

# Two-Tier Communication for UAV-Enabled Massive IoT Systems: Performance Analysis and Joint Design of Trajectory and Resource Allocation

Zhuo Sun, Zhiqiang Wei, Nan Yang, and Xiangyun Zhou

**Abstract**—In this paper, we propose a two-tier communication strategy to facilitate data collection in unmanned aerial vehicle (UAV)-enabled massive Internet of Things (IoT) systems through introducing ground access points (APs) to serve between the UAV and IoT devices. In the first tier of our proposed strategy, all IoT devices transmit their packets to their local APs via a multi-channel ALOHA-based random access scheme, while in the second tier, APs deliver their aggregated data to the UAV through coordinated time division multiple access. Thus, our introduced APs not only liberate the UAV from the potential massive IoT congestion but also facilitate the design of UAV's trajectory based on the location of APs. To examine the performance of our strategy, we propose a tractable framework to analyze the average system throughput. We reveal that the average two-tier throughput of each AP monotonically increases with its maximum achievable throughput in the second tier, while the increasing slope becomes steeper with a higher traffic load mean in the first tier. Then, we formulate the joint design of UAV's trajectory and resource allocation as a non-convex optimization problem to maximize the average system throughput while considering the heterogeneous quality of service requirement of each AP. To solve this problem, a low-complexity iterative algorithm is devised based on successive convex approximation. Numerical results demonstrate the substantial average system throughput gain achieved by our proposed strategy and design in the context of massive access, compared to the baseline schemes in the literature.

**Index Terms**—UAV communications, massive Internet of Things, performance analysis, trajectory design, resource allocation.

## I. INTRODUCTION

Driven by the proliferation of Internet of Things (IoT) devices, massive IoT is envisioned to revolutionize our daily life via building a networked world in the next generation wireless network and beyond. In the future IoT systems, the number of connected devices is expected to experience an explosive growth and reach the order of billions [1, 2]. To achieve this expectation, it is of paramount significance to efficiently accommodate a massive number of devices in the future IoT paradigm.

The work of Z. Sun, N. Yang, and X. Zhou was supported by the Australian Research Council's Discovery Project (DP180104062). (Corresponding author: Zhiqiang Wei)

Z. Sun, N. Yang, and X. Zhou are with the Research School of Electrical, Energy and Materials Engineering, Australian National University, Canberra, ACT 2600, Australia (Email: nan.yang@anu.edu.au).

Z. Wei is with the School of Electrical Engineering and Telecommunications, the University of New South Wales, Sydney, NSW 2052, Australia (Email: zhiqiang.wei@unsw.edu.au).

One of the key enablers in supporting massive IoT is the design of appropriate massive access schemes [3]. In [4, 5], the grant-free random access schemes were proposed, where the devices directly transmit their unique pilot sequences following with data packets without a grant procedure and the base station (BS) detects the active devices based on its received preambles. As a result, both the intolerantly high access latency and the prohibitively large signaling overhead can be significantly reduced in the context of massive access. However, due to the massive number of devices, the use of non-orthogonal pilot sequences makes it challenging to detect the active devices and to estimate their channels for grant-free systems. By exploiting the sporadic traffic pattern of IoT devices, the compressed sensing (CS) technique can be employed for the joint device activity detection and channel estimation [5]. In [6], the authors proposed a transmission control scheme to further improve the performance of device activity detection and channel estimation. Besides, the non-orthogonal multiple access (NOMA) technique has also been introduced for massive IoT in [7, 8]. In [7], the massive number of IoT devices were spatially clustered and the NOMA was applied within each cluster. In [8], the authors proposed a fully non-orthogonal communication scheme in the context of massive access, where the non-orthogonal pilot sequences were used to estimate the channel and the estimated channel was used for the design of spatial beamforming for interference cancellation. Despite the fruitful results in the literature, the performance of massive access systems is still restrained by the devices with weak communication links and the massive IoT congestion.

The application of unmanned aerial vehicles (UAVs) serves as a promising candidate for massive IoT, owing to the UAV's high mobility and the line-of-sight (LOS) dominated air-ground channels. In fact, the UAV has been widely exploited to data collection in wireless IoT systems [9–15]. In [9–11], the UAV-enabled IoT techniques were reviewed, which include those for data collection and information dissemination. [13] studied the UAV-enabled data collection system to minimize the UAV's flight time by jointly optimizing the data collection intervals, the UAV's speed, and the devices' transmit power. However, [13] considered a one-dimensional (1D) system where all devices are located on a straight line. [14] further extended the 1D system with a two-dimensional (2D) distribution of devices and UAVs. In [15], the maximum energy consumption of all devices was minimized for the UAV-enabled data collection system, by jointly optimizing the

UAV's trajectory and the wake-up scheduling for the devices.

An important limitation encountered in the aforementioned studies lies in their assumption that the number of devices supported by the UAV is relatively limited. Thus, all the devices can be effectively coordinated to perform their transmissions, e.g. through the time division multiple access (TDMA) and frequency division multiple access (FDMA). However, when we consider the massive connectivity scenario, a large number of IoT devices need to be supported by the UAV but their traffic is typically sporadic [16–18]. Such a massive number of IoT devices result in an unbearably large coordination overhead. This hinders the operation of coordinated multiple access, in particular for the UAV-enabled system with limited cruise time. To avoid the large coordination overhead, uncoordinated multiple access becomes attractive for the massive and sporadic connectivity system, such as ALOHA and its enhanced variant, coded slotted ALOHA (CSA) [19, 20]. Unfortunately, without coordination, a large number of collisions may occur in ALOHA and CSA, which significantly reduces the connectivity efficiency. Furthermore, as the number of IoT devices significantly increases, it becomes impractical to gather and save the location of all IoT devices. This implies that it becomes almost impossible to design the UAV's trajectory based on the location of massive IoT devices for improving the system performance, such as the existing studies [12–15]. As a result, there is an emerging need for the deployment of a new network architecture to fully unleash the potential of UAV-enabled massive IoT systems.

To meet this emerging need, some existing studies proposed a hierarchical network architecture for the massive connectivity scenario, such as [21–24]. In [21], a two-hop cluster access scheme based on slotted ALOHA was proposed and a criterion was introduced for selecting the cluster head. Then [22] extended this scheme to a multi-hop scheme and proposed a hierarchical network to aggregate data. In [23], a coverage probability-based optimal data aggregation scheme was designed for the hierarchical network to minimize the average total energy expenditure per unit area per unit time. In [24], the interference and coverage performance were characterized for the large-scale data aggregation system, where the data aggregators deploy both resource scheduling and data delivery to the base stations (BSs). We note that these studies mainly considered static ground BSs but did not take into account the UAV's mobility.

In this paper, we propose a two-tier communication strategy with the aid of ground access points (APs) for UAV-enabled massive IoT systems. The transmission performance in the proposed strategy is analyzed and the joint trajectory and resource allocation design problem is studied to further improve the system performance. The main contributions of this work are summarized as follows:

- 1) We propose to introduce multiple ground APs to form a two-tier network architecture which helps the transmission between the UAV and massive ground IoT devices. In the first tier, the IoT devices transmit their packets to their local APs, once they are active, via a multi-channel ALOHA-based random access scheme, which avoids the laborious coordination overhead. The

successfully detected packets are stored in the buffer of APs, which in turn becomes the traffic load for the AP to communicate with the UAV in the second tier. In this tier, the APs deliver their aggregated data to the UAV in a coordinated fashion, i.e., using TDMA. Such a multi-AP aided architecture significantly reduces the number of direct connections from IoT devices to the UAV, thus decreasing the congestion and improving the connectivity efficiency. Furthermore, the exploitation of APs allows the UAV to communicate to a limited number of APs, instead of massive IoT devices. This implies that based on the location of APs, the UAV's trajectory can be efficiently designed to further enhance the system performance.

- 2) To characterize the performance and to unveil useful insights into the practical implementation of our proposed strategy, we analyze the average system throughput achieved by our strategy in the considered UAV-enabled massive IoT system. Specifically, in order to analyze the performance of random access from IoT devices to APs in the first tier, we leverage the stochastic geometry to obtain the signal-to-interference ratio (SIR) distribution for active IoT devices and derive the probability that the packet from IoT devices can be successfully detected by its serving AP. Based on this packet detection probability, the traffic load distribution of each AP is obtained, which represents the number of its collected bits and is found to follow a Poisson distribution. For the second tier, we derive the maximum achievable throughput from each AP to the UAV as a function of the UAV's trajectory and the resource allocation algorithm. We then derive the average system throughput that connects the performance of both tiers. Our analysis provides insights into the interplay between the traffic load imposed by the first tier and the maximum achievable throughput supported by the second tier, as well as their joint impact on the average system throughput.
- 3) Based on the obtained analytical results, we formulate the joint design on the UAV's trajectory and resource allocation as a non-convex optimization problem to maximize the system throughput while taking into account the heterogeneous quality of service (QoS) requirement of each AP. Since the formulated problem is non-convex which is generally intractable, we propose an iterative algorithm based on the successive convex approximation (SCA) technique to achieve a suboptimal solution. Extensive numerical results are presented to verify our performance analysis and to evaluate the superior average system throughput achieved by our proposed strategy and joint design, compared to the baseline schemes without AP or UAV's trajectory design.

The remainder of this paper is organized as follows. In Section II, we introduce the proposed two-tier communication strategy for UAV-enabled massive IoT systems. The performance analysis of the proposed strategy is provided in Section III. In Section IV, we formulate an optimization problem to maximize the average system throughput and propose an

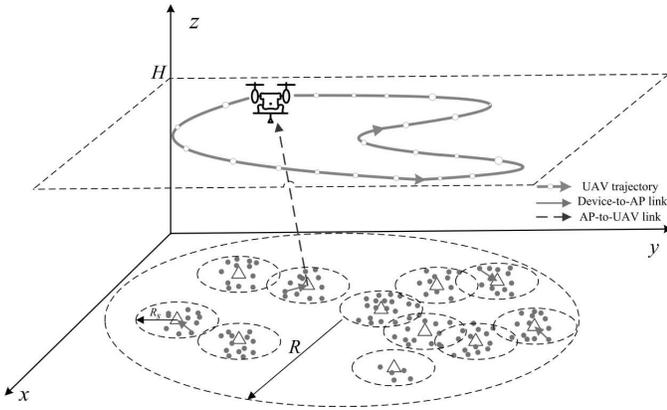


Fig. 1. Illustration of our proposed two-tier communication strategy for UAV-enabled massive IoTs. The triangles represent APs and the circles represent IoT devices.

iterative algorithm to achieve a suboptimal solution of the formulated problem. Section V provides numerical results to validate the effectiveness of our analysis and design. The paper is concluded in Section VI.

Notations used in this paper are listed as follows. Bold-face lower case letters are reserved for vectors.  $(\cdot)^T$  denotes the transpose of a vector;  $|\cdot|$  denotes the absolute value of a complex scalar and  $\|\cdot\|$  denotes the Euclidean norm of a vector;  $\Pr(\cdot)$  denotes the probability of a random event occurring and  $\mathbb{E}_X[\cdot]$  denotes the expectation taken over a random variable (RV)  $X$ .  $f_X(\cdot)$  denotes the probability density function (PDF) of a RV  $X$ .  $\text{Pois}(\lambda)$  represents the Poisson distribution with a mean of  $\lambda$ . The circularly symmetric complex Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$  is denoted by  $\mathcal{CN}(\mu, \sigma^2)$ . Unless otherwise stated, the notations used throughout the paper are summarized in Table I.

## II. TWO-TIER COMMUNICATION FOR UAV-ENABLED MASSIVE IoT SYSTEMS

We propose a two-tier UAV communication strategy where a single UAV serves as an aerial BS to collect data from a massive number of ground IoT devices in a disc of a radius  $R$ , as shown in Fig. 1. In our proposed strategy, the two-tier architecture is formed by introducing  $M$  ground APs<sup>1</sup> to help the transmission between the IoT devices and the UAV. We assume that all the IoT devices are located within  $M$  clusters. For each cluster, the IoT devices are spatially distributed according to a homogeneous Poisson point process (PPP) with density<sup>2</sup>  $\lambda_d$  within a disc of a radius  $R_s$ , where the location of this disc center is arbitrary but fixed. We also assume that the AP is located at the center of its serving IoT device cluster, where the location of AP  $m$  is denoted by  $\mathbf{w}_m \in \mathbb{R}^{2 \times 1}$ , and its service zone is the disc corresponding to

<sup>1</sup>Note that, the concept of ground APs has been exploited for wireless sensor networks in [25–27]. However, they mainly considered that the sensors transmit data to a static ground base station and employed multiple APs as the cluster heads to reduce the sensors' energy consumption.

