# Learning and Uncertainty-Exploited Directional Antenna Control for Robust Long-Distance and Broad-Band Aerial Communication

Mushuang Liu, *Student Member, IEEE*, Yan Wan, *Senior Member, IEEE*, Songwei Li, *Student Member, IEEE*, Frank L. Lewis, *Fellow, IEEE*, and Shengli Fu, *Senior Member, IEEE*

***Abstract*—Aerial communication using directional antennas (ACDA) is a promising solution to enable long-distance and broad-band unmanned aerial vehicle (UAV)-to-UAV networking. The automatic alignment of directional antennas allows the transmission energy to focus in certain direction and significantly extends the communication range and rejects interference. Robust automatic alignment of directional antennas is not easy to achieve, considering practical issues such as the limited on-board sensing devices due to the physical constraints of UAV payload and power supplies, uncertain and varying UAV movement patterns, and unstable GPS and unknown communication environments. In this paper, we develop reinforcement learning (RL)-based online antenna control solutions for the ACDA system to conquer these challenges. The control solution adopts an uncertain UAV mobility modeling and intention estimation framework to capture and predict the uncertain intentions of UAV maneuvers and hence permit robust tracking. To account for an unstable GPS environment, the control solution features a learning of communication channel models to provide additional measurement signals in GPS-denied settings. A novel stochastic optimal control solution for nonlinear random switching dynamics is developed that integrates RL, an effective uncertainty evaluation method called multivariate probabilistic collocation method (MPCM), and unscented Kalman Filter (UKF). Simulation studies are conducted to illustrate and validate the proposed solutions.**

***Index Terms*—Learning control, random switching systems, uncertainty quantification, UAV communication, directional antennas.**

## I. Introduction

UNMANNED Aerial Vehicles (UAVs) have been widely used in civilian and commercial applications including emergency response, connectivity service, intelligent transportation, precision agriculture, among others [2]–[4]. Aerial communication among UAVs is expected to play an indispensable role in these applications when multiple UAVs are involved [5]–[7]. In applications such as emergency response and remote infrastructure health monitoring, the long-distance and broad-band UAV-to-UAV communication capability is desired.

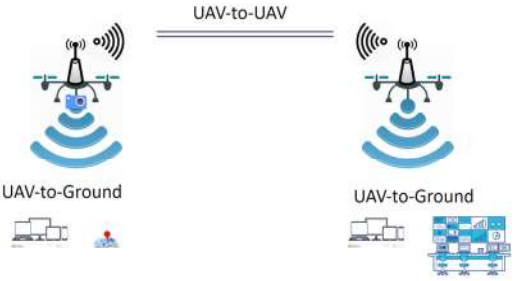



Fig. 1. Illustration of the broadband long-distance communication infrastructure using controllable UAV-carried directional antennas [10].

To enable long-distance and broad-band UAV-to-UAV communication, the aerial communication using directional antennas (ACDA) has been developed as a promising solution [8]–[13]. Through using directional antennas that focus the transmission energy in certain direction, ACDA significantly extends the communication distance and rejects interference, compared to omni-drectional antenna based solutions. With ACDA, UAVs-carried communication infrastructures can be quickly deployed to deliver Wi-Fi services from the air, through which high-rate data such as monitoring streams from remote locations can be transmitted in real-time (see Fig. 1). The detailed design prototype and hardware components of this ACDA system are described in [10].

A critical component of the ACDA system is the automatic alignment of directional antennas to maximize the communication performance. Each UAV in the ACDA system carries a rotational plate mounted with a directional antenna [11], which is controlled to align with the directional antenna carried by the other UAV. Robust automatic alignment of directional antennas is not easy to achieve, considering practical issues such as the limited on-board sensing devices due to the physical constraints of UAV payload and power supplies, uncertain and varying UAV mobility, and unstable GPS and unknown communication environments.

There are two general design configurations of the ACDA system, depending on whether the communication channel used

Manuscript received July 30, 2019; accepted October 6, 2019. Date of publication November 6, 2019; date of current version January 15, 2020. This work was supported in part by the National Science Foundation under Grants 1730675, 1714519, and 1839804 and in part by the Office of Naval Research under ONR Grant N00014-18-1-2221. This article was presented in part at the IEEE Vehicle Technology Conference, Honolulu, HI, September 2019 [1]. The review of this article was coordinated by Prof. S. Coleri Ergen. *(Corresponding author: Yan Wan.)*

M. Liu, Y. Wan, S. Li, and F. L. Lewis are with the Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76019 USA (e-mail: mushuang.liu@mavs.uta.edu; yan.wan@uta.edu; songwei.li@mavs.uta.edu; lewis@uta.edu).

S. Fu is with the Department of Electrical Engineering, University of North Texas, Denton, TX 76207 USA (e-mail: shengli.fu@unt.edu).

Digital Object Identifier 10.1109/TVT.2019.2951721

for antenna control is omni or not. The first configuration uses a directional antenna-equipped broad-band channel for the transmission of application-oriented data (e.g., real-time video streams), and an additional low-rate omni-directional communication channel for control and command data. In [11], omni-directional antennas are used to transmit the GPS information of the remote UAV for the alignment of antennas. This configuration simplifies the antennas controller design, as the control channel still functions even if the directional antennas are not in alignment. However, the omni-directional control channel suffers from practical issues such as interference and dissipation over a long communication distance [14], [15]. As such, in this paper, we aim to design the ACDA system using the second configuration where the high-rate application data and low-rate control and command data share the same channel equipped with directional antennas.

Although more practical, this solution that removes the additional control and command channel introduces more challenges to the robustness of antennas control. As control and command data cannot be transmitted if the directional communication channel fails, the antenna control system needs to robustly lock and track the other directional antennas, once the communication channel is established initially. To do that, we develop an uncertain UAV mobility modeling and intention estimation framework to capture and predict the uncertain intentions of the remote UAV's maneuvers. Predictive intentions for robot-robot and human-robot collaborations have been studied in e.g., [16]–[18]. Most of these studies assume that an agent's intention can be described and modeled in a deterministic and predictable form [16]–[18]. This is not suitable for UAVs considering their highly flexible and random movement patterns. Probabilistic intentions and their estimation have also been studied in e.g., [19], [20], using stochastic models such as Markov chain and Baysien networks. In this paper, we use random mobility models (RMMs) [21]–[23], and in particular, the smooth turn (ST) UAV RMM [24], [25] to more realistically capture the uncertain mobility intentions of UAVs. RMMs are a class of random switching models that capture the statistics of random moving objects. The intelligence on RMMs is exploited in this paper to facilitate robust tracking.

In indoor and many emergency scenarios, GPS signals may be unstable considering environmental disturbances and blockages. In GPS unstable or denied environment, we need additional measurement signals for antenna control. Received Signal Strength Indicator (RSSI), a communication performance indicator, is a promising measurement signal for ACDA, as it can be measured from ACDA self-equipped directional antennas, and does not require additional localization sensors to be carried by UAVs. In [26], we adopted the RSSI of directional antennas, to compensate unstable GPS signals, under the assumption that the communication environment is perfect. In particular, GPS and directional Wi-Fi RSSI based fusion algorithms were developed to estimate the other UAV's location, which is used to align the headings of directional antennas. However, in an imperfect communication environment, the effects of reflection, refraction and absorption by buildings, obstacles, and interference sources can distort the strongest signal directions. In this case, simply aligning directional antennas using their GPS locations may not lead to the best communication performance (see experimental studies in [10], [27]). In this paper, we develop a distributed antenna control solution for the goal of maximizing the communication performance, instead of using location-based antenna heading alignment. The solution learns directional Wi-Fi channel models online and provides RSSI as not only alternative measurement signals, but also the goal function for antenna control, in GPS-denied settings.

Our antenna control adopts a novel stochastic optimal control approach that integrates Reinforcement Learning (RL) for online optimal control, Multivariate Probabilistic Collocation Method (MPCM) for effective uncertainty evaluation, and Unscented Kalman Filter (UKF) for nonlinear state estimation. On the aspect of optimal control, RL has been developed in [28], [29] for deterministic system dynamics. Paper [30] developed the stochastic optimal control solution that integrates MPCM and RL for systems modulated by uncertain parameters, and paper [31] applied this solution for an air traffic management problem subject to uncertain weather conditions. In this paper, we study the stochastic optimal control problem for broad random switching systems. On the aspect of estimation, nonlinear system estimation methods such as Extended Kalman Filter (EKF) and UKF have been widely used typically for known and deterministic systems corrupted with additive noises, but not random switching RMMs. In this paper we develop a new stochastic optimal control solution for systems that involve nonlinear random switching RMMs and limited measurements, by integrating UKF, RL and MPCM.

