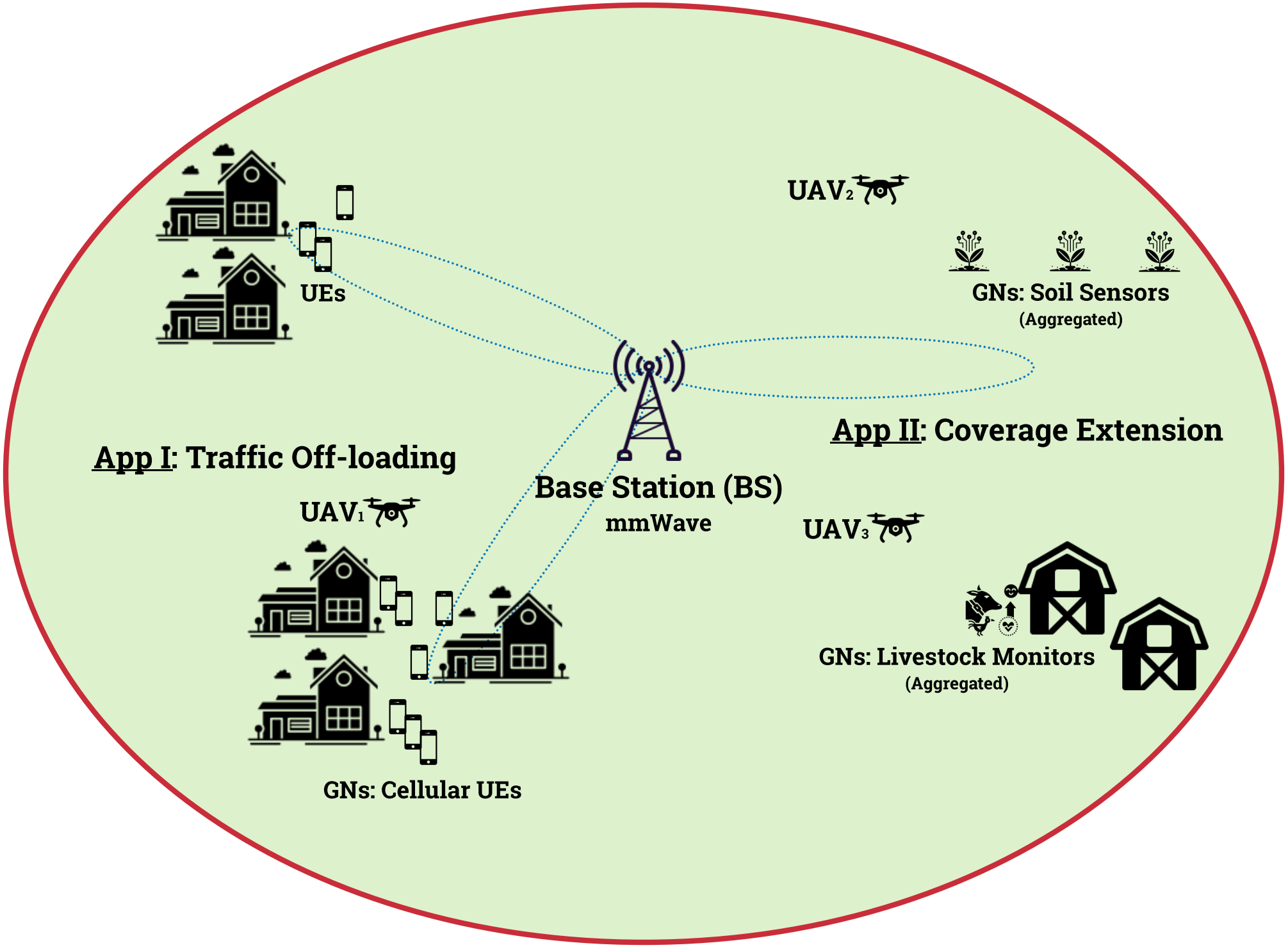
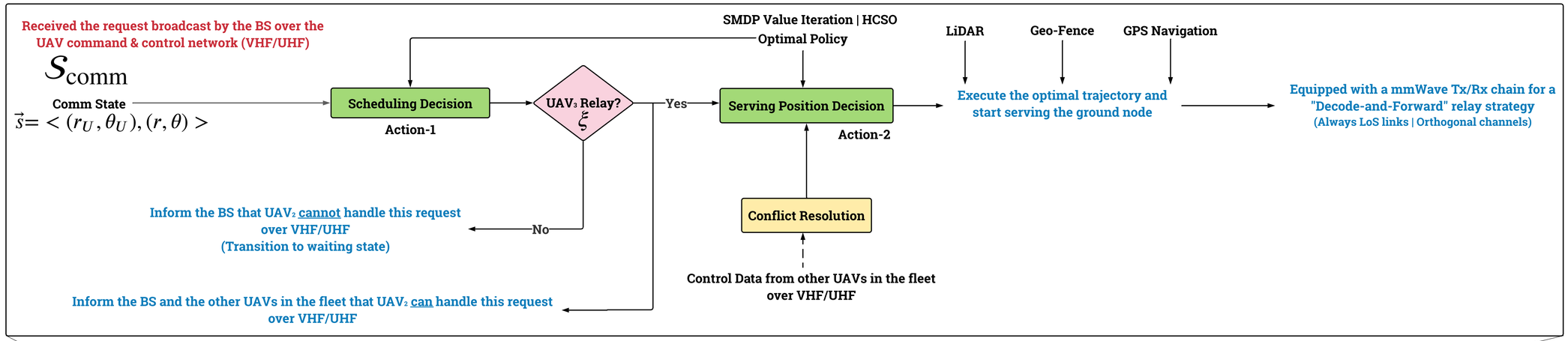
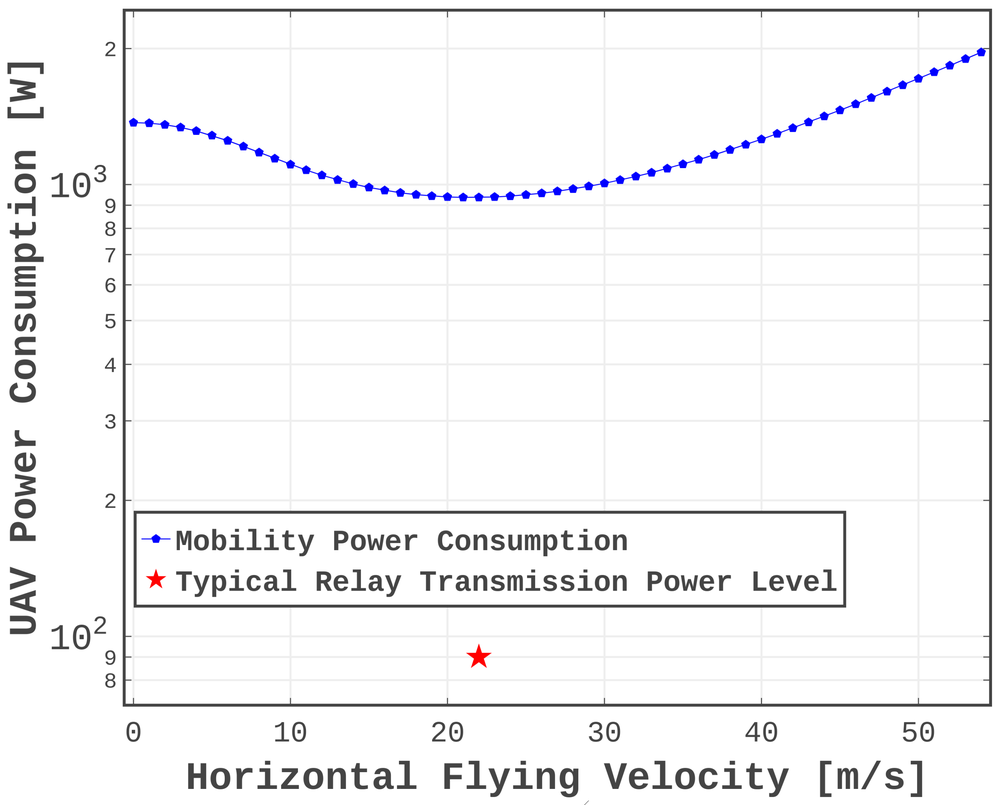
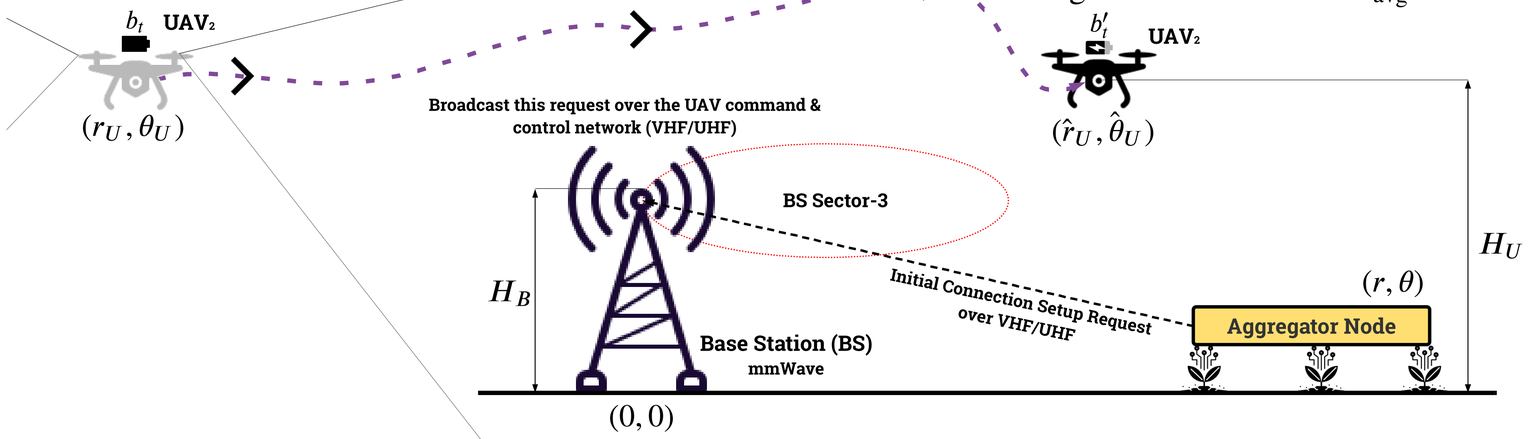
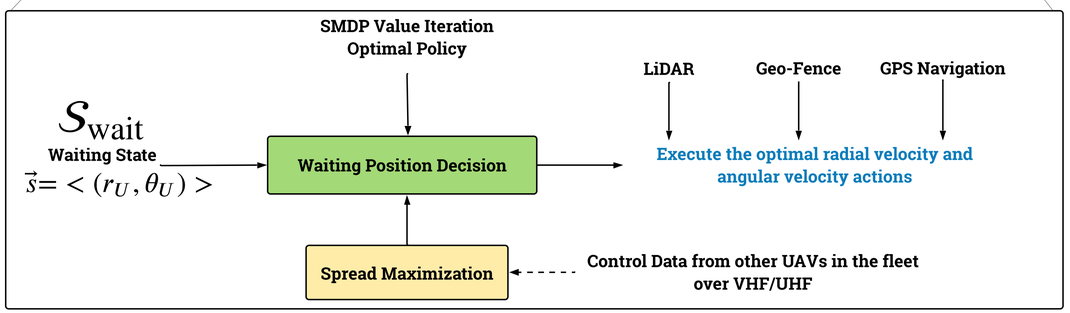

UAV2
UEs
GNs: Soil Sensors
(Aggregated)
App II: Coverage Extension
App I: Traffic Off-loading
Base Station (BS)
mmWave
UAV1
UAV3
GNs: Livestock Monitors
(Aggregated)
GNs: Cellular UEs







Sample Scenario Evaluation: UAV for Coverage Extension | Aggregated Soil Sensors



Until Convergence (k)

$$|g_k - g_{k-1}| < \epsilon_{DI}; \bar{E}_k - P_{\text{avg}}\bar{T}_k < \epsilon_{PF}; \nu_k|\bar{E}_k - P_{\text{avg}}\bar{T}_k| < \epsilon_{CS}$$

$$O^*(s), \forall s \in \mathcal{S}_{\text{wait}} | U^*(s), \forall s \in \mathcal{S}_{\text{comm}}$$

TensorFlow Constrained Angular Velocity Optimization

θ_c^*

$\forall s \in \mathcal{S}_{\text{wait}}$

$\ell_\nu^*(s; \upsilon_r)$

$E^*(s; \upsilon_r, \theta_c^*)$

$T^*(s; \upsilon_r, \theta_c^*)$

Waiting State Cost Evaluations

Tensor CSO ↔ Trajectory and Velocity Waypoints Improvement via Localized Randomization

$\mathbf{q}_U^*$

Comm State Cost Evaluations

$\forall s \in \mathcal{S}_{\text{comm}}$

$\ell_\nu^*(s; \xi^*, (\hat{r}_U, \hat{\theta}_U))$

$\xi^* E(s; (\hat{r}_U, \hat{\theta}_U), \mathbf{q}_U^*)$

$\xi^* T(s; (\hat{r}_U, \hat{\theta}_U), \mathbf{q}_U^*)$

Concurrent Threads

Until Convergence (i)

$$H(s) = V_{i+1}(s) - V_i(s);\ \max_{s \in \mathcal{S}} H(s) - \min_{s \in \mathcal{S}} H(s) < \delta$$

Projected Subgradient Ascent

ν_k

Cost Evaluation (Lagrangian | Energy | Delay)

Concurrent Threads

SMDP Value Iteration

Adaptive Multiscale Scheduling and Trajectory Design for Power-Constrained UAV Relays

Matthew Bliss and Nicolò Michelusi

Abstract

This paper investigates adaptive trajectory and communication scheduling design for an UAV relaying random data traffic generated by ground nodes to a base station. The goal is to minimize the expected average communication delay to serve requests, subject to an average UAV mobility power constraint. The problem is cast as a semi-Markov decision process with a multiscale structure, which is optimized efficiently: in the outer decision, UAV radial velocities for waiting phases and end radii for communication phases optimize the average long-term delay-power trade-off; given outer decisions, inner decisions greedily minimize instantaneous delay-power costs, yielding optimal angular velocities in waiting states, and relay strategies and UAV trajectories in communication states; given large-scale propagation conditions and small-scale fading statistics, rate adaptation maximizes throughput, leveraging air-to-ground channel features. A novel hierarchical framework using competitive swarm optimization (CSO) is designed to optimize trajectories, demonstrating 21$\times$ faster computational speeds than successive convex approximation and 42% improved accuracy over non-hierarchical CSO. Simulations demonstrate that intelligent adaptive design exploiting UAV mobility features, e.g., helicopter translational lift, reduces the average communication delay and UAV mobility power consumption by 51% and 6%, respectively, against an optimal hovering strategy and by 3% and 14%, respectively, against a delay-minimizing heuristic.

