation since the group variances $\boldsymbol{\Sigma}_l$ are all equal. To showcase
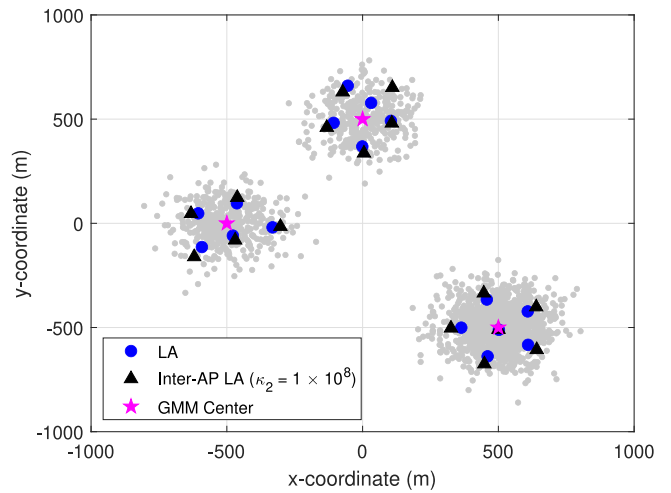

FIGURE 1. AP locations after convergence of the Lloyd and Inter-AP Lloyd algorithms for $\kappa_2 = 1 \times 10^8$ with $M = 16$ under GMM-1.
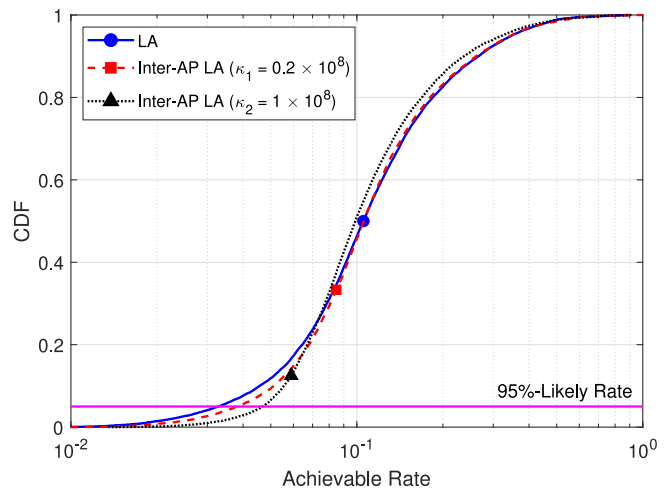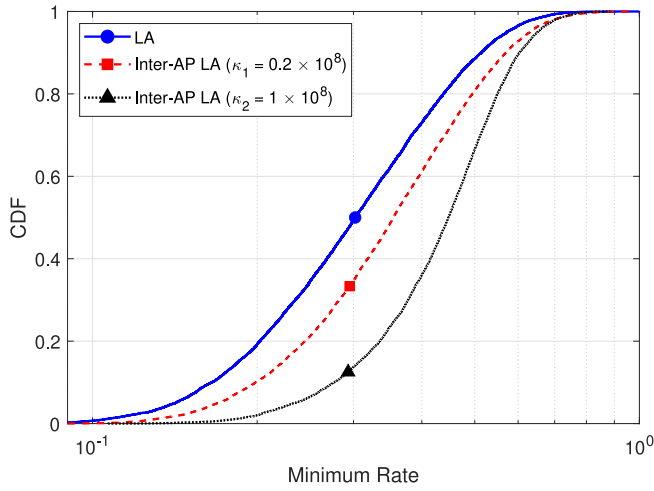

FIGURE 2. CDF plots of normalized per-user achievable rate for the Lloyd and Inter-AP Lloyd algorithms for $\kappa_1 = 0.2 \times 10^8$ and $\kappa_2 = 1 \times 10^8$ with $M = 16$ under GMM-1.

the Inter-AP Lloyd algorithm, we use two distinct trade-off factors $\kappa_1 = 0.2 \times 10^8$ and $\kappa_2 = 1 \times 10^8$. Fig. 1 shows a user realization as well as the final AP locations obtained using the Lloyd and Inter-AP Lloyd algorithms. Note that we only show the final placement for $\kappa_2 = 1 \times 10^8$ for simplicity. The arrangement of the AP locations of the Inter-AP Lloyd algorithm are spread apart more than of the Lloyd algorithm and increases with higher κ. This is primarily due to the influence of the inter-AP distances term in the distortion function of the former, resulting in the APs being pushed away so that ICI is reduced, especially for those users that are near the cell boundaries. Now, in Fig. 2 we plot the CDF curves of the per-user achievable rate for the two values of trade-off factor $\kappa_1 = 0.2 \times 10^8$ and $\kappa_2 = 1 \times 10^8$. For comparison, we focus on the rate value of the worst 5% of the users, the 95%-likely rate, which mainly corresponds to those users near the cell boundaries, affected mostly by ICI. It is clear

FIGURE 3. CDF plots of normalized minimum rate for the Lloyd and Inter-AP Lloyd algorithms for $\kappa_1 = 0.2 \times 10^8$ and $\kappa_2 = 1 \times 10^8$ with $M = 16$ under GMM-1.

FIGURE 4. AP locations obtained after HAPPA with $M_f = 8$ T-APs and $M_u = 8$ UAV-APs (hybrid network) and Inter-AP Lloyd algorithm with $M = 16$ APs (fully flexible network) under GMM-2.

from the figure that the rates of these users are improved using the proposed Inter-AP Lloyd algorithm. On the other hand, users that are closer to the GMM centers, experience a reduction in desirable signal strength as the APs are pushed away, thus suffering a slight loss in capacity. This effect could be intuitively understood from examining desirable-signal/interference-signal trade-off induced by κ from (22), with the degree of trade-off depending on κ. Using the rate quantities measured by the value at which the 95%-likely rate line intersects the CDF curves, we find that κ_1 results in a 16.07% and κ_2 in a 42.75% improvement, establishing that an increase in the value of κ results in an increase in the 95%-likely rate. In Fig. 3, we examine the minimum rates guaranteed by the network when using both algorithms. Minimum rate is defined as the smallest among all the rates achieved by the served users during a transmission time interval. It is clearly shown that the minimum rate of the network that employs the Inter-AP Lloyd algorithm is always superior to that when the Lloyd algorithm is used. Similar to the previous figure, the performance gap increases as κ is assigned higher values. Through multiple experiments, we have observed that a low κ value $< 0.1 \times 10^8$ results in a performance similar to the Lloyd algorithm, while a value $> 2 \times 10^8$ results in convergence issues. Hence, the choice of κ is an important criterion in using the Inter-AP Lloyd algorithm. We can think of κ as a network design parameter, which allows for controlling the emphasis placed on the signal and the interference powers.