The contributions of this paper are summarized as follows.

1) *The design configuration of the ACDA system using pure directional antennas.* In this ACDA system, the high-rate application data and low-rate control and command data share the same communication channel equipped with directional antennas. This design is more practical compared to the previously developed ACDA systems, which use both directional and omni-directional antennas.
2) *RL-based antenna control.* This solution learns directional channel models online and provides RSSI as not only alternative measurement signals, but also the goal function for antenna control, in GPS-denied settings. In addition, this solution does not require a known and perfect communication channel, which was assumed in the previously developed ACDA systems.
3) *Stochastic UAV intention modeling.* We use RMMs to capture the highly flexible and random movement patterns of UAVs, and develop an online model estimation framework to capture and predict the uncertain intentions of the remote UAV's maneuvers.
4) *Real-time state estimation for random switching systems.* Agents' states in general random switching systems are usually regarded as unpredictable. This solution makes the best prediction out of the agents' intentions coded in the statistics of the agents' random maneuvers to analyze and further predict the agents' future behaviors.
5) *Real-time distributed optimal control for random switching systems.* This solution integrates RL and MPCM to provide an effective online optimal control solution for agents moving with random switching system models.

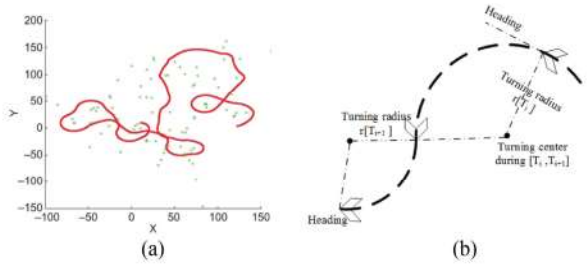



Fig. 2. Illustration of the ST RMM: (a) UAV trajectory ensemble (red curve). Green spots are the randomly chosen turning centers [24]; (b) maneuver selection and switching.

The remainder of this paper is organized as follows. In Section II, we describe the ACDA system shown in Fig. 1, including both system and measurement models. The antenna control problem is also formulated. In Section III, we develop the RL based stochastic optimal control solutions. In Section IV, an uncertain intention estimation method is provided to estimate the random variables of the remote UAV's uncertain maneuvers. In Section V, simulation studies are conducted, and Section VI concludes this paper.

## II. Modeling and Problem Formulation

In this section, we first describe the ACDA system model, including the UAV RMM and directional antenna dynamics. We then describe the GPS and RSSI measurement models. The antenna control problems are then formulated.

### A. System Models

We consider two UAVs independently fly in a low-altitude airspace at approximately the same height to fulfill their missions such as search and rescue (see Fig. 1). The same altitude assumption is reasonable because that 1) the range of flight altitudes for small UAVs is very limited [32]; and 2) the optimal flight altitudes to maximize coverage are proved to be the same for UAVs of the same type [33]–[35]. On each UAV, a tunable plate attached with a directional antenna is installed and driven by a gear motor [11]. To establish a long-range air-to-air communication channel to transmit both application data (e.g., surveillance videos) and control and command data, the channel performance needs to be maximized.

*1) UAV Random Mobility Model:* We use the smooth turn (ST) mobility model ([24], [25]) to capture the uncertain intentions of UAVs executing surveillance-like missions (see Fig. 2). The random maneuvers described by a ST mobility model work as follows. At randomly selected time points $T_0^i$, $T_1^i$, $T_2^i$, ..., where $0 = T_0^i < T_1^i < \cdots$, UAV $i$ selects a point in the airspace along the line perpendicular to its current heading direction, and then circles around it until the UAV chooses another turning center. The perpendicularity guarantees smooth trajectories [24]. The time duration for UAV $i$ to maintain its current maneuver $\tau_i[T_j^i] = T_{j+1}^i - T_j^i$ follows a memoryless exponential distribution [36].

$$f_\tau(\tau_i[T_j^i]) = \lambda_i e^{-\lambda_i \tau_i[T_j^i]}, \tag{1}$$

where $1/\lambda_i$ is the mean of $\tau_i[T_j^i]$. The velocity $v_i[T_j^i]$ follows a uniform distribution with the minimum and maximum velocity constraints $v_{i,min} < v_i[T_j^i] < v_{i,max}$,

$$f_v(v_i[T_j^i]) = \frac{1}{v_{i,\max} - v_{i,\min}}. \tag{2}$$

The inverse of the turning radius $\frac{1}{r_i[T_j^i]}$ follows the zero-mean Gaussian distribution with variance $\sigma_i^2$,

$$f_r\left(\frac{1}{r_i[T_j^i]}\right) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{1}{2r\sigma_i^2}}. \tag{3}$$

Denote the position of UAV $i$ along $x$ and $y$ axes at time instant $k$ as $x_i[k]$ and $y_i[k]$ respectively. The dynamics of UAV $i$ (denote as $f_i(.)$) following the ST uncertain maneuvering intentions are described as

$$\begin{aligned} x_i[k+1] &= x_i[k] + v_i[k]\cos(\phi_i[k])\delta, \\ y_i[k+1] &= y_i[k] + v_i[k]\sin(\phi_i[k])\delta, \\ \phi_i[k+1] &= \phi_i[k] + \omega_i[k]\delta, \end{aligned} \tag{4}$$

where $\delta$ is the sampling period, $\phi_i[k]$ and $\omega_i[k]$ are the heading angle and angular velocity at time instant $k$, and

$$\omega_i[k] = \frac{v_i[k]}{r_i[k]}. \tag{5}$$

Note that the ST RMM is a random switching model composed of two types of random variables [37]. Type 1 random variables, $v_i[k]$ and $r_i[k]$, describe the characteristics for each maneuver.

$$v_i[k] = \begin{cases} v_i[T_j^i], & \text{if } \exists j \in [0,1,2,..), k = T_j^i \\ v_i[k-1], & \text{if } \forall j = 0,1,2,.., k \neq T_j^i \end{cases} \tag{6}$$

$$r_i[k] = \begin{cases} r_i[T_j^i], & \text{if } \exists j \in [0,1,2,..), k = T_j^i \\ r_i[k-1], & \text{if } \forall j = 0,1,2,.., k \neq T_j^i \end{cases} \tag{7}$$

The maneuvers' random switching behavior is governed by the type 2 random variable, $\tau_i[T_j^i]$, which describes how often the switching of type 1 random variables occurs.

The two groups of uncertain maneuvers for the UAVs $(v_1[T_j^1], r_1[T_j^1], \tau_1[T_j^1])$ and $(v_2[T_j^2], r_2[T_j^2], \tau_2[T_j^2])$ are independent, as UAV mobility is application-specific, and is not constrained from the communication mission.

*2) Directional Antennas Dynamics:* The directional antenna installed on each UAV autonomously adjusts its heading angle to establish a robust communication channel between the two UAVs. For UAV $i$, the heading angle dynamics of its directional antennas is described as

$$\theta_i[k+1] = \theta_i[k] + (\omega_i'[k] + \omega_i[k])\delta, \tag{8}$$

where $\theta_i$ is the heading angle of antennas $i$, and $\omega_i'$ is the angular velocity of antennas $i$ due to its heading control. Note that the change of $\theta_i$ is caused by both the control of antenna $i$, $\omega_i'$, and the movement of UAV $i$, $\omega_i$.

### B. Measurement Models

We consider two measurement signals for the ACDA system, GPS and RSSI.

*1) GPS Measurement:* If GPS is available, the measured GPS signal of UAV $i$ is denoted as $\mathbf{z}_{G,i}(k)$,

$$\mathbf{z}_{G,i}[k] = \mathbf{H}_G(k)\mathbf{x}_i[k] + \varpi_{G,i}[k], \tag{9}$$

where $\mathbf{H}_G = [1,0,0,0;0,1,0,0]$ is the measurement matrix, $\mathbf{x}_i[k] = [x_i[k], y_i[k], \phi_i[k], \theta_i[k]]^T$ is the system state of UAV $i$, and $\varpi_{G,i}$ is the white Gaussian noise with zero mean and covariance $\mathbf{R}_{G,i}$. GPS signals can be transmitted through the air-to-air communication channel to assist with the control of directional antennas. Denote the relation between the GPS signal and system state as $h_{G,i}$, i.e., $\mathbf{z}_{G,i}[k] = h_{G,i}(\mathbf{x}_i[k])$.