Index Terms

UAV communication, rotary-wing UAV, trajectory optimization, competitive swarm optimization

I. Introduction

Recently, research into unmanned aerial vehicles (UAVs) operating in wireless networks has surged, thanks to their unique advantages over terrestrial base stations (BSs) in terms of mobility, maneuverability, and enhanced line-of-sight (LoS) [2]–[6]. The potential to exploit UAV mobility

A preliminary version of this manuscript appeared at IEEE ICC 2020 [1].

Bliss and Michelusi are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA; emails: {blissm,michelus}@purdue.edu.

Part of this work has been supported by NSF CNS-1642982.

for wireless networking applications is vast, and UAVs acting as aerial BSs or relays promise increased cellular coverage, reduced communication delay, and improved energy efficiency [3].

Although demonstrating improvements to wireless networks, the design of UAV deployment strategies faces many challenges [2], [3], due to the intricate coupling between UAV mobility and air-to-ground (A2G) channel propagation. Already, works such as [7]–[9] included onboard energy constraints, and hence mission times, inherent to many low-altitude platforms [2]. Additionally, UAV path planning presents analytical challenges, due to the large design space, i.e., time-varying UAV and user positions, communication scheduling, channel quality, and quality of service constraints [3]. To address these challenges, most prior research focused on:

- Non-adaptive optimization, i.e., data traffic is known in advance [7], [8], [10], [11];
- Static hovering strategies that incorporate randomness [9], [12]–[19];
- Approaches based on direct application of reinforcement learning (RL) [20]–[22].

However, non-adaptive optimization is unrealistic; in practice, wireless networks exhibit random data traffic, which requires adaptive trajectory and communications. On the other hand, static hovering deployments [18], [19] 1) do not fully exploit UAV mobility to improve the channel conditions proactively; and 2) they suffer from increased UAV power consumption, due to *helicopter translational lift* [23], which dictates that flying at moderate forward speed is more energy-efficient than hovering. RL-based methods, despite not inherently being black-box approaches, are overwhelmingly used in this way and have thus far failed to leverage the coupling between communications and trajectory optimization to achieve more scalable design.

To address these open design challenges, in this paper we consider an UAV serving as a relay for randomly generated uplink transmission requests between densely deployed ground nodes (GNs) in a cell and a BS. We formulate an adaptive UAV trajectory and communication scheduling design so as to minimize the expected average communication delay to serve the requests, subject to an average UAV mobility power constraint. Our design leverages UAV mobility to simultaneously minimize communication delay and reduce UAV power consumption. To address the computational challenges of dynamic programming formulations, we leverage the coupling between channel propagations and UAV trajectory design via a *multiscale decision framework* based on semi-Markov decision processes (SMDPs), which decouples *outer*-decisions affecting long-term dynamics from short-term optimization of trajectory and communications (*inner*-decisions), and rate adaptation to A2G propagation features.

To optimize trajectories, we develop a competitive swarm optimizer (CSO) [24] embedded in

a *novel* hierarchical framework, which we term hierarchical CSO (HCSO): rather than applying CSO directly to design high resolution trajectories, we solve a sequence of CSO problems with increasing resolution, in which the lower resolution trajectory of the previous CSO stage is used to initialize the next CSO stage at increased resolution. By doing so, HCSO achieves both scalability and improved accuracy in trajectory design, over both conventional CSO and successive convex approximation (SCA), as we will show numerically in Sec. VI. Numerical simulations demonstrate that the intelligent adaptive communication and trajectory design reduces both the average communication delay and UAV mobility power consumption by 51% and 6%, respectively, with respect to an optimal hovering strategy, and by 3% and 14%, respectively, with respect to a heuristic delay minimization scheme.

A. Related Work

Works that optimize static UAV hovering positions consider sources of randomness [12]–[14], but are sub-optimal in terms of both UAV endurance [8], [25] and communication performance, since they do not proactively improve channel conditions. In our work, we will show that intelligent exploitation of UAV mobility enhances *both* communications *and* UAV power usage.

Alternatively, in non-adaptive trajectory design, data traffic is *deterministic*, such that the UAV charts an optimized trajectory and communication schedule before its deployment, as done in [7], [8]. Yet, transmission requests often arrive randomly and cannot be known in advance. In this case, the design of UAV trajectories and communications is much more challenging, since both must adapt to uncertain system dynamics. In fact, adaptive strategies must account for the various realizations that occur due to random data traffic. Even for one realization, the number of UAV trajectories to follow is infinite and intractable to optimize without approximation. These trajectory optimization problems are often path-discretized [8] and lack convex problem structure. A common approach has been SCA, as implemented in [8] and recognized as a standard tool [3]. However, SCA has major drawbacks: 1) its time consumption prohibits efficient real-time trajectory design to accommodate uncertain dynamics, as we will show in Sec. VI; 2) it requires closed-form achievable rate expressions to develop convex approximations, which may not be feasible in realistic A2G propagation environments, or when rate adaptation is used to maximize throughput, as done in this paper.

To overcome these challenges, in this work we design trajectories via CSO [24]. In contrast to particle swarm optimization (PSO)–which updates particles in accordance with each particle's

personal best and the swarm's best performance [26]–CSO invokes pairwise competition between particles, permitting the winners to advance to the subsequent iteration and the loser particles to learn from the winners. As [24] shows, CSO yields better performance than PSO on many benchmark large-scale optimization problems. Unlike SCA which relies on the existence of convex approximations of the objective and constraints, CSO does not depend on the specific problem structure to work effectively. To further scale CSO to higher dimensional trajectory design problems, we embed it into a hierarchical framework (HCSO) which solves trajectories of higher and higher resolution to achieve accuracy and scalability. PSO was used in [27]–[32], but either optimized static UAV hovering positions [27]–[31] or imposed a specific trajectory structure, e.g., a circular path [32]. In this work, we propose HCSO to optimize a series of waypoints and speeds that determine a trajectory, with no other path structures imposed.

Works performing RL-based UAV trajectory optimization consider unknown system dynamics, e.g., natural obstructions or data traffic [20]–[22]. Although RL-based schemes can handle random system dynamics, formulations often use a black-box approach, in which problem structures, that could be leveraged to reduce state and action spaces and speed up learning, are largely set aside [20], [33]. In this work, we *exploit* problem structure by employing a multiscale framework, in which the outer decisions remain oblivious to the inner optimization, greatly reducing the action space for the decision-making framework. This multiscale structure, cast as a SMDP in this paper, could be naturally extended to RL and help reduce learning time.

To address the aforementioned challenges, our previous work [34] investigated the design of UAV trajectory and communications subject to random data traffic from two ground nodes, but did not consider the UAV mobility power. The model was extended in our work [1] to account for the UAV power constraint and relaying from densely deployed GNs to a BS. However, UAV relaying periods utilized a *fly-hover-communicate* strategy and the channel model followed a simplified LoS model, which we extend in this paper to account for more complex flying trajectories and realistic A2G propagation models.

B. Summary of Contributions

To summarize, the contributions of this paper are as follows:

- We demonstrate that the optimal trajectory and communication strategy exhibits a multiscale structure that can be efficiently optimized: in the outer decision, the UAV radial velocity for the *waiting* phase and end radius for the *communication* phase are selected to minimize the average

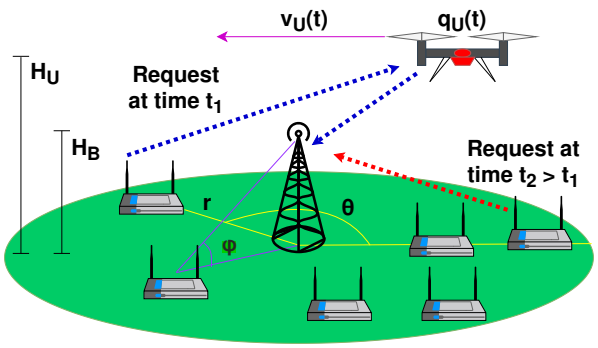


Fig. 1: System model.

long-term delay-power tradeoff; given outer decisions, the inner decisions greedily minimize the instantaneous delay-power cost; finally, rate adaptation is used to maximize throughput.

- We develop a value iteration algorithm to solve the outer decisions as an average cost-per-stage SMDP, and leverage sufficient statistics in the decision-making process.
- We develop a *hierarchical* CSO framework to optimize the UAV trajectories of the communication phases in the inner decisions, and show $21\times$ faster computational speed than SCA, and 42% improvement to the objective function over conventional CSO.
- We propose a computationally efficient method for rate adaptation that optimizes throughput by leveraging large-scale pathloss conditions and small-scale fading statistics.

The remainder of the paper is organized as follows. In Sec II, we introduce the system model; in Sec. III we formulate the optimization goal and formalize the problem as a semi-Markov decision process; in Sec. IV, we present the multiscale optimization approach; in Sec. V, we formulate the trajectory optimization using HCSO; in Sec. VI, we provide the numerical results, followed by concluding remarks in Sec. VII.

II. System Model

Consider the scenario of Fig. 1, where densely deployed ground nodes (GNs) in a circular cell of radius a on the flat ground surface (x-y plane) generate random data traffic to be transmitted to a base station (BS) located in the center of the cell at height H_B. The GNs are distributed uniformly in the circular cell with density λ_G [GNs/m^2]. Due to poor pathloss conditions experienced by GNs farther away from the BS (e.g., low-power sensors or IoT devices), a rotary-wing UAV is deployed as a mobile relay, flying at a fixed height H_U: GNs transmit their data payloads either directly to the BS or using the UAV as a relay.[1] Let $\mathbf{q}_U(t) = (r_U(t), \theta_U(t))$ be the

[1]This framework can be extended to both uplink and downlink transmissions, by creating an additional state differentiating the two types of data traffic. When a downlink request is generated, the BS may choose to either 1) transmit the downlink data payload directly to the destination GN or 2) relay it through the UAV.

polar coordinate of the UAV projected onto the x-y plane, with $r_U(t) \in \mathbb{R}_+$ and $\theta_U(t) \in [0, 2\pi)$ denoting the UAV's radius and angle with respect to the BS at time t, and let $v_U(t)$ be its instantaneous forward flying speed at time t, as depicted in Fig. 1.

A. Communication Model

Each GN generates uplink transmission requests of L bits, according to a Poisson process with rate λ_P [requests/GN/sec]. Thus overall, uplink transmission requests arrive in time according to a Poisson process with rate $\lambda{=}\lambda_G{\cdot}\lambda_P$ [requests/sec/m^2]. Since a new request is uniformly distributed in the cell area, its angular coordinate θ is uniform in $[0, 2\pi]$, and its radial coordinate pdf is $f_R(r) = \frac{2r}{a^2}\mathbb{I}(r \leq a)$, where $\mathbb{I}(\cdot)$ denotes the indicator function.

Upon receiving an uplink transmission request, the GN will either 1) transmit directly to the BS, or 2) relay through the UAV using a two-phase decode-and-forward (D&F) protocol (see Sec. III), where the UAV, while following a trajectory part of the design, first receives the entire data payload from the GN, and subsequently forwards it to the BS. Transmissions occur over one of k orthogonal sub-channels, each of bandwidth B (Hz). Moreover, the UAV can handle only one transmission request at a time, whereas the BS can handle multiple simultaneous transmissions over the remaining sub-channels.[2]

For practical implementation, a reliable and robust control channel is used to coordinate data transmissions between the BS, GNs, and UAV over long-distances [35], but may not be suitable to transmit large data payloads, hence the need for an UAV acting as a relay. Such coordination occurs quickly relative to data communication, hence we neglect control delay. When a GN generates an uplink transmission request, it transmits a control packet to the BS indicating the need for uplink transmission, as well as its physical location. The BS then queries the UAV, which responds with its current position. With the GN and UAV positions obtained, the BS makes a scheduling decision part of our design (Sec. III): if direct transmission is chosen, the BS instructs the GN to begin direct transmission; otherwise, the BS computes the UAV's trajectory to relay the uplink request. This trajectory, encoded as a sequence of waypoints traversed by the UAV and speeds (see Sec. V), is sent as a control message to the UAV, and the UAV instructs the GN to begin the D&F protocol. As a result of the different scheduling decisions, the GN$\rightarrow$BS, GN$\rightarrow$UAV, and the UAV$\rightarrow$BS links must be considered, as described next.

[2]Practically, k is finite, so that the BS state can be modeled as a $M/G/k$ queue. We assume that k is large enough so that the probability of queue overflow is small.

B. A2G Channel Model

For a generic link, we denote the channel coefficient as $h \triangleq \sqrt{\beta} g$, where β captures the large-scale channel variations, and g with $\mathbb{E}\left[|g|^2\right]=1$ is the small-scale fading component. We model the large-scale component as $\beta=\beta_{\mathrm{LoS}}(d) \triangleq \beta_0 d^{-\alpha}$ for line-of-sight (LoS) and $\beta=\beta_{\mathrm{NLoS}}(d) \triangleq \kappa\beta_0 d^{-\tilde{\alpha}}$ for non-LoS (NLoS) links [8], where β_0 is the pathloss referenced at a distance of 1 meter, $2 \leq \alpha \leq \tilde{\alpha}$ are the pathloss exponents, $\kappa \in (0,1]$ captures the additional NLoS attenuation, and d is the Euclidean distance between transmitter and receiver [8]. We model the LoS/NLoS probability as a function of the elevation angle $\varphi \in (0, 90^o]$ (see Fig. 1) as [17]

$$P_{\mathrm{LoS}}(\varphi) = \frac{1}{1 + z_1 \cdot \exp\left(-z_2\left[\varphi - z_1\right]\right)}, \quad P_{\mathrm{NLoS}}(\varphi) = 1 - P_{\mathrm{LoS}}(\varphi), \tag{1}$$

where z_1 and z_2 are parameters specific to the propagation environment, e.g., rural, suburban, urban, etc. For the UAV$\rightarrow$BS link, we assume that the UAV and BS are elevated high enough such that $z_1=0$ and $P_{\mathrm{LoS}}(\varphi) = 1$ [36].

The distribution of the small-scale fading component g also depends on the LoS/NLoS link state [37]. Specifically, for a LoS link, we model g as Rician with a φ-dependent K-factor

$$K(\varphi) = k_1 \cdot \exp\left(k_2 \cdot \varphi\right), \tag{2}$$

where coefficients k_1 and k_2 are determined by the propagation environment [36], and the deterministic LoS component's phase is compensated at the receiver. For a NLoS link, we model g as Rayleigh [37] (Rician with $K = 0$). Given h, the link capacity [bps] is $C(h) = B \cdot \log_2\left(1 + \frac{|h|^2 P}{\sigma^2 \Gamma}\right)$, where P is the transmission power, σ^2 is the receiver noise power, and Γ is the SNR gap to the theoretical capacity [36]. Other sources of signal degradation, such as the Doppler effect, are well-compensated at the receiver, using techniques developed in [38].