<sup>2</sup>We assume that all IoT devices are static in this paper. When considering the mobility of IoT devices, the density of each cluster varies across subframes. This has a great impact on the system performance analysis and the design of UAV's trajectory, which will be studied in our future work.

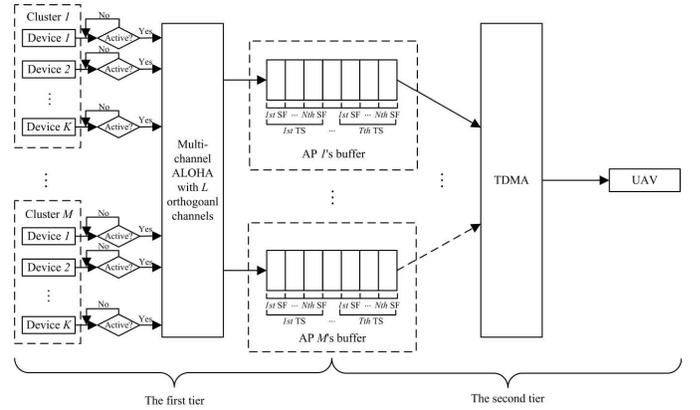


Fig. 2. Data transmission in the proposed two-tier architecture, in which  $K$  denotes the random number of IoT devices in each cluster following a Poisson distribution with a mean  $\lambda_d \pi R_s^2$ , and SF and TS denote subframe and time slot for presentation simplicity, respectively.

this cluster. We also assume that the IoT devices, APs, and the UAV are equipped with a single antenna<sup>3</sup>. The UAV is cruising within a given area in the sky with a constant altitude of  $H$ , to collect data from ground IoT devices. The total cruising period consists of  $T$  time slots with  $\delta_t$  denoting the elemental slot duration. Therefore, the cruising trajectory can be denoted by a sequence  $\{\mathbf{q}[t]\}_{t=1}^T$ , where  $\mathbf{q}[t] = [x[t], y[t]]^T$  denotes the ground projected coordinate of the UAV during time slot  $t$ . Within each time slot, we assume that the distance between the UAV and each AP is invariant since this distance is much longer compared to the displacement of the UAV during  $\delta_t$  [11]. Thus, during time slot  $t$ , the distance between the UAV and AP  $m$  is given by  $d_m^{UA}[t] = \|\mathbf{q}[t] - \mathbf{w}_m\|$ . In addition, we have  $\mathbf{q}[1] = \mathbf{q}[T]$  as the UAV is cruising within an area.

The data transmission from each IoT device to the UAV is achieved by the proposed two-tier architecture, as shown in Fig. 2. We assume that the communications in the first tier and the second tier employ orthogonal frequency bands<sup>4</sup> with corresponding bandwidth of  $B_1$  and  $B_2$ , respectively, and  $B_1 + B_2 = B$ . As a result, the communications in the first tier and in the second tier can be established simultaneously without suffering from the inter-tier interference. In the first tier, we assume that the AP keeps collecting and storing the packets from IoT devices within one UAV cruising period. In the second tier, the AP transmits its aggregated data within the previous cruising period to the UAV. We note that the proposed two-tier strategy is suitable for delay-tolerable data collection systems, where one data packet may bear a delay ranging

<sup>3</sup>It is noted that the APs may have multiple antennas to support the massive number of IoT devices in practice, while the IoT devices and the UAV usually have a single antenna due to the limitation on the physical size, weight, and deployment cost [28, 29]. It is intuitive that equipping multiple antennas at the APs has a great impact on the design of massive access and the system performance, e.g., the design of new beamforming strategy and the analysis of SIR. The performance analysis and joint design of UAV's trajectory and resource allocation become challenging tasks for the case with multi-antenna APs, which will be considered in our future work.

<sup>4</sup>In fact, the serving IoT devices of the non-scheduled APs can utilize the total bandwidth  $B$  to increase the number of subchannels and to improve the first-tier performance. However, such bandwidth utilization can result in the mutual interference between the two tiers and accordingly reduce the average system throughput.

TABLE I  
NOTATIONS FOR MAIN SYSTEM PARAMETERS.

Notations	Physical meaning	Notations	Physical meaning
$M$	Number of APs	$T$	Number of time slots in a cruising period
$A_m$	AP $m$	$\alpha$	Path-loss exponent
$R$	Radius of considered area	$D_k$	IoT device $k$
$\beta_0$	Channel power gain at reference distance	$\delta_t$	Slot duration
$D$	Packet length	$N_0$	Power spectrum density of thermal noise
$R_s$	Radius of each AP's serving disc	$C_l$	Subchannel $l$
$S_t$	Time slot $t$	$h_{k,m,l}$	Small-scale channel fading from $D_k$ to $A_m$ over $C_l$
$\mathbf{w}_m$	Location of AP $m$	$I_{k,m,l}^{\text{intra}}$	Intra-cluster interference power of $D_k$ at $A_m$ over $C_l$
$\delta_n$	Subframe duration	$\gamma_{k,m,l}$	SIR of $D_k$ at $A_m$ over $C_l$
$\lambda_d$	Density of IoT devices	$p_a$	Activity probability
$\mathbf{q}[t]$	UAV's ground projected coordinate in $S_t$	$L$	Number of subchannels
$I_{k,m,l}^{\text{inter}}$	Inter-cluster interference power of $D_k$ at $A_m$ over $C_l$	$R_{k,m,l}$	Achievable data rate from $D_k$ to $A_m$ over $C_l$
$B_1$	Bandwidth for the first tier	$\Delta_f$	Bandwidth of each subchannel
$R_0$	Target data rate of IoT device	$x_m[t]$	Scheduling indicator of $A_m$ in $S_t$
$B_2$	Bandwidth for the second tier	$r_{k,m}$	Distance between IoT device $k$ to AP $m$
$p_m[t]$	Transmit power of $A_m$ in $S_t$	$g_m[t]$	Channel coefficient from $A_m$ to the UAV in $S_t$
$H$	Height of UAV's trajectory	$R_m[t]$	Achievable data rate of $A_m$ in $S_t$
$U_m$	Maximum achievable throughput of $A_m$	$N_m$	Number of collected packets in buffer of $A_m$
$\eta$	Detection threshold	$P_{\text{Det}}[m]$	Packet detection probability at $A_m$
$\Theta_m$	Average two-tier throughput of $A_m$	$\Theta$	Average system throughput
$p_{\text{max}}$	AP's maximum transmit power	$V_{\text{max}}$	UAV's maximum flying speed
$N$	Number of subframes in a time slot	$\tau_m$	Minimum required average two-tier throughput of $A_m$

from 1 to  $2T - 1$  time slots<sup>5</sup>. We next detail the transmission processes in the first tier and second tier.

#### A. First Tier Transmission

To facilitate the data collection in the first tier, each time slot is further divided into  $N$  subframes where the duration of each subframe is  $\delta_n$ . In each subframe, every IoT device becomes active to transmit one packet of  $D$  bits, according to an activity probability  $p_a$ . The activities among all the IoT devices and the activities of each IoT device in different subframes during different time slots are assumed to be identical and independent distributed<sup>6</sup> (i.i.d.). All active IoT devices transmit their packets to their local APs according to the multi-channel ALOHA-based random access scheme [30–32]. In particular, all the IoT devices in all clusters share  $L$  orthogonal subchannels. Each subchannel has a fixed bandwidth of  $\Delta_f$  with  $B_1 = L\Delta_f$ . When an IoT device becomes active in a subframe during a time slot, it randomly selects one subchannel to transmit its packet to the AP. Due to uncoordinated subchannel selection, more than one IoT device may select the same subchannel at the same time. Let us denote  $\mathcal{A}_{m,l}[t, n]$  as the set of active IoT devices in cluster  $m$  transmitting packets through the  $l$ -th subchannel in subframe  $n$  during time slot  $t$ . Thus, the received signal at AP  $m$  over the  $l$ -th subchannel in subframe  $n$  during time slot  $t$  is given by

$$y_{m,l}^{\text{AP}}[t, n] = \sqrt{P_t r_{k,m}^{-\alpha} \beta_0 h_{k,m,l}[t, n] s_k[t, n]} + \sum_{k' \in \mathcal{A}_{m,l}[t, n] \setminus \{k\}} \sqrt{P_t r_{k',m}^{-\alpha} \beta_0 h_{k',m,l}[t, n] s_{k'}[t, n]}$$

<sup>5</sup>To reduce the data collection delay, one AP can collect and transmit data packets within the same cruising period. However, this collection and transmission need a complicated and dynamic packet management protocol. Thus, it will be considered in our future work.

<sup>6</sup>In this paper, we consider the homogeneous IoT devices with i.i.d. activities in different subframes during different time slots. The proposed scheme and performance analysis can be easily extended to the case where all the IoT devices have independent but non-identical activity probabilities.

$$+ \sum_{\substack{k'' \in \mathcal{A}_{m',l}[t, n], \\ m' \in \{1, \dots, M\} \setminus \{m\}}} \sqrt{P_t r_{k'',m'}^{-\alpha} \beta_0 h_{k'',m',l}[t, n] s_{k''}[t, n]} + n_{m,l}[t, n], \quad (1)$$

where the three terms on the right-hand side correspond to the desired signal, the intra-cluster interference signal, and the inter-cluster interference signal, respectively,  $P_t$  is the transmit power of each IoT device,  $r_{k,m}$  is the distance from IoT device  $k$  to AP  $m$ ,  $\alpha$  is the path-loss exponent,  $\beta_0$  is the channel power gain at the reference distance of 1 meter,  $h_{k,m,l}[t, n]$  is the small-scale Rayleigh fading coefficient from IoT device  $k$  to AP  $m$  over the  $l$ -th subchannel in subframe  $n$  during time slot  $t$ , with  $h_{k,m,l}[t, n] \sim \mathcal{CN}(0, 1)$ ,  $s_k[t, n]$  is the transmitted signal of IoT device  $k$  in subframe  $n$  during time slot  $t$  with a unit power, i.e.,  $\mathbb{E}[|s_k[t, n]|^2] = 1$ , and  $n_{m,l}[t, n]$  is the thermal noise, with the power  $\sigma_1^2$ , at AP  $m$  over the  $l$ -th subchannel in subframe  $n$  during time slot  $t$ , i.e.,  $n_{m,l}[t, n] \sim \mathcal{CN}(0, \sigma_1^2)$ . Here, we assume a quasi-static fading channel where  $h_{k,m,l}[t, n]$  remains constant in one subframe but changes independently from subframe to subframe.