*2) RSSI Measurement:* RSSI measures the signal power received from the transmitting antenna [38], and hence is an important indicator of communication channel performance. In the ACDA system, RSSI is affected by the relative positions of two UAVs that carry these directional antennas, headings, field radiation patterns of these antennas, and also communication environment. Denote the measured RSSI signal as $z_R[k]$, the relation between the RSSI signal, $z_R[k]$, and the system states, $\mathbf{x}[k]$ ($\mathbf{x}[k] = [\mathbf{x}_1^T[k], \mathbf{x}_2^T[k]]^T$), as $h_R(.)$, i.e., $z_R[k] = h_R(\mathbf{x}[k])$, then $z_R[k]$ is given by the Friis free space equation [38]:

$$\begin{aligned} z_R[k] = {} & P_{t|dBm}[k] + 20\log_{10}(\lambda) - 20\log_{10}(4\pi) \\ & - 20\log_{10}(d[k]) + G_{l|dBi}[k] + \varpi_R[k], \end{aligned} \tag{10}$$

where $P_{t|dBm}[k]$ is the transmitted signal power, $\lambda$ is the wavelength, and $d[k]$ is the distance between the two UAVs at time $k$, i.e., $d[k] = \sqrt{(x_1[k]-x_2[k])^2 + (y_1[k]-y_2[k])^2}$. $G_{l|dBi}[k]$ is the sum of gains at both the transmitting and receiving sides [39]. The Ubiquiti NanoStation loco M5 directional antennas [40] that we use in the ACDA system is modeled based on the filed pattern of the end-fire array antennas [41],

$$\begin{aligned} & G_{l|dBi}[k] \\ & = (G^{\max}_{t|dBi} - G^{\min}_{t|dBi}) \\ & \quad \times \sin\frac{\pi}{2n}\frac{\sin\left(\frac{n}{2}(k_a d_a(\cos(\gamma_t[k]-\theta_t[k]))-1)-\frac{\pi}{n}\right)}{\sin\left(\frac{1}{2}(k_a d_a(\cos(\gamma_t[k]-\theta_t[k]))-1)-\frac{\pi}{n}\right)} \\ & \quad + (G^{\max}_{r|dBi} - G^{\min}_{r|dBi}) \\ & \quad \times \sin\frac{\pi}{2n}\frac{\sin\left(\frac{n}{2}(k_a d_a(\cos(\gamma_r[k]-\theta_r[k]))-1)-\frac{\pi}{n}\right)}{\sin\left(\frac{1}{2}(k_a d_a(\cos(\gamma_r[k]-\theta_r[k]))-1)-\frac{\pi}{n}\right)} \\ & \quad + G^{\min}_{t|dBi} + G^{\min}_{r|dBi}, \end{aligned} \tag{11}$$

where $G^{\max}_{t|dBi}$, $G^{\min}_{t|dBi}$, and $G^{\max}_{r|dBi}$, $G^{\min}_{r|dBi}$ are the maximum and minimum gains of transmitting and receiving antennas. $k_a$ is the wave number, and $k_a = \frac{2\pi}{\lambda}$. $n$ and $d_a$ are design parameters of the directional antenna. $\theta_t[k]$ and $\theta_r[k]$ are the heading angles of the transmitting and receiving antennas at time $k$, respectively. $\gamma_t[k]$ and $\gamma_r[k]$ are the heading angles of the transmitting and receiving antennas corresponding to the maximal $G_l$ at time $k$, respectively.

The parameters $G^{\max}_{t|dBi}$, $G^{\min}_{t|dBi}$, $G^{\max}_{r|dBi}$, and $G^{\min}_{r|dBi}$, can be obtained from the antenna's datasheet. In ACDA, the two directional antennas are of the same type, and hence $G^{\max}_{t|dBi} = G^{\max}_{r|dBi}$, and $G^{\min}_{t|dBi} = G^{\min}_{r|dBi}$. In an imperfect environment these parameters in $G_{l|dBi}[k]$ can be environment-specific.

Similarly, in a perfect communication environment, $\gamma_t[k]$ and $\gamma_r[k]$ are achieved when the two antennas are aligned [26]. Affected by the impact of imperfect environment, such as blockages, the desired heading angles can be captured as

$$\gamma_r[k] = \arctan\frac{y_t[k]-y_r[k]}{x_t[k]-x_r[k]} + \theta_{r_{env}}, \tag{12}$$

$$\gamma_t[k] = \arctan\frac{y_r[k]-y_t[k]}{x_r[k]-x_t[k]} + \theta_{t_{env}}, \tag{13}$$

where $(x_t[k], y_t[k])$ and $(x_r[k], y_r[k])$ are the positions of UAVs that carry the transmitting and receiving antennas respectively, and $\theta_{r_{env}}$ and $\theta_{t_{env}}$ are environment-specific shift angles at the receiver and transmitter sides. $\theta_{r_{env}}$ and $\theta_{t_{env}}$ are zeros in a perfect environment.

### C. Problem Formulation

We aim to design the angular Velocity of each directional antenna to maximize the expected RSSI performance of ACDA over a look-ahead window. The RSSI model (as described in Equations (10)–(13)) contains unknown environment-specific parameters ($G^{\max}_{t|dBi}$, $G^{\min}_{t|dBi}$, $\theta_{t_{env}}$ and $\theta_{r_{env}}$), and the UAV dynamics contain uncertain parameters $(v_1[k], r_1[k], \tau_1[T_j^1], v_2[k], r_2[k], \tau_2[T_j^2])$.

Here we formulate the problem as stochastic optimal control. Mathematically, considering the random switching system dynamics described in Equations (4)–(8), the optimal control policy $\mathbf{u}[k]$ is sought to maximize the expected value function, which is the summation of the predicted RSSI signals over a look-ahead window, i.e.,

$$V(\mathbf{x}[k]) = E\left\{\sum_{l=k}^{k+N} \alpha^{l-k} z_R[l](\mathbf{x}[l], \mathbf{u}[k])\right\}, \tag{14}$$

where $\mathbf{x}[k]$ is the global state, $\mathbf{x}[k] = [\mathbf{x}_1^T[k], \mathbf{x}_2^T[k]]^T$. $\mathbf{u}[k]$ is the control input, $\mathbf{u}[k] = [u_1[k], u_2[k]]^T$, $u_i[k] = [\omega_i'[k]]$. $z_R[l]$ is the RSSI signal at time $l$, and $\alpha \in (0,1]$ is a discount factor. Note that the control is decentralized, in the sense that each antenna finds its own optimal control policy, with the assumption that the other antenna adopts its optimal control policy. Each UAV only needs to learn its own environment-specific parameters ($G^{\max}_{t|dBi}$, $G^{\min}_{t|dBi}$, and $\theta_{t_{env}}/\theta_{r_{env}}$) to find its optimal control policy.

In the rest of this article, we develop the control solution for the local UAV, or UAV 1. The control solution for the remote UAV, or UAV 2 is designed in the same manner.

## III. Reinforcement Learning Based Stochastic Optimal Control for ACDA

In this section, we develop new online solutions to solve the stochastic optimal control problem for the ACDA system

described in Section II-C. The solution integrates the uncertainty sampling method MPCM, the adaptive optimal control method RL, and the nonlinear estimation method UKF, to address the challenges including nonlinear and random switching dynamics, unknown RSSI model, limited measurements of system outputs, and online time requirement to derive optimal solutions for random switching systems.

In Section III-A, we describe the solution when GPS is available but the RSSI model is unknown. Online stochastic optimal control solutions are derived and the environment-specific RSSI model is learned. Section III-B further develops online solutions in both GPS-available and GPS-denied environments, with the learned environment-specific RSSI model.

### A. Stochastic Optimal Control With Unknown RSSI

To develop a decentralized optimal control solution that maximizes the value function (Equation (14)) for the nonlinear random switching ACDA dynamics with unknown RSSI model and limited measurements, two main steps are involved: 1) state estimation, and 2) adaptive optimal controller design.

*1) State Estimation:* The states of both local and remote UAVs need to be estimated. For the local UAV, the trajectory-specific maneuvers ($v_1[k]$, $r_1[k]$, and $\tau_1[T_j^i]$) are known locally, and hence, the local-system states ($x_1[k], y_1[k], \phi_1[k]$) can be estimated utilizing UKF as described in [26, Section 3.1]. We do not repeat the process here to save the space.

For the remote UAV that has random switching dynamics, the RMM-related maneuvers ($v_2[k]$, $r_2[k]$, and $\tau_2[T_j^2]$) are unknown to the local UAV, and hence the remote UAV's states ($x_2[k], y_2[k], \phi_2[k]$) can not be directly estimated using the existing filtering type of methods. We design a new estimation algorithm for the nonlinear and random switching dynamics. Here, a subset of $\mathrm{x}_2[k]$, i.e., $[x_2[k], y_2[k], \phi_2[k]]$ is needed for this estimation, and we use $\mathrm{x}_2[k]$ to represent this subset to simplify presentation, when it does not cause confusion.