Experiment 2: Now, for the proposed HAPPA, we evaluate the effects of AP placement on the throughput performance of the hybrid network as we add additional UAV-APs under the GMM-2 configuration. For all cases, we set step-size $\delta = 0.5$ and trade-off factor $\kappa = 1 \times 10^8$. As a benchmark result, we calculate the performance of the system which has $M_f = 8$ T-APs and no UAV-APs, the fixed network, which is labeled as $\langle M_f = 8, M_u = 0 \rangle$. Then, we
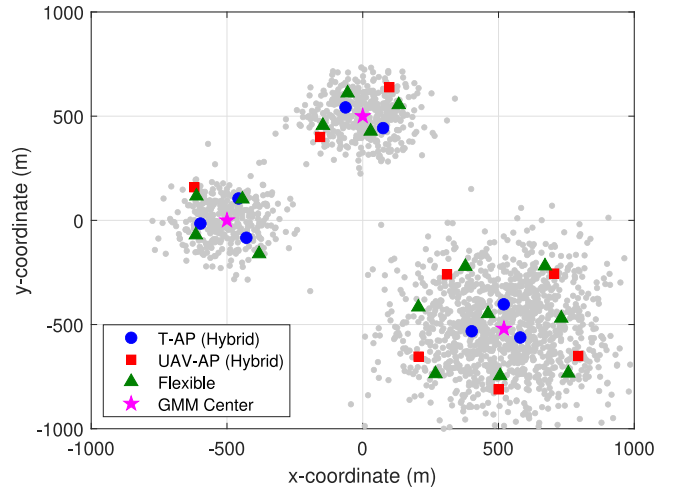
add 2, 4, 6, and 8 UAV-APs to the fixed network, resulting in up to $M = 16$ APs, the hybrid networks, labelled as $\langle M_f = 8, M_u = 2 \rangle$, $\langle M_f = 8, M_u = 4 \rangle$, $\langle M_f = 8, M_u = 6 \rangle$, and $\langle M_f = 8, M_u = 8 \rangle$, respectively. Finally, we also evaluate the performances of the system in the above networks when all M APs are UAV-APs, the fully flexible networks, labelled as $\langle M_f = 0, M_u = 10 \rangle$, $\langle M_f = 0, M_u = 12 \rangle$, $\langle M_f = 0, M_u = 14 \rangle$, and $\langle M_f = 0, M_u = 16 \rangle$, respectively. Note that the added UAV-APs are initialized using the proposed initialization algorithm. We start by showing in Fig. 4 the AP locations of fixed network $\langle M_f = 8, M_u = 0 \rangle$, the final UAV-AP locations obtained from HAPPA in the hybrid network $\langle M_f = 8, M_u = 8 \rangle$, and the final AP locations in the corresponding fully flexible network $\langle M_f = 0, M_u = 16 \rangle$ (where the Inter-AP Lloyd algorithm is run). Knowing that the T-APs were optimally placed for GMM-1, the UAV-APs position themselves according to the current GMM-2 configuration, in order to retrieve the capacity lost due to the new user density. Clearly, in the fully flexible network $\langle M_f = 0, M_u = 16 \rangle$, all APs are positioned optimally for GMM-2. For performance comparison, we calculate the sum rate in each network and the results are plotted in Fig. 5. The sum rate curve for the fixed network $\langle M_f = 8, M_u = 0 \rangle$ clearly performs the worst. As we add UAV-APs to the system, the rate curves move to the right showing an increase in system sum rate. Thus, we demonstrate that, in this system, the HAPPA operating on a hybrid network efficiently takes care of the ICI even when the *degree of hybridity* $M_u/M \leq 50\%$. This implies that even when there are more T-APs than UAV-APs, i.e., less flexibility to place APs, significant improvements in system capacity are obtained. The curve for the fully flexible network $\langle M_f = 0, M_u = 16 \rangle$ is also shown and only slightly outperforms the hybrid network $\langle M_f = 8, M_u = 8 \rangle$. In terms of the 95%-likely rate, we measure the percentage improvement of each hybrid network relative to the fixed network and the values are tabulated in column 3 of Table 2. Furthermore,
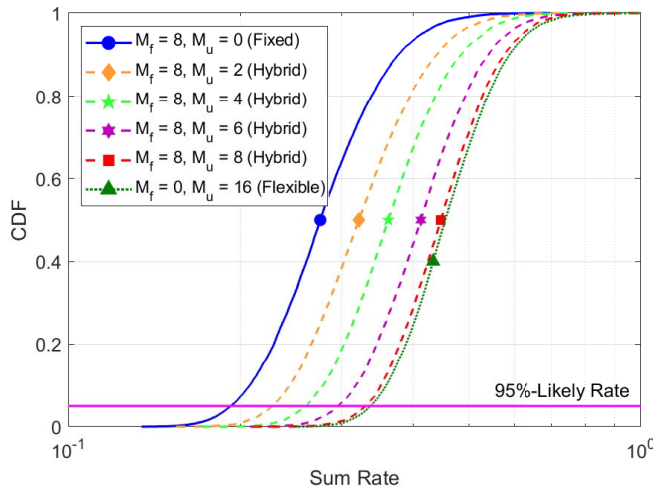
FIGURE 5. CDF plots of normalized sum rate for fixed $\langle M_f = 8, M_u = 0\rangle$, hybrid $\langle M_f = 8, M_u = 2, 4, 6, 8\rangle$, and fully flexible $\langle M_f = 0, M_u = 16\rangle$ networks under GMM-2.

TABLE 2. Hybrid network configuration, degree of hybridity, and 95%-likely sum rate improvement ratios under GMM-2.

Hybrid Network	$\frac{M_u}{M}$ [%]	$\frac{\text{Hybrid}}{\text{Fixed}}$ [%]	$\frac{\text{Flexible}}{\text{Hybrid}}$ [%]
$\langle M_f = 8, M_u = 2\rangle$	20	18.10	6.65
$\langle M_f = 8, M_u = 4\rangle$	33.33	35.61	4.84
$\langle M_f = 8, M_u = 6\rangle$	42.86	53.67	4.13
$\langle M_f = 8, M_u = 8\rangle$	50	71.92	2.02

TABLE 3. Hybrid network configuration, degree of hybridity, and 95%-likely sum rate improvement ratios under GMM-3.