It can be seen that the transmission from one IoT device may be interfered by the concurrent transmissions from other active IoT devices who select the same subchannel. In this work, such interference is referred to as the intra-cluster interference if the interfering IoT devices are located in the same cluster, or the inter-cluster interference if the interfering IoT devices are located in other clusters. Without loss of generality, we consider an arbitrary active IoT device  $k$  of cluster  $m$  and its served AP  $m$ ,  $m \in \{1, \dots, M\}$ , as the typical IoT device and the typical AP, respectively, in order to characterize the impacts of both the intra-cluster interference and the inter-cluster interference on the detection probability of the transmitted packet from IoT device  $k$  to AP  $m$ . In a specific subframe  $n$  during time slot  $t$ , the typical IoT device  $k$  is assumed to be active and selects the  $l$ -th subchannel for its data transmission to the typical AP  $m$ . Furthermore, due to uncoordinated concurrent transmissions, the aggregate

interference power at the AP dominates the thermal noise power. Therefore, an interference-limited system is considered by ignoring the thermal noise and the SIR is adopted to evaluate the detection probability<sup>7</sup>. The SIR for detecting the packet of IoT device  $k$  at AP  $m$  in subframe  $n$  during time slot  $t$  is given by

$$\gamma_{k,m,l}[t,n] = \frac{S_{k,m,l}[t,n]}{I_{k,m,l}^{\text{intra}}[t,n] + I_{k,m,l}^{\text{inter}}[t,n]}, \quad (2)$$

with

$$S_{k,m,l}[t,n] = \beta_0^2 P_t |h_{k,m,l}[t,n]|^2 r_{k,m}^{-\alpha}, \quad (3)$$

$$I_{k,m,l}^{\text{intra}}[t,n] = \sum_{k' \in \mathcal{A}_{m,l}[t,n] \setminus \{k\}} \beta_0^2 P_t |h_{k',m,l}[t,n]|^2 r_{k',m}^{-\alpha}, \quad (4)$$

and

$$I_{k,m,l}^{\text{inter}}[t,n] = \sum_{\substack{k'' \in \mathcal{A}_{m',l}[t,n], \\ m' \in \{1, \dots, M\} \setminus \{m\}}} \beta_0^2 P_t |h_{k'',m,l}[t,n]|^2 r_{k'',m}^{-\alpha}, \quad (5)$$

where  $S_{k,m,l}[t,n]$  is the desired received signal power,  $I_{k,m,l}^{\text{intra}}[t,n]$  is the aggregate intra-cluster interference power, and  $I_{k,m,l}^{\text{inter}}[t,n]$  is the aggregate inter-cluster interference power. Accordingly, the achievable data rate from IoT device  $k$  to its served AP  $m$  is given by

$$R_{k,m,l}[t,n] = \Delta_f \log_2(1 + \gamma_{k,m,l}[t,n]). \quad (6)$$

We assume the IoT device is transmitting its data with a fixed target rate of  $R_0$ , where  $R_0 = D/\delta_n$ . If the achievable data rate is larger than or equal to the target rate  $R_0$ , the packet transmitted by IoT device  $k$  in subframe  $n$  during time slot  $t$  can be successfully detected by its served AP  $m$ . In this case, the detected packet would be stored in the buffer of AP  $m$  until being delivered to the UAV. On the other hand, if the packet is not successfully detected due to the severe interference, it would be dropped and thus makes no contribution to the system throughput. Due to the randomness in the channel fading and subchannel selection, the successful detection of a packet at the AP happens opportunistically. This motivates our analysis of the *packet detection probability* in Section III.

### B. Second Tier Transmission

In the second tier, the UAV periodically cruises within a given area and collects packets from all the APs in a coordinated manner, i.e., the TDMA. Then, the UAV decodes the received packets from APs and stores the decoded packets into a data buffer of sufficiently large size. Afterwards, the UAV offloads the stored packets to the central control station, when completing its flying task<sup>8</sup>. Let us denote a binary variable  $x_m[t]$  as the AP scheduling indicator. If AP  $m$  is scheduled to transmit its packet to the UAV during time slot

$t$ , we have  $x_m[t] = 1$ ; otherwise,  $x_m[t] = 0$ . Furthermore, we assume that at most one AP is scheduled in each time slot, i.e.,  $\sum_{m=1}^M x_m[t] \leq 1$ . Thus, the received signal at the UAV during time slot  $t$  is given by

$$y^{\text{UAV}}[t] = \sum_{m=1}^M \sqrt{p_m[t]} x_m[t] g_m[t] s_m[t] + n[t], \quad (7)$$

where  $p_m[t]$  is the transmit power of AP  $m$  during time slot  $t$ ,  $g_m[t]$  is the channel coefficient from AP  $m$  to the UAV,  $s_m[t]$  is the transmitted signal from AP  $m$  during time slot  $t$  with a unit power<sup>9</sup>, i.e.,  $\mathbb{E}[|s_m[t]|^2] = 1$ , and  $n[t]$  is the thermal noise, with the power  $\sigma_2^2 = N_0 B_2$ , at the UAV, where  $N_0$  is the power spectrum density of the thermal noise and  $B_2$ , as aforementioned, is the bandwidth occupied by the AP in the second tier. Finally, we assume that  $g_m[t]$  is dominated by the LoS link<sup>10</sup> [11, 35], given by

$$g_m[t] = \frac{\beta_0}{\sqrt{H^2 + \|\mathbf{q}[t] - \mathbf{w}_m\|^2}}. \quad (8)$$

We note that due to the adoption of a LoS channel model between the AP and the UAV, each AP can predict its channel gain to the UAV and obtain the maximum achievable throughput.

Thanks to the adopted TDMA scheme and the orthogonal frequency bands employed in the proposed two-tier strategy, the transmission from each AP to the UAV enjoys an interference-free channel. As such, the achievable data rate of AP  $m$  during time slot  $t$  is obtained as

$$R_m[t] = B_2 x_m[t] \log_2 \left( 1 + \frac{p_m[t] \beta_0^2}{B_2 N_0 (H^2 + \|\mathbf{q}[t] - \mathbf{w}_m\|^2)} \right). \quad (9)$$

Now, we define the *maximum achievable throughput* as the maximum number of bits that can be supported from AP  $m$  to the UAV during one cruising period of  $T$  time slots, i.e.,

$$U_m(\mathbf{q}[t], x_m[t], p_m[t], B_2) = \sum_{t=1}^T R_m[t] \delta_t, \quad (10)$$

which is a function of the UAV's trajectory and resource allocation strategy.

As previously mentioned, each AP needs to deliver all the collected packets during the previous  $T$  time slots to the UAV. We denote the number of collected packets in the buffer of AP  $m$  as  $N_m$ , whose distribution will be analyzed in Section III. Here, we adopt the first-come-first-serve queuing model for the stored packets at each AP. Given the packet size of  $D$  bits, there are totally  $DN_m$  bits to be delivered by AP  $m$  to the UAV in the second tier, which is referred to as the *traffic load*

<sup>9</sup>Without causing notation confusion,  $s_k[t,n]$  denotes transmitted signal from IoT device  $k$  in subframe  $n$  during time slot  $t$  in the first tier and  $s_m[t]$  denotes the transmitted signal from AP  $m$  during time slot  $n$  in the second tier.

<sup>10</sup>According to the field measurements [33, 34], the LoS channel model gives a reasonable approximation of practical channels in a rural environment, when the UAV's flight height is 100 meters and the radius of the service disc is 350 meters. Besides, as the first attempt to introducing APs for the UAV-enabled massive IoT system, adopting the simple yet reasonably accurate LoS channel model can provide better insights into the performance and the design of the considered system.

<sup>7</sup>In fact, the obtained results can be easily generalized to the case including the thermal noise, which provides no significant difference.

<sup>8</sup>In this work, we focus on the packet collection process from IoT devices to the UAV, since this is the performance bottleneck of the considered massive access scenario. The process of offloading the collected data at the UAV to the central control station is not considered in this work.

of AP  $m$ . Once AP  $m$  is scheduled to communicate with the UAV during a time slot, it selects the data from its buffer based on its maximum achievable throughput  $U_m$  and transmits to the UAV. Otherwise, AP  $m$  only sends the earliest received  $U_m$  bits of its total traffic load to the UAV and drops the remaining bits so as to avoid the potential buffer overflow. Due to the adoption of a coordinated access scheme, we assume that the transmitted data can be successfully detected at the UAV in the second tier. Hence, if  $DN_m < U_m$ , AP  $m$  can successfully transmit all its traffic load  $DN_m$  to the UAV. Otherwise, AP  $m$  only sends  $U_m$  bits of its total traffic load to the UAV. Based on this, we can derive the average two-tier throughput of each AP and optimize the average system throughput through designing the UAV's trajectory and resource allocation strategy in Section III and Section IV, respectively.

### III. PERFORMANCE ANALYSIS

In this section, we first analyze the statistics of the number of successfully detected packets at each AP, i.e., *packet detection probability*, in the first tier. Then, we derive the average two-tier throughput of each AP in the second tier, which serves as a building block for the joint design of trajectory and resource allocation of the UAV in Section IV.

#### A. Analysis of First Tier Transmission

Based on (6), we obtain the detection threshold from the target data rate of each IoT device as  $\eta = 2^{\frac{R_0}{\Delta f}} - 1$ . Accordingly, the packet detection probability at AP  $m$  is written as

$$\begin{aligned} P_{\text{Det}}[m] &= \Pr\left(\frac{S_{k,m}}{I_{k,m}^{\text{intra}} + I_{k,m}^{\text{inter}}} > \eta\right) \\ &= \Pr\left(|h_{k,m}|^2 > (\zeta_{k,m} I_{k,m}^{\text{intra}} + \zeta_{k,m} I_{k,m}^{\text{inter}})\right) \\ &= \mathbb{E}_{r_{k,m}} \left[ \mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right] \right. \\ &\quad \left. \times \mathbb{E}_{I_{k,m}^{\text{inter}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{inter}}) \right] \right], \quad (11) \end{aligned}$$

where  $\zeta_{k,m} = \frac{\eta r_{k,m}^\alpha}{\beta_0^2 P_t}$ . It is noted that the packet detection probability is same for different subframes, time slots, as well as subchannels. Therefore, the subchannel index  $l$ , the time slot index  $t$ , and the subframe index  $n$  are omitted from (11) for the simplicity of notation. Furthermore, the third step in (11) is obtained based on the fact that  $|h_{k,m}|^2 \sim \exp(1)$  and the independence between  $I_{k,m}^{\text{intra}}$  and  $I_{k,m}^{\text{inter}}$ . Additionally, compared to the number of detected packets from an individual IoT device, we are more interested in the total number of detected packets at AP  $m$  since this total number determines the traffic load for the second-tier. To this end, we derive the packet detection probability at AP  $m$  in (11), through taking the expectation over the distance distribution of IoT devices.

We observe from (11) that the packet detection probability  $P_{\text{Det}}[m]$  is determined by the Laplace transform of intra-cluster interference  $I_{k,m}^{\text{intra}}$ , i.e.,  $\mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right]$ , and the Laplace transform of inter-cluster interference  $I_{k,m}^{\text{inter}}$ , i.e.,  $\mathbb{E}_{I_{k,m}^{\text{inter}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{inter}}) \right]$ . In the following, we obtain

the Laplace transform of  $I_{k,m}^{\text{intra}}$  and  $I_{k,m}^{\text{inter}}$  and present them in Lemma 1 and Lemma 2, respectively.

*Lemma 1:* Within cluster  $m$ , the Laplace transform of intra-cluster interference experienced by the packet transmission from the typical IoT device to AP  $m$  is given by

$$\begin{aligned} &\mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right] \\ &= \exp \left[ -\frac{p_a \lambda_d \pi R_s^2}{L} \left( \frac{\xi_{k,m}}{\xi_{k,m} + 1} + \frac{\xi_{k,m}}{(1 + \frac{2}{\alpha})(\xi_{k,m} + 1)^2} \right. \right. \\ &\quad \left. \left. \times {}_2F_1 \left( 1, 2; \frac{2}{\alpha} + 2; \frac{1}{\xi_{k,m} + 1} \right) \right) \right], \quad (12) \end{aligned}$$

where  $\xi_{k,m} = \eta r_{k,m}^\alpha R_s^{-\alpha}$  and  ${}_2F_1(a, b; c; x)$  is the Gauss hypergeometric function [36, Def. (9.100)].

*Proof:* Please refer to Appendix A. ■

We highlight that the Gauss hypergeometric function  ${}_2F_1(a, b; c; x)$  used in (12) is readily available in standard mathematical packages, such as Mathematica. Thus, the expression for  $\mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right]$  is practically closed-form by computing  ${}_2F_1(a, b; c; x)$  via Mathematica.

Apart from the intra-cluster interference, the packet transmitted from the typical IoT device is also affected by the inter-cluster interference, due to the random selection of subchannels for the devices in all the other clusters. We next derive the Laplace transform of inter-cluster interference  $I_{k,m}^{\text{inter}}$  in Lemma 2.

*Lemma 2:* Within cluster  $m$ , the Laplace transform of inter-cluster interference experienced by the packet transmission from the typical IoT device to AP  $m$  is given by

$$\begin{aligned} &\mathbb{E}_{I_{k,m}^{\text{inter}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{inter}}) \right] \\ &= \prod_{m' \in \{1, \dots, M\} \setminus \{m\}} \exp \left[ -\frac{p_a \lambda_d}{L} \int_0^{R_s} \int_0^{2\pi} \frac{\eta r_{k,m}^\alpha r_{k'',m'}^\alpha}{\eta r_{k,m}^\alpha + \varpi^{\frac{\alpha}{2}}} d\theta dr_{k'',m'} \right], \quad (13) \end{aligned}$$

where  $\varpi = r_{k'',m'}^2 + r_{mm'}^2 - 2r_{k'',m'} r_{mm'} \cos(\theta - \theta_{mm'})$ ,  $r_{k'',m'}$  is the distance from IoT device  $k''$  to its serving AP  $m'$ , following the distribution  $f_{r_{k'',m'}}(x) = \frac{2x}{R_s^2}$ , and  $r_{mm'}$  and  $\theta_{mm'}$  are the distance and the relative angle between AP  $m$  and AP  $m'$ , obtained from the given location of AP  $m$  and AP  $m'$ , respectively.