As the large-scale components are expected to vary slowly relative to the rate of acquisition of channel state information (CSI), we assume that the large-scale CSI, (β, K), is known throughout the data communication process;[3] however, small-scale fading cannot be compensated, possibly causing outage. Given (β, K) and a transmission rate of Υ [bps], we thus define the outage probability $P_{\mathrm{out}}(\Upsilon, \beta, K) \triangleq \mathbb{P}(C(\sqrt{\beta} g) < \Upsilon | \beta, K) = \mathbb{P}\left(|g|^2 < u(\Upsilon, \beta)\right)$, where we have defined $u(\Upsilon, \beta) \triangleq \sigma^2\Gamma(2^{\Upsilon/B}-1)/(\beta P)$. Since $2(K+1)|g|^2$ has non-central χ^2 distribution with 2 degrees

[3]Note that β and K vary due to temporal variations in the LoS/NLoS condition, in the transmitter-receiver distance d, and in the elevation angle φ.

of freedom and non-centrality parameter $2K$, this is computed as

$$P_{\text{out}}(\Upsilon, \beta, K) = 1 - Q_1\left(\sqrt{2K}, \sqrt{2(K+1)u(\Upsilon, \beta)}\right), \tag{3}$$

where $Q_1(\cdot,\cdot)$ is the standard Marcum Q-function [36]. Note that when $K = 0$ (NLoS link), the function specializes to $Q_1\left(0, \sqrt{2u(\Upsilon, \beta)}\right) = \exp(-u(\Upsilon, \beta))$. We assume that the small-scale fading is averaged out in the temporal and spatial dimensions, yielding the expected throughput

$$R(\Upsilon, \beta, K) = \Upsilon \cdot (1 - P_{\text{out}}(\Upsilon, \beta, K)) = \Upsilon \cdot Q_1\left(\sqrt{2K}, \sqrt{2(K+1)u(\Upsilon, \beta)}\right). \tag{4}$$

In our model, we permit rate adaptation at the transmitter based on large-scale CSI (β, K), coordinated through the control channel via CSI feedback, i.e., $\Upsilon^*(\beta, K) \triangleq \arg\max_{\Upsilon \geq 0} R(\Upsilon, \beta, K)$. To solve this problem, let $Z \triangleq \sqrt{\frac{2\beta P}{\sigma^2 \Gamma} u(\Upsilon, \beta)}$, so that $\Upsilon = B \log_2\left(1 + \frac{1}{2}Z^2\right) \triangleq f(Z)$. It follows that $\Upsilon^*(\beta, K) = f(Z^*(\beta, K))$, where $Z^*(\beta, K) \triangleq \arg\min_{Z \geq 0} -\ln f(Z) - \ln Q_1\left(\sqrt{2K}, \sqrt{\frac{(K+1)\beta P}{\sigma^2 \Gamma}} Z\right)$, which is a convex optimization problem (see Appendix A). Therefore, $Z^*(\beta, K)$ can be found efficiently using a bisection method. Upon determining the optimal transmission rate $\Upsilon^*(\beta, K)$, we define the optimized throughput as $R^*(\beta, K) \triangleq R(\Upsilon^*(\beta, K), \beta, K)$.

Assuming that the LoS/NLoS conditions are averaged out in the temporal and spatial dimensions, we finally compute the average link throughput as

$$\bar{R}(d, \varphi) \triangleq P_{\text{LoS}}(\varphi) \cdot R^*(\beta_{\text{LoS}}(d), K(\varphi)) + P_{\text{NLoS}}(\varphi) \cdot R^*(\beta_{\text{NLoS}}(d), 0), \tag{5}$$

which is then specialized to the three distinct communication links by expressing the transmission powers, environment-specific parameters z_1, z_2, k_1, and k_2, the large-scale parameters (β, K) and LoS/NLoS probabilities in (1) based on the spatial configuration, i.e., transmitter-receiver distance and elevation angle. Specifically, for the GN→BS link, we let $\bar{R}_{GB}(r)$ be the throughput with the GN in position (r, θ), computed by setting the GN-BS distance as $d = \sqrt{H_B^2 + r^2}$ and the elevation angle as $\varphi = \sin^{-1}(H_B/d)$ in (5). Similarly, for the GN→UAV link, we let $\bar{R}_{GU}(r_{GU})$ be the throughput when the GN-UAV distance (projected onto the x-y plane) is r_{GU}, computed by setting the GN-UAV Euclidean distance as $d = \sqrt{r_{GU}^2 + H_U^2}$ and the elevation angle as $\varphi = \sin^{-1}(H_U/d)$ in (5). Finally, for the UAV→BS link, we let $\bar{R}_{UB}(r_{UB})$ be the throughput when the UAV-BS distance (projected onto the x-y plane) is r_{UB}, computed by setting the GN-UAV Euclidean distance as $d = \sqrt{r_{UB}^2 + (H_U - H_B)^2}$ and the elevation angle as $\varphi = \sin^{-1}((H_U - H_B)/d_{UB})$ in (5), with $z_1 = 0$ to enforce the LoS condition.

C. UAV Mobility Power Model

Let $P_U(t) = P_c(t) + P_{\text{mob}}(v_U(t))$ be the instantaneous UAV power consumption, including the total communication power $P_c(t)$ and the forward flight mobility power $P_{\text{mob}}(v_U(t))$, a generally non-convex function of the UAV horizontal flying speed $v_U(t)$, vehicle drag coefficient, and total weight (e.g., vehicle battery and added equipment) [25]. However, as the communication power (order of 1-10W) is usually dwarfed by the UAV mobility power (order of 200-1kW [8]), in this paper we neglect $P_c(t)$ and approximate $P_U(t) \approx P_{\text{mob}}(v_U(t))$, modeled as [8]

$$P_{\text{mob}}(V) = P_1 \left(1 + \frac{3V^2}{U_{\text{tip}}^2}\right) + P_2 \left(\sqrt{1 + \frac{V^4}{4v_0^4}} - \frac{V^2}{2v_0^2}\right)^{1/2} + P_3 V^3, \tag{6}$$

where P_i are scaling constants, U_{tip} is the rotor blade tip speed, and v_0 is the mean rotor induced velocity while hovering (see [8] for details). This model is validated by real UAV deployments [25], and an example power vs. speed curve is found in [8], for reference. Therein, hovering requires 1.37kW, whereas flying at the power-minimizing speed of 22m/s only consumes 0.94kW. As we will show, a nonzero power-minimizing flying speed can be exploited to simultaneously reduce communication delay, UAV power consumption, and enhance its endurance.

III. Optimization Goal and SMDP Formulation

The goal is to design the UAV trajectory and scheduling decisions, so as to minimize the average delay to serve random GN requests, under an average UAV power constraint. At any time, the system is in one of the following phases. In the *waiting phase*, no GN requests are being served by the UAV. When a new GN request arrives, say from position (r, θ), the system transitions to the *request scheduling phase*, where the BS decides if the GN transmits its data payload directly to the BS, or relays through the UAV. Using direct transmission, the system immediately re-enters the *waiting phase*, as the UAV remains free to serve other transmission requests; otherwise, the system enters the *UAV relay phase*, where the GN relays its data payload through the UAV using the D&F protocol described in this section; upon termination, the system re-enters the *waiting phase*. As the BS can accommodate simultaneous transmissions (see Sec. II-A), new requests received during an *UAV relay phase* are directly served by the BS.

We define a *decision interval* as the time duration spanning the start of a *waiting phase*, the subsequent *request scheduling phase* when a GN request is received, until the system re-enters the *waiting phase* after scheduling a direct transmission to the BS, or following the *UAV relay*

phase. Consider the uth such decision interval; let $\Delta_u^{(w)}$ be the time to wait for a new request, and $\Delta_u^{(s)}$ be the time to serve it, either through the BS (denoted with the scheduling decision ξ_u=0) or through the UAV (ξ_u=1). Then, the uth decision interval duration is Δ_u=$\Delta_u^{(w)}$+$\xi_u\Delta_u^{(s)}$, since the *waiting phase* follows immediately after scheduling a direct transmission to the BS.

We now formulate the average communication delay and UAV mobility power. Let $N_u \geq 0$ be the number of additional requests received during the UAV relay phase of the uth decision period, $\Delta_{u,i}^{(bs)},\ i=1,\ldots,N_u$ be the communication delays to serve these requests directly by the BS. Let E_u be the UAV mobility energy expended during the uth decision interval. Let M_t be the total number of decision intervals completed up to time t. Then, we define the expected average communication delay per GN request under a given policy μ that defines the request scheduling, communication strategy, and UAV trajectory (formally defined later), with the UAV starting from the geometric center $\mathbf{q}_U(0)$=$(0,0)$,[4] as the total communication delay accrued until time t, over the total number of requests served until time t; similarly, we define the expected average UAV power as the total energy consumption, over the total duration of the M_t decision intervals up to time t; mathematically,

$$\bar{D}_\mu \triangleq \lim_{t\to\infty} \mathbb{E}_\mu\left[\frac{\frac{1}{M_t}\sum_{u=1}^{M_t}(\Delta_u^{(s)}+\xi_u\sum_{i=1}^{N_u}\Delta_{u,i}^{(bs)})}{\frac{1}{M_t}\sum_{u=1}^{M_t}(1+\xi_u N_u)}\right],\ \bar{P}_\mu \triangleq \lim_{t\to\infty}\mathbb{E}_\mu\left[\frac{\frac{1}{M_t}\sum_{u=1}^{M_t}E_u}{\frac{1}{M_t}\sum_{u=1}^{M_t}\Delta_u}\right]. \tag{7}$$

Note that $\bar{D}_\mu$ in (7) captures the delays of all request scenarios, i.e., those relayed through the UAV ($\xi_u=1$), those transmitted directly to the BS ($\xi_u=0$), as well as the N_u additional requests served directly by the BS during the UAV relay phase. To simplify these expressions, let

$$\{\bar{E}_\mu;\bar{T}_\mu;\bar{N}_\mu;\bar{W}_\mu\} \triangleq \lim_{t\to\infty}\mathbb{E}_\mu\left[\frac{1}{M_t}\sum_{u=1}^{M_t}\left\{E_u\ ;\ \Delta_u\ ;\ 1+\xi_u N_u\ ;\ \Delta_u^{(s)}+\xi_u\sum_{i=1}^{N_u}\Delta_{u,i}^{(bs)}\right\}\right] \tag{8}$$

be the expected average UAV energy expenditure ($\bar{E}_\mu$), duration ($\bar{T}_\mu$), number of requests served ($\bar{N}_\mu$), and total communication delays accrued ($\bar{W}_\mu$) per decision interval, under policy μ. From Little's Law [39], we can then express $\bar{P}_\mu=\bar{E}_\mu/\bar{T}_\mu$ and $\bar{D}_\mu=\bar{W}_\mu/\bar{N}_\mu$. The goal is to solve[5]

$$\bar{D}_\mu^* = \min_\mu \frac{\bar{W}_\mu}{\bar{N}_\mu}\ \text{ s.t. }\frac{\bar{E}_\mu}{\bar{T}_\mu}\leq P_{\text{avg}}, \tag{9}$$

with optimal policy μ^*. Note that the delay to serve a request from a GN in position (r,θ) by

[4]In the following, expectations $\mathbb{E}_\mu[\cdot]$ will be implicitly assumed to be conditioned on $\mathbf{q}_U(0)=(0,0)$.

[5]Without loss of generality, we assume $P_{\text{avg}}\geq\min_{V\geq 0}P_{\text{mob}}(V)\triangleq P_{\min}$, since any $P_{\text{avg}}<P_{\min}$ is not feasible.

direct transmission to the BS is $L/\bar{R}_{GB}(r)$. Then, if all requests are served directly by the BS, the expected delay is obtained by computing the expectation with respect to the radial coordinate,

$$\bar{\Delta}_{BS} \triangleq \int_0^a \frac{L}{\bar{R}_{GB}(r)} f_R(r) \mathrm{d}r. \tag{10}$$

Clearly, the optimal policy is such that $\bar{D}_\mu^* \leq \bar{\Delta}_{BS}$ (in fact, a feasible policy is obtained by having all requests served directly by the BS with the UAV flying at the power minimizing speed). This suggests that optimization of the UAV trajectory and scheduling decisions is key to improving the delay performance to serve requests. Note the inherent complexity that is required to solve $\bar{D}_\mu^*$: as the policy varies, the delay metric changes both the numerator and denominator terms, which cannot be solved with standard dynamic programming algorithms. We now propose an alternative optimization criterion with reduced complexity that upper- and lower-bounds the average delay $\bar{D}_\mu$, based on Lemma 1, whose proof is provided in Appendix B.

Lemma 1. *Let $\bar{W}_\mu^{(s)}$ be the expected average delay of requests for which a scheduling decision has been made (opposed to those served directly by the BS during the UAV relay phase),*

$$\bar{W}_\mu^{(s)} \triangleq \lim_{t\to\infty} \mathbb{E}_\mu \left[\frac{1}{M_t} \sum_{u=1}^{M_t} \Delta_u^{(s)} \right]. \tag{11}$$

Then, under any policy μ such that $\bar{W}_\mu^{(s)} \leq \bar{\Delta}_{BS}$ (including the optimal μ^*), we have that

$$\bar{W}_\mu^{(s)} \leq \bar{D}_\mu \leq \bar{W}_\mu^{(s)} \frac{1+\Lambda\bar{\Delta}_{BS}}{1+\Lambda\bar{W}_\mu^{(s)}} \leq \bar{\Delta}_{BS}, \tag{12}$$

where $\Lambda \triangleq \lambda\pi a^2$ is the overall request arrival rate over the circular cell.

Remarkably, the lower and upper bounds of $\bar{D}_\mu$ are both increasing functions of $\bar{W}_\mu^{(s)}$. This observation motivates us to look at the alternative optimization problem

$$\min_\mu \bar{W}_\mu^{(s)} \text{ s.t. } \bar{E}_\mu - P_{\text{avg}}\bar{T}_\mu \leq 0, \tag{13}$$

which will be the focus of the following analysis. To solve it, we use the Lagrangian formulation

$$g(\nu) = \min_\mu \left(\bar{W}_\mu^{(s)} + \nu(\bar{E}_\mu - P_{\text{avg}}\bar{T}_\mu) \right) = \min_\mu \lim_{t\to\infty} \mathbb{E}_\mu \left[\frac{1}{M_t} \sum_{u=1}^{M_t} \left(\Delta_u^{(s)} + \nu(E_u - P_{\text{avg}}\Delta_u) \right) \right], \tag{14}$$

where ν is the dual variable, optimized by solving $\max_{\nu\geq 0} g(\nu)$. We now demonstrate that for a given $\nu \geq 0$, (14) can be cast as a semi-Markov decision process (SMDP) and solved with dynamic programming tools. We first characterize the states, actions, and policy of this SMDP.