Hybrid Network	$\frac{M_u}{M}$ [%]	$\frac{\text{Hybrid}}{\text{Fixed}}$ [%]	$\frac{\text{Flexible}}{\text{Hybrid}}$ [%]
$\langle M_f = 8, M_u = 2\rangle$	20	23.42	7.99
$\langle M_f = 8, M_u = 4\rangle$	33.33	44.28	6.57
$\langle M_f = 8, M_u = 6\rangle$	42.86	70.44	6.20
$\langle M_f = 8, M_u = 8\rangle$	50	93.63	5.65

we compare the performance of each hybrid network to its corresponding fully flexible network for the various degrees of hybridity (e.g., comparison of $\langle M_f = 8, M_u = 2\rangle$ with $\langle M_f = 0, M_u = 10\rangle$), as tabulated in column 4 of Table 2. The performance gap between the hybrid and fully flexible networks is only 6.65% when the degree of hybridity is 20% and culminates in a minimal 2.02% gap when the hybridity is 50%. Note that in order to determine the minimum and sufficient number of UAV-APs required for the system to achieve a threshold sum rate η_{th}, we perform

$$U_{\min} = \arg\min_{M_u} S^{5\%}(M_f, M_u) \text{ s.t. } S^{5\%}(M_f, M_u) \geq \eta_{\text{th}}, \quad (37)$$

where $S^{5\%}(M_f, M_u)$ is the 95%-likely sum rate when M_u UAV-APs are added to M_f T-APs. The minimum number of UAV-APs can be obtained by sequential or bisectional search.

Experiment 3: In this experiment, we conduct a similar sum rate performance analysis of HAPPA as in experiment 2,
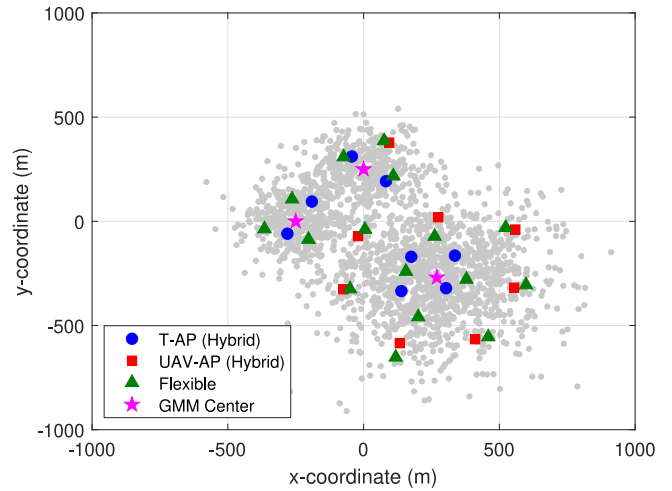

FIGURE 6. AP locations obtained after HAPPA with $M_f = 8$ T-APs and $M_u = 8$ UAV-APs (hybrid network) and Inter-AP Lloyd algorithm with $M = 16$ APs (fully flexible network) under GMM-3.
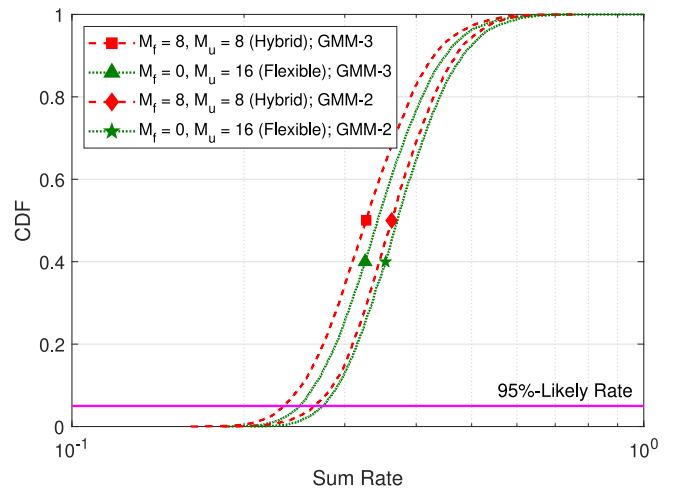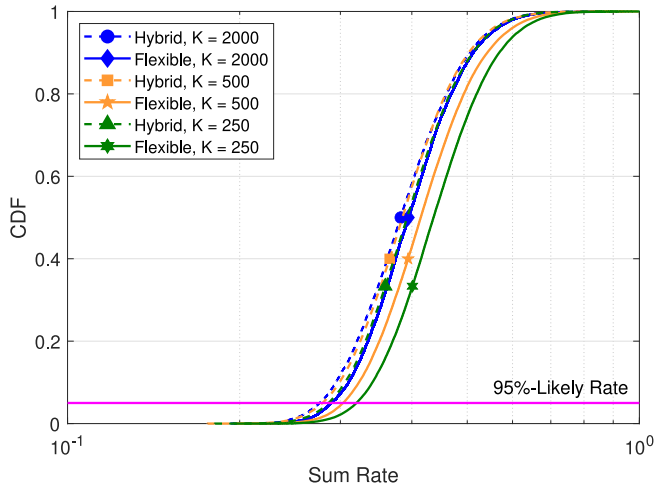

FIGURE 7. CDF plots of normalized sum rate for hybrid $\langle M_f = 8, M_u = 8\rangle$ and fully flexible $\langle M_f = 0, M_u = 16\rangle$ networks under GMM-3 and GMM-2.

however, under the compact GMM-3 user configuration. Fig. 6 shows the AP locations of the fixed network $\langle M_f = 8, M_u = 0\rangle$ and the resulting UAV-AP locations for the hybrid network $\langle M_f = 8, M_u = 8\rangle$ and the fully flexible network $\langle M_f = 0, M_u = 16\rangle$. For the hybrid network, the UAV-APs position themselves around the existing T-APs, which are suboptimally placed, to increase system capacity and the fully flexible network has the UAV-APs positioned optimally for GMM-3. In this scenario, we compare the sum rates achieved by the fixed network, the fully flexible network, and the hybrid networks with varying degrees of hybridity $\langle M_f = 8, M_u = 2\rangle$, $\langle M_f = 8, M_u = 4\rangle$, $\langle M_f = 8, M_u = 6\rangle$, and $\langle M_f = 8, M_u = 8\rangle$. For the sake of visual clarity, in Fig. 7, we focus on the comparison between the sum rate curves of the 50% hybrid network $\langle M_f = 8, M_u = 8\rangle$ and of the fully flexible network $\langle M_f = 0, M_u = 16\rangle$, since in the same figure, we also plot the corresponding sum rate
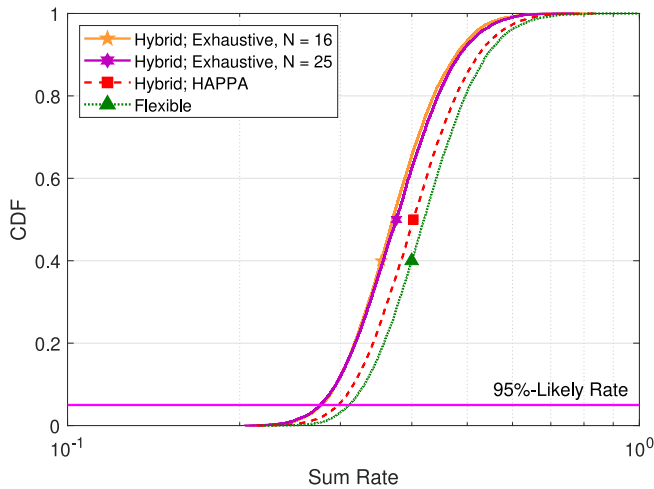
FIGURE 8. CDF plots of normalized sum rate for hybrid $\langle M_f = 8, M_u = 8 \rangle$ and fully flexible $\langle M_f = 0, M_u = 16 \rangle$ networks for K = 2000, 500, and 250 under GMM-3.