Denote the switching behavior of the remote UAV at time $k$ as $s[k]$. $s[k] = 1$ or 0 represent the current maneuver switches at time $k$ or not. Considering the two possible switching behaviors, the expected conditional current state of UAV 2 given the previous state, $\mathrm{x}_2[k-1]$, can be derived as

$$\begin{aligned} &E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1]) \\ &\quad = E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1], s[k-1]=0)P(s[k-1]=0) \\ &\qquad + E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1], s[k-1]=1)P(s[k-1]=1). \end{aligned} \tag{15}$$

When $s[k-1] = 0$, the remote UAV remains its previous maneuvers $v_2[k-1]$ and $r_2[k-1]$, and thus the expected system state $E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1], s[k-1]=0)$ can be estimated from the system dynamics $f(\mathrm{x}_2[k-1], v_2[k-1], r_2[k-1])$. When $s[k-1] = 1$, UAV 2 selects new maneuvers from the two random variables $v_2[T_j^2]$ and $r_2[T_j^2]$. In this case, the estimation of the system state $E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1], s[k-1]=1)$ involves uncertainty evaluation, which is typically solved by the Monte Carlo method, too slow to be used for real-time control. Here we use a multivariate probabilistic collocation method (MPCM) [42] to effectively evaluate the uncertainty. MPCM accurately evaluates the output mean of a system mapping subject to uncertain input parameters, by smartly selecting a limited number of sample points according to the Gaussian Quadrature rules. The main property of MPCM is described in the following lemma. Please refer to [42] for the detailed MPCM design procedure.

*Lemma 1:* [42, Theorem 2] Consider a system $G$ modulated by $m$ independent uncertain parameters, $a_i$, where $i \in \{1, ..m\}$,

$$G(a_1, \ldots, a_m) = \sum_{j_1=0}^{2n_1-1} \sum_{j_2=0}^{2n_2-1} \cdots \sum_{j_m=0}^{2n_m-1} \psi_{j_1,\ldots,j_m} \prod_{i=1}^{m} a_i^{j_i}, \tag{16}$$

where $a_i$ has a degree up to $2n_i - 1$. $n_i$ is a positive integer for any $i$. $\psi_{j_1,\ldots,j_m} \in \mathbb{R}$ are the coefficients. Each uncertain parameter $a_i$ follows an independent pdf $f_{A_i}(a_i)$. The MPCM approximates $G(a_1, \ldots a_m)$ with the following low-order mapping

$$G'(a_1, \ldots, a_m) = \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} \cdots \sum_{j_m=0}^{n_m-1} \Omega_{j_1,\ldots,j_m} \prod_{i=1}^{m} a_i^{j_i}, \tag{17}$$

with $E[G(a_1, \ldots, a_m)] = E[G'(a_1, \ldots, a_m)]$, where $\Omega_{j_1,\ldots,j_m} \in \mathbb{R}$ are coefficients. MPCM reduces the number of simulations from $2^m \prod_{i=1}^m n_i$ to $\prod_{i=1}^m n_i$.

Define a system mapping subject to uncertain input parameters $v_2[T_j^2]$ and $r_2[T_j^2]$: $G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1]) = f(\mathrm{x}_2[k-1], v_2[T_j^2], r_2[T_j^2])$. When $s[k-1] = 1$, the expected current state $E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1], s[k-1]=1)$ can be estimated from the mean output of the system mapping $G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])$, i.e., $E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1], s[k-1]=1) = E[G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])]$, using MPCM according to Lemma 1 and paper [42]. Under the assumption that the two uncertain parameters $v_2[T_j^2]$ and $r_2[T_j^2]$ have degrees up to $2n_1 - 1$ and $2n_2 - 1$ respectively, $G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])$ has the following form,

$$\begin{aligned} &G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1]) \\ &\quad = \sum_{j_1=0}^{2n_1-1} \sum_{j_2=0}^{2n_2-1} \psi_{j_1,j_2}(\mathrm{x}_2[k-1]) v_2^{j_1}[T_j^2] r_2^{j_2}[T_j^2]. \end{aligned} \tag{18}$$

According to Lemma 1, the output mean of this system mapping can be estimated from the output of a reduced-order mapping $G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])$, i.e., $E[G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])] = E[G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])]$, where the reduced mapping $G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])$ has the following form

$$\begin{aligned} &G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1]) \\ &\quad = \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} \Omega_{j_1,j_2}(\mathrm{x}_2[k-1]) v_2^{j_1}[T_j^2] r_2^{j_2}[T_j^2]. \end{aligned} \tag{19}$$

The coefficients $\Omega_{j_1,j_2}(\mathrm{x}_2[k-1])$ and output mean $E[G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])]$ are obtained using the evaluated outputs $G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])$ at each selected simulation point according to the procedures described in [42, Section II-B].

*Theorem 1:* Given the previous state $\mathrm{x}_2[k-1]$ of the remote UAV 2, the expected current state $E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1])$ is estimated by the local UAV 1 as

$$\begin{aligned} &E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1]) \\ &\quad = P_2 E[G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])] \\ &\qquad + (1-P_2) f(\mathrm{x}_2[k-1], v_2[k-1], r_2[k-1]), \end{aligned} \tag{20}$$

where $P_2$ is the switching probability of the remote UAV's maneuver at each time instant, $P_2 = \lambda_2 \delta$.

*Proof:* Let us first find the switching probability $P_i$. Since the time duration for UAV $i$ to maintain its current maneuver $\tau_i[T_j^i]$ follows exponential distribution as described in Equation (1), $P_i$ can be approximated from its exponential distribution as

$$P_i = \lambda_i \delta. \tag{21}$$

With the switching probability and the defined system mapping $G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])$, Equation (15) can be further written as

$$\begin{aligned} &E(\mathrm{x}_2[k]|\mathrm{x}_2[k-1]) \\ &\quad = P_2 E[G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])] \\ &\qquad + (1-P_2) f(\mathrm{x}_2[k-1], v_2[k-1], r_2[k-1]). \end{aligned} \tag{22}$$

Since $E[G_2(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])] = E[G_2'(v_2[T_j^2], r_2[T_j^2], \mathrm{x}_2[k-1])]$ according to Lemma 1 and Equations (18) and (19), Theorem 1 is derived naturally by combining Equations (18), (19), (22) and Lemma 1. ■

Theorem 1 provides a general approach to estimate the expected system state of a random switching system with computational efficiency, given the previous system state. Here we use this state estimation approach with UKF to estimate the state of the remote UAV from the measurement $\mathrm{z}_{G,2}[k]$. In particular, we integrate MPCM and UKF for a 5-step state estimation procedure. ***Steps 1*** and ***2*** select initial conditions and MPCM points to initialize ***Steps 3–5***; ***Step 3*** and ***4*** find the state estimators when the switching behavior $s[k-1] = 0$ and 1 respectively; ***Step 5*** finds the expected state by integrating the two estimators found in ***Steps 3*** and ***4***.

***Step 1. Initialize:*** Select initial conditions $\hat{\mathrm{x}}_2[0]$ and $\mathbf{P}[0]$ to initialize the system.

***Step 2. Select MPCM points:*** $n_1 n_2$ MPCM simulation point pairs are selected for the random variables $v_2[T_j^2]$ and $r_2[T_j^2]$ according to the MPCM procedure [42, Section II]. Denote the selected MPCM point pairs as $(\mathcal{V}_{j_1}[T_j^2], \mathcal{R}_{j_2}[T_j^2])$, where $j_1 \in \{0, \ldots, n_1 - 1\}$ and $j_2 \in \{0, \ldots, n_2 - 1\}$).

***Step 3. Estimate system state when*** $\mathrm{s}[\mathrm{k}-1] = 0$***:*** When $s[k-1] = 0$, the remote UAV does not change its maneuver, and hence the conditional expected current state $E(\mathrm{x}_2[k]|\hat{\mathrm{x}}_2[k-1], s[k-1] = 0, \mathrm{z}_{G,2}[k])$ can be estimated using UKF as described in sub-steps (a)–(d).