States: The state is defined by the UAV position, with values from the state space $\mathcal{Q}_{\text{UAV}} \triangleq \mathbb{R}_+ \times [0, 2\pi)$, and the position of an uplink transmission request, with values from the state space $\mathcal{Q}_{\text{GN}} \triangleq [0, a] \times [0, 2\pi)$ (both expressed as polar coordinates). The state space is then $\mathcal{S} = \mathcal{S}_{\text{wait}} \cup \mathcal{S}_{\text{comm}}$, where $\mathcal{S}_{\text{wait}} = \mathcal{Q}_{\text{UAV}} \times \{w_0\}$ (with $w_0 \triangleq (-1, -1)$ indicating no active transmission request) is the set of *waiting states* and $\mathcal{S}_{\text{comm}} = \mathcal{Q}_{\text{UAV}} \times \mathcal{Q}_{\text{GN}}$ is the set of *communication* states.

Key to the definition of the SMDP is how the system is sampled in time to define Markov dynamics in the evolution of the sampled states, as specified below. Accordingly, we define the actions available in each state, the transition probabilities, along with the duration $T(s,\mathbf{a})$, UAV energy usage $E(s,\mathbf{a})$, and communication delay $\Delta(s,\mathbf{a})$ metrics accrued in state s under action $\mathbf{a}$.

Waiting states' actions, transitions and metrics: If the UAV is in the *waiting state* $s_n = (r_U, \theta_U, w_0) \in \mathcal{S}_{\text{wait}}$ at time t, i.e., it is in position $\mathbf{q}_U(t) = (r_U, \theta_U)$ with no active requests, then the actions available are to move the UAV with radial (referred to the outward direction) and angular (referred counter-clockwise) velocity components (v_r, θ_c), over an arbitrarily small duration $\Delta_0 \ll 1/\Lambda$. Under a maximum speed constraint $V_{\max}$, the action space is then

$$\mathcal{A}_{\text{wait}}(r_U) \triangleq \left\{ (v_r, \theta_c) \in \mathbb{R}^2 \;\middle|\; \sqrt{v_r^2 + r_U^2 \cdot \theta_c^2} \leq V_{\max} \right\}, \tag{15}$$

where $v_U = \sqrt{v_r^2 + r_U^2 \cdot \theta_c^2}$ is the speed expressed with respect to polar coordinates. Upon choosing action $\mathbf{a} = (v_r, \theta_c) \in \mathcal{A}_{\text{wait}}(r_U)$, the communication delay is $\Delta(s; \mathbf{a}) = 0$, since there is no ongoing communication; the duration of a waiting state visit is $T(s; \mathbf{a}) = \Delta_0$, during which the UAV uses an amount of energy $E(s; \mathbf{a}) = \Delta_0 P_{\text{mob}}(v_U)$ to move at speed v_U.

The new state is then sampled at time $t + \Delta_0$, with the UAV moved to the new position $\mathbf{q}_U(t+\Delta_0) \approx (r_U, \theta_U) + (v_r, \theta_c)\Delta_0$. With probability $e^{-\Lambda\Delta_0}$, no new request is received in the time interval $[t, t + \Delta_0]$, so that the new state is a waiting state. Otherwise, a new request is received from a GN in position (r, θ), so that the new state is a communication state. Thus, the transition probability from the waiting state $s_n = (r_U, \theta_U, w_0)$ under action $\mathbf{a}_n = (v_r, \theta_c) \in \mathcal{A}_{\text{wait}}(r_U)$ is

$$\mathbb{P}(s_{n+1} = (r_U + v_r\Delta_0, \theta_U + \theta_c\Delta_0, w_0) | s_n, \mathbf{a}_n) = e^{-\Lambda\Delta_0}, \tag{16}$$

$$\mathbb{P}(s_{n+1} \in (r_U + v_r\Delta_0, \theta_U + \theta_c\Delta_0, \mathcal{F}) | s_n, \mathbf{a}_n) = \left[A(\mathcal{F}) \cdot (1 - e^{-\Lambda\Delta_0})\right] / \pi a^2, \ \forall \mathcal{F} \subseteq \mathcal{Q}_{\text{GN}}, \tag{17}$$

where $A(\mathcal{F})$ is the area of region $\mathcal{F}$, since requests are uniformly distributed in the cell.

Communication states' actions, transitions and metrics: Upon reaching a communication state $s_n = (r_U, \theta_U, r, \theta) \in \mathcal{S}_{\text{comm}}$ at time t, the system must serve a GN request at position (r, θ).

The BS first determines the scheduling decision $\xi\in\{0,1\}$. With $\xi=0$, the GN transmits directly to the BS, and the next state is sampled immediately after the decision, so that it is the waiting state $s_{n+1}=(r_U,\theta_U,w_0)\in\mathcal{S}_{\text{wait}}$ with probability 1. With $\xi=1$, the UAV follows a trajectory starting from its current position $\mathbf{q}_U(t)=(r_U,\theta_U)$ which allows the D&F protocol, as described next.

In the first phase of the D&F protocol, of duration t_p, the GN transmits its payload to the UAV; in the second phase of duration $\Delta-t_p$, the UAV relays the data payload to the BS. Assuming a *move-and-transmit* strategy (see [8]), the trajectory and durations t_p and $\Delta-t_p$ must satisfy

$$\int_0^{t_p}\bar{R}_{GU}(r_{GU}(t+\eta))\mathrm{d}\eta\geq L,\ \int_{t_p}^{\Delta}\bar{R}_{UB}(r_{UB}(t+\eta))\mathrm{d}\eta\geq L, \tag{C.1}$$

i.e., the entire payload is transmitted to the UAV and then to the BS, so that the total communication delay is Δ. Let $\mathcal{Q}_{r,\theta}\big(r_U,\theta_U\to\hat{r}_U,\hat{\theta}_U\big)$ be the set of feasible UAV trajectories starting in (r_U,θ_U), terminating in $(\hat{r}_U,\hat{\theta}_U)$, and serving a GN located at (r,θ) via the D&F protocol, i.e.,

$$\begin{aligned}\mathcal{Q}_{r,\theta}\big(r_U,\theta_U\to\hat{r}_U,\hat{\theta}_U\big)=\Big\{\mathbf{q}_U:[0,\Delta]\mapsto\mathbb{R}_+\times[0,2\pi):\ &\text{C.1},\\ &v_U(\eta)\leq V_{\max},\ \forall\eta\in[0,\Delta], &\text{(C.2)}\\ &\mathbf{q}_U(0)=(r_U,\theta_U),\mathbf{q}_U(\Delta)=(\hat{r}_U,\hat{\theta}_U),\ \exists\Delta\geq0,\exists\ 0\leq t_p\leq\Delta\Big\}; &\text{(C.3)}\end{aligned}$$

in this definition, C.1 reflects the data payload constraints, C.2 the maximum speed constraint, and C.3 the trajectory constraints. Then, the action space in state $(r_U,\theta_U,r,\theta)\in\mathcal{S}_{\text{comm}}$ when $\xi=1$ is the set $\mathcal{Q}_{r,\theta}(r_U,\theta_U)\triangleq\cup_{\hat{\theta}_U\in[0,2\pi)}\cup_{\hat{r}_U\in\mathbb{R}_+}\mathcal{Q}_{r,\theta}\big(r_U,\theta_U\to\hat{r}_U,\hat{\theta}_U\big)$ of feasible trajectories starting in (r_U,θ_U) that serve the GN via the D&F protocol. The total action space of the *communication states* $s=(r_U,\theta_U,r,\theta)\in\mathcal{S}_{\text{comm}}$ is then $\mathcal{A}_{\text{comm}}(r_U,\theta_U,r,\theta)\triangleq\{0\}\cup\mathcal{Q}_{r,\theta}(r_U,\theta_U)$, where the point $\{0\}$ reflects the option $\xi=0$, i.e., the GN transmits directly to the BS.

When $\xi=0$ (direct transmission), the cost metrics are computed as

$$\Delta(s;0)=\frac{L}{\bar{R}_{GB}(r)},\ E(s;0)=0,\ T(s;0)=0, \tag{18}$$

since direct transmissions occur at throughput $\bar{R}_{GB}(r)$, and the system moves immediately to the waiting state $(r_U,\theta_U,w_0)\in\mathcal{S}_{\text{wait}}$, resulting in the action duration and energy expenditure being 0. Instead, when $\xi=1$ (D&F protocol), and the trajectory $\mathbf{q}_U$ is selected, of duration Δ,

$$\Delta(s;\mathbf{a})=\Delta,\ E(s;\mathbf{a})=\int_0^{\Delta}P_{\text{mob}}\left(v_U(\eta)\right)\mathrm{d}\eta,\ T(s;\mathbf{a})=\Delta. \tag{19}$$

Upon completing the D&F protocol, the UAV enters the *waiting phase* again, which becomes the new state sampled at time $t+\Delta$. Specifically, under action $\mathbf{q}_U \in \mathcal{Q}_{r,\theta}(r_U, \theta_U \to \hat{r}_U, \hat{\theta}_U)$, the state transitions from $s_n = (r_U, \theta_U, r, \theta)$ to $s_{n+1} = (\hat{r}_U, \hat{\theta}_U, w_0)$ with probability 1.

Policy μ: For *waiting* states $(r_U, \theta_U, w_0) \in \mathcal{S}_{\text{wait}}$, the policy selects a velocity (v_r, θ_c) from the action space defined in (15), i.e., $\mu(r_U, \theta_U, w_0) \in \mathcal{A}_{\text{wait}}(r_U)$. Likewise, for *communication* states $(r_U, \theta_U, r, \theta) \in \mathcal{S}_{\text{comm}}$, the policy selects the scheduling decision $\xi \in \{0, 1\}$ and, if $\xi = 1$, the trajectory followed in the D&F protocol, $\mu(r_U, \theta_U, r, \theta) \in \mathcal{A}_{\text{comm}}(r_U, \theta_U, r, \theta)$.

With a *stationary policy* μ defined, the Lagrangian metric $L_\mu^{(\nu)} \triangleq \bar{W}_\mu^{(s)} + \nu(\bar{E}_\mu - P_{\text{avg}}\bar{T}_\mu)$ in (14) is reformulated using Little's Law [39] in terms of the SMDP steady-state probabilities as

$$L_\mu^{(\nu)} = \lim_{N\to\infty} \mathbb{E}_\mu \left[\frac{\frac{1}{N}\sum_{n=0}^{N-1} \ell_\nu(s_n; \mu(s_n))}{\frac{1}{N}\sum_{n=0}^{N-1} \mathbb{I}(s_n \in \mathcal{S}_{\text{comm}})} \right] = \frac{1}{\pi_{\text{comm}}} \int_{\mathcal{S}} \Pi_\mu(s) \ell_\nu(s; \mu(s)) \mathrm{d}s, \tag{20}$$

where $\Pi_\mu(s)$ is the steady-state pdf of the SMDP being in a state s under policy μ, $\pi_{\text{comm}} = \int_{\mathcal{S}_{\text{comm}}} \Pi_\mu(s) \mathrm{d}s$ is the steady-state probability that the UAV is in the communication phase in the SMDP, whose expression is provided in Appendix C, and $\ell_\nu(s; \mathbf{a}) \triangleq \Delta(s; \mathbf{a}) + \nu\big(E(s; \mathbf{a}) - P_{\text{avg}} T(s; \mathbf{a})\big)$ is the overall Lagrangian metric in state s under action $\mathbf{a}$. To see (20), note that $\sum_{n=0}^{N-1} \ell_\nu(s_n; \mu(s_n))$ is the total Lagrangian cost accrued during the first N SMDP stages, and $\sum_{n=0}^{N-1} \mathbb{I}(s_n \in \mathcal{S}_{\text{comm}})$ is the number of communication states encountered in the SMDP; since a new decision interval is initiated after a communication state, this is equal to the number of decision intervals. Therefore, after taking the expectation and the limit $N \to \infty$, $L_\mu^{(\nu)}$ represents the expected Lagrangian cost per decision interval, as expressed in (14). The right hand expression in (20) follows by noticing that the SMDP achieves a steady-state behavior when $N \to \infty$.

We now specialize the Lagrangian metric $\ell_\nu(s; \mathbf{a})$. Specifically, for waiting states,

$$\ell_\nu(r_U, \theta_U, w_0; v_r, \theta_c) = \nu\Big(P_{\text{mob}}\Big(\sqrt{v_r^2 + r_U^2 \cdot \theta_c^2}\Big) - P_{\text{avg}}\Big)\Delta_0; \tag{21}$$

for *communication* states under action $\xi = 0$, we use (18) to obtain $\ell_\nu(r_U, \theta_U, r, \theta; 0) = L/\bar{R}_{GB}(r)$; for *communication* states under action $\xi = 1$ with trajectory $\mathbf{q}_U$ of duration Δ, we obtain

$$\ell_\nu(r_U, \theta_U, r, \theta; \mathbf{q}_U) \triangleq (1 - \nu P_{\text{avg}})\Delta + \nu \int_0^\Delta P_{\text{mob}}\left(v_U(\eta)\right) \mathrm{d}\eta. \tag{22}$$

The minimization problem of (14) can then be expressed as the *average cost-per-stage problem*

$$g(\nu) = \frac{1}{\pi_{\text{comm}}} \min_\mu \int_{\mathcal{S}} \Pi_\mu(s) \ell_\nu(s; \mu(s)) \mathrm{d}s, \tag{23}$$

solvable through standard dynamic programming approaches, after discretization of the state space, followed by the dual maximization $\max_{\nu\geq 0} g(\nu)$.

IV. State Space Simplification and Multiscale Decomposition of Policy μ

Since GN transmission requests are uniformly distributed in the circular cell with the BS in the center, the UAV radius information is a sufficient statistic in decision making for a *waiting* state (r_U, θ_U, w_0), which can be thus expressed as $s = (r_U, w_0) \in \mathcal{S}_{\text{wait}}$. Likewise, for a *communication* state $(r_U, \theta_U, r, \theta)$, only the UAV radius, GN request radius, and the angle $\psi \in [0, 2\pi)$ *between* them suffice to characterize the state. Thus *communication* states can be compactly represented as $s = (r_U, r, \psi) \in \mathcal{S}_{\text{comm}}$. A consequence of these sufficient statistics for decision making is that the policy affects the SMDP state transitions (hence steady-state behavior) only through the UAV radial velocity v_r in the *waiting* states and the UAV trajectory's target end radius position $\hat{r}_U$ in the *communication* states. On the other hand, the angular velocity θ_c in the *waiting* states and the UAV trajectory's target end angular coordinate $\hat{\theta}_U$ in the *communication* states do not influence state dynamics, but only the instantaneous Lagrangian metric ℓ_ν.