*Proof:* Please refer to Appendix B. ■

We note that the Laplace transform of  $I_{k,m}^{\text{inter}}$  in (13) is obtained via assuming that the location of IoT devices in different clusters follows a homogeneous PPP. We further clarify that (13) can be directly applied to a more general scenario where the location of IoT devices in different clusters follows independent but non-identical homogeneous PPPs. Specifically, the result for this scenario can be obtained by using different IoT device densities for different clusters in (13).

Based on the definition of the packet detection probability in (11) and the Laplace transforms of  $I_{k,m}^{\text{intra}}$  and  $I_{k,m}^{\text{inter}}$ , given in Lemma 1 and Lemma 2, respectively, we now derive the packet detection probability,  $P_{\text{Det}}[m]$ , and present it in Theorem 1.

$$\begin{aligned} P_{\text{Det}}[m] = \frac{2}{R_s^2} \int_0^{R_s} \exp \left[ -\frac{p_a \lambda_d \pi R_s^2}{L} \left( \frac{\xi_{k,m}}{\xi_{k,m} + 1} + \frac{\xi_{k,m}}{(1 + \frac{2}{\alpha})(\xi_{k,m} + 1)^2} {}_2F_1 \left( 1, 2; \frac{2}{\alpha} + 2; \frac{1}{\xi_{k,m} + 1} \right) \right. \right. \\ \left. \left. + \sum_{m' \in \{1, \dots, M\} \setminus \{m\}} \frac{1}{\pi R_s^2} \int_0^{R_s} \int_0^{2\pi} \frac{\eta r_{k,m}^\alpha r_{k',m'}'}{\eta r_{k,m}^\alpha + \tilde{\omega}^{\frac{\alpha}{2}}} d\theta dr_{k',m'}' \right) \right] r_{k,m} dr_{k,m}. \end{aligned} \quad (14)$$

*Theorem 1:* For cluster  $m$ , the probability that the packet from an IoT device can be successfully detected by its serving AP  $m$ , i.e., the packet detection probability, is given by (14).

*Proof:* We express (11) as

$$\begin{aligned} P_{\text{Det}}[m] = \int_0^{R_s} \left[ \mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right] \right. \\ \left. \times \mathbb{E}_{I_{k,m}^{\text{inter}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{inter}}) \right] \right] f_{r_{k,m}}(r_{k,m}) dr_{k,m}. \end{aligned} \quad (15)$$

By substituting the Laplace transform of intra-cluster interference  $I_{k,m}^{\text{intra}}$ , derived in (12), and the Laplace transform of inter-cluster interference  $I_{k,m}^{\text{inter}}$ , derived in (13), into (15), and using the PDF of  $r_{k,m}$  given by  $f_{r_{k,m}}(r_{k,m}) = \frac{2r_{k,m}}{R_s^2}$ , we obtain the packet detection probability,  $P_{\text{Det}}[m]$ , as (14). ■

By observing (14), we make three remarks as follows. First, the packet detection probability,  $P_{\text{Det}}[m]$ , is a decreasing function of  $p_a$  and  $\lambda_d$ . This indicates that increasing the activity probability  $p_a$  or the IoT device density  $\lambda_d$  in a cluster results in a larger interference, thereby decreasing the packet detection probability at the AP in the first tier. Second,  $P_{\text{Det}}[m]$  is an increasing function of  $L$ . This indicates that using more orthogonal subchannels in the first tier reduces the congestion in each subchannel and thus increases the packet detection probability at the AP. Third,  $P_{\text{Det}}[m]$  is affected by  $r_{mm'}$  and  $\theta_{mm'}$ . This indicates that the packet detection probability for AP  $m$  depends on its relative distance and angle with respect to (w.r.t.) all the other APs. Thus, each AP has a unique packet detection probability. We note that while (14) involves two nested integrals, the integral interval is finite and thus, the integrals are not difficult to be numerically solved.

When considering the special case with  $\alpha = 2$ , we derive the packet detection probability,  $P_{\text{Det}}^{\alpha=2}[m]$ , as

$$\begin{aligned} P_{\text{Det}}^{\alpha=2}[m] = \frac{2}{R_s^2} \\ \times \int_0^{R_s} \left( \prod_{m=1}^M \frac{2\eta r_{k,m}^2}{\sqrt{\tilde{\omega}^2 + 4\eta r_{mm'}^2 r_{k,m}^2 + \tilde{\omega}}} \right)^{\varrho r_{k,m}^2} r_{k,m} dr_{k,m}. \end{aligned} \quad (16)$$

where  $\varrho = \frac{1}{L} p_a \lambda_d \pi \eta$  and  $\tilde{\omega} = \eta r_{k,m}^2 + R_s^2 - r_{mm'}^2$ .

*Proof:* Based on (15), the packet detection probability for the special case with  $\alpha = 2$  can be expressed as

$$\begin{aligned} P_{\text{Det}}^{\alpha=2}[m] = \int_0^{R_s} \left[ \mathbb{E}_{I_{k,m}^{\text{intra}}}^{\alpha=2} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right] \right. \\ \left. \times \mathbb{E}_{I_{k,m}^{\text{inter}}}^{\alpha=2} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{inter}}) \right] \right] f_{r_{k,m}}(r_{k,m}) dr_{k,m} \end{aligned}$$

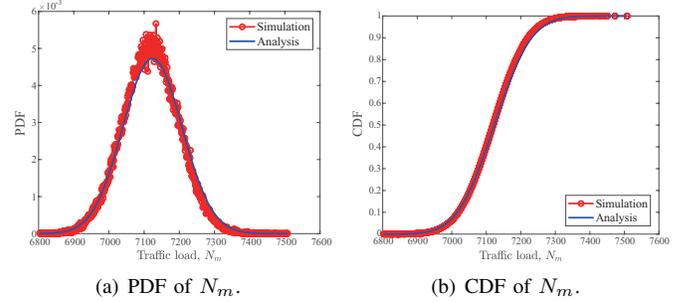


Fig. 3. The CDF and PDF of AP  $m$ 's traffic load  $N_m$ .

$$\begin{aligned} = \int_0^{R_s} \frac{2r_{k,m}}{R_s^2} \exp \left[ -\frac{p_a \lambda_d}{L} \pi \eta \left( \frac{r_{k,m}^2}{\chi_1} + \frac{r_{k,m}^2}{2\chi_1^2} {}_2F_1 \left( 1, 2; 3; \frac{1}{\chi_1} \right) \right. \right. \\ \left. \left. + r_{k,m}^2 \sum_{m' \neq m}^M \int_0^{R_s} \frac{2r_{k',m'}}{\sqrt{r_{k',m'}^4 + \chi_2^2 + 2\chi_3 r_{k',m'}^2}} dr_{k',m'}' \right) \right] dr_{k,m} \\ = \int_0^{R_s} \exp \left[ -\varrho r_{k,m}^2 \left( \frac{1}{\chi_1} + \frac{1}{\chi_1^2} \int_0^1 \frac{x}{1 - \frac{x}{\chi_1}} dx \right. \right. \\ \left. \left. + \ln \left( \frac{\sqrt{\tilde{\omega}^2 + 4\eta r_{mm'}^2 r_{k,m}^2 + \tilde{\omega}}}{2\eta r_{k,m}^2} \right) \right) \right] \frac{2r_{k,m}}{R_s^2} dr_{k,m}, \end{aligned} \quad (17)$$

where  $\chi_1 = \eta r_{k,m}^2 R_s^2 + 1$ ,  $\chi_2 = \eta r_{k,m}^2 + r_{mm'}^2$  and  $\chi_3 = \eta r_{k,m}^2 - r_{mm'}^2$ . We note that the second step is obtained from [36, Eq. (3.645)] and the third step is obtained via using the Euler integral representation [37] of  ${}_2F_1(a, b; c; x)$ , i.e.,  ${}_2F_1(a, b; c; x) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-xt)^a} dt$ . By solving the inner integral in (17), according to [36, Eq. (2.261)], we obtain the packet detection probability for  $\alpha = 2$ . ■

It can be seen from (17) that the intra-cluster interference and the inter-cluster interference are expressed together, which is due to the closed-form expressions for  $\mathbb{E}_{I_{k,m}^{\text{intra}}}^{\alpha=2} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right]$  and  $\mathbb{E}_{I_{k,m}^{\text{inter}}}^{\alpha=2} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{inter}}) \right]$ , as well as the fact of  $r_{mm} = 0$ .

Next, we prove that the number of detected packets at AP  $m$  in a subframe follows a Poisson distribution and present it in Theorem 2.

*Theorem 2:* The number of detected packets at AP  $m$  in subframe  $n$  during time slot  $t$ , denoted by  $N_m[t, n]$ , can be modeled as a Poisson RV. Thus, the total number of detected packets at AP  $m$  across  $N$  subframes during  $T$  time slots, denoted by  $N_m$ , is also a Poisson RV. Mathematically, it is given by

$$N_m = \sum_{t=1}^T \sum_{n=1}^N N_m[t, n] \sim \text{Pois}(\lambda_m), \quad (18)$$

where  $\lambda_m = NTP_{\text{Det}}[m]p_a\lambda_d\pi R_s^2$ .

*Proof:* Since the active IoT device is considered to transmit only one packet in a subframe, the number of transmitted packets is the same as the number of active IoT devices in a subframe within a cluster, which is a Poisson RV with the mean of  $p_a\lambda_d\pi R_s^2$ . On average, one packet can be detected by AP  $m$  with a probability  $P_{\text{Det}}[m]$ . As a result, the number of detected packets at AP  $m$  in subframe  $n$  during time slot  $t$ , denoted by  $N_m[t, n]$ , can be modeled as a Poisson variable with mean of  $P_{\text{Det}}[m]p_a\lambda_d\pi R_s^2$ . Since the number of detected packets  $N_m[t, n]$  is i.i.d. among different subframes and different time slots, the total number of detected packets at AP  $m$  during  $T$  time slots is a Poisson RV with the mean of  $\lambda_m = NTP_{\text{Det}}[m]p_a\lambda_d\pi R_s^2$ , which completes the proof. ■

We note that all APs face heterogeneous traffic load distribution for the second tier, since different APs have different packet detection probabilities. This will be exploited for the trajectory design in Section IV. To demonstrate the correctness of Theorem 2, we plot the analytical and simulation results for the PDF and the cumulative distribution function (CDF) of  $N_m$  in Fig. 3(a) and Fig. 3(b), respectively. It can be seen that the analytical results well match the simulation results in both figures, which verifies our statement in (18).

### B. Analysis of Second Tier Transmission

As previously mentioned, we assume that the transmitted data can be successfully detected at the UAV in the second tier, owing to the adoption of a coordinated access scheme. Specifically, when the maximum achievable throughput from AP  $m$  to the UAV is larger than or equal to the traffic load of AP  $m$ , i.e.,  $U_m \geq DN_m$ , AP  $m$  transmits all the aggregated data to the UAV. Otherwise, AP  $m$  can only select  $U_m$  bits from its buffer and transmit to the UAV. Thus, we derive the *average system throughput* as

$$\Theta = \sum_{m=1}^M \Theta_m, \quad (19)$$

where  $\Theta_m$  denotes the average two-tier throughput of AP  $m$ . Mathematically,  $\Theta_m$  is derived as

$$\begin{aligned} \Theta_m &= \mathbb{E}_{N_m} [DN_m \Pr(DN_m \leq U_m) + U_m \Pr(DN_m > U_m)] \\ &= D\lambda_m Q(\lfloor \kappa \rfloor, \lambda_m) + U_m P(\lfloor \kappa \rfloor + 1, \lambda_m), \end{aligned} \quad (20)$$

where  $\kappa = \frac{U_m}{D}$ ,  $\lfloor \cdot \rfloor$  is the floor function,  $\Gamma(\cdot, \cdot)$  is upper incomplete Gamma function, and  $\Gamma(\cdot)$  is Gamma function, and  $Q(a, \lambda)$  and  $P(a, \lambda)$  are the regularized upper and regularized lower incomplete Gamma functions, defined as  $Q(a, \lambda) = \frac{\Gamma(a, \lambda)}{\Gamma(a)}$  and  $P(a, \lambda) = 1 - \frac{\Gamma(a, \lambda)}{\Gamma(a)}$ , respectively. Here, as  $Q(a, \lambda)$  refers to the CDF of a discrete Poisson variable  $N_m$  and  $U_m$  is continuous, the floor function is applied in (20) to ensure that the first parameter of  $Q(a, \lambda)$ , i.e.,  $a$ , is an integer. Furthermore, we note that the throughput  $\Theta_m$  in (20) is a pairwise linear function of the continuous variable  $U_m$ , since the floor function is involved.