curves for the 50% hybrid and fully flexible networks we obtained for the GMM-2, in experiment 2. The comparison of the 95%-likely rate values for all the considered degrees of hybridity are tabulated in Table 3. As expected, the fixed network performs the worst and performance substantially improves as the degree of hybridity increases. Notice that the fully flexible network only performs slightly better than the hybrid network even for a small degree of hybridity of 20%, which results in a small gap of nearly 8%. The gap shrinks to 5.65% for a 50% hybrid network. In Fig. 7, it is worth mentioning that in comparison to the rates yielded by GMM-2, the ones of GMM-3 are lower due to the increased ICI arising from the cells being relatively closer to one another. Quantitatively, relative to the corresponding curves yielded by the GMM-2, there is a 11.57% and 9.14% reduction in the 95%-likely values for the 50% hybrid and for the fully flexible networks, respectively, due to the GMM-3. Note that the effects of increased ICI are not only observed when comparing distinct GMMs but also within one GMM. This is observed in column 4 of Table 3, where the percentage improvements of a fully flexible network is determined over its hybrid counterpart for GMM-3. In summary, the proposed HAPPA is demonstrated to significantly improve sum rates in both cases when the GMM groups are far apart (lower overall ICI) and very close together (higher overall ICI). Thus, we can reasonably conclude that for user configurations in between the two extremes shown above, HAPPA will still be able to place APs optimally in order to recover the capacity lost as a result of changes in the user configuration, in the presence of ICI.

Experiment 4: In this experiment, we investigate the effects that the number of users in GMM-3 has on the performance of HAPPA. In this regard, we plot in Fig. 8, the sum rate curves achieved by a $\langle M_f = 8, M_u = 8 \rangle$ hybrid network and the corresponding $\langle M_f = 0, M_u = 16 \rangle$ fully flexible network when K = 2000, K = 500, and K = 250.

Two observations can be made from this figure. First, the sum rate performances of both the hybrid and fully flexible networks increase as the number of users is reduced. Second, the performance gap between hybrid and the fully flexible networks increases as the number of users is decreased. This follows from the fact that as the total number of users K reduces, the proportion of users that are far from the serving AP in each cell also reduces, thereby shifting the overall rate curve to the right. This also increases the importance of flexibility, therefore favoring the deployment that has more UAV-APs. Quantitatively, we calculate the performance gaps in terms of the 95%-likely rate. Among the hybrid networks, we observe a 1.20% and 3.74% gap for the curves when K = 500 and K = 250, respectively, over the curve when K = 2000. The values for the fully flexible networks are 4.82% and 10.58%, respectively. Further, the performance increase of the fully flexible networks over their hybrid counterparts starts at 5.65% when K = 2000, increases to 8.65% when K = 500, and to a higher 11.81% when K = 250.

Experiment 5. In this experiment, we compare the proposed HAPPA with the conventional exhaustive search. In particular, we perform the comparison of HAPPA with an exhaustive search algorithm to find the UAV-AP positions for the 50% hybrid network $\langle M_f = 8, M_u = 8 \rangle$ with GMM-2. The exhaustive search process is described as follows. We first split the geographical area under consideration into N grid points. Given the 8 T-APs, we have to determine the positions of the 8 UAV-APs at the grid points that generates the best 95%-likely sum rate performance. Since it is not beneficial for two APs to be colocated at the same grid point, the total number of combinations to distinctly place the 8 UAV-APs among N grid points is the binomial coefficient $^N C_8$. Note that N is chosen such that both HAPPA and the exhaustive search have similar implementation complexities for a fair performance comparison. The overall complexity order of HAPPA is $\mathcal{O}(KMI_SI_L)$ where $I_S = 5$ and $I_L = 15$ are average values obtained from conducting a large number of trials. As such, the complexity order for K = 2000 users and total M = 16 APs is 2.4×10^6. On the other hand, the exhaustive search algorithm has an overall complexity order of $KM \times^N C_8$, since for each combination, it is necessary to associate each of the K users to the M APs. Thus, N = 16 APs yields a complexity order of 4.12×10^6 for the exhaustive search, which is similar to that of HAPPA. With similar complexities ensuring a fair performance comparison, we then conduct the exhaustive search and generate the sum rate curve for the best combination, which is shown in Fig. 9 below. Clearly, it is observed that the performance of the exhaustive search with N = 16 grid points is inferior (the 95%-likely value is 4.22% lower) to HAPPA. To further investigate, we increased the number of grid points to N = 25, resulting in a complexity order of $KM \times^{25} C_8 = 3.46 \times 10^{10}$, which is clearly several orders of magnitude higher than that of HAPPA. The sum rate performance corresponding to N = 25 shown in Fig. 9

FIGURE 9. CDF plots of normalized sum rate for hybrid network $\langle M_f = 8, M_u = 8 \rangle$ from exhaustive search, hybrid network $\langle M_f = 8, M_u = 8 \rangle$ from HAPPA, and fully flexible network $\langle M_f = 0, M_u = 16 \rangle$ under GMM-2.
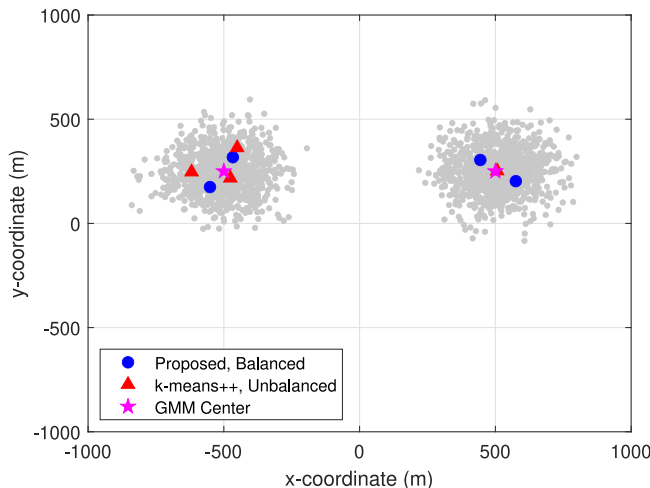
FIGURE 10. Representative example showing the final AP locations for the proposed and k-means++ initialization schemes for $M = 4$ under GMM-3. The proposed scheme always results in a balanced allocation while the k-means++ scheme may result in a balanced (not shown) or an unbalanced allocation, whose likelihood is given in Table 4.