*(a) Select Sigma Points:* $2n+1$ symmetric weighted sigma points are selected from $\hat{\mathrm{x}}_2[k-1]$, the estimator of $\mathrm{x}_2[k-1]$.

$$\mathcal{X}_0[k-1] = \hat{\mathrm{x}}_2[k-1],$$

and for $i = 1, 2, \ldots n$,

$$\begin{aligned} \mathcal{X}_i[k-1] &= \hat{\mathrm{x}}_2[k-1] + \sqrt{(n+\kappa)\mathbf{P}[k-1]}_i, \\ \mathcal{X}_{i+n}[k-1] &= \hat{\mathrm{x}}_2[k-1] - \sqrt{(n+\kappa)\mathbf{P}[k-1]}_i, \end{aligned}$$

where $\mathbf{P}[k-1]_i$ is the $i$th column of the error covariance matrix of $\hat{\mathrm{x}}_2[k-1]$, $n$ is the states' dimension, and $n = 3$ here for the remote UAV system. The weights associated with the selected sigma points are $W_0 = \frac{\kappa}{n+\kappa}$, $W_i = \frac{1}{2(n+\kappa)}$, and $W_{i+n} = \frac{1}{2(n+\kappa)}$ respectively. $\kappa$ is a scaling parameter usually set to 0 in the general case or set to $3-n$ in the Gaussian case to capture the fourth-order moment correctly [43], [44].

*(b) State Prediction:* The system state can be predicted by instantiating each of the sigma points through the system dynamics $f_2(.)$ described in Equation (4).

$$\mathcal{X}_l[k|k-1] = f_2(\mathcal{X}_l[k-1], r_2[k-1], v_2[k-1]).$$

Then the priori state estimation can be approximated as a weighted sample mean

$$\hat{\mathrm{x}}_2[k|k-1] = \sum_{l=0}^{2n} W_l(\mathcal{X}_l[k|k-1]).$$

The corresponding covariance matrix is calculated as

$$\begin{aligned} \mathbf{P}[k|k-1] = \sum_{l=0}^{2n} &W_l(\mathcal{X}_l[k|k-1] - \hat{\mathrm{x}}_2[k|k-1]) \\ &\times (\mathcal{X}_l[k|k-1] - \hat{\mathrm{x}}_2[k|k-1])^T. \end{aligned}$$

*(c) Measurement Prediction:* $2n+1$ sigma points are selected from $\hat{\mathrm{x}}_2[k|k-1]$ with the error covariance $\mathbf{P}[k|k-1]$.

$$\begin{aligned} \mathcal{X}_0[k|k-1] &= \hat{\mathrm{x}}_2[k|k-1], \\ \mathcal{X}_i[k|k-1] &= \hat{\mathrm{x}}_2[k|k-1] + \sqrt{(n+\kappa)\mathbf{P}[k|k-1]}_i, \\ \mathcal{X}_{i+n}[k|k-1] &= \hat{\mathrm{x}}_2[k|k-1] - \sqrt{(n+\kappa)\mathbf{P}[k|k-1]}_i, \end{aligned}$$

with the weights $W_0, W_i$ and $W_{i+n}$ respectively.

The GPS measurement is then predicted by instantiating each of the prediction points through the measurement model $h_{G,2}$ described in Equation (9),

$$\begin{aligned} \mathcal{Z}_l[k|k-1] &= h_{G,2}(\mathcal{X}_l[k|k-1]), \\ \hat{\mathrm{z}}_{G,2}[k|k-1] &= \sum_{l=0}^{2n} W_l(\mathcal{Z}_l[k|k-1]). \end{aligned}$$

Correspondingly, the measurement covariance matrix and cross correlation matrix are determined by

$$\mathbf{P}_{ZZ}[k|k-1] = \sum_{l=0}^{2n} W_l(\mathcal{Z}_l[k|k-1] - \hat{\mathbf{z}}_{G,2}[k|k-1]) \times (\mathcal{Z}_l[k|k-1]] - \hat{\mathbf{z}}_{G,2}[k|k-1]])^T + \mathbf{R}_{G,2},$$

$$\mathbf{P}_{XZ}[k|k-1] = \sum_{l=0}^{2n} W_l(\mathcal{X}_l[k|k-1] - \hat{\mathbf{x}}_2[k|k-1]) \times (\mathcal{Z}_l[k|k-1] - \hat{\mathbf{z}}_{G,2}[k|k-1])^T.$$

*(d) Kalman Gain Update:* The Kalman gain is then updated using the covariance information,

$$\mathcal{K} = \mathbf{P}_{ZZ}[k|k-1]\mathbf{P}_{XZ}^{-1}[k|k-1].$$

The estimated state and covariance are thus derived as

$$E(\mathbf{x}_2[k]|\hat{\mathbf{x}}_2[k-1], \mathbf{z}_{G,2}[k], s[k-1]=0]) = \hat{\mathbf{x}}_2[k|k-1] + \mathcal{K}(\mathbf{z}_{G,2}[k] - \hat{\mathbf{z}}_{G,2}[k|k-1]),$$

$$E(\mathbf{P}[k]|\mathbf{P}[k-1], \mathbf{z}_{G,2}[k], s[k-1]=0) = \mathbf{P}[k|k-1] - \mathcal{K}\mathbf{P}_{ZZ}[k|k-1]\mathcal{K}^T.$$

***Step 4. Estimate system state when*** $\mathbf{s}[\mathbf{k}-1]=1$***:*** When $s[k-1]=1$, the remote UAV changes its maneuvers according to the random variables $v_2[T_j^2]$ and $r_2[T_j^2]$. With the MPCM points selected in ***Step 2***, $(\mathcal{V}_{j_1}[T_j^2], \mathcal{R}_{j_2}[T_j^2])$, the expected state $E(\mathbf{x}_2[k]|\hat{\mathbf{x}}_2[k-1], \mathbf{z}_{G,2}[k], s[k-1]=1)$ and covariance $E(\mathbf{P}[k]|\mathbf{P}[k-1], \mathbf{z}_{G,2}[k], s[k-1]=1)$ can be estimated using the following three sub-steps (a)-(c) that integrate MPCM and UKF.

*(a) Estimate system state at each selected MPCM point:* The system state is estimated at each selected MPCM point $(\mathcal{V}_{j_1}[T_j^2], \mathcal{R}_{j_2}[T_j^2])$ by conducting the UKF procedures shown in ***Step 3, (a)-(d)***. Denote the estimated state from UKF at each MPCM point as $\hat{\mathbf{x}}_{j_1,j_2}[k]$ with the covariance $\mathbf{P}_{j_1,j_2}[k]$.

*(b) Find the reduced polynomial mappings:* Define the system mappings $G_x(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])$ and $G_P(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])$ as the relationships between the expected system state and covariance with the random variables $v_2[T_j^2]$ and $r_2[T_j^2]$ respectively. According to Lemma 1, the mean outputs of the two system mappings can be estimated from the outputs of the reduced-order mappings $G'_x(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])$ and $G'_P(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])$ respectively, following the MPCM procedures [42, Section II].

$$G'_x(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2]) = \sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1} \Omega_{X_{j_1,j_2}}(\hat{\mathbf{x}}_2[k-1]) v_2^{j_1}[T_j^2] r_2^{j_2}[T_j^2],$$

$$G'_P(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2]) = \sum_{j_1=0}^{n_1-1}\sum_{j_2=0}^{n_2-1} \Omega_{\mathbf{P}_{j_1,j_2}}(\hat{\mathbf{x}}_2[k-1]) v_2^{j_1}[T_j^2] r_2^{j_2}[T_j^2].$$

The coefficients $\Omega_{X_{j_1,j_2}}$ and $\Omega_{\mathbf{P}_{j_1,j_2}}$ and mean outputs can be obtained using the evaluated outputs $G'_x(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])$ and $G'_P(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])$ at each selected MPCM point, according to the procedures in [42, Section II-B].

*(c) Find the expected system state and covariance:* The expected state and covariance are then found from the system mapping according to Lemma 1 and the MPCM design procedures [42] as

$$E(\mathbf{x}_2[k]|\hat{\mathbf{x}}_2[k-1], \mathbf{z}_{G,2}[k], s[k-1]=1) = E[G'_x(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])],$$

$$E(\mathbf{P}[k]|\mathbf{P}[k-1], \mathbf{z}_{G,2}[k], s[k-1]=1) = E[G'_P(\hat{\mathbf{x}}_2[k-1], v_2[T_j^2], r_2[T_j^2])].$$

***Step 5. Estimate the expected system state:*** The estimated state and covariance are derived according to Theorem 1.

$$\begin{aligned} &E(\mathbf{x}_2[k]|\hat{\mathbf{x}}_2[k \quad 1], \mathbf{z}_{G,2}[k]) \\ &= P_2 E(\mathbf{x}_2[k]|\hat{\mathbf{x}}_2[k-1], \mathbf{z}_{G,2}[k], s[k-1]=1]) \\ &\quad + (1-P_2)E(\mathbf{x}_2[k]|\hat{\mathbf{x}}_2[k-1], \mathbf{z}_{G,2}[k], s[k-1]=0]), \end{aligned} \tag{23}$$

$$\begin{aligned} &E(\mathbf{P}[k]|\mathbf{P}[k-1], \mathbf{z}_{G,2}[k]) \\ &= P_2 E(\mathbf{P}[k]|\mathbf{P}[k-1], \mathbf{z}_{G,2}[k], s[k-1]=1]) \\ &\quad + (1-P_2)E(\mathbf{P}[k]|\mathbf{P}[k-1], \mathbf{z}_{G,2}[k], s[k-1]=0]). \end{aligned} \tag{24}$$

As such, the estimate of $\mathbf{x}_2[k]$ is $\hat{\mathbf{x}}_2[k] = E(\mathbf{x}_2[k]|\hat{\mathbf{x}}_2[k-1], \mathbf{z}_{G,2}[k])$, and the expected error covariance is $\mathbf{P}[k] = E(\mathbf{P}[k]|\mathbf{P}[k-1], \mathbf{z}_{G,2}[k])$.