We find from (20) that the average two-tier throughput of AP  $m$  is determined by both the mean of its traffic

load,  $D\lambda_m$ , and the maximum achievable throughput from AP  $m$  to the UAV,  $U_m$ . In particular, when the maximum achievable throughput overwhelms the mean of its traffic load, i.e.,  $\Pr(DN_m \leq U_m) \rightarrow 1$  and  $\Pr(DN_m > U_m) \rightarrow 0$ ,  $\Theta_m$  is dominated by  $D\lambda_m$ . Alternatively, when the considered system is traffic demanding, due to the massive IoT of the first tier, i.e.,  $\Pr(DN_m > U_m) \rightarrow 1$  and  $\Pr(DN_m \leq U_m) \rightarrow 0$ ,  $\Theta_m$  is mainly determined by  $U_m$ . Additionally, we prove that  $\Theta_m$  is a strictly monotonically increasing and concave function of  $U_m$ , and present it in the following lemma.

*Lemma 3:* The average two-tier throughput of AP  $m$ ,  $\Theta_m$ , is a strictly monotonically increasing and concave function w.r.t. the maximum achievable throughput from AP  $m$  to the UAV,  $U_m$ . This reveals that for a fixed mean of traffic load, i.e., fixed  $D\lambda_m$ , the benefit of increasing  $U_m$  is firstly profound and then vanishes with  $U_m \gg \lambda_m$ .

*Proof:* Based on (20), we obtain the derivative of  $\Theta_m$  w.r.t.  $U_m$  as

$$\begin{aligned} \frac{d\Theta_m}{dU_m} &= \frac{\lambda_m^{\lfloor \kappa \rfloor} \exp(-\lambda_m)}{(\lfloor \kappa \rfloor - 1)!} \sum_{k=1}^{\infty} \delta(\lfloor \kappa \rfloor - k) + P(\lfloor \kappa \rfloor + 1, \lambda_m) \\ &\quad - U_m \frac{\lambda_m^{\lfloor \kappa \rfloor} \exp(-\lambda_m)}{D(\lfloor \kappa \rfloor)!} \sum_{k=1}^{\infty} \delta(\lfloor \kappa \rfloor - k) \\ &= P(\lfloor \kappa \rfloor + 1, \lambda_m). \end{aligned} \quad (21)$$

Since  $Q(a, \lambda) > 0$  with an integer input  $a$  is the CDF of a Poisson RV, its derivative w.r.t.  $a$  equals to the PDF of a Poisson RV, which leads to the first step in (21). Moreover, since  $P(a, \lambda) > 0$  for all positive  $a$  and  $\lambda$ , we obtain  $\frac{d\Theta_m}{dU_m} > 0$ . This implies that  $\Theta_m$  is a strictly monotonically increasing function of  $U_m$ . In addition, we derive the second order derivative of  $\Theta_m$  w.r.t.  $U_m$  as

$$\begin{aligned} \frac{d^2\Theta_m}{dU_m^2} &= \frac{dP(\lfloor \kappa \rfloor + 1, \lambda_m)}{dU_m} \\ &= -\frac{\lambda_m^{\lfloor \kappa \rfloor} \exp(-\lambda_m)}{D(\lfloor \kappa \rfloor)!} \sum_{k=1}^{\infty} \delta(\lfloor \kappa \rfloor - k) < 0. \end{aligned} \quad (22)$$

This indicates that  $\Theta_m$  in (20) is a concave function of  $U_m$ , which completes the proof. ■

We next investigate the impact of the mean of traffic load,  $D\lambda_m$ , on the increasing slope of the average two-tier throughput w.r.t. maximum achievable throughput. Specifically, we take the derivative of  $P(\lfloor \kappa \rfloor + 1, \lambda_m)$  in (21) w.r.t.  $\lambda_m$ , which gives

$$\frac{dP(\lfloor \kappa \rfloor + 1, \lambda_m)}{d\lambda_m} = \frac{e^{-\lambda_m} \lambda_m^{\lfloor \kappa \rfloor}}{\Gamma(\lfloor \kappa \rfloor + 1)} > 0. \quad (23)$$

We see that when  $D\lambda_m$  becomes larger, the increasing slope of  $\Theta_m$  w.r.t.  $U_m$  becomes steeper. As such, for the AP with a larger mean of traffic load, increasing its maximum achievable throughput to the UAV is more beneficial to improve its average two-tier throughput, which can be achieved by navigating the UAV closer to and allocating more time slots to this AP.

Now we note the interesting trade-off for the performance between the first tier and the second tier when  $B = B_1 + B_2$  is given. On one hand, the larger  $B_1$  allows more subchannels in the first tier, which leads to an increasing packet detection

probability,  $P_{\text{Det}}[m]$ , and a higher mean of traffic load,  $D\lambda_m$ , which may in turn increase the average system throughput. On the other hand, a larger  $B_1$  results in a smaller  $B_2$  and hence less maximum achievable throughput is produced in the second tier, which may decrease the average system throughput. This interesting trade-off calls for an optimal bandwidth allocation between the two tiers to maximize the average system throughput.

#### IV. JOINT DESIGN OF UAV'S TRAJECTORY AND RESOURCE ALLOCATION

In this section, we first formulate the joint design of the UAV's trajectory and resource allocation to maximize the average system throughput. We then develop an iterative algorithm to produce a suboptimal solution to the formulated optimization problem.

Based on the analytical results derived in Section III, we formulate the joint design to maximize the average system throughput as

$$\begin{aligned} & \max_{\mathbf{q}[t], x_m[t], p_m[t], B_1, B_2} \sum_{m=1}^M \Theta_m \quad (24) \\ \text{s.t.} \quad & \text{C1: } x_m[t] \in \{0, 1\}, \forall m, t, \\ & \text{C2: } \sum_{m=1}^M x_m[t] \leq 1, \forall t, \\ & \text{C3: } p_m[t] \geq 0, \forall m, t, \\ & \text{C4: } p_m[t] \leq p_{\max}, \forall m, t, \\ & \text{C5: } \|\mathbf{q}[t] - \mathbf{q}[t-1]\| \leq \delta_t V_{\max}, t \in \{2, \dots, T\}, \\ & \text{C6: } \mathbf{q}[1] = \mathbf{q}[T], \\ & \text{C7: } \Theta_m(U_m) \geq \tau_m, \forall m. \\ & \text{C8: } B_1 + B_2 \leq B. \end{aligned}$$

In (24), Constraint C2 means that within one cruising period, there is at most one AP communicating with the UAV during each time slot to avoid the interference among APs. Constraint C4 guarantees that the transmit power of each AP during each time slot cannot exceed its maximum transmit power  $p_{\max}$ , which is in general limited by the power amplifier at the AP. Constraint C5 is imposed to limit the maximum replacement of the UAV with its maximum flying speed  $V_{\max}$ . Constraint C6 denotes that the initial location of the UAV is exactly the final location of the UAV, i.e., it is cruising in the considered area within a period of  $T$  time slots. In Constraint C7,  $\tau_m$  is a unique minimum required average two-tier throughput for AP  $m$  within the cruising period, which is introduced to guarantee the APs' heterogeneous quality of service (QoS) requirements. At the same time, introducing the minimum required average two-tier throughput for each AP implies that the UAV cannot keep visiting several same APs during the cruising period. Constraint C8 limits the total available bandwidth for bandwidth allocation between the two tiers. We note that the problem formulated in (24) is a mixed-integer non-convex optimization, such that its globally optimal solution is generally difficult to obtain. In particular, the non-convexity arises from the binary constraint in C1.

More importantly, the average two-tier throughput in (20) is a complicated non-convex function w.r.t. the optimization variables  $\mathbf{q}[t]$ ,  $x_m[t]$ ,  $p_m[t]$ ,  $B_1$ , and  $B_2$ , which is a major obstacle to solve the problem in (24).

It is known that the average two-tier throughput of AP  $m$ ,  $\Theta_m$ , in the objective function of (24) monotonically increases when  $U_m$  increases, as proved in Lemma 3, and the maximum achievable throughput  $U_m$  is an increasing function of the transmit power, according to (10). Thus, for any given value of other decision variables, the maximum transmit power is the optimal solution to maximize the average system throughput. It follows that the variable  $p_m[t]$  can be set as  $p_{\max}$ , i.e.,  $p_m[t] = p_{\max}, \forall m, t$ , in the following. Moreover, due to the complicated expression for  $P_{\text{Det}}[m]$  w.r.t.  $B_1 = L\Delta_f$ , as analyzed in (16), it is non-trivial to allocate the bandwidth between the two tiers through convex optimization. Fortunately, we observe that Constraint C8 holds with equality at the optimal point since the performance of the first tier and the performance of the second tier increases with  $B_1$  and  $B_2$ , respectively. Thus, the bandwidth allocation can be optimized via one-dimensional exhaustive search. Of course, we note that  $B_1$  can only take discrete values to ensure an integer number of subchannels  $L$ , due to  $B_1 = L\Delta_f$  with a fixed bandwidth  $\Delta_f$ . Additionally, we define a slack variable  $d_m[t] = (H^2 + \|\mathbf{q}[t] - \mathbf{w}_m\|^2)^{-1} > 0$  to facilitate our design of UAV's trajectory. Accordingly, given  $p_m[t] = p_{\max}$  and the optimal bandwidth allocation,  $B_1^*$  and  $B_2^* = B - B_1^*$ , we equivalently transform the formulated problem as

$$\max_{\mathbf{q}[t], x_m[t], d_m[t]} \sum_{m=1}^M \Theta_m(U_m(d_m[t], x_m[t] | p_{\max}, B_2^*) | B_1^*) \quad (25)$$

s.t. C1, C2, C5, C6,

$$\text{C7: } U_m(d_m[t], x_m[t] | p_{\max}, B_2^*) \geq \Theta_m^{-1}(\tau_m), \forall m,$$

$$\text{C9: } H^2 + \|\mathbf{q}[t] - \mathbf{w}_m\|^2 \leq \frac{1}{d_m[t]}, \forall m, t,$$

where  $U_m(d_m[t], x_m[t] | p_{\max}, B_2^*)$  in the objective function and Constraint C7 is obtained as

$$\begin{aligned} & U_m(d_m[t], x_m[t] | p_{\max}, B_2^*) \\ & = \delta_t B_2^* \sum_{t=1}^T x_m[t] \log_2 \left( 1 + \frac{p_{\max} \beta_0^2 d_m[t]}{B_2^* N_0 x_m[t]} \right). \quad (26) \end{aligned}$$

Recall that  $\Theta_m$  is a strictly monotonically increasing function of  $U_m$ . In other words, there is a one-to-one mapping between  $\Theta_m$  and  $U_m$  and thus, Constraint C7 can be equivalently rewritten as  $U_m \geq \Theta_m^{-1}(\tau_m), \forall m$ . Furthermore, we observe that  $U_m$  in (26) is a concave function w.r.t.  $d_m[t]$  and  $x_m[t]$ . Hence, the objective function is a concave function w.r.t.  $d_m[t]$  and  $x_m[t]$  since  $\frac{\partial \Theta_m}{\partial U_m} > 0$  and  $\frac{\partial^2 \Theta_m}{\partial U_m^2} \leq 0$  [38].

Now the major obstacle to solve (25) is the binary Constraint C1 spanning a disjoint feasible solution set. To overcome this obstacle, we rewrite Constraint C1 as two constraints, given by

$$\text{C1a: } x_m[t] \in [0, 1], \forall m, t, \quad (27)$$

$$\text{C1b: } \sum_{m=1}^M \sum_{t=1}^T x_m[t] - \sum_{m=1}^M \sum_{t=1}^T (x_m[t])^2 \leq 0. \quad (28)$$

Although  $x_m[t]$  is relaxed to be real between zero and one in C1a, the additional constraint C1b guarantees that  $x_m[t]$  must be zero or one, which is commonly adopted for user scheduling design in the literature [39, 40].