is also significantly lower (the 95%-likely value is 3.88% lower) than that of HAPPA and is only slightly better than that of $N = 16$. This shows that in spite of similar or even with significantly lower complexity, our proposed Lloyd-type algorithm, i.e., HAPPA, results in a placement of UAV-APs that yields a better sum rate performance in comparison to the exhaustive search method.

Experiment 6. In this experiment, we compare the proposed initialization scheme with the popular k-means++ initialization [58]. For this comparison, we consider a user distribution GMM-4, which has parameters $L = 2$, $\boldsymbol{\mu}_1 = [-0.5, 0]^T$, $\boldsymbol{\mu}_2 = [0.5, 0]^T$, $\sigma_1 = \sigma_2 = 100$, and $p_1 = p_2 = 0.5$, and is shown in Fig. 10. In order to show how our proposed initialization method described in Section VI-C

is advantageous over the k-means++ initialization scheme, we present the following three points of comparison.

Firstly, we examine the distribution of the AP locations between the GMM groups such that the number of users being served by each AP is considered. Since GMM-4 has equal mixture component probability (and hence, an equal number of users) and variances for both groups, the proposed scheme allocates an equal number of APs to both groups, i.e., if $M = 4$, then each group is allocated 2 APs. This balanced allocation, which is henceforth represented as $(2, 2)$, is true for any realization of the user distribution. In contrast, the k-means++ scheme does not always result in a balanced $(2, 2)$ allocation. Through multiple realizations, it is observed to result in unbalanced $(3, 1)$ and $(1, 3)$ allocations as well. Examples of the allocations are illustrated in Fig. 10, where we show the final AP locations obtained using the Inter-AP Lloyd algorithm, resulting in a balanced $(2, 2)$ allocation for the proposed scheme and an unbalanced $(3, 1)$ allocation for the k-means++ scheme. In a balanced allocation, the division of the $K = 2000$ users is nearly even among the $M = 4$ APs, resulting in each AP serving about 500 users. This is in contrast to an unbalanced $(1, 3)$ or $(3, 1)$ allocation where one AP alone would serve nearly 1000 users while the other 3 APs would serve about 333 users each. In Table 4, we present the results of 1000 realizations of initializations and the percentage of balanced and unbalanced allocations are noted for $M = 4$, 6, and 8. Note that for $M = 6$, the possible unbalanced allocations would be $(4, 2)$, $(2, 4)$, $(1, 5)$, and $(5, 1)$, while for $M = 8$, they would be $(5, 3)$, $(3, 5)$, $(6, 2)$, $(2, 6)$, $(7, 1)$, and $(1, 7)$. Table 4 shows us that while the proposed scheme always results in a balanced allocation, the k-means++ scheme results in unbalanced allocations 30.1% of the time when $M = 4$ and increases with M up to 44.5% when $M = 8$.

Secondly, to observe the effects of balanced and unbalanced AP allocations on the rate performance, we plot one curve for the proposed scheme and one curve each for the balanced and unbalanced allocations arising from the k-means++ scheme. We use the same user distribution as before with $M = 4$ APs and $K = 250$ users, and in Fig. 11, we show the CDF of the minimum rates achieved in each case. It is clear that regardless of the scheme, the balanced allocation results in the same performance. The unbalanced allocation yielded by the k-means++ scheme, however, performs considerably worse. This performance loss is due to the fact that one AP in a GMM group has to serve all the users within the group. As such, this AP has a higher average user-AP distance than the APs in the balanced allocation case, and contributes primarily to the lower achievable rates. Henceforth, we can conclude that the proposed initialization scheme is preferable to the k-means++ scheme since the latter has an incidence of unbalanced AP allocations, which in turn results in reduced achievable rates.

Finally, we demonstrate that the benefits of the proposed scheme is not only limited to rate improvements, but also reduced net distance traveled by the APs over time as the user

TABLE 4. Percentage of unbalanced allocations under the proposed and k-means++ initialization schemes.

M	Proposed [%]	k-means++ [%]
4	0	30.1
6	0	39.6
8	0	44.5
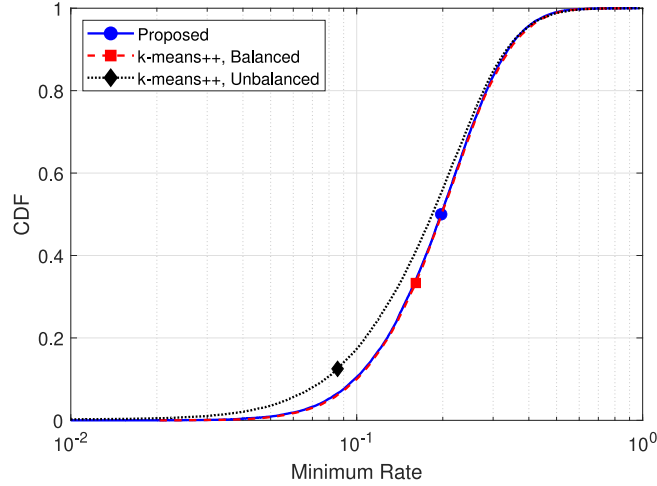
FIGURE 11. CDF plots of normalized minimum rate for the proposed and k-means++ balanced and unbalanced allocations with $M = 4$ and $K = 250$ under GMM-4.

realization changes and the APs need to update their positions. We again use GMM-4 with $K = 250$ and $M = 4$ and consider two user realizations, at time instances T_1 and T_2. Both initialization schemes are used at T_1 and the Inter-AP Lloyd algorithm is applied to place the APs. Then, using the final AP positions at T_1 as initial positions for T_2, the placement algorithm is again applied. Accordingly, we conduct 1000 trials to compare the distances traversed by the APs from T_1 to T_2 in both schemes. It is found that the average distance traveled by the APs following the proposed scheme is 168.16 m, which is smaller compared to the averaged traveled distance of 175.08 m, yielded by the k-means++ scheme. The sole contributors to the higher traversed distances in the k-means++ scheme are the unbalanced allocations. Note that with smaller traversed distances, practical implementation issues such as the potential for collisions, UAV battery drainage, cell handovers, as well as signalling between NC and UAV-APs are reduced. Moreover, the benefits of the proposed scheme are also observed with the traditional Lloyd algorithm and HAPPA.