With this observation, let $O(s) \triangleq v_r \in [-V_{\max}, V_{\max}]$ define the *radial velocity policy* of the *waiting* states $s \in \mathcal{S}_{\text{wait}}$, specifying the radial velocity component of a waiting action $\mathbf{a} = (v_r, \theta_c) \in \mathcal{A}_{\text{wait}}(r_U)$; let $U(s) \triangleq \hat{r}_U \in [0, a]$ define the *next radius position policy* of the *communication* states $s \in \mathcal{S}_{\text{comm}}$, specifying the end radius position of a scheduling and communication action $\mathbf{a} \in \mathcal{A}_{\text{comm}}(s)$. Under this decomposition, O and U constitute the *outer decisions* made by the SMDP and are the only actions to affect the steady-state distribution, denoted as $\Pi_{O,U}$ under the *outer policy* (O, U).

Thus, the optimization problem (23) can be restated as

$$g(\nu) = \frac{1}{\pi_{\text{comm}}} \min_{O,U} \Big[\int_{\mathcal{S}_{\text{wait}}} \Pi_{O,U}(s) \ell_\nu^*(s; O(s)) \mathrm{d}s + \int_{\mathcal{S}_{\text{comm}}} \Pi_{O,U}(s) \ell_\nu^*(s; U(s)) \mathrm{d}s \Big], \tag{24}$$

where ℓ_ν^* is the Lagrangian metric optimized with respect to the *inner action* components not specified by O and U. In particular, for waiting states $s = (r_U, w_0)$ and radial velocity $O(s) = v_r$, the inner optimization is performed with respect to the UAV angular velocity θ_c, i.e.,

$$\ell_\nu^*(s; v_r) = \min_{\theta_c} \nu \left(P_{\text{mob}} \left(\sqrt{v_r^2 + r_U^2 \cdot \theta_c^2} \right) - P_{\text{avg}} \right) \Delta_0 \ \text{ s.t. } \ \sqrt{v_r^2 + r_U^2 \cdot \theta_c^2} \leq V_{\max}. \tag{25}$$

Since $\nu \geq 0$, $\Delta_0 > 0$, and P_{avg} are constant, the optimizer θ_c^* is the angular velocity minimizing the UAV power consumption for a given UAV radial velocity v_r and radius r_U, solvable through

exhaustive search. For communication states $s = (r_U, r, \psi)$, $\ell^*_\nu(s; \hat{r}_U)$ is determined by optimizing over the scheduling decision $\xi \in \{0, 1\}$ and, if $\xi = 1$, the trajectory $\mathbf{q}_U$ followed by the UAV, terminating at the radial position $\hat{r}_U$. To formalize it, let $\ell^*_\nu(s; \hat{r}_U, \xi)$ denote the optimized metric as a function of $\xi \in \{0, 1\}$. For $\xi = 1$ (D&F protocol),

$$\ell^*_\nu(s; \hat{r}_U, 1) = \min_{\Delta, \mathbf{q}_U, t_p} (1 - \nu P_{\text{avg}})\Delta + \nu \int_0^\Delta P_{\text{mob}} \left(\sqrt{r'_U(\eta)^2 + r_U^2(\eta) \cdot \theta'_U(\eta)^2} \right) \mathrm{d}\eta \tag{26}$$

$$\text{s.t. C.1, C.2, } \mathbf{q}_U(0) = (r_U, 0), \|\mathbf{q}_U(\Delta)\|_2 = \hat{r}_U, \tag{27}$$

where C.1 reflects the data payload constraints of the D&F protocol, C.2 is the maximum UAV speed constraint, and (27) enforces the UAV trajectory starting position and end radius. For $\xi = 0$ (direct transmission to the BS, hence $\hat{r}_U = r_U$[6]), using (18) we obtain $\ell^*_\nu(s; r_U, 0) = L/\bar{R}_{GB}(r)$, hence $\ell^*_\nu(s; r_U)$ is obtained by further minimizing over the scheduling decision $\xi \in \{0, 1\}$, yielding

$$\ell^*_\nu(s; \hat{r}_U) = \min_{\xi \in \{0,1\}} \{\ell^*_\nu(s; r_U, \xi)\} \cdot \mathbb{I}(\hat{r}_U = r_U) + \ell^*_\nu(s; \hat{r}_U, 1) \cdot \mathbb{I}(\hat{r}_U \neq r_U). \tag{28}$$

Thus, if the outer decision selects $U(s) = r_U$, the inner scheduling decision $\xi \in \{0, 1\}$ is obtained by greedily minimizing a cost metric that trades off communication delay and energy consumption, i.e., direct transmission to the BS occurs if $\ell^*_\nu(s; r_U, 0) < \ell^*_\nu(s; r_U, 1)$. Otherwise, the UAV relays using the D&F protocol, and the inner decision UAV trajectory greedily minimizes the delay-energy tradeoff, terminating at the target radius of the outer decision, $U(s) = \hat{r}_U$.

The optimization problem (26) will be the focus of Sec. V. We now describe Algorithm 1, which designs the outer-policy and computes the average cost-per-stage metric, $g(\nu)$, along with the average energy- and time-per-stage metrics for a given ν, by solving problem (24) via value iteration, and Algorithm 2, which solves the dual maximization $\max_{\nu \geq 0} g(\nu)$ via projected subgradient ascent [40].

Specifically, in Algorithm 1:[7] lines 2 and 3 compute the inner Lagrangian cost metric optimized with respect to the inner actions, along with the energy and time cost metrics, for all states and outer actions; line 6 computes the value iteration update for waiting states, which accounts for the transition to the waiting state $(r_U + v_r \cdot \Delta_0, w_0)$ (if no transmission request is received, w.p.

[6]Note that with direct transmission to the BS the system immediately transitions to the waiting state (r_U, w_0) with the UAV at distance r_U from the center, hence $\hat{r}_U = r_U$ in this case.

[7]Note that, although the pseudo-algorithm is defined on continuous state and action spaces, in a practical implementation these spaces are discretized, so that integrals are replaced by sums.

Algorithm 1 Value Iteration: $(O^*, U^*, g(\nu), \bar{E}, \bar{T}) = \mathrm{VITER}(\nu)$

1: **init:** i=0; the value function $V_i(s)$=0, total energy cost $E_i(s)$=0, and total time cost $T_i(s)$=0, for all states $s \in \mathcal{S}$; stopping criterion δ.
2: $\forall s \in \mathcal{S}_{\mathrm{wait}}, \forall v_r \in [-V_{\max}, V_{\max}]$: calculate $\ell^*_\nu(s; v_r)$ as in (25), with minimizer θ^*_c; compute energy cost $e^*(s; v_r) = E(s; v_r, \theta^*_c)$ and time cost $t^*(s; v_r) = T(s; v_r, \theta^*_c)$.
3: $\forall s \in \mathcal{S}_{\mathrm{comm}}, \forall \hat{r}_U \in [0, a]$: calculate $\ell^*_\nu(s; \hat{r}_U)$ as in (28) and (26), with minimizer $(\xi^*, \hat{\theta}^*_U, \mathbf{q}^*_U, \Delta^*)$; compute energy cost $e^*(s; \hat{r}_U) = \xi^* E(s; \hat{r}_U, \hat{\theta}^*_U, \mathbf{q}^*_U)$ and time cost $t^*(s; \hat{r}_U) = \xi^* T(s; \hat{r}_U, \hat{\theta}^*_U, \mathbf{q}^*_U)$ as in (19).
4: **repeat**
5: **for** each $s = (r_U, w_0) \in \mathcal{S}_{\mathrm{wait}}$ **do**
6: $V_{i+1}(s) \leftarrow \min_{v_r \in [-V_{\max}, V_{\max}]} \big[\ell^*_\nu(s; v_r) + e^{-\Lambda\Delta_0} \cdot V_i(r_U + v_r \cdot \Delta_0, w_0)$
$+ \big(1 - e^{-\Lambda\Delta_0}\big) \cdot \int_0^{2\pi} \frac{1}{2\pi} \int_0^a \frac{2r'}{a^2} \cdot V_i(r_U + v_r \cdot \Delta_0, r', \psi') \mathrm{d}r' \mathrm{d}\psi' \big]$ and $O_{i+1}(s)$ is the $\arg\min$.
7: Update energy cost: $E_{i+1}(s) \leftarrow e^*(s; O_{i+1}(s)) + e^{-\Lambda\Delta_0} \cdot E_i(r_U + O_{i+1}(s) \cdot \Delta_0, w_0)$
$+ \big(1 - e^{-\Lambda\Delta_0}\big) \cdot \int_0^{2\pi} \frac{1}{2\pi} \int_0^a \frac{2r'}{a^2} \cdot E_i(r_U + O_{i+1}(s) \cdot \Delta_0, r', \psi') \mathrm{d}r' \mathrm{d}\psi'$.
8: Update time cost: $T_{i+1}(s) \leftarrow t^*(s; O_{i+1}(s)) + e^{-\Lambda\Delta_0} \cdot T_i(r_U + O_{i+1}(s) \cdot \Delta_0, w_0)$
$+ \big(1 - e^{-\Lambda\Delta_0}\big) \cdot \int_0^{2\pi} \frac{1}{2\pi} \int_0^a \frac{2r'}{a^2} \cdot T_i(r_U + O_{i+1}(s) \cdot \Delta_0, r', \psi') \mathrm{d}r' \mathrm{d}\psi'$.
9: **end for**
10: **for** each $s = (r_U, r, \psi) \in \mathcal{S}_{\mathrm{comm}}$ **do**
11: $V_{i+1}(s) \leftarrow \min_{\hat{r}_U \in [0,a]} [\ell^*_\nu(s; \hat{r}_U) + V_i(\hat{r}_U, w_0)]$ and $U_{i+1}(s)$ is the $\arg\min$.
12: Update energy cost: $E_{i+1}(s) \leftarrow e^*(s; U_{i+1}(s)) + E_i(U_{i+1}(s), w_0)$;
13: Update time cost: $T_{i+1}(s) \leftarrow t^*(s; U_{i+1}(s)) + T_i(U_{i+1}(s), w_0)$.
14: **end for**
15: $\forall s$, calculate the stopping criterion metric $H(s) = V_{i+1}(s) - V_i(s)$; $i \leftarrow i + 1$.
16: **until** $\max_{s \in \mathcal{S}} H(s) - \min_{s \in \mathcal{S}} H(s) < \delta$.
17: Approximate $\big[g(\nu); \bar{E}; \bar{T}\big] \approx \frac{1}{\pi_{\mathrm{comm}}} \cdot \frac{1}{i} \left[V_i(s); E_i(s); T_i(s)\right]$, for some arbitrary $s \in \mathcal{S}$.
18: **return** $O^*(s) = O_i(s), \forall s \in \mathcal{S}_{\mathrm{wait}}$, $U^*(s) = U_i(s), \forall s \in \mathcal{S}_{\mathrm{comm}}$, $g(\nu)$, $\bar{E}$, and $\bar{T}$.

$e^{-\Lambda\Delta_0}$), and to a communication state $(r_U + v_r \cdot \Delta_0, r', \psi')$ (if a transmission request is received in position r', ψ', with pdf $\frac{1}{2\pi}\frac{2r'}{a^2}$); line 11 computes the value iteration update for communication states, which, for end radius $\hat{r}_U$, accounts for the transition to the waiting state $(\hat{r}_U, w_0)$, w.p. 1; lines 7,8, 12 and 13 update the corresponding total energy and time costs; line 17 estimates the values of the average cost-, energy-, and time-per-stage metrics.

In Algorithm 2, line 1 initializes the dual variable and a sequence of step-sizes used for projected dual subgradient ascent; line 3 calls the value iteration of Algorithm 1 using the current dual variable value, ν_k, and outputs the optimal radial velocity (for waiting states) and next radius position (for communication states) policies, as well as the average cost-, energy-, and time-per-stage metrics; line 4 monitors convergence, checking the change in the average cost-per-stage value, as well as primal feasibility and complementary slackness, with respect to

Algorithm 2 Projected Subgradient Ascent Algorithm

1: **init:** k=0; dual variable $\nu_0 \geq 0$; step-size $\{\rho_j = \rho_0/(j+1), j \geq 0\}$; $g_{-1} = \infty$.
2: **for** $k = 0, 1, \ldots$ **do**
3: Determine $(O_k^*, U_k^*, g_k, \bar{E}_k, \bar{T}_k) = \text{VITER}(\nu_k)$ via Algorithm 1.
4: **if** $|g_k - g_{k-1}| < \epsilon_{DI}$; $\bar{E}_k - P_{\text{avg}}\bar{T}_k < \epsilon_{PF}$; $\nu_k|\bar{E}_k - P_{\text{avg}}\bar{T}_k| < \epsilon_{CS}$ **then**
5: **return:** optimal outer policy (O_k^*, U_k^*);
6: **else**
7: Update $\nu_{k+1} = \max\left\{\nu_k + \rho_k \cdot \left(\bar{E}_k - P_{\text{avg}}\bar{T}_k\right), 0\right\}$; $k \leftarrow k+1$.
8: **end if**
9: **end for**

the average UAV power constraint; line 7 updates the value of the dual variable in the direction of its subgradient and projects the value to the non-negative range to ensure dual feasibility.

V. Trajectory Optimization via Hierarchical Competitive Swarm Optimization

In this section, we design the UAV trajectory during the D&F protocol by solving (26). The literature solves problems like (26) with SCA (see [7], [8]), in which the non-convex objective and constraints are approximated by convex functions around a local point; the approximated problem is solved with convex optimization methods, the minimizers become the next local point, convex approximations are made about the new point, and the process repeats until convergence. However, SCA has major drawbacks: 1) it is computationally slow, as shown numerically in Sec. VI; 2) its data payload constraints require closed-form expressions, which are not available in our model due to the complex propagation environment and rate adaptation. In fact, the numerical simulations of SCA for solving (26), as seen in Fig. 2a of Sec. VI, are peformed for $z_1 = 0$ and $K = \infty$ (deterministic LoS), so as to yield closed-form throughput expressions.[8]

To overcome the challenges of SCA, we propose an optimization framework based on CSO [24]. Since the design space involves optimizing over infinitely many variables $(r_U(t), \theta_U(t))$, we simplify the continuous UAV trajectory into a finite sequence of waypoints connected by straight lines at constant speed. However, a direct application of CSO suffers from poor convergence (see Fig. 2a in Sec. VI), because solution spaces of problems tend to exponentially increase with the dimensionality [41]. We address this weakness by proposing a *Hierarchical*-CSO (HCSO) algorithm, in which a sequence of CSO problems is solved: initially CSO produces a low-resolution trajectory; the optimized trajectory is then used to create a higher-resolution trajectory

[8]Note that our SCA formulation follows [8], but requires a final end *radius* constraint in place of an end coordinate constraint.

via interpolation, which is further optimized with CSO; and so on. This sequence of CSO problems generate trajectories of increasing resolution and helps alleviate the poor performance achieved when optimizing the high-resolution trajectory directly, as demonstrated in Fig. 2a.