With the aforementioned manipulation, the non-convexity of the problem in (25) now arises from Constraints C1b and C9 only. To address this, we employ an iterative algorithm based on the SCA technique [39, 41] to achieve a suboptimal solution for the joint design of UAV's trajectory and resource allocation. In particular, given a feasible solution,  $d_m^{\text{iter}}[t]$  and  $x_m^{\text{iter}}[t]$  in the  $\text{iter}$ -th iteration, we have

$$\max_{\mathbf{q}[t], x_m[t], d_m[t]} \sum_{m=1}^M \Theta_m (U_m(d_m[t], x_m[t] | p_{\max}, B_2^*) | B_1^*) \quad (29)$$

s.t. C1a, C2, C5, C6, C7,

$$\begin{aligned} \widetilde{\text{C1b}}: & \sum_{m=1}^M \sum_{t=1}^T x_m[t] - \sum_{m=1}^M \sum_{t=1}^T (x_m^{\text{iter}}[t])^2 \\ & + 2 \sum_{m=1}^M \sum_{t=1}^T x_m^{\text{iter}}[t] (x_m[t] - x_m^{\text{iter}}[t]) \leq 0, \\ \widetilde{\text{C9}}: & H^2 + \|\mathbf{q}[t] - \mathbf{w}_m\|^2 - \frac{1}{d_m^{\text{iter}}[t]} \\ & + \frac{d_m[t] - d_m^{\text{iter}}[t]}{(d_m^{\text{iter}}[t])^2} \leq 0, \forall m, t, \end{aligned}$$

where constraints  $\widetilde{\text{C1b}}$  and  $\widetilde{\text{C9}}$  are obtained via the first order Taylor expansion on their non-convex terms in their left hand side functions. As constraints  $\widetilde{\text{C1b}}$  and  $\widetilde{\text{C9}}$  span a smaller feasible solution set, compared to that of C1b and C9, the problem in (29) provides a lower bound on the problem in (25).

We find that the transformed problem in (29) is a convex optimization problem given  $d_m^{\text{iter}}[t]$  and  $x_m^{\text{iter}}[t]$ , which can then be solved efficiently by standard convex problem solvers, such as CVX [42]. Therefore, based on the SCA technique [39, 41], we propose an iterative algorithm to tighten the lower bound obtained from solving (29), which is summarized in **Algorithm 1**. In this algorithm, the initial cruising trajectory  $\mathbf{q}^1[t]$  is set as a circular trajectory in the considered serving area as shown in Fig. 8 in Section V and thus,  $d_m^1[t] = (H^2 + \|\mathbf{q}^1[t] - \mathbf{w}_m\|^2)^{-1}$ ,  $\forall m, t$ . Moreover, the AP scheduling variables are initialized by selecting the AP closest to the UAV during each time slot, i.e.,  $x_m^1[t] = \arg\max_m d_m^1[t]$ ,  $\forall t$ . In the  $\text{iter}$ -th iteration, an intermediate solution  $(\mathbf{q}^{\text{iter}+1}[t], d_m^{\text{iter}+1}[t], x_m^{\text{iter}+1}[t])$  is obtained by solving the problem in (29) with given  $(\mathbf{q}^{\text{iter}}[t], d_m^{\text{iter}}[t], x_m^{\text{iter}}[t])$ . This solution is then used for updating the problem in (29) for the next iteration. The algorithm terminates when the maximum iteration number is reached, i.e.,  $\text{iter} = \text{iter}_{\max}$ , or the change of optimization variables between adjacent iterations becomes smaller than a given convergence tolerance  $\epsilon$ . As proved in [43], the proposed iterative trajectory and resource allocation

### Algorithm 1 Iterative Trajectory and Resource Allocation Algorithm

- 1: **Initialization**  
Initialize the convergence tolerance  $\epsilon$ , the maximum number of iterations  $\text{iter}_{\max}$ , the iteration index  $\text{iter} = 1$ , and the initial feasible solution  $(\mathbf{q}^{\text{iter}}[t], d_m^{\text{iter}}[t], x_m^{\text{iter}}[t])$ .
- 2: **repeat**
- 3: Solve (29) for a given  $(\mathbf{q}^{\text{iter}}[t], d_m^{\text{iter}}[t], x_m^{\text{iter}}[t])$  to obtain an intermediate solution  $(\mathbf{q}^{\text{iter}+1}[t], d_m^{\text{iter}+1}[t], x_m^{\text{iter}+1}[t])$ .
- 4: Set  $\text{iter} = \text{iter} + 1$ .
- 5: **until**  $\text{iter} = \text{iter}_{\max}$  or  $\sum_{m=1}^M \sum_{t=1}^T (|x_m^{\text{iter}+1}[t] - x_m^{\text{iter}}[t]| + |d_m^{\text{iter}+1}[t] - d_m^{\text{iter}}[t]|) \leq \epsilon$ .
- 6: Return the solution  $(d_m^* [t], x_m^* [t], \mathbf{q}^* [t]) = (d_m^{\text{iter}} [t], x_m^{\text{iter}} [t], \mathbf{q}^{\text{iter}} [t])$ .

TABLE II  
PARAMETER VALUES FOR SIMULATIONS [14, 24, 44].

Notations	Simulation value	Notations	Simulation value
$R$	350 m	$N_0$	-169 dBm/Hz
$M$	10	$T$	30-60
$R_s$	50 m	$\tau_m$	100T bits
$D$	10 bits	$N_0$	-169 dBm/Hz
$p_{\max}$	15 dBm		

algorithm is guaranteed to converge to a stationary point with a polynomial time computational complexity.

## V. NUMERICAL RESULTS

In this section, we present the numerical results to verify our analysis and to evaluate the performance of our proposed design. We consider a UAV-enabled massive IoT system deployed in a rural environment, where little blockage or scattering obstacles exist. In the simulations, we set the UAV's height as  $H = 100$  m, the channel power gain at reference distance 1 m as  $\beta_0 = -50$  dBW, and the maximum flying speed as  $V_{\max} = 30$  m/s [35, 45, 46]. The time slot length is  $\delta_t = 1$  s and each time slot is further divided into  $N = 100$  subframes [35, 47]. We assume that the subchannel bandwidth is  $\Delta_f = 1$  kHz and the target rate of each IoT device is  $R_0 = 1$  kbps, thereby leading to the spectral efficiency of 1 bit/s/Hz as that in [48]. Unless otherwise stated, the parameter values are given in Table II. To demonstrate the performance gain of our proposed design, we consider three baseline schemes (BaSs) for comparison, described as follows:

- BaS 1: One-tier system with a fixed ground BS located in the center of the considered area. All the IoT devices directly communicate with the ground BS through a multi-channel ALOHA-based random access scheme using the total bandwidth  $B$  and the total time duration  $T\delta_t$ .
- BaS 2: One-tier system with a mobile UAV BS. All the IoT devices directly communicate with the UAV through a multi-channel ALOHA-based random access scheme using the total bandwidth  $B$  and the total time duration  $T\delta_t$ . In this system, a circular UAV's cruising trajectory is adopted.
- BaS 3: Our proposed two-tier system but with a circular UAV's cruising trajectory.

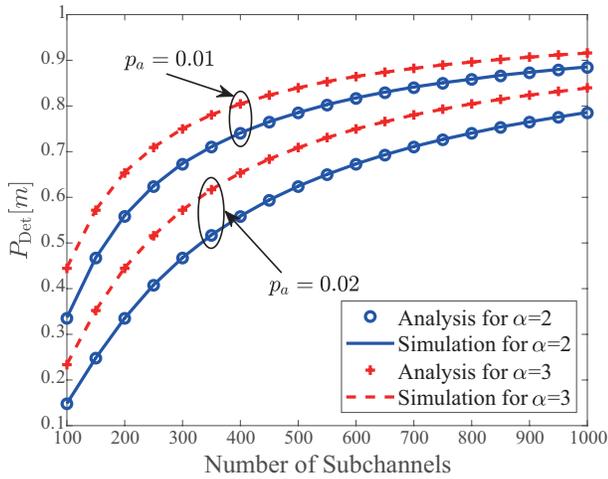


Fig. 4. Analysis and simulation for the packet detection probability at  $P_{\text{Det}}[m]$ .

### A. Detection Probability

Fig. 4 plots the packet detection probability,  $P_{\text{Det}}[m]$ , versus the number of available subchannels,  $L$ , in the first tier. From this figure, we first observe that our derived analytical results tightly match the simulation results, which demonstrates the accuracy of our performance analysis. Second, we observe that  $P_{\text{Det}}[m]$  increases with  $L$ , particularly in the low to medium regime of  $L$ . This is because that when there are a low number of available subchannels, multiple IoT devices have a very high chance to select the same subchannel. Thus, the intra-cluster and inter-cluster interferences are severe when  $L$  is in the low to medium regime, which implies that  $L$  profoundly affects  $P_{\text{Det}}[m]$ . Third, we observe that a larger path-loss exponent  $\alpha$  leads to an increase in  $P_{\text{Det}}[m]$ , due to the reduced intra-cluster and inter-cluster interference caused by a larger  $\alpha$ . Fourth, we observe that when the active probability  $p_a$  increases,  $P_{\text{Det}}[m]$  decreases. This is due to the fact that for a larger  $p_a$ , more IoT devices become active, which introduces more severe interference.

### B. Average System Throughput

Fig. 5 plots the average system throughput versus the density of IoT devices in each cluster,  $\lambda_d$ . First, we observe that when  $\lambda_d$  increases, the average system throughput of the two-tier systems, i.e., our design and BaS 3, increases while that of the one-tier systems, i.e., BaS 1 and BaS 2, decreases. This observation is not surprising, since for the one-tier systems, a higher density of IoT devices causes more severe interference. For the two-tier systems, introducing APs between IoT devices and the UAV effectively reduces the potential congestion on the same subchannel in the first tier. Although increasing the density of IoT devices generates more severe interference at the AP within each cluster, our proposed bandwidth allocation between two tiers allocates more subchannels to the first tier for accommodating additional IoT devices. Second, we observe that our design achieves a much higher average system throughput than BaS 3, and BaS 1 achieves a higher average system throughput than BaS 2. This is because that

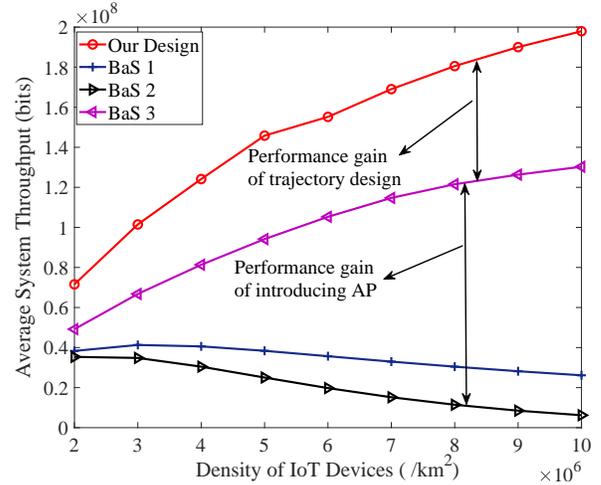


Fig. 5. Average system throughput (bits) versus the density of IoT devices in each cluster for  $B = 1$  MHz and  $T = 50$ .

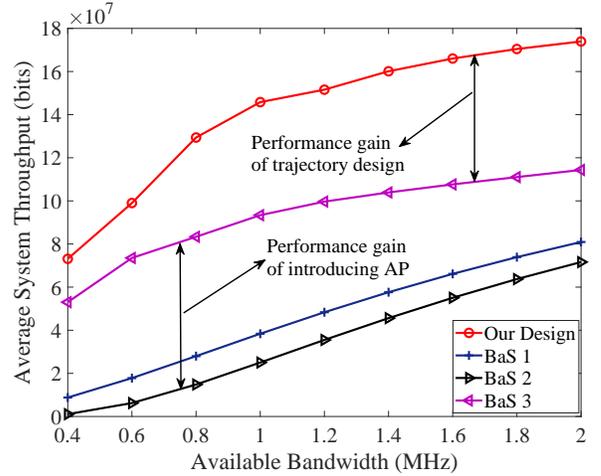


Fig. 6. Average system throughput (bits) versus total available bandwidth for  $\lambda_d = 5 \times 10^6$  /km<sup>2</sup> and  $T = 50$ .

for the two-tier system, our proposed design exploits the heterogeneous traffic load distribution among different clusters (or equivalently, APs) through designing the cruising trajectory of the UAV, while BaS 3 does not. For the one-tier system, each IoT device enjoys a LoS channel to the mobile UAV in the sky in BaS 2; therefore, BaS 2 suffers from much stronger interference than BaS 1 when the density of IoT devices increases.