*Remark 1:* The above estimation procedure integrates UKF and MPCM to provide a novel and efficient estimation method for nonlinear random switching systems. Note that the ST RMM involves three random variables: $\tau_i[T_j^i]$, $v[T_j^i]$, and $r_i[T_j^i]$. In the UKF estimation procedure, $\tau_i[T_j^i]$ plays a role in determining the switching probability $P_i$ as described in Equations (21), (23), and (24). $v[T_j^i]$ and $r_i[T_j^i]$ are random maneuvers, and play roles in the random maneuver sampling procedure (i.e., Step 2) and future state prediction procedure when $s[k-1]=1$ (i.e., Step 4). Note that if the remote UAV's previous maneuver information ($v_2[k-1]$ and $\omega_2[k-1]$) is unavailable, an additional estimation step is needed before processing Step 3. In particular, $v_2[k-1]$ and $\omega_2[k-1]$ need to be estimated from two consecutive previous states $\hat{\mathbf{x}}_2[k-1]$ and $\hat{\mathbf{x}}_2[k-2]$ as

$$\hat{v}_2[k-1] = \sqrt{\hat{v}_{2x}^2[k-1] + \hat{v}_{2y}^2[k-1]},$$

$$\hat{\omega}_2[k-1] = (\hat{\theta}_2[k-1] - \hat{\theta}_2[k-2])/\delta,$$

where $\hat{v}_{2x}[k-1]$ and $\hat{v}_{2y}[k-1]$ are the estimated velocities along the $x$ and $y$ axes respectively, $\hat{v}_{2x}[k-1] = (\hat{x}_2[k-1] - \hat{x}_2[k-2])/\delta$, and $\hat{v}_{2y}[k-1] = (\hat{y}_2[k-1] - \hat{y}_2[k-2])/\delta$.

*2) Adaptive Optimal Control:* An online adaptive optimal controller is designed to maximize the expected value function (14) with the estimated system state. The existence and uniqueness of the optimal control policy is guaranteed here because of properties of the RSSI model (10)-(13) (shown in [10, Fig. 17]). In particular, to maximize $z_R[k]$, one needs to find $\theta_t[k]$ and $\theta_r[k]$ to maximize $G_{l|dBi}[k]$ as described in Equation (10). $G_{l|dBi}[k]$

is maximized when the heading angles of the two directional antennas are selected as $\theta_t[k] = \gamma_t[k]$ and $\theta_r[k] = \gamma_r[k]$ respectively, where $\gamma_t[k]$ and $\gamma_r[k]$ are uniquely determined by the positions of two UAVs and environment-related shift angles as shown in Equations (12) and (13).

Because the uncertain parameters are independent from the system state at time $k$, the value function for UAV 1 can be further rewritten as

$$\begin{aligned} V_1(\mathbf{x}[k]) &= E[\sum_{l=k}^{k+N} \alpha^{l-k} z_R[l](\mathbf{x}[l], u_1[k], u_2^*[k])] \\ &= E[z_R[k](\mathbf{x}[k], u_1[k], u_2^*[k]) \\ &\quad + \sum_{l=k+1}^{k+N} \alpha^{l-k} z_R[l](\mathbf{x}[l], u_1[k], u_2^*[k])]. \end{aligned} \tag{25}$$

where $u_2^*[k]$ is the optimal control policy of UAV 2.

The above equation can be solved backward-in-time using dynamic programming, or forward-in-time using RL [28], [29]. Here we use RL, in particular, the policy iteration method, to find the optimal control policy by iteratively conducting two steps: policy evaluation and policy improvement. The policy evaluation step is designed to solve the value function $V_1(\mathbf{x}[k])$ using Equation (25), given the current control policy. The policy improvement step is designed to find the best control policy to maximize the value function. The two steps are conducted iteratively until convergence.

***Policy Evaluation***

$$\begin{aligned} V_{1,j+1}(\mathbf{x}[k]) &= E[z_R[k](\mathbf{x}[k], u_{1,j}[k], u_2^*[k]) \\ &\quad + \sum_{l=k+1}^{k+N} \alpha^{l-k} z_{j,R}[l](\mathbf{x}[l], u_{1,j}[k], u_2^*[k])] \end{aligned} \tag{26}$$

***Policy Improvement***

$$\begin{aligned} u_{1,j+1}(\mathbf{x}[k]) = \arg\max_{u_{1,j}[k]} & E[z_R[k](\mathbf{x}[k], u_{1,j}[k], u_2^*[k]) \\ & + \sum_{l=k+1}^{k+N} \alpha^{l-k} z_{j+1,R}[l](\mathbf{x}[l], u_{1,j}[k], u_2^*[k])] \end{aligned} \tag{27}$$

where $j$ is the iteration step index, and $z_{j,R}[l](\mathbf{x}[l], u_{1,j}[k], u_2^*[k])$ is the RSSI model with parameters learned in the $j$th iteration step.

Note that Equation (26) involves three unknown parameters for the environment-specific RSSI model ($G_{t|dBi}^{\max}$, $G_{t|dBi}^{\min}$, and $\theta_{t_{env}}$), which need to be learned. In particular, for each iteration $j$, three time steps ($k$, $k+1$ and $k+2$) are needed to come up with three equations to iteratively solve for the three parameters. To solve the nonlinear equations, the Newton's method [45] is utilized here. Newton's method is a root-finding algorithm that iteratively finds better approximations to the roots of a real-valued function. To calculate the value function $V_{1,j+1}(\mathbf{x}[k])$ at each time step (Equation (26)), the uncertainty evaluation method needs to be utilized. To reduce the computational cost, we use the MPCM method here. In particular, define a system mapping $G_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])$ as the relationship between the value function and the random variables $v_2[T_j^2]$ and $r_2[T_j^2]$, i.e., $G_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2]) = z_R[k](\mathbf{x}[k], u_1[k], u_2^*[k]) + \sum_{l=k+1}^{k+N} \alpha^{l-k} z_R[l](\mathbf{x}[l], u_1[k], u_2^*[k])$. Then the value function $V_1(\mathbf{x}[k])$ can be estimated by evaluating the mean output of the system mapping using MPCM: $V_1(\mathbf{x}[k]) = E[G_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. According to Lemma 1, the output mean of this system mapping can be obtained using the evaluated outputs of a reduced-order mapping $G'_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])$ at each selected MPCM point, according to the procedures described in [42, Section II-B].

*Theorem 2:* Consider the random switching system shown in Equation (4), with the value function given by Equation (14). Given the current system state $\mathbf{x}[k]$, the optimal control policy is the solution found by applying the policy iteration of RL and approximating the value function using MPCM as shown in Equations (26) and (27).