In Sec. V-A, we formulate the meta-heuristic UAV trajectory to be solved with HCSO and convert the constrained problem into an *unconstrained* one, so as to make the problem tractable for the CSO algorithm [24]. In Sec. V-B, we design the hierarchical framework for HCSO, and specifically show how high-dimensional trajectories leverage information from low-dimensional ones to improve the solution accuracy.

A. Formulation of the Meta-Heuristic UAV Trajectory

To define the UAV meta-heuristic trajectory, we use Cartesian coordinates instead of polar coordinates, for easier evaluation of the data payload constraints. For a state $s = (r_U, r, \psi) \in \mathcal{S}_{\text{comm}}$ and a *next radius position* $U(s) = \hat{r}_U$, we encode a trajectory as a sequence of waypoints and UAV speeds between two consecutive waypoints. Specifically, we define a meta-heuristic trajectory of $\mathbf{q}_U(t)$ by a sequence of M trajectory segments, $\{\mathbf{x}_m\}_{m=0}^{M} = \{(x_m, y_m)\}_{m=0}^{M}$. The UAV trajectory begins at $\mathbf{x}_0$ and ends at $\mathbf{x}_M$; the UAV traverses each straight trajectory segment $\Psi_m \triangleq \mathbf{x}_{m+1} - \mathbf{x}_m$ with speed $v_m \in [V_{\text{low}}, V_{\max}]$, with minimum speed $V_{\text{low}} \ll V_{\max}$ introduced to ensure well-defined durations; the sequences of trajectory waypoints $\mathbf{p} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_M]^T$ and speeds $\mathbf{v} \triangleq [v_0, \ldots, v_{M-1}]^T$ are the optimization variables. For M trajectory segments, the first and second $M/2$ segments correspond to the two-phase D&F protocol. Because the throughput equations are angular-invariant, without loss of generality, the initial UAV location is set to $\mathbf{x}_0 = (r_U, 0)$ and the GN request location to $\mathbf{x}_G \triangleq (r\cos\psi, r\sin\psi)$.

Since the number of bits communicated (see (C.1)) during each trajectory segment cannot be computed in closed-form, we approximate them for each trajectory segment numerically. Specifically, between subsequent waypoints, $\mathbf{x}_m$ and $\mathbf{x}_{m+1}$ with speed v_m, we generate a sequence of n_{res} evenly-spaced points with a sufficiently high resolution, i.e., $\{\mathbf{x}_k^{\text{new}}\}_{k=0}^{n_{\text{res}}}$; the time durations between the sequence of points, $\{t_k^{\text{new}}\}_{k=0}^{n_{\text{res}}-1}$, and expected throughput (corresponding to D&F protocol) at each point, $\{R_k^{\text{new}}\}_{k=0}^{n_{\text{res}}-1}$, then determine the number of bits communicated for the mth trajectory segment by calculating the lower Riemann sum, given by $F_m \triangleq \sum_{k=0}^{n_{\text{res}}-1} t_k^{\text{new}} \cdot R_k^{\text{new}}$.

With the number of communicated bits evaluated for each trajectory segment, we present the

heuristic problem of $\ell_\nu^*(s;U(s))$ to be solved with HCSO. The problem in standard form is

$$(\mathbf{P}.\mathbf{0}) \quad \min_{\mathbf{p},\mathbf{v}\in[V_{\rm low},V_{\rm max}]^M} (1-\nu P_{\rm avg}) \sum_{m=0}^{M-1} \frac{\|\Psi_m\|_2}{v_m} + \nu \sum_{m=0}^{M-1} \frac{\|\Psi_m\|_2}{v_m} P_{\rm mob}(v_m) \tag{29}$$

$$\text{s.t. } h_i(\mathbf{p},\mathbf{v}) \triangleq L - \sum_{m=\frac{M}{2}(i-1)}^{\frac{M}{2}(i-1)+\frac{M}{2}-1} F_m \leq 0, \;\; i=1,2, \tag{30}$$

$$\mathbf{x}_0 = (r_U, 0), \; \mathbf{x}_M = \hat{r}_U \frac{\mathbf{x}_{M-1}}{\|\mathbf{x}_{M-1}\|_2} \cdot \mathbb{I}(\|\mathbf{x}_{M-1}\|_2 \neq 0) + (\hat{r}_U, 0) \cdot \mathbb{I}(\|\mathbf{x}_{M-1}\|_2 = 0), \tag{31}$$

where $h_i(\mathbf{p},\mathbf{v}) \leq 0$ ensure the data payload constraints and (31) ensures the end radius constraint is satisfied by projecting the waypoint $\mathbf{x}_{M-1}$ to the nearest point with radius $\hat{r}_U$, and $\|\Psi_m\|_2/v_m$ is the duration of the mth trajectory segment. Thus, the final waypoint variable $\mathbf{x}_M$ is eliminated, and $\mathbf{x}_0$ is fixed.

As introduced in Sec. I, we solve $(\mathbf{P}.\mathbf{0})$ with HCSO, as it promises to provide results efficiently. The HCSO method is executed by converting the constrained problem $(\mathbf{P}.\mathbf{0})$ into an *unconstrained* one and solving a sequence of unconstrained problems with CSO. We remove the data payload constraints by adding them to the objective function as penalties for constraint violations by enforcing a particular solution: if the UAV does not receive (or transmit) its data payload at the end of either phase, then it must hover at the current position until it receives (or transmits) the data payload. This penalized objective function is

$$\hat{f}(\mathbf{p},\mathbf{v}) \triangleq (1-\nu P_{\rm avg}) \sum_{m=0}^{M-1} \frac{\|\Psi_m\|_2}{v_m} + \nu \sum_{m=0}^{M-1} \frac{\|\Psi_m\|_2}{v_m} P_{\rm mob}(v_m) \tag{32}$$
$$+ (1-\nu P_{\rm avg})(\hat{t}_{P,1} + \hat{t}_{P,2}) + \nu(\hat{E}_{P,1} + \hat{E}_{P,2}),$$

where the time delay penalties for hovering in the first and second phases, respectively, are

$$\hat{t}_{P,1} \triangleq h_1(\mathbf{p},\mathbf{v}) / \bar{R}_{GU}\left(\|\mathbf{x}_{M/2} - \mathbf{x}_G\|_2\right) \cdot \mathbb{I}\left(h_1(\mathbf{p},\mathbf{v}) > 0\right), \tag{33}$$

$$\hat{t}_{P,2} \triangleq h_2(\mathbf{p},\mathbf{v}) / \bar{R}_{UB}\left(\|\mathbf{x}_M\|_2\right) \cdot \mathbb{I}(h_2(\mathbf{p},\mathbf{v}) > 0), \tag{34}$$

and the energy penalties for hovering, respectively, are $[\hat{E}_{P,1}; \hat{E}_{P,2}] \triangleq P_{\rm mob}(0) \cdot [\hat{t}_{P,1}; \hat{t}_{P,2}]$.

B. Solution of ($\mathbf{P}.\mathbf{0}$) via the HCSO Method

To solve the problem with HCSO, we begin by randomly initializing a population (swarm) of M-segment trajectory solutions (particles). Specifically, in the first call to the CSO algorithm

[24] (see Algorithm 4) inside the HCSO framework, we initialize N trajectory particles $\mathbf{p}_{1:N} \triangleq \mathbf{p}_1, \ldots, \mathbf{p}_N$, with particle velocities $\mathbf{u}_{1:N} \triangleq \mathbf{u}_1, \ldots, \mathbf{u}_N$, and N UAV speed particles $\mathbf{v}_{1:N} \triangleq \mathbf{v}_1, \ldots, \mathbf{v}_N$, with particle velocities $\mathbf{w}_{1:N} \triangleq \mathbf{w}_1, \ldots, \mathbf{w}_N$, imposing no trajectory structure in the first iteration. After the first CSO iteration, the resulting trajectory is interpolated to form a reference $2M$-segment trajectory, $\tilde{\mathbf{p}} = [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_{M-1}]^T$ and $\tilde{\mathbf{v}} = [\tilde{v}_0, \ldots, \tilde{v}_{M-1}]^T$, with $M \leftarrow 2M$, so as to increase the resolution for the next iteration. The new swarm size is then reduced, $N \leftarrow N - N_{\text{red}}$, to lower the computational burden of CSO, and a new swarm of trajectories is generated randomly in a *neighborhood* around the reference trajectory. More precisely, the mth waypoint, $\mathbf{x}_m = (x_m, y_m)$, of a new trajectory is generated according to

$$\mathbf{x}_m = \tilde{\mathbf{x}}_m + (\chi_m, \zeta_m), \tag{35}$$

where the localized randomness in the new waypoints is modeled using zero-mean Gaussian random variables, $\chi_m, \zeta_m \sim \mathcal{N}\left(0, \varsigma\left(\|\tilde{\mathbf{x}}_{m+1} - \tilde{\mathbf{x}}_m\|^2 + \|\tilde{\mathbf{x}}_{m-1} - \tilde{\mathbf{x}}_m\|^2\right)\right)$, where $\varsigma > 0$ is a scaling factor, and where the variance is determined by the variation between neighboring reference trajectory waypoints, selected due to the empirical observation that areas with clustered UAV waypoints are the regions where the objective function (32) is sensitive to large variations. Likewise, the UAV speed of the mth new trajectory segment is generated by

$$v_m = [\tilde{v}_m + \varkappa_m]^{[V_{\text{low}}, V_{\text{max}}]}, \tag{36}$$

where $\varkappa_m \sim \mathcal{N}\left(0, \varepsilon\left(V_{\text{max}} - V_{\text{low}}\right)^2\right)$ induces randomness in the new speed, $\varepsilon > 0$ is a scaling factor, and $[\cdot]^{[V_{\text{low}}, V_{\text{max}}]} = \max\left\{\min\left\{\cdot, V_{\text{max}}\right\}, V_{\text{low}}\right\}$ projects speed v_m to the feasible set. The variance is chosen due to the observation that the UAV speeds exhibit consistent convergence with CSO much faster than the trajectory waypoints and are less sensitive to random initializations.

The HCSO framework is shown in Algorithm 3. Line 1 initializes N random M-segment UAV trajectories; line 3 obtains an M-segment trajectory with CSO [24] (Algorithm 4); line 4 doubles the trajectory resolution and interpolates the CSO output to a $2M$-segment reference trajectory; lines 8 and 11 generate random trajectories in a neighborhood of the reference trajectory, $(\tilde{\mathbf{p}}, \tilde{\mathbf{v}})$. The process repeats until the trajectory resolution equals a desired value.

The CSO algorithm follows [24] and is shown in Algorithm 4. During the kth iteration inside CSO, the N particles are randomly grouped into $N/2$ pairwise competitions. For both pair members, the penalized objective function in (32) is calculated; the winner of the competition

Algorithm 3 HCSO Algorithm for $\ell_\nu^*(s; U(s))$

Randomly initialize UAV waypoints $\mathbf{p}_{1:N}$ and speeds $\mathbf{v}_{1:N}$.
while $M \leq M_{\max}$ **do**
  Obtain M-segment trajectory: $(\mathbf{p}^*, \mathbf{v}^*) = \mathrm{CSO}(\mathbf{p}_{1:N}, \mathbf{v}_{1:N}, N, M)$ (see Algorithm 4).
  Increase $M \leftarrow 2M$; interpolate to form reference trajectory: $(\tilde{\mathbf{p}}, \tilde{\mathbf{v}}) = \mathrm{interp}(\mathbf{p}^*, \mathbf{v}^*, M)$.
  Reduce swarm size $N \leftarrow N - N_{\mathrm{red}}$.
  for $n_1 = 1, 2, \ldots, N$ **do**
    for $n_2 = 1, 2, \ldots, M-1$ **do**
      Extract $\tilde{\mathbf{x}}_{n_2} = (\tilde{x}_{n_2}, \tilde{y}_{n_2})$ from reference $\tilde{\mathbf{p}}$ and generate $\mathbf{x}_{n_2}$ of $\mathbf{p}_{n_1}$ via (35).
    end for
    for $n_3 = 0, \ldots, M-1$ **do**
      Extract $\tilde{v}_{n_3}$ from reference $\tilde{\mathbf{v}}$ and generate v_{n_3} of $\mathbf{v}_{n_1}$ via (36).
    end for
    Update new swarm particle $\mathbf{p}_{n_1} \leftarrow [\mathbf{x}_1, \ldots, \mathbf{x}_{M-1}]^T$ and $\mathbf{v}_{n_1} \leftarrow [v_0, \ldots, v_{M-1}]^T$.
  end for
end while