VII. CONCLUSION

In this paper, we have addressed the AP placement problem in the small-cell uplink paradigm for hybrid network composed of terrestrial APs (T-APs) and AP-enabled UAVs (UAV-APs) in the presence of inter-cell interference (ICI). We accounted for ICI through signal-to-generated-interference-plus-noise ratio (SGINR) instead of signal-to-interference-plus-noise ratio (SINR) and arrived at the Inter-AP Lloyd algorithm for fully flexible networks.

It generated up to a 42.75% increase in 95%-likely achievable rate over the Lloyd algorithm. To account for the loss in capacity due to change in the user density, we devised a UAV-AP placement algorithm called the Hybrid AP Placement Algorithm (HAPPA). We showed that HAPPA, even with a small proportion of UAV-APs to T-APs, exhibited significant sum rate increase (up to 93.63%) over fixed networks and close to the ideal performance (as little as 2.02% difference) of fully flexible networks. Finally, we derived an initialization method for the Lloyd or any Lloyd-type algorithm, applicable to the Gaussian mixture model of user distribution. When compared to the popular k-means++ method, the proposed method always offered a relatively balanced (similar number of users served by each AP) allocation, in contrast to the unbalanced allocation that was possible in the k-means++ method at least 30% of the time, which resulted in a lower minimum rate over the balanced allocation.

APPENDIX A
PROOF OF (19)
Consider the term

$$||\mathbf{p} - \mathbf{q}_{m'}||^2 = ||\underbrace{\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}}_{\mathbf{y}} + \underbrace{\mathbf{q}_{\mathcal{E}(\mathbf{p})} - \mathbf{q}_{m'}}_{\mathbf{x}}||^2. \quad (38)$$

From the conventional cell planning rules, it is clear that the distance between a user and its serving AP, denoted by $\mathbf{y}$, is smaller than the distance between the serving AP and any other interfering AP, denoted by $\mathbf{x}$. Hence, we can write

$$||\mathbf{p} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}|| \leq ||\mathbf{q}_{\mathcal{E}(\mathbf{p})} - \mathbf{q}_{m'}|| \Rightarrow ||\mathbf{y}|| \leq ||\mathbf{x}||. \quad (39)$$

Each interfering AP m' can be classified into whether it is an immediate neighbor $\mathcal{IN}$ of AP $\mathcal{E}(\mathbf{p})$ or not, and correspondingly obtain the following relations

$$||\mathbf{x}|| \geq ||\mathbf{y}||, \forall m' \in \mathcal{IN}(\mathcal{E}(\mathbf{p})), \quad m' \neq \mathcal{E}(\mathbf{p}),$$
$$||\mathbf{x}|| \gg ||\mathbf{y}||, \forall m' \notin \mathcal{IN}(\mathcal{E}(\mathbf{p})), \quad m' \neq \mathcal{E}(\mathbf{p}). \quad (40)$$

At this stage, we make the assumption that $||\mathbf{x}|| \gg ||\mathbf{y}||$ holds true $\forall m' \neq \mathcal{E}(\mathbf{p})$ and we simplify the term in (38) as

$$||\mathbf{x} + \mathbf{y}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2||\mathbf{x}||||\mathbf{y}|| \cos\theta,$$
$$= ||\mathbf{x}||^2 \left(1 + \frac{||\mathbf{y}||^2}{||\mathbf{x}||^2} + \frac{||\mathbf{y}|| \cos\theta}{||\mathbf{x}||} \right),$$
$$\approx ||\mathbf{x}||^2. \quad (41)$$

This relation also holds true for $||\mathbf{p} - \mathbf{q}_{m'}||^\gamma$ with values of γ other than $\gamma = 2$. Hence, rearranging and taking the expectation gives us

$$\mathbb{E}_{\mathbf{p}} \left\{ \frac{1}{||\mathbf{p} - \mathbf{q}_{m'}||^\gamma} \right\} \approx \frac{1}{||\mathbf{q}_{m'} - \mathbf{q}_{\mathcal{E}(\mathbf{p})}||^\gamma}. \quad (42)$$

APPENDIX B
PROOF OF (35)
The bit allocation problem and the associated solution can be found in [53, Ch. 8] for the scalar (random variable) case.

We extend this procedure to random vectors of dimension k and then derive the specific case for $k = 2$.

If we have N random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ that are to be quantized, then the overall distortion can be defined as

$$D = \sum_{i=1}^{N} g_i W_i(b_i), \qquad (43)$$

where $W_i(b_i)$ is the distortion incurred when $\mathbf{X}_i$ is quantized using b_i bits and g_i's are some non-negative weights. From [53], $W_i(b_i)$ for k-dimensional vectors when the high-resolution approximation is applied is written as

$$W_i(b_i) = \frac{k}{k+2}\left(\frac{2\pi^{\frac{k}{2}}}{k\Gamma\left(\frac{k}{2}\right)}\right)^{-\frac{2}{k}}$$
$$\times \left\{\int_{-\infty}^{\infty}\left[f_{\mathbf{X}_i}(\mathbf{x})\right]^{\frac{k}{k+2}}d\mathbf{x}\right\}^{\frac{k+2}{k}} 2^{-\frac{2b_i}{k}}. \qquad (44)$$

where $\Gamma(\cdot)$ is the gamma function. We now use this expression to derive the bit allocation result for vectors of dimension k.

Lemma 1: Consider N k-dimensional random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ of zero mean and covariance $\boldsymbol{\Sigma}_i$, and whose distributions $f_{\mathbf{X}_i}(\mathbf{x})$ are known for $i = 1, 2, \ldots, N$. Let $b_i = \log_2 q_i$ be the number of bits required in order to design a quantizer for $\mathbf{X}_i$ with q_i quantization levels. With B being the total number of bits available for the N quantizers (the bit budget), and the distortion defined as in (44), the number of bits to be allocated to achieve optimal quantization, called the optimal bit allocation, is given by

$$b_i = \bar{b} + \frac{k}{2}\log_2\frac{h_i}{H} + \frac{k}{2}\log_2\frac{g_i}{G}, \qquad (45)$$

where:

$$\bar{b} = \frac{B}{N}, \qquad \epsilon(k) = \frac{k}{k+2}\left(\frac{2\pi^{\frac{k}{2}}}{k\Gamma\left(\frac{k}{2}\right)}\right)^{-\frac{2}{k}},$$

$$h_i = \epsilon(k)\left\{\int_{-\infty}^{\infty}\left[f_{\mathbf{X}_i}(\mathbf{x})\right]^{\frac{k}{k+2}}d\mathbf{x}\right\}^{\frac{k+2}{k}} = \epsilon(k)\left\|f_{\mathbf{X}_i}(\mathbf{x})\right\|_{\frac{k}{k+2}},$$

$$H = \left(\prod_{i=1}^{N} h_i\right)^{\frac{1}{N}}, \qquad G = \left(\prod_{i=1}^{N} g_i\right)^{\frac{1}{N}}. \qquad (46)$$