*Proof:* Denote the optimal policies derived by evaluating the mean outputs of the reduced order mapping $G'_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])$ and the original value function $G_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])$ as $u_1'^*$ and $u_1^*$ respectively, i.e., $u_1'^* = \text{argmax}_{u_1} E[G'_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$, and $u_1^* = \text{argmax}_{u_1} E[G_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. To prove this theorem, we need to prove that $u_1'^* = u_1^*$, i.e., $\text{argmax}_{u_1} E[G'_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] = \text{argmax}_{u_1} E[G_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. This is equivalent to proving the following two statements: a) $\nexists u_1'^* \neq u_1^*$ such that $E[G'_{V_1}(\mathbf{x}[k], u_1'^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] > E[G_{V_1}(\mathbf{x}[k], u_1^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$, and b) $\nexists u_1'^* \neq u_1^*$ such that $E[G'_{V_1}(\mathbf{x}[k], u_1'^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] < E[G_{V_1}(\mathbf{x}[k], u_1^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. Here we use a contradiction approach to prove the above two statements. To prove the first statement, we assume there exists $u_1'^* \neq u_1^*$ such that $E[G'_{V_1}(\mathbf{x}[k], u_1'^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] > E[G_{V_1}(\mathbf{x}[k], u_1^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. According to Lemma 1, we have

$$\begin{aligned} &E[G'_{V_1}(\mathbf{x}[k], u_1'^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] \\ &= E[G_{V_1}(\mathbf{x}[k], u_1'^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] \\ &> E[G_{V_1}(\mathbf{x}[k], u_1^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])], \end{aligned}$$

which violates the assumption that $u_1^* = \text{argmax}_{u_1} E[G_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. Similarly, to prove that the second statement, we assume there exists $u_1'^* \neq u_1^*$ such that $E[G'_{V_1}(\mathbf{x}[k], u_1'^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] < E[G_{V_1}(\mathbf{x}[k], u_1^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. According to Lemma 1, we have

$$\begin{aligned} &E[G_{V_1}(\mathbf{x}[k], u_1^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] \\ &= E[G'_{V_1}(\mathbf{x}[k], u_1^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])] \\ &> E[G'_{V_1}(\mathbf{x}[k], u_1'^*[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])], \end{aligned}$$

which violates the assumption that $u_1'^* = \text{argmax}_{u_1} E[G'_{V_1}(\mathbf{x}[k], u_1[k], u_2^*[k], v_2[T_j^2], r_2[T_j^2])]$. As such, both statements a) and b) are true, and the result $u_1'^* = u_1^*$ is derived naturally. ■

*Theorem 3:* Consider the random switching system described in Equation (4). Given the current system state $\mathbf{x}[k]$, the optimal policy found by the decentralized control algorithm (shown in Section III-A2) maximizes the global value function described in Equation (14).

*Proof:* Denote the global optimal control policy that maximizes the value function described in Equation (14) as $(u_{1,g}^*[k], u_{2,g}^*[k])$. We need to show that $u_1^*[k] = u_{1,g}^*[k]$ and $u_2^*[k] = u_{2,g}^*[k]$. According to Theorem 2, $u_1^*[k]$ is the optimal solution to Equation (14) under the assumption that $u_2[k] = u_2^*[k]$. The global optimal control policy $u_{1,g}^*[k]$ can be regarded as the decentralized optimal solution with the assumption that $u_2[k] = u_{2,g}^*[k]$. We show that for each time $k$, the optimal solution of UAV 1 is unique for any given $u_2[k]$.

Note that given any heading angle of the transmitting antenna $\theta_t[k]$, the optimal heading angle of the receiving antenna is $\gamma_r[k]$ to maximize $G_{l|dBi}[k]$ in Equation (11). The desired heading angle $\gamma_r[k]$, which is described in Equation (12), is decided uniquely by the positions of the two UAVs and the environment, instead of the transmitting antennas' heading angle. In such cases, we have $\text{argmax}_{u_1[k]} z_R[k](\mathbf{x}[k], u_1[k], u_2^*[k]) = \text{argmax}_{u_1[k]} z_R[k](\mathbf{x}[k], u_1[k], u_{2,g}^*[k])$, and $\text{argmax}_{u_2[k]} z_R[k](\mathbf{x}[k], u_1^*[k], u_2[k]) = \text{argmax}_{u_2[k]} z_R[k](\mathbf{x}[k], u_{1,g}^*[k], u_2[k])$, which lead to the result that $u_1^*[k] = u_{1,g}^*[k]$ and $u_2^*[k] = u_{2,g}^*[k]$. The proof is completed. ■

### B. Using the Learned RSSI Model in Both GPS-Available and GPS-Denied Environments

With the learned RSSI model, the optimal solution can then be obtained in both GPS-available and GPS-denied environments. In a GPS-denied environment, the RSSI is the only measurement signal. In this case, the optimal control solution can be found following a similar procedure as shown in Section III-A, by replacing $\mathbf{z}_{G,2}[k]$ and $h_{G,2}$ with $z_R[k]$ and $h_R$. In the GPS-available environment, GPS and RSSI measurements can be fused to estimate the system states, using a fuzzy-logic based fusion algorithm [26] to improve the reliability. The details are not repeated here.

*Remark 2:* Note that RSSI is often calibrated for localization, in order to correct the environmental effects [46], [47]. This calibration can be captured by a calibrated propagation constant [46], which is environment-related and is usually found by conducting experiments in the testing area prior to implementing the localization algorithm. In our study, this calibrated propagation constant is captured by the environment-related parameters in the RSSI model, i.e., $G_{t|dBi}^{\max}$, $G_{t|dBi}^{\min}$, and $\theta_{t_{env}}/\theta_{r_{env}}$. In other words, the learning process we proposed in this paper, which learns the environment-related parameters, can be regarded as an RSSI calibration process in the literature. With the learned parameters, the RSSI model is calibrated and then used in the antenna alignment algorithm.

*Remark 3:* The above distributed antenna control solution assumes a pair of UAVs in the ACDA system. When multiple UAVs are involved, the communications among UAVs can be realized using controllable multi-sector directional antennas or phased array antennas [48], [49]. In such cases, the communication network can be regarded as a collection of UAV pairs. As such, the study on the communication link between a pair of UAVs developed in this paper is an important building block for a network of more than two UAVs.

## IV. Remote UAV Uncertain Intention Estimation

In this section, we provide an online uncertain intention estimation method to estimate the characteristics of the remote UAV's uncertain maneuver intentions. The estimation procedure includes two major steps adopted from [50]: 1) estimation of the trajectory-specific maneuvers at each time instant, and 2) estimation of the pdfs of uncertain variables: $f_v(v_2[T_j^2])$, $f_r(r_2[T_j^2])$, and $f_\tau(\tau_2[T_j^2])$. The uncertain intention estimation solution provided in [50] is offline. We here enhance it to an online process to reduce computational costs.

### A. Estimation of Trajectory-Specific Maneuvers

We develop two estimation procedures to estimate the two types of random variables in the ST RMM respectively.

*1) Estimation of type 1 Random Variables:* Type 1 random variables (i.e., velocity $v_2[T_j^2]$ and turn radius $r_2[T_j^2]$) describe the movement characteristics of each maneuver, and can be estimated from the system states. Given two consecutive system states $(\mathbf{x}_2[k-1] = (x_2[k-1], y_2[k-1], \theta_2[k-1])$ and $\mathbf{x}_2[k] = (x_2[k], y_2[k], \theta_2[k]))$, the type 1 random variables in the remote UAV system are estimated as

$$\hat{v}_2[k] = \sqrt{\hat{v}_{2x}^2[k] + \hat{v}_{2y}^2[k]},$$

$$\hat{r}_2[k] = \frac{\hat{v}_2[k]}{\hat{\omega}_2[k]},$$

where $\hat{\omega}_2[k]$ is the estimated angular velocity, and $\hat{\omega}_2[k] = (\theta_2[k] - \theta_2[k-1])/\delta$. $\hat{v}_{2x}[k]$ and $\hat{v}_{2y}[k]$ are the estimated velocities in $x$ and $y$ axes respectively, $\hat{v}_{2x}[k] = (\mathbf{x}[k] - x[k-1])/\delta$ and $\hat{v}_{2y}[k] = (y[k] - y[k-1])/\delta$.

*2) Estimation of type 2 Random Variable:* The type 2 random variable (i.e., travel time $\tau_2[T_j^2]$) describes how often the maneuvers are switched, and thus is estimated from the change of type 1 random variables. Different from [50] which uses the change of turn radius to find the length of each travel time, we here use the change of the angular velocity $\omega_2[T_j^2]$, which is affected by velocity $v_2[T_j^2]$ and turn radius $r_2[T_j^2]$. Therefore, to estimate $\tau_2[T_j^2]$, we scan the angular velocity $\omega_2[k]$ from $k = T_j^2$ at each time instant, until the change of $\omega_2[k]$ exceeds a threshold $\omega_2^{thrd}$. The travel time interval at $T_j^2$ is estimated as $\hat{\tau_2}[T_j^2] = k - T_j^2$. The determination of $\omega_2^{thrd}$ has a significant impact on the estimation performance. In general, a smaller threshold improves the estimation accuracy but decreases the predictability of the underlining model. Please refer to [50] for the detailed discussion about the threshold selection.