Algorithm 4 CSO Algorithm for ($\mathbf{P}.0$): $(\mathbf{p}^*, \mathbf{v}^*) = \mathrm{CSO}\left(\mathbf{p}_{1:N}(0), \mathbf{v}_{1:N}(0), N, M\right)$

init: $k = 0$ and particle velocities $\mathbf{u}_{1:N}(0)$ and $\mathbf{w}_{1:N}(0)$ for $\mathbf{p}_{1:N}(0)$ and $\mathbf{v}_{1:N}(0)$, respectively.
repeat
  Shuffle particle indices with a random permutation: $t(i) : \{1, \ldots, N\} \mapsto \{1, \ldots, N\}$.
  for $j = 1, 3, 5, \ldots, N-1$ **do**
    if $\hat{f}(\mathbf{p}_{t(j)}(k), \mathbf{v}_{t(j)}(k)) \leq \hat{f}(\mathbf{p}_{t(j+1)}(k), \mathbf{v}_{t(j+1)}(k))$ **then**
      $(\mathbf{p}^w, \mathbf{v}^w) = (\mathbf{p}_{t(j)}(k), \mathbf{v}_{t(j)}(k))$; $(\mathbf{p}^l, \mathbf{v}^l) = (\mathbf{p}_{t(j+1)}(k), \mathbf{v}_{t(j+1)}(k))$
      $j_{\mathrm{win}} = t(j)$; $j_{\mathrm{los}} = t(j+1)$.
    else
      $(\mathbf{p}^w, \mathbf{v}^w) = (\mathbf{p}_{t(j+1)}(k), \mathbf{v}_{t(j+1)}(k))$; $(\mathbf{p}^l, \mathbf{v}^l) = (\mathbf{p}_{t(j)}(k), \mathbf{v}_{t(j)}(k))$
      $j_{\mathrm{win}} = t(j+1)$; $j_{\mathrm{los}} = t(j)$.
    end if
    Pass the winner to next iteration: $(\mathbf{p}_{j_{\mathrm{win}}}(k+1), \mathbf{v}_{j_{\mathrm{win}}}(k+1)) \leftarrow (\mathbf{p}^w, \mathbf{v}^w)$.
    Update loser particle velocities $(\mathbf{u}_{j_{\mathrm{los}}}(k+1), \mathbf{w}_{j_{\mathrm{los}}}(k+1))$ using equations (37)–(38).
    Update loser particles $(\mathbf{p}_{j_{\mathrm{los}}}(k+1), \mathbf{v}_{j_{\mathrm{los}}}(k+1))$ using equation (39).
  end for
  $k \leftarrow k+1$.
until maximum number of cost function evaluations performed.
return $(\mathbf{p}^*, \mathbf{v}^*) = \mathrm{argmin}\{\hat{f}(\mathbf{p}_i(k), \mathbf{v}_i(k)),\ \forall i = 1, \ldots, N\}$

is passed into the $(k+1)$th iteration of CSO, and the loser is modified by learning from the winner. Let $j_{\mathrm{win}}, j_{\mathrm{los}} \in \{1, \ldots, N\}$ denote the indices of the winner and loser from the swarm during the jth pairwise competition of the kth iteration of CSO, and denote the winner and loser

particles by $(\mathbf{p}^w, \mathbf{v}^w)$ and $(\mathbf{p}^l, \mathbf{v}^l)$, respectively. The loser particle velocities are updated as

$$\mathbf{u}_{j_{\rm los}}(k+1) = r_{j,1}(k)\mathbf{u}^l + r_{j,2}(k)\left(\mathbf{p}^w - \mathbf{p}^l\right) + \omega \cdot r_{j,3}(k)\left(\bar{\mathbf{p}}(k) - \mathbf{p}^l\right), \tag{37}$$

$$\mathbf{w}_{j_{\rm los}}(k+1) = r_{j,1}(k)\mathbf{w}^l + r_{j,2}(k)\left(\mathbf{v}^w - \mathbf{v}^l\right) + \omega \cdot r_{j,3}(k)\left(\bar{\mathbf{v}}(k) - \mathbf{v}^l\right), \tag{38}$$

and the losing particle is updated into the $(k+1)$th CSO iteration as

$$\mathbf{p}_{j_{\rm los}}(k+1) = \mathbf{p}^l + \mathbf{u}_{j_{\rm los}}(k+1), \;\; \mathbf{v}_{j_{\rm los}}(k+1) = \left[\mathbf{v}^l + \mathbf{w}_{j_{\rm los}}(k+1)\right]^{[V_{\rm low}, V_{\rm max}]}, \tag{39}$$

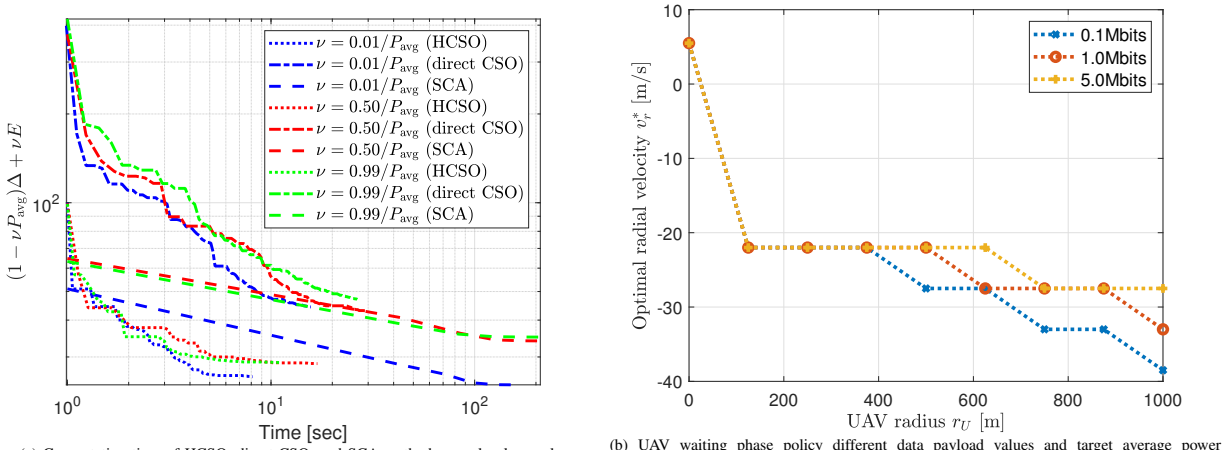
where $r_{j,1}(k), r_{j,2}(k), r_{j,3}(k)$ are independent uniform random variables on the interval $[0,1]$, $\bar{\mathbf{p}}(k)$ and $\bar{\mathbf{v}}(k)$ are the global means of the UAV waypoint and speed particles in the kth CSO iteration, and ω is a scale parameter for the degree of influence of $\bar{\mathbf{p}}(k)$ and $\bar{\mathbf{v}}(k)$.

VI. Numerical Results

For the simulation, we use a channel bandwidth of B=1MHz, NLoS attenuation constant κ=0.2, 1-meter reference SNRs $\frac{\beta_0 P}{\sigma^2 \Gamma} \triangleq \gamma_{GB}$=$\gamma_{GU}$=$\gamma_{UB}$=40dB, pathloss exponents α=$\tilde{\alpha}$=2 for all links, Rician K-factor parameters k_1=1 and k_2=$\frac{\ln(100)}{90}$ ($0 \leq K \leq$ 20dB) for all links, LoS probability parameters z_1=9.61 and z_2=0.16 for all links (urban environment [17]) except the UAV→BS link, where z_1=0, i.e., $P_{\rm LoS}$=1, UAV height H_U=120m, BS antenna height H_B=60m, maximum UAV speed $V_{\max}$=55m/s, cell radius a=1000m, and total Poisson rate Λ=0.0085req/sec. The UAV power consumption model uses the relationship in (6) and the parameters in [8]. We solve an approximation of problem (24) by discretizing the state and action spaces and applying linearly-interpolated value iteration. The dual variable yields meaningful results for $\nu \in [0, 1/P_{\rm avg}]$, thus by solving the problem for 20 values in the interval, an approximation to the maximum, $g(\nu^*)$, emerges. We discretize the states with $N_{\rm sp}$=9 equispaced radii values, one GN in the center of the cell, $K_{\rm sp}$=3 equispaced angular GN positions in the next radius, $2K_{\rm sp}$ in the next, and so on, so that the GN distribution is approximately uniform-circular. We use $R_{\rm sp}$=21 equispaced *radial velocity* actions, $v_r \in \{-V_{\max}, \ldots, , V_{\max}\}$. For *next radius position* actions, we use the same $N_{\rm sp}$ radii values. Lastly, Δ_0 is chosen to satisfy $e^{-\Lambda\Delta_0}$=0.93, so that it is unlikely to receive two or more requests in Δ_0 seconds.

A. Comparison of HCSO and SCA Optimization Schemes

To compare HCSO to SCA and non-hierarchical CSO for solving the inner Lagrangian cost metric, $\ell^*_\nu(s; U(s))$, we choose state s=$(800, 500, \pi/4)$[m,m,rad], end radius $U(s)$=700m, target



(a) Computation time of HCSO, direct CSO, and SCA methods on a log-log scale.

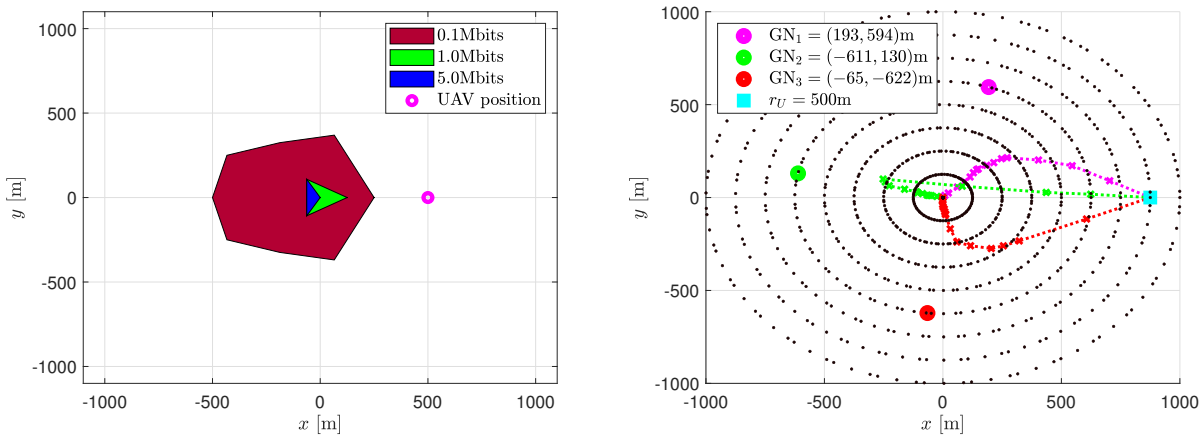
(b) UAV waiting phase policy different data payload values and target average power P_{avg}=1.1kW.

Fig. 2: HCSO/SCA comparison (a) and optimal UAV waiting phase trajectory policy (b).

average power P_{avg}=1.1kW, and data payload 1Mbits. For HCSO (Algorithm 3), we make 3 calls to CSO, doubling the trajectory resolution each iteration as $M_{\text{HCSO}}(i)=2^{i+1}$, for $i=1,2,3$; we decrease the swarm size after each CSO call as $N_{\text{HCSO}}(i)=180-20\cdot i$, for $i=1,2,3$. Whereas SCA approximates the number of bits communicated in the mth trajectory segment by using the throughput at waypoint $\mathbf{x}_m$ and time duration t_m [8], our HCSO formulation evaluates this integral more finely by interpolating points between waypoints (with resolution of 20m). Thus for SCA, we use $M_{\text{SCA}}=100$ trajectory segments so as to compensate the discretization error. Numerical simulations of SCA are peformed with $z_1=0$ and $K=\infty$ (deterministic LoS), in order to yield closed-form throughput expressions, and hence converge to different values than HCSO. In Fig. 2a, we observe that the HCSO realizations converge up to $21\times$ faster in time. Thus, HCSO is preferable to SCA for time efficiency, considering that 8829 realizations must be solved with the SMDP discretization for fixed ν and P_{avg}. Finally, Fig. 2a shows that HCSO finds solutions that reduce the objective function value by up to 42%, when compared to non-hierarchical CSO, which supports the improved convergence properties of our proposed hierarchical formulation.

B. Optimal Policy Structure

To observe the optimal *waiting* phase behavior, we fix P_{avg}=1.1kW and vary the data payload value. Fig. 2b shows that the UAV moves to an optimal flying radius ($\leq$100m from the center) during *waiting* phases to address two considerations: 1) to be well-positioned for future transmission requests; 2) to fly at the power-minimizing speed so as to reduce UAV energy consumption. Near the cell edge, the UAV radial speed increases for decreasing data payload, as smaller data packets relaying through the UAV complete transmission faster than large ones, thus the UAV



(a) Map of GNs that transmit directly to the BS under the optimal policy for P_{avg}=1.1kW. (b) UAV trajectories for 0.1Mbit data payload, r=625m, r_U=875m, and P_{avg}=1.4kW.

Fig. 3: Optimal UAV waiting phase trajectory (a) and communication strategy (b).

is less likely to spend time following trajectories that place it near the cell edge and can afford the additional power consumption required to fly at high speeds when this is the case.

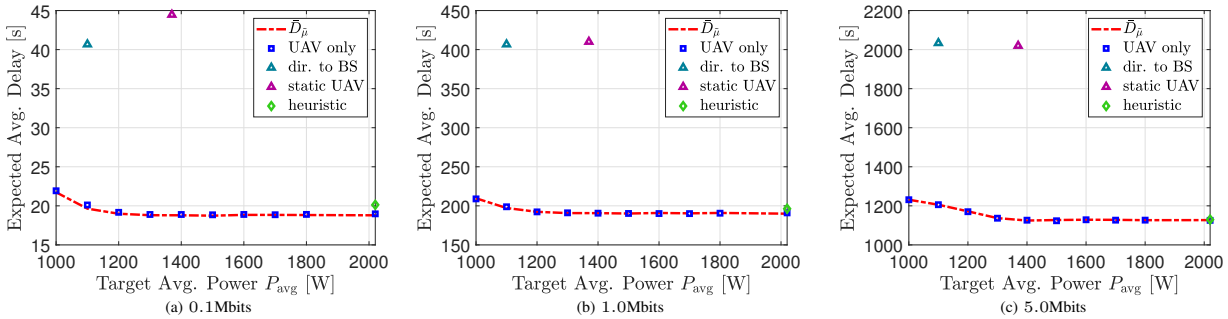
Fig. 3a shows which GNs transmit directly to the BS under the optimal policy $\tilde{\mu}$ during the *communication* phases. The target average power is P_{avg}=1.1kW, and the UAV radius is r_U=500m. One fifth of the GNs transmit directly to the BS for small data payloads, e.g., 0.1Mbits, and as the data payload grows large (1.0 and 5.0Mbits), the benefit of relaying through the UAV and utilizing its mobility increases, so that the majority of GNs relay through the UAV.

Next, we observe the UAV trajectory in *communication* phases for several realizations. In Fig. 3b, we fix the target average power P_{avg}=1.4kW, initial UAV radius r_U=875m, GN transmission request radius r=625m, and data payload value 0.1Mbits, and show the optimal trajectories for various UAV-GN angles. In all cases, the UAV flies at a constant speed near $V_{\max}$=55m/s. From the initial position, the UAV does not fly directly toward the GN of interest, due to considering both phases of D&F protocol. For the various realizations, the target end radius of the UAV is near the center of the cell ($\hat{r}_U \leq 100$m) to minimize the long-term delay-power tradeoff.

C. Optimal Performance

To analyze the communication delay performance, we simulate over various data payloads and target average power constraints, P_{avg}, to observe a range of behavior. The expected average delay is shown in Figs. 4a–4c. Note that the expected average delay, $\bar{D}_{\tilde{\mu}}$, is not directly accessible through the value iteration analysis, so we apply the optimal policy $\tilde{\mu}$ in a simulation, sampling a sequence of several thousand random transmission requests, in order to obtain it. Using the transmission request sequence, we compare $\bar{D}_{\tilde{\mu}}$ against *four* heuristics, explained as follows:

(1) *UAV only*: all actions remain the same, except GNs may not directly transmit to the BS



(a) 0.1Mbits (b) 1.0Mbits (c) 5.0Mbits

Fig. 4: Expected average communication delay vs. P_{avg} for optimal policy $\tilde{\mu}$, various heuristics, and continuous-space simulation.

(2) *Direct to BS*: no UAV is deployed; all GNs transmit directly to the BS.[9]

(3) *Static UAV*: the UAV hovers at the optimized radius r_U^*=459m; GN transmission requests may directly transmit to the BS or relay through the UAV so as to minimize delay.