Proof: Using the above notation in (44), we have

$$W_i(b_i) = h_i 2^{-\frac{2b_i}{k}}. \qquad (47)$$

To minimize the overall distortion in (43) constrained by the bit budget, the objective function is

$$J = \sum_{i=1}^{N} g_i h_i 2^{-\frac{2b_i}{k}} + \lambda\left(\sum_{i=1}^{N} b_i - B\right), \qquad (48)$$

which, on differentiating, we get

$$\frac{\partial J}{\partial b_i} = 0 \Rightarrow \lambda = \frac{2g_i h_i(\ln 2)}{k} 2^{-\frac{2b_i}{k}}$$

$$\frac{\partial J}{\partial \lambda} = 0 \Rightarrow \sum_{i=1}^{N} b_i = B. \qquad (49)$$

Solving this set of equations for b_i yields the required result. $\blacksquare$

Noting that h_i is the norm of the pdf, we determine the closed-form expression when $\mathbf{X}_i$ is a multivariate Gaussian, as is the case in the GMM distribution.

Lemma 2: For the multivariate k-dimensional Gaussian

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}\exp\left(\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), (50)$$

the η-norm defined by

$$\|f_{\mathbf{X}}(\mathbf{x})\|_\eta = \left\{\int_{-\infty}^{\infty}\left[f_{\mathbf{X}}(\mathbf{x})\right]^\eta d\mathbf{x}\right\}^{\frac{1}{\eta}}, \qquad (51)$$

is given as

$$\|f_{\mathbf{X}}(\mathbf{x})\|_\eta = \left(\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}\right)^{\left(\frac{1}{\eta}-1\right)}\left(\frac{1}{\eta^{\frac{k}{2}}}\right)^{\frac{1}{\eta}}. \qquad (52)$$

Proof:

$$\|f_{\mathbf{X}}(\mathbf{x})\|_\eta$$
$$= \frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}\left\{\int_{-\infty}^{\infty}\exp\left(\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\eta\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)d\mathbf{x}\right\}^{\frac{1}{\eta}}$$
$$= \frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}\left\{\sqrt{(2\pi)^k\left|\frac{1}{\eta}\boldsymbol{\Sigma}\right|}\right\}^{\frac{1}{\eta}}$$
$$= \left(\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}\right)^{\left(\frac{1}{\eta}-1\right)}\left(\frac{1}{\eta^{\frac{k}{2}}}\right)^{\frac{1}{\eta}}. \qquad (53)$$

$\blacksquare$

Theorem 1: With M APs (the AP budget), the number of APs allocated to the L groups of a GMM with parameters p_l, $\boldsymbol{\mu}_l$, and $\boldsymbol{\Sigma}_l$, and K_l being the number of users in group l, $l = 1, 2, \ldots, L$, is

$$u_l = \bar{u} + \log_2\frac{h_l}{H} + \log_2\frac{g_l}{G}, \qquad (54)$$

where

$$\bar{u} = \frac{M}{L} \qquad h_l = 4\sqrt{|\boldsymbol{\Sigma}_l|} \qquad g_l = K_l$$

$$H = \left(\prod_{l=1}^{L} h_l\right)^{\frac{1}{L}} \qquad G = \left(\prod_{l=1}^{L} g_l\right)^{\frac{1}{L}}. \qquad (55)$$

Proof: Since the users are distributed in the $\mathbb{R}^2$ plane, we have $k = 2$. The AP budget is M, number of groups is L, and the distortion weights g_l correspond to the mixture component weights p_l, which in turn correspond to the number of users K_l. Since each GMM group l has $f_{\mathbf{X}_l}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$, from Lemma 2 we obtain

$$h_l = \frac{1}{2\pi}\left\|f_{\mathbf{X}_l}(\mathbf{x})\right\|_{\frac{1}{2}} = 4\sqrt{|\boldsymbol{\Sigma}_l|}. \qquad (56)$$

Making these parallels and simplifications in (45) of Lemma 1, we obtain the desired expression. $\blacksquare$

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] 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.

[3] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[4] 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.

[5] G. P. Villardi, K. Ishizu, and F. Kojima, "Reducing the codeword search complexity of FDD moderately large MIMO beamforming systems," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 273–287, Jan. 2019.

[6] 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.

[7] 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.

[8] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming with limited intercell coordination," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 728–738, Feb. 2011.

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

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

[12] R. Rogalin *et al.*, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1815–1831, Apr. 2014.

[13] 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.

[14] 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.

[15] 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. 50th Asilomar Conf. Signals Syst. Comput.*, Nov. 2016, pp. 203–207.

[16] 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, Jul. 2017.

[17] D. Gesbert, S. Hanly, H. Huang, S. S. 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.

[18] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[19] S. T. Abraha, D. F. Castellana, X. Liang, A. Ngoma, 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, pp. 1–3.

[20] S. Karimi-Bidhendi, J. Guo, and H. Jafarkhani, "Using quantization to deploy heterogeneous nodes in two-tier wireless sensor networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1502–1506.

[21] A. Garcia-Rodriguez, G. Geraci, D. Lopez-Perez, L. G. Giordano, M. Ding, and E. Bjornson, "The essential guide to realizing 5G-connected UAVs with massive MIMO," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 84–90, Dec. 2019.

[22] U. Challita, W. Saad, and C. Bettstetter, "Interference management for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2125–2140, Apr. 2019.

[23] W. Mei and R. Zhang, "Aerial-ground interference mitigation for cellular-connected UAV," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 167–173, Feb. 2021.

[24] 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.

[25] B. Galkin, J. Kibilda, and L. A. DaSilva, "Deployment of UAV-mounted access points according to spatial user locations in two-tier cellular networks," in *Proc. Wireless Days (WD)*, Mar. 2016, pp. 1–6.

[26] R. I. Bor-Yaliniz, A. El-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–5.

[27] 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.

[28] M. Alzenad, A. El-Keyi, F. Lagum, and H. Yanikomeroglu, "3-D placement of an unmanned aerial vehicle base station (UAV-BS) for energy-efficient maximal coverage," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 434–437, Aug. 2017.