### B. Estimation of Pdfs of Uncertain Intention Variables

The pdfs of uncertain intention variables in the remote UAV system can be estimated from the trajectory-specific maneuvers. In particular, assuming that the random variables $v_2[T_j^2]$, $\frac{1}{r_2[T_j^2]}$, and $\tau_2[T_j^2]$ follow the uniform, Gaussian, and Poisson distributions respectively, then the parameters in the distributions: $v_{2\min}$ and $v_{2\max}$ (minimum and maximum velocity constraints), $\mu_2$ and $\sigma_2$ (mean and variance of $\frac{1}{r_2[T_j^2]}$), and $\lambda_2$ (expected value of $\tau_2[T_j^2]$), can be estimated from the following three steps.

*Step 1. Estimate the velocity pdf:* Denote the expectation and variance of velocity as $\mu_v$ and $\sigma_v^2$ respectively. $\mu_v$ and $\sigma_v^2$ can be estimated recursively as

$$\begin{aligned}\hat{\mu}_v[k] &= \frac{1}{k}\sum_{j=1}^{k}\hat{v}_2[j] = \frac{1}{k}\left(\sum_{j=1}^{k-1}\hat{v}_2[j] + \hat{v}_2[k]\right)\\ &= \frac{1}{k}\left((k-1)\hat{\mu}_v[k-1] + \hat{v}_2[k]\right)\\ &= \frac{k-1}{k}\hat{\mu}_v[k-1] + \frac{1}{k}\hat{v}_2[k], \end{aligned} \tag{28}$$

$$\begin{aligned}\hat{\sigma}_v^2[k] &= \frac{1}{k-1}\sum_{j=1}^{k}(\hat{v}_2[j] - \hat{\mu}_v[k])^2\\ &= \frac{1}{k-1}\left(\sum_{j=1}^{k-1}(\hat{v}_2[j] - \hat{\mu}_v[k])^2 + (\hat{v}_2[k] - \hat{\mu}_v[k])^2\right)\\ &= \frac{1}{k-1}\left((k-2)\hat{\sigma}_v^2[k-1] + (\hat{v}_2[k] - \hat{\mu}_v[k])^2\right)\\ &= \frac{k-2}{k-1}\hat{\sigma}_v^2[k-1] + \frac{1}{k-1}(\hat{v}_2[k] - \hat{\mu}_v[k])^2. \end{aligned} \tag{29}$$

*Remark 4:* Note that the sample mean of a random variable (i.e., $\frac{1}{k}\sum_{j=1}^{k}\hat{v}_2[j]$) is the minimum variance unbiased estimator (MVUE), and also, is the maximum likelihood estimator to $\mu_v$ [51]. To estimate $\sigma_v^2$, here we use the unbiased estimator ($\frac{1}{k-1}\sum_{j=1}^{k}(\hat{v}_2[j] - \hat{\mu}_v[k])^2$). The performance of the online estimation algorithm is as good as the offline solution proposed in [50] in terms of estimation accuracy. The equivalence of the two algorithms is shown in Equations (28) and (29). Here we enhance the offline method to an online process to reduce the computational costs. The offline method needs to reuse all previous data in the UAV uncertain intention estimation whenever new data arrives, while the online method only utilizes the newest data.

From the relation between $v_{2\min}$, $v_{2\max}$ and $\mu_v$, $\sigma_v^2$, the parameters in the velocity's pdf ($v_{2\min}$, $v_{2\max}$) can be estimated as

$$\begin{aligned}\hat{v}_{2\min}[k] &= \hat{\mu}_v[k] - \sqrt{3}\hat{\sigma}_v[k]\\ &= \frac{k-1}{k}\hat{\mu}_v[k-1] + \frac{1}{k}\hat{v}_2[k]\\ &\quad - \sqrt{\frac{3(k-2)}{k-1}\hat{\sigma}_v^2[k-1] + \frac{3}{k-1}(\hat{v}_2[k] - \hat{\mu}_v)^2}, \end{aligned} \tag{30}$$

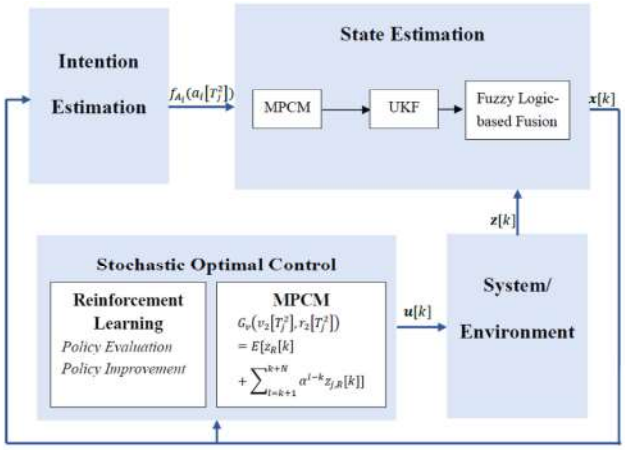



Fig. 3. Illustration of the proposed algorithm.

$$\begin{aligned}\hat{v}_{2\max}[k] &= \hat{\mu}_v[k] + \sqrt{3}\hat{\sigma}_v[k]\\ &= \frac{k-1}{k}\hat{\mu}_v[k-1] + \frac{1}{k}\hat{v}_2[k]\\ &\quad + \sqrt{\frac{3(k-2)}{k-1}\hat{\sigma}_v^2[k-1] + \frac{3}{k-1}(\hat{v}_2[k] - \hat{\mu}_v)^2}. \end{aligned} \tag{31}$$

*Step 2. Estimate the radius pdf:* The parameters in the radius pdf ($\mu_2$ and $\sigma_2^2$) are estimated recursively using $\hat{r}_2[k]$ following a similar procedure as described in Equations (28) and (29)

$$\hat{\mu}_2[k] = \frac{k-1}{k}\hat{\mu}_2[k-1] + \frac{1}{k}\frac{1}{\hat{r}_2[k]}, \tag{32}$$

$$\hat{\sigma}_2^2[k] = \frac{k-2}{k-1}\hat{\sigma}_2^2[k-1] + \frac{1}{k-1}\left(\frac{1}{\hat{r}_2}[k] - \hat{\mu}_2[k]\right)^2. \tag{33}$$

*Step 3. Estimate the travel time pdf:* $\lambda_2$ is the only parameter in the Poisson distribution, and can be estimated recursively from the mean of $\hat{\tau}_2[T_j^2]$ as

$$\hat{\lambda}_2[j] = \frac{j\hat{\lambda}_2[j-1]}{j - 1 + \hat{\lambda}_2[j-1]\hat{\tau}_2[T_j^2]}. \tag{34}$$

*Remark 5:* The uncertain intention estimation procedure can be implemented together with the stochastic optimal control procedure described in Section III. The overall algorithm structure is described in Fig. 3 . We also note that because the uncertain intention is estimated from the system states, which are estimated from the measurements, it is suggested to conduct the uncertain intention estimation procedure in a GPS-available environment, which helps to improve the reliability of the estimated system states.

## V. Simulation Studies

In this section, we conduct simulation studies to illustrate and validate the results and algorithms developed in this paper. Two UAVs move in a 2-D airspace following the ST RMM independently. Two directional antennas of the same type are

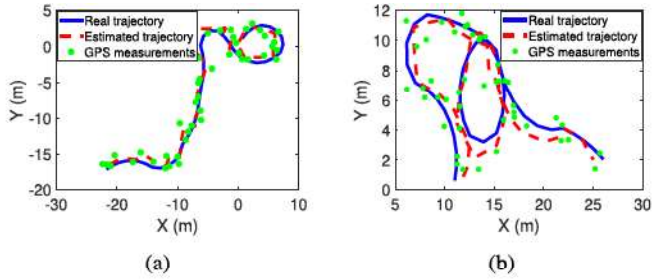



Fig. 4. (a) Trajectories of UAV 1, and (b) Trajectories of UAV 2. The blue solid curves are real trajectories, red dotted curves are estimated trajectories, and green dots are GPS measurements.

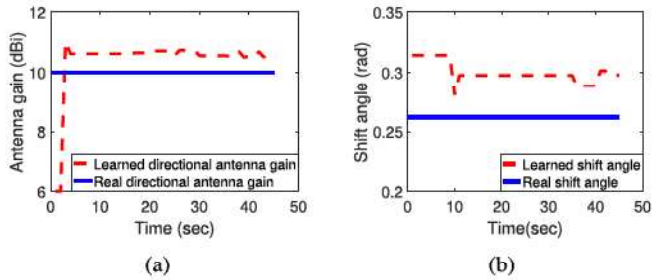



Fig. 5. Learned environment-specific (a) maximum directional antenna gain ($G^{\max}_{t|dBm}$), and (b) shift angle ($\theta_{env}$) in the RSSI model. The blue solid lines