Fig. 6 plots the average system throughput versus the total available bandwidth,  $B$ . We first observe that for our design and three BaSs, the larger  $B$  leads to the higher average system throughput. Second, we observe that BaS 3 achieves a substantial throughput gain relative to BaS 1 and BaS 2, especially when the total bandwidth is limited. This is because that in the one-tier systems, congestion occurs very frequently due to the limited number of subchannels. In the two-tier systems, introducing APs between IoT devices and the UAV significantly liberates the UAV from the congestion and yields a higher average system throughput. Third, we observe that for the two-tier systems, the performance gain of our

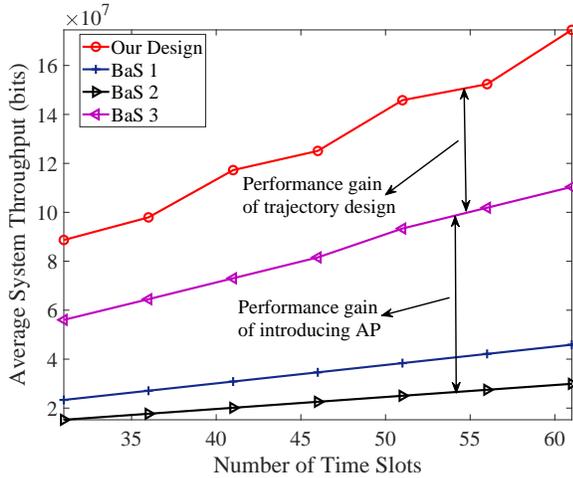


Fig. 7. Average system throughput (bits) versus the number of time slots within one cruising period for  $\lambda_d = 5 \times 10^6$  /km<sup>2</sup> and  $B = 1$  MHz.

proposed design over BaS 3 increases with larger  $B$ . This is because our design of UAV’s trajectory provides a higher channel power gain compared to BaS 3, which enables a more efficient exploitation of the spectrum resource. Therefore, more maximum achievable throughput can be achieved in the second tier, which improves the average system throughput. In addition to this reason, since a smaller bandwidth is required due to the better channel condition in the second tier, there are more available bandwidth saved for the multi-channel ALOHA in the first tier. As a result, a higher packet detection probability can be achieved at APs, which is also beneficial to improve the average system throughput. Fourth, the increasing slope w.r.t. the total bandwidth for the two-tier systems, i.e., our proposed design and BaS 3, is first higher and then lower, while that of the one-tier systems, i.e., BaS 1 and BaS 2, is almost constant. This is because that the traffic congestion in the two-tier systems is much less than that in the one-tier systems, due to the deployment of APs. Thus, when the total bandwidth increases from small to medium, this increase efficiently deals with the relatively low traffic congestion in the two-tier systems, leading to a profound improvement in the average system throughput. When the total bandwidth becomes large, the performance of the first tier is already good such that the increasing bandwidth only brings a limited improvement in the average system throughput.

Fig. 7 plots the average system throughput versus the number of time slots within one cruising period,  $T$ . First, we observe that for our design and three BaSs, the increasing  $T$  improves the average system throughput, since the UAV or ground BS can collect more data when there are more time slots. Second, we observe that our design achieves the best performance compared to other BaSs, which demonstrates the effectiveness of introducing APs and our proposed design of UAV’s trajectory and resource allocation. Third, we observe that the increasing slope of our design w.r.t.  $T$  becomes slightly higher when  $T$  increases. This shows the benefit of the additional degrees of freedom provided by the UAV’s trajectory design. Fourth, we observe that the performance of

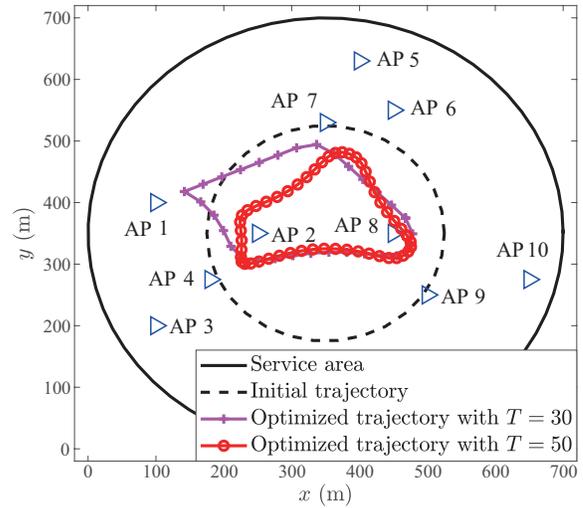


Fig. 8. Geometry distribution of APs and the trajectory of UAV.

BaS 1, BaS 2, and BaS 3 almost linearly increases with  $T$ , since the system is static with increasing  $T$ . Fifth, we observe that the increasing slope of BaS 3 is much higher than those of BaS 1 and BaS 2, which demonstrates the performance gain of introducing APs.

### C. UAV’s trajectory

Fig. 8 plots the optimized trajectories with  $T = 30$  and  $T = 50$  revealed by our proposed design to examine the impact of  $T$  on the UAV’s trajectory. When  $T = 30$ , the number of time slots is relatively limited. Thus, the UAV’s trajectory design is stringent and it is difficult to provide a high maximum achievable throughput in the second tier. It follows that the UAV prefers to fly close to the AP having a higher traffic load demand, in order to achieve a higher average system throughput, such as AP 1 in Fig. 8. When  $T$  increases from 30 to 50, the UAV’s trajectory design becomes more flexible to provide a sufficiently large maximum achievable throughput in the second tier. As proved in Lemma 3, the average two-tier throughput of AP  $m$ ,  $\Theta_m$ , is a concave function w.r.t. its maximum achievable throughput to the UAV,  $U_m$ . Therefore, further increasing  $U_m$  for a sufficiently large  $U_m$  does not bring a significant gain in the average two-tier throughput of AP  $m$ . It can be seen that the UAV with  $T = 50$  prefers to cruise to each set of APs, rather than to the specific AP with a high traffic load, as shown for the case with  $T = 30$ .

## VI. CONCLUSIONS

This paper proposed a two-tier UAV communication strategy where multiple APs are introduced to help a UAV for data collection from massive ground IoT devices. In such strategy, a multi-channel ALOHA-based random access scheme was adopted in the first tier to accommodate the massive IoT devices within each cluster, while APs were required to transmit their aggregated data to the UAV through TDMA in the second tier. The packet detection probability and the traffic load distribution at each AP were analyzed to evaluate the performance of the first tier transmission. Moreover, the

average system throughput was derived for the proposed strategy based on the maximum achievable throughput in the second tier transmission. We found that the average two-tier throughput of each AP monotonically increases with its maximum achievable throughput in the second tier, while the increasing slope becomes steeper with a higher traffic load mean in the first tier. Furthermore, the joint design of UAV's trajectory and resource allocation was formulated as a non-convex optimization problem and a low complexity iterative algorithm was developed to achieve a suboptimal solution to the formulated joint design problem. With extensive numerical results, we demonstrated the accuracy of our proposed analysis and identified the substantial performance gain of introducing APs and our design in the context of massive access, compared to the baseline schemes.

#### APPENDIX A PROOF OF LEMMA 1

According to the independent thinning property of stochastic geometry [49], for each cluster the location of active IoT devices can be modeled as a homogeneous PPP with the density  $p_a \lambda_d$ . Moreover, since the active IoT devices select the subchannels randomly and uniformly, the location of active IoT devices which select the same subchannel can be modeled as a homogeneous PPP with the density  $\frac{p_a \lambda_d}{L}$ , denoted by  $\Phi_m$  for cluster  $m$ . Thus, we exploit the probability generating functional (PGFL) of homogeneous PPP and derive the Laplace transform of  $I_{\text{intra}}$  as

$$\begin{aligned} & \mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right] \\ &= \mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp \left( -\eta r_{k,m}^\alpha \sum_{k' \in \Phi_m \setminus \{k\}} |h_{k',m}|^2 r_{k',m}^{-\alpha} \right) \right] \\ &= \mathbb{E}_{|h_{k',m}|^2, \Phi_m} \left[ \prod_{k' \in \Phi_m} \exp \left( -\eta |h_{k',m}|^2 r_{k,m}^\alpha r_{k',m}^{-\alpha} \right) \right]. \end{aligned} \quad (30)$$

By defining  $V = |h_{k',m}|^2$  which has the PDF given by  $f_V(v) = \exp(-v)$ , we express the expectation in the last line of (30) as

$$\exp \left[ -\frac{p_a \lambda_d}{L} \int_0^\infty \ell f_V(v) dv \right], \quad (31)$$

where  $\ell$  is given by and derived as

$$\begin{aligned} \ell &= \int_0^{2\pi} \int_0^{R_s} \left( 1 - \exp \left( -\eta v r_{k,m}^\alpha r_{k',m}^{-\alpha} \right) \right) r_{k',m} dr_{k',m} d\theta \\ &= \pi R_s^2 \left( 1 - \exp(-\eta v r_{k,m}^\alpha R_s^{-\alpha}) \right) \\ &\quad + \pi \eta^{\frac{2}{\alpha}} r_{k,m}^2 v^{\frac{2}{\alpha}} \Gamma \left( 1 - \frac{2}{\alpha}, \eta r_{k,m}^\alpha R_s^{-\alpha} v \right). \end{aligned} \quad (32)$$

where  $\zeta_{k,m}$  is given below (12). Substituting (32) into (30) and solving the resultant integral, we obtain the Laplace transform of  $I_{\text{intra}}$  as

$$\begin{aligned} & \mathbb{E}_{I_{k,m}^{\text{intra}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{intra}}) \right] \\ &= \exp \left[ -\frac{p_a \lambda_d}{L} \pi \left( R_s^2 - \frac{R_s^2}{\eta r_{k,m}^\alpha R_s^{-\alpha} + 1} \right) \right] \end{aligned}$$

$$\begin{aligned} & + \frac{\eta^{\frac{2}{\alpha}} r_{k,m}^2 (\eta r_{k,m}^\alpha R_s^{-\alpha})^{1-\frac{2}{\alpha}}}{\left( 1 + \frac{2}{\alpha} \right) (\eta r_{k,m}^\alpha R_s^{-\alpha} + 1)^2} \\ & \times {}_2F_1 \left( 1, 2; \frac{2}{\alpha} + 2; \frac{1}{\eta r_{k,m}^\alpha R_s^{-\alpha} + 1} \right) \Bigg]. \end{aligned} \quad (33)$$

By rearranging (33), we obtain (12), which completes the proof.