(4) *Heuristic*: scheduling decisions greedily minimize the delay incurred to serve transmission requests, without considering long-term dynamics; if the UAV is chosen for relaying, it executes a trajectory to minimize the communication delay; the UAV hovers statically after communication phases until receiving a new request.

Across the range of data payloads and target average powers, the optimal policy simulation reduces the average communication delay by up to 54% when compared to the *direct to BS* heuristic. The UAV consumes 1.37kW to hover following the *static UAV* scheme; whereas for the optimal policy with P_{avg}=1.1kW, the UAV consumes 1.29, 1.36, and 1.26kW for the 0.1, 1.0, 5.0Mbits data payloads, respectively, reducing the average communication delay by 51%, 51%, and 41%, respectively, thus showing that optimally exploiting UAV mobility can reduce both communication delay and power consumption. Next, the optimal policy slightly reduces the communication delay (up to 3%) with respect to the *UAV only* heuristic, which does not adaptively schedule transmission requests but exploits UAV mobility. The UAV power consumption is unconstrained in the heuristic scheme, but fails to exploit UAV mobility for long-term goals. For example, in the 0.1Mbits case (Fig. 4a), the UAV consumes 1.5kW on average for mobility following the *heuristic* scheme, yet the optimal strategy reduces both communication delay and power consumption for target average powers $P_{\text{avg}}\geq$1.1kW. When P_{avg}=1.1kW, for instance, the optimal strategy reduces the average communication delay and UAV mobility power by 3% and 14%, respectively. Note from Figs. 4a–4c: GNs performing direct transmission to the BS suffer poor performance for large data payloads, because far-away GNs experience harsh propagation

[9]Note that the UAV consumes 0W in this policy, but the data points in Figs. 4a–4c are at $P_{\text{avg}} = 1.1$kW to fit in the axes.

conditions. For example, a GN at the cell edge has an average LoS probability of $P_{\rm LoS}$=0.02.

VII. Conclusions

In this paper, we studied the adaptive trajectory and communication scheduling design of a power-constrained UAV serving as a relay for random data traffic between GNs and a BS. The problem was cast as a SMDP, and we showed that it exhibits a multiscale structure that can be efficiently optimized. Overall, we showed that by optimally exploiting realistic features of UAV mobility and A2G propagation, and by adaptively scheduling transmission requests, both the expected average communication delay and UAV mobility power consumption can be reduced when compared to an optimal hovering deployment and a heuristic delay minimization scheme.

Appendix A: Convexity of $-\ln f(z) - \ln Q_1\left(a_1, a_2 z\right)$

To show this, note that $Q_1\left(a_1, y\right)$ is log-concave in $y\geq 0$ [42], so that $-\ln Q_1\left(a_1, a_2 z\right)$ is convex in $z \geq 0$. We now show that $[-\ln f(z)]$ is convex, i.e., $\frac{\mathrm{d}^2}{\mathrm{d}z^2}[-\ln f(z)]\geq 0, \forall z\geq 0$. We have

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}[-\ln f(z)] \propto 2z^2 + \left(z^2-2\right)\ln\left(1+z^2/2\right) \triangleq h(z).$$

$h(z)\geq 0$ holds trivially for $z\geq\sqrt{2}$, and it can be shown to hold for $z \leq \sqrt{2}$ by using $\ln(1+x) \leq x$.

Appendix B: Proof of Lemma 1

Consider $\bar{W}_\mu$ in (8). If ξ_u=1, then additional requests received during the UAV relay phase are served directly by the BS, with delay $L/\bar{R}_{GB}(r)$ for a GN in position (r,θ). Hence, the expected average communication delay to serve these additional requests is $\mathbb{E}[\Delta_{u,i}^{(bs)}]=\bar{\Delta}_{BS}$, yielding

$$\bar{W}_\mu = \lim_{t\to\infty} \mathbb{E}_\mu\left[\frac{1}{M_t}\sum_{u=1}^{M_t}[\Delta_u^{(s)} + \xi_u N_u \bar{\Delta}_{BS}]\right] = \bar{\Delta}_{BS}(\bar{N}_\mu - 1) + \bar{W}_\mu^{(s)}, \tag{40}$$

where the second step uses the definition of $\bar{N}_\mu$ in (8). Then, it follows that

$$\bar{D}_\mu = \frac{\bar{W}_\mu}{\bar{N}_\mu} = \frac{1}{\bar{N}_\mu}\bar{W}_\mu^{(s)} + \left(1 - \frac{1}{\bar{N}_\mu}\right)\bar{\Delta}_{BS}. \tag{41}$$

Let μ_{BS} be the policy such that the UAV flies at the power-minimizing speed $V_{\min}$, and all requests are served by the BS. This policy is feasible (it uses minimum power) and its average delay is $\bar{D}_{\mu_{BS}} = \bar{\Delta}_{BS}$, since all requests are served directly by the BS. Therefore, under the optimal policy we must have $\bar{D}_\mu^* \leq \bar{D}_{\mu_{BS}} = \bar{\Delta}_{BS}$. Using (41) and the fact that $\bar{N}_\mu \geq 1$, this

implies that $\bar{W}_{\mu^*}^{(s)} \leq \bar{\Delta}_{BS}$. Let μ be any policy that satisfies $\bar{W}_{\mu}^{(s)} \leq \bar{\Delta}_{BS}$ (note that any other policy is suboptimal). Under such policy, (41) along with $\bar{N}_\mu \geq 1$, implies $\bar{W}_{\mu}^{(s)} \leq \bar{D}_\mu \leq \bar{\Delta}_{BS}$.

Moreover, if $\xi_u = 1$, then N_u requests are received during the *UAV relay phase* of duration $\Delta_u^{(s)}$; since these requests are received with an overall rate Λ, we find that $\mathbb{E}[N_u|\Delta_u^{(s)}] = \Delta_u^{(s)}\Lambda$, so that using the bound $\xi_u \leq 1$, we can express

$$\bar{N}_\mu = \lim_{t\to\infty} \mathbb{E}_\mu \left[\frac{1}{M_t} \sum_{u=1}^{M_t} (1 + \xi_u \Delta_u^{(s)} \Lambda) \right] \leq 1 + \Lambda \bar{W}_\mu^{(s)}, \tag{42}$$

with strict equality if the UAV always serves requests. Under any policy μ such that $\bar{W}_{\mu}^{(s)} \leq \bar{\Delta}_{BS}$, we can then bound as in (12), where the second inequality follows from $\bar{W}_{\mu}^{(s)} \leq \bar{\Delta}_{BS}$.

Appendix C: Evaluation of $\pi_{\rm comm}$

Lemma 2. *Let $\pi_{\rm comm}=\int_{\mathcal{S}_{\rm comm}} \Pi_\mu(s) {\rm d}s$ and $\pi_{\rm wait}=1-\pi_{\rm comm}$ be the steady-state probabilities that the UAV is in the waiting and communication phases in the SMDP, respectively. We have that*

$$\pi_{\rm wait} = \frac{1}{2 - e^{-\Lambda\Delta_0}},\ \pi_{\rm comm} = \frac{1 - e^{-\Lambda\Delta_0}}{2 - e^{-\Lambda\Delta_0}}. \tag{43}$$

Proof. Let $p_{ww}=e^{-\Lambda\Delta_0}$ and $p_{cw}=1$ be the probabilities of remaining in the waiting phase (ww; if no request is received, the SMDP remains in the waiting state) and of moving from a communication to a waiting state (cw; a waiting state follows a communication state) in one SMDP state transition, respectively. Therefore, $\pi_{\rm wait}$ and $\pi_{\rm comm}$ satisfy the stationary equations

$$\begin{cases} \pi_{\rm wait} = \pi_{\rm wait} p_{ww} + \pi_{\rm comm} p_{cw} = e^{-\Lambda\Delta_0}\pi_{\rm wait} + \pi_{\rm comm}, \\ \pi_{\rm comm} = \pi_{\rm wait}(1 - p_{ww}) + \pi_{\rm comm}(1 - p_{cw}) = (1 - e^{-\Lambda\Delta_0})\pi_{\rm wait}, \\ \pi_{\rm wait} + \pi_{\rm comm} = 1, \end{cases}$$

whose solution is given in the statement of the lemma. ■

References

[1] M. Bliss and N. Michelusi, "Power-constrained trajectory optimization for wireless uav relays with random requests," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[2] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L. G. Giordano, A. Garcia-Rodriguez, and J. Yuan, "Survey on uav cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[3] M. Mozaffari, W. Saad, M. Bennis, Y. Nam, and M. Debbah, "A tutorial on uavs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2019.

[4] L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in uav communication networks," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 1123–1152, Secondquarter 2016.

[5] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, May 2016.

[6] Q. Wu, L. Liu, and R. Zhang, "Fundamental Trade-offs in Communication and Trajectory Design for UAV-Enabled Wireless Network," *IEEE Wireless Communications*, vol. 26, pp. 36–44, 02 2019.

[7] Y. Zeng and R. Zhang, "Energy-efficient uav communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, June 2017.

[8] Y. Zeng, J. Xu, and R. Zhang, "Energy Minimization for Wireless Communication With Rotary-Wing UAV," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, April 2019.

[9] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Optimal transport theory for power-efficient deployment of unmanned aerial vehicles," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[10] S. Zhang, Y. Zeng, and R. Zhang, "Cellular-enabled uav communication: A connectivity-constrained trajectory optimization perspective," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 2580–2604, March 2019.

[11] F. Jiang and A. L. Swindlehurst, "Optimization of uav heading for the ground-to-air uplink," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 5, pp. 993–1005, June 2012.

[12] S. ur Rahman, G. Kim, Y. Cho, and A. Khan, "Positioning of uavs for throughput maximization in software-defined disaster area uav communication networks," *Journal of Communications and Networks*, vol. 20, no. 5, pp. 452–463, 2018.

[13] O. Esrafilian, R. Gangula, and D. Gesbert, "Uav-relay placement with unknown user locations and channel parameters," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 1075–1079.

[14] X. Li, D. Guo, H. Yin, and G. Wei, "Drone-assisted public safety wireless broadband network," in *2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2015, pp. 323–328.

[15] A. Fotouhi, M. Ding, and M. Hassan, "Dynamic base station repositioning to improve performance of drone small cells," in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–6.

[16] Y. Chen, N. Zhao, Z. Ding, and M. Alouini, "Multiple uavs as relays: Multi-hop single link versus multiple dual-hop links," *IEEE Transactions on Wireless Communications*, vol. 17, no. 9, pp. 6348–6359, Sep. 2018.

[17] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP Altitude for Maximum Coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.

[18] Q. Zhang, M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Machine learning for predictive on-demand deployment of uavs for wireless communications," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[19] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (uavs) for energy-efficient internet of things communications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7574–7589, 2017.

[20] V. Saxena, J. Jaldén, and H. Klessig, "Optimal uav base station trajectories using flow-level models for reinforcement learning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 4, pp. 1101–1112, 2019.

[21] J. Hu, H. Zhang, and L. Song, "Reinforcement learning for decentralized trajectory design in cellular uav networks with sense-and-send protocol," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6177–6189, 2019.

[22] A. M. Koushik, F. Hu, and S. Kumar, "Deep Q -learning-based node positioning for throughput-optimal communications in dynamic uav swarm network," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 554–566, 2019.

[23] W. Johnson, *Rotorcraft aeromechanics*, ser. Cambridge aerospace series ; 36. Cambridge: Cambridge University Press, 2013.

[24] R. Cheng and Y. Jin, "A competitive swarm optimizer for large scale optimization," *IEEE Transactions on Cybernetics*, vol. 45, no. 2, pp. 191–204, 2015.

[25] M.-H. Hwang, H.-R. Cha, and S. Jung, "Practical endurance estimation for minimizing energy consumption of multirotor unmanned aerial vehicles," *Energies*, vol. 11, no. 9, 2018.

[26] M. Clerc, *Particle swarm optimization*, ser. ISTE. London ; Newport Beach: ISTE, 2006.

[27] H. Shakhatreh, A. Khreishah, A. Alsarhan, I. Khalil, A. Sawalmeh, and N. S. Othman, "Efficient 3d placement of a uav using particle swarm optimization," in *2017 8th International Conference on Information and Communication Systems (ICICS)*, 2017, pp. 258–263.

[28] Z. Yuheng, Z. Liyan, and L. Chunpeng, "3-d deployment optimization of uavs based on particle swarm algorithm," in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, 2019, pp. 954–957.

[29] J. Cho and J. Kim, "Performance comparison of heuristic algorithms for uav deployment with low power consumption," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 2018, pp. 1067–1069.

[30] S. A. Hadiwardoyo, C. T. Calafate, J. Cano, K. Krinkin, D. Klionskiy, E. HernÃ¡ndez-Orallo, and P. Manzoni, "Optimizing uav-to-car communications in 3d environments through dynamic uav positioning," in *2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*, 2019, pp. 1–8.

[31] A. H. Kulaç, A. Akarsu, and T. Girici, "Optimal deployment of uav's for communication," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018, pp. 1–4.

[32] R. K. Patra and P. Muthuchidambaranathan, "Optimisation of spectrum and energy efficiency in uav-enabled mobile relaying using bisection and pso method," in *2018 3rd International Conference for Convergence in Technology (I2CT)*, 2018, pp. 1–7.

[33] H. Bayerlein, P. De Kerret, and D. Gesbert, "Trajectory Optimization for Autonomous Flying Base Station via Reinforcement Learning," in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[34] M. Bliss and N. Michelusi, "Trajectory optimization for rotary-wing uavs in wireless networks with random requests," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[35] A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, "A study of lora: Long range & low power networks for the internet of things," *Sensors*, vol. 16, no. 9, 2016. [Online]. Available: https://www.mdpi.com/1424-8220/16/9/1466

[36] C. You and R. Zhang, "3d trajectory optimization in rician fading for uav-enabled data harvesting," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 3192–3207, 2019.

[37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press, 2005.

[38] R. Essaadali and A. Kouki, "A new simple unmanned aerial vehicle doppler effect rf reducing technique," in *MILCOM 2016 - 2016 IEEE Military Communications Conference*, 2016, pp. 1179–1183.

[39] J. D. C. Little and S. Graves, *Little's Law*, 07 2008, pp. 81–100.

[40] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *lecture notes of EE392o, Stanford University, Autumn Quarter*, vol. 2004, pp. 2004–2005, 2003.

[41] P. Yang, K. Tang, and X. Yao, "Turning high-dimensional optimization into computationally expensive optimization," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 1, pp. 143–156, 2018.

[42] Y. Sun, A. Baricz, and S. Zhou, "On the monotonicity, log-concavity, and tight bounds of the generalized marcum and nuttall q-functions," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1166–1186, 2010.