[29] 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.

[30] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1046–1061, May 2017.

[31] J. Lyu, Y. Zeng, and R. Zhang, "UAV-aided offloading for cellular hotspot," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3988–4001, Jun. 2018.

[32] L. Zhang, Q. Fan, and N. Ansari, "3-D drone-base-station placement with in-band full-duplex communications," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1902–1905, Sep. 2018.

[33] Y. Huang, W. Mei, J. Xu, L. Qiu, and R. Zhang, "Cognitive UAV communication via joint maneuver and power control," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7872–7888, Nov. 2019.

[34] J. Qin, Z. Wei, C. Qiu, and Z. Feng, "Edge-prior placement algorithm for UAV-mounted base stations," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.

[35] C.-C. Lai, C.-T. Chen, and L.-C. Wang, "On-demand density-aware UAV base station 3D placement for arbitrarily distributed users with guaranteed data rates," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 913–916, Jun. 2019.

[36] X. Guo, B. Li, K. Liu, and R. Zhang, "Joint placement optimization and RNC in UAV-based wireless multicast networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

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

[38] 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.

[39] Y. Hu, M. Chen, W. Saad, H. V. Poor, and S. Cui, "Distributed multi-agent meta learning for trajectory design in wireless drone networks," 2020, [Online]. Available: arXiv:2012.03158.

[40] Q. Zhang, W. Saad, M. Bennis, X. Lu, M. Debbah, and W. Zuo, "Predictive deployment of UAV base stations in wireless networks: Machine learning meets contract theory," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 637–652, Jan. 2021.

[41] E. Nayebi and B. D. Rao, "Access point location design in cell-free massive MIMO systems," in *Proc. 52nd Asilomar Conf. Signals Syst. Comput.*, Oct. 2018, pp. 985–989.

[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," 2021, [Online]. Available: arXiv:2011.02591.

[43] H. Yamamura, K. Kaneda, and Y. Mizobata, "Communication problems after the great east Japan earthquake of 2011," *Disaster Med. Public Health Preparedness*, vol. 8, no. 4, pp. 293–296, Aug. 2014.

[44] K. Hill. (Oct. 2019). *Wildfires, Power Shut-Offs Impact Cellular Networks in California*. [Online]. Available: https://www.rcrwireless.com/20191031/network-infrastructure/wildfires-power-shut-offs-impact-cellular-networks-in-california

[45] G. P. Villardi, G. T. F. de Abreu, and H. Harada, "TV white space technology: Interference in portable cognitive emergency network," *IEEE Veh. Technol. Mag.*, vol. 7, no. 2, pp. 47–53, Jun. 2012.

[46] G. R. Gopal and B. D. Rao, "Throughput and delay driven access point placement," in *Proc. 53rd Asilomar Conf. Signals Syst. Comput.*, Nov. 2019, pp. 1010–1014.

[47] E. Nayebi, "TDD massive MIMO systems: Channel estimation, power optimization, and access point location design," Ph.D. dissertation, Dept. Electr. Eng. (Commun. Theory Syst.), Univ. California, San Diego, CA, USA, 2018.

[48] J. Lyu and R. Zhang, "Network-connected UAV: 3-D system modeling and coverage performance analysis," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 7048–7060, Aug. 2019.

[49] E. Yanmaz, S. Hayat, J. Scherer, and C. Bettstetter, "Experimental performance analysis of two-hop aerial 802.11 networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 3118–3123.

[50] N. Ahmed, S. S. Kanhere, and S. Jha, "On the importance of link characterization for aerial wireless sensor networks," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 52–57, May 2016.

[51] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.

[52] E. Ostlin, H.-J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2735–2747, Jul. 2010.

[53] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer Acad. Publ., 1991.

[54] M. Sadek, A. Tarighat, and A. H. Sayed, "Active antenna selection in multiuser MIMO communications," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1498–1510, Apr. 2007.

[55] B. O. Lee, H. W. Je, O.-S. Shin, and K. B. Lee, "A novel uplink MIMO transmission scheme in a multicell environment," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 4981–4987, Oct. 2009.

[56] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[57] S. Hur, B. C. Jung, and B. D. Rao, "Sum rate enhancement by maximizing SGINR in an opportunistic interference alignment scheme," in *Proc. 45th Asilomar Conf. Signals Syst. Comput.*, Nov. 2011, pp. 354–358.

[58] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discr. Algorithms*, Jan. 2007, pp. 1027–1035.

BHASKAR D. RAO (Fellow, IEEE) is currently a Distinguished Professor with the Department of Electrical and Computer Engineering, University of California San Diego, San Diego, CA, USA, where he is the holder of the Ericsson Endowed Chair of Wireless Access Networks. His research interests include digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and biomedical signal processing. He received the 2016 IEEE Signal Processing Society Technical Achievement Award.

GABRIEL PORTO VILLARDI (Senior Member, IEEE) received the B.E. degree in electrical engineering with an emphasis on telecommunications from the Federal Center of Technological Education of Rio de Janeiro, Brazil, in 2002, and the M.E. and Ph.D. degrees in physics and electrical and computer engineering as a Japanese Government (Mombukagakusho) Scholar from Yokohama National University, Japan, in 2006 and 2009, respectively. In 2009, he joined the National Institute of Information and Communications Technology, where he is currently a Senior Researcher with the Wireless Systems Laboratory, Yokosuka Research Park, Japan. From 1999 to 2000, he received the Coordination for the Improvement of Higher Education Personnel/Institute of International Education Scholarship to pursue his studies with Clemson University, Clemson, SC, USA. From March 2019 to December 2020, he was a Visiting Scholar with the Qualcomm Institute of Calit2, University of California San Diego, La Jolla, CA, USA. His current research interests include wireless communications theory, statistical modeling, multiple-input multiple-output, cognitive radios, and beyond 5th generation wireless systems. From 2014 to 2015, he was the Secretary of the IEEE 802.22b Task Group on Enhancements for Broadband Services and Monitoring Applications in TVWS, until the standard completion, and the IEEE 802.22 Working Group on Wireless Regional Area Networks in TVWS, from 2014 to 2016. He is member of the IEEE-SA, and has actively contributed to the IEEE 802.22, IEEE 802.11, IEEE 802.15, IEEE 802.19, and IEEE 1900.6 standardization working groups.

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. degree 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, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, under the supervision of Prof. Bhaskar D. Rao. His previous experience is on routing in device-to-device communications. His current research interests include wireless communication theory, signal processing for wireless applications, multiple-input-multiple-output small-cell and cell-free systems, vector quantization, access point location design, and beyond 5th generation wireless systems.