#### APPENDIX B PROOF OF LEMMA 2

In the considered system, all IoT devices are spatially distributed in a multi-cluster form. For each cluster, the location of the IoT devices are assumed to follow an i.i.d. homogeneous PPP, which is centered at the serving AP within a finite area. Then, the distance from IoT device  $k''$  of cluster  $m'$ ,  $m' \in \{1, \dots, M\} \setminus \{m\}$ , to the typical AP  $m$  is given by  $r_{k'',m} = \sqrt{\varpi}$ , where  $\varpi$  is given below (13). Moreover, within cluster  $m'$ , the location of interfering IoT devices for the typical IoT device is modeled as a homogeneous PPP with the density  $\frac{p_a \lambda_d}{L}$ , denoted by  $\Phi_{m'}$ . Thus, according to the PGFL of the homogeneous PPP, the Laplace transform of  $I_{k,m}^{\text{inter}}$  can be derived as

$$\begin{aligned} & \mathbb{E}_{I_{k,m}^{\text{inter}}} \left[ \exp(-\zeta_{k,m} I_{k,m}^{\text{inter}}) \right] \\ &= \mathbb{E}_{I_{k,m}^{\text{inter}}} \left[ \exp \left( -\eta r_{k,m}^\alpha \sum_{k'' \in \Phi_{m'}, m' \in \{1, \dots, M\} \setminus \{m\}} |h_{k'',m}|^2 r_{k'',m}^{-\alpha} \right) \right] \\ &= \prod_{m' \in \{1, \dots, M\} \setminus \{m\}} \mathbb{E}_{|h_{k'',m}|^2, \Phi_{m'}} \left[ \prod_{k'' \in \Phi_{m'}} \exp \left( -\eta r_{k,m}^\alpha |h_{k'',m}|^2 r_{k'',m}^{-\alpha} \right) \right] \end{aligned} \quad (34)$$

By defining  $\tilde{V} = |h_{k'',m}|^2$  which has the PDF given by  $f_{\tilde{V}}(\tilde{v}) = \exp(-\tilde{v})$ , we express the expectation in the last line of (34) as

$$\exp \left[ -\frac{p_a \lambda_d}{L} \int_0^{R_s} r_{k'',m'} \int_0^{2\pi} \tilde{\ell} d\theta dr_{k'',m'} \right], \quad (35)$$

where  $\tilde{\ell}$  is given by and derived as

$$\tilde{\ell} = \int_0^\infty \left( 1 - \exp \left( \frac{-\eta r_{k,m}^\alpha \tilde{v}}{\varpi^{\frac{\alpha}{2}}} \right) \right) f_{\tilde{V}}(\tilde{v}) d\tilde{v} = \frac{\eta r_{k,m}^\alpha}{\eta r_{k,m}^\alpha + \varpi^{\frac{\alpha}{2}}}. \quad (36)$$

Finally, we substitute (36) into (35) and then into (34) to obtain the Laplace transform of  $I_{\text{inter}}$  as (13), thus completing the proof.

#### REFERENCES

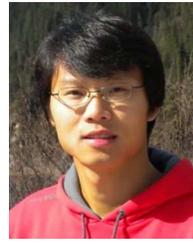
- [1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, May 2020.
- [2] J. Bradley, C. Reberger, A. Dixit, and V. Gupta, "The Internet of Everything global private sector economic analysis," CISCO, Tech. Rep., 2013.
- [3] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Select. Areas Commun.*, 2020, [Online]: <https://arxiv.org/pdf/2002.03491.pdf>.
- [4] A. T. Abebe and C. G. Kang, "Comprehensive grant-free random access for massive & low latency communication," in *Proc. IEEE Intern. Commun. Conf.*, 2017, pp. 1–6.

- [5] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, Sep. 2018.
- [6] Z. Sun, Z. Wei, L. Yang, J. Yuan, X. Cheng, and L. Wan, "Exploiting transmission control for joint user identification and channel estimation in massive connectivity," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6311–6326, Sep. 2019.
- [7] X. Chen, Z. Zhang, C. Zhong, and D. W. K. Ng, "Exploiting multiple-antenna techniques for non-orthogonal multiple access," *IEEE J. Select. Areas Commun.*, vol. 35, no. 10, pp. 2207–2220, Oct. 2017.
- [8] X. Chen, Z. Zhang, C. Zhong, R. Jia, and D. W. K. Ng, "Fully non-orthogonal communication for massive access," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1717–1731, Apr. 2018.
- [9] M. Dong, K. Ota, M. Lin, Z. Tang, S. Du, and H. Zhu, "UAV-assisted data gathering in wireless sensor networks," *J. Supercomput.*, vol. 70, no. 3, pp. 1142–1155, Dec. 2014.
- [10] N. H. Motlagh, T. Taleb, and O. Arouk, "Low-altitude unmanned aerial vehicles-based Internet of Things services: Comprehensive survey and future perspectives," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 899–922, Dec. 2016.
- [11] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *Proc. of the IEEE*, vol. 107, no. 12, pp. 2327–2375, Dec. 2019.
- [12] Y. Sun, D. Xu, D. W. K. Ng, L. Dai, and R. Schober, "Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4281–4298, Jun. 2019.
- [13] J. Gong, T. Chang, C. Shen, and X. Chen, "Flight time minimization of UAV for data collection over wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 36, no. 9, pp. 1942–1954, Sep. 2018.
- [14] C. Zhan, Y. Zeng, and R. Zhang, "Energy-efficient data collection in UAV enabled wireless sensor network," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 328–331, Jun. 2018.
- [15] D. Yang, Q. Wu, Y. Zeng, and R. Zhang, "Energy tradeoff in ground-to-UAV communication via trajectory design," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6721–6726, Jul. 2018.
- [16] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.
- [17] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive Non-Orthogonal Multiple Access for cellular IoT: Potentials and Limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.
- [18] L. Liu and W. Yu, "Massive connectivity with massive MIMO-part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.
- [19] E. Paolini, G. Liva, and M. Chiani, "Coded Slotted ALOHA: A graph-based method for uncoordinated multiple access," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6815–6832, Dec. 2015.
- [20] Z. Sun, Y. Xie, J. Yuan, and T. Yang, "Coded slotted ALOHA for erasure channels: Design and throughput analysis," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4817–4830, Nov. 2017.
- [21] Z. Zhao, S. S. Szyszkwicz, T. Beitelmal, and H. Yanikomeroglu, "Spatial clustering in slotted ALOHA two-hop random access for machine type communication," in *Proc. IEEE Global Commun. Conf.*, Dec. 2016, pp. 1–6.
- [22] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, Feb. 2017.
- [23] D. Malak, H. S. Dhillon, and J. G. Andrews, "Optimizing data aggregation for uplink machine-to-machine communication networks," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1274–1290, Mar. 2016.
- [24] J. Guo, S. Durrani, X. Zhou, and H. Yanikomeroglu, "Massive machine type communication with data aggregation and resource scheduling," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 4012–4026, Sep. 2017.
- [25] C. Li, M. Ye, G. Chen, and J. Wu, "An energy-efficient unequal clustering mechanism for wireless sensor networks," in *IEEE Intern. Conf. on Mobile Adhoc and Sensor Systems*, 2005, pp. 8–604.
- [26] Z. Zhou, S. Zhou, S. Cui, and J. Cui, "Energy-efficient cooperative communication in a clustered wireless sensor network," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3618–3628, Nov. 2008.
- [27] L. Q. Q. Zhu, and M. Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks," *Comput. Commun.*, vol. 29, no. 12, pp. 2230–2237, Aug. 2006.
- [28] X. Chen, R. Jia, and D. W. K. Ng, "The application of relay to massive non-orthogonal multiple access," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5168–5180, Nov. 2018.
- [29] Y. Zeng, Q. Wu, and R. Zhang, "Accessing from the sky: A tutorial on UAV communications for 5G and beyond," *arXiv preprint arXiv:1903.05289*, 2019.
- [30] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts. Mag.*, vol. 19, no. 3, pp. 1527–1550, Third Quarter 2017.
- [31] Z. Zhang and Y. J. Liu, "Multichannel ALOHA data networks for personal communications services (PCS)," in *Proc. IEEE Global Commun. Conf.*, Dec. 1992, pp. 21–25.
- [32] Z. Sun, Y. Xie, J. Yuan, and T. Yang, "Coded slotted ALOHA for erasure channels: Design and throughput analysis," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4817–4830, Nov. 2017.
- [33] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
- [34] X. Lin, V. Jainanarayana, S. D. Muruganathan, S. Gao, H. Asplund, H. Maattanen, M. Bergstrom, S. Euler, and Y. . E. Wang, "The sky is not the limit: LTE for unmanned aerial vehicles," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 204–210, Apr. 2018.
- [35] C. Zhan and Y. Zeng, "Completion time minimization for multi-UAV-enabled data collection," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4859–4872, Oct. 2019.
- [36] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Francisco, CA: Academic Press, 2007.
- [37] G. E. Andrews, R. Askey, and R. Roy, *Special Functions (Encyclopedia of Mathematics and its Applications)*, 1st ed. Cambridge, UK: Cambridge Univ. Press, 2001.
- [38] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [39] Z. Wei, D. W. K. Ng, J. Yuan, and H. M. Wang, "Optimal resource allocation for power-efficient MC-NOMA with imperfect channel state information," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3944–3961, Sep. 2017.
- [40] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1077–1091, Mar. 2017.
- [41] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "Multi-beam NOMA for hybrid mmwave systems," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1705–1719, Feb. 2019.
- [42] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <http://cvxr.com/cvx>, Mar. 2014.
- [43] N. Vucic, S. Shi, and M. Schubert, "DC programming approach for resource allocation in wireless networks," in *Proc. Int. Symp. Model. Optim. Mobile Ad Hoc Wireless Netw.*, May 2010, pp. 380–386.
- [44] "Evolved universal terrestrial radio access: Further advancements for E-UTRA physical layer aspects," 3GPP TR 36.814, Tech. Rep., 2010.
- [45] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329–2345, Apr. 2019.
- [46] J. Zhang, Y. Zeng, and R. Zhang, "UAV-enabled radio access network: Multi-mode communication and trajectory design," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5269–5284, Oct. 2018.
- [47] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (UAVs) for energy-efficient internet of things communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7574–7589, Nov. 2017.
- [48] Y. . E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A primer on 3GPP narrowband internet of things," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 117–123, Mar. 2017.
- [49] M. Haenggi, *Stochastic Geometry for Wireless Networks*, 1st ed. Cambridge, UK: Cambridge Univ. Press, 2012.



**Zhuo Sun** (S'15–M'19) received the B.E. and M.E. degrees from the Northwestern Polytechnical University (NPU), Xi'an, China, in 2012 and 2015, respectively, and the Ph.D. degree in electrical engineering and telecommunications from the University of New South Wales, Sydney, Australia, in 2019. She is currently a Post-Doctoral Research Fellow with the Research School of Electrical, Energy and Materials Engineering at the Australian National University. Her current research interests include massive connectivity for machine-to-machine communications, grant-free multiple access, and statistical signal processing.

communications, grant-free multiple access, and statistical signal processing.



**Xiangyun Zhou** (SM'17) is an Associate Professor at the Australian National University (ANU). He received the Ph.D. degree from ANU in 2010. His research interests are in the fields of communication theory and wireless networks. He has served as an Editor for various IEEE journals, including IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS and IEEE COMMUNICATIONS LETTERS. He served as a guest editor for IEEE COMMUNICATIONS MAGAZINE's feature

topic on wireless physical layer security in 2015. He also served as symposium/track and workshop co-chairs for major IEEE conferences. He was the chair of the ACT Chapter of the IEEE Communications Society and Signal Processing Society from 2013 to 2014. He is a recipient of the Best Paper Award at ICC'11 and IEEE ComSoc Asia-Pacific Outstanding Paper Award in 2016. He was named the Best Young Researcher in the Asia-Pacific Region in 2017 by IEEE ComSoc Asia-Pacific Board.



**Zhiqiang Wei** (S'16–M'19) received the B.E. degree in information engineering from Northwestern Polytechnical University (NPU), Xian, China, in 2012, and the Ph.D. degree in electrical engineering and telecommunications from the University of New South Wales, Sydney, Australia, in 2019. His current research interests include statistics and array signal processing, non-orthogonal multiple access, millimeter wave communications, and resource allocation design. He is currently a Postdoctoral Research Fellow with the University of New South Wales. He

received the Best Paper Award from the IEEE International Conference on Communications (ICC), in 2018.



**Nan Yang** (S'09–M'11–SM'18) received the B.S. degree in electronics from China Agricultural University in 2005, and the M.S. and Ph.D. degrees in electronic engineering from the Beijing Institute of Technology in 2007 and 2011, respectively. He has been with the Research School of Electrical, Energy and Materials Engineering at the Australian National University since July 2014, where he currently works as a Senior Lecturer. Prior to this, he was a Postdoctoral Research Fellow at the University of New South Wales (2012–2014) and a

Postdoctoral Research Fellow at the Commonwealth Scientific and Industrial Research Organization (2010–2012). He received the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2014 and the Best Paper Awards from the IEEE GlobeCOM 2016 and the IEEE VTC 2013-Spring. He also received the Top Editor Award from the Transactions on Emerging Telecommunications Technologies, the Exemplary Reviewer Awards from the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, and IEEE COMMUNICATIONS LETTERS, and the Top Reviewer Award from the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2012 to 2019. He is currently serving in the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES. His research interests include millimeter wave and terahertz communications, ultra-reliable low latency communications, cyber-physical security, massive multi-antenna systems, and molecular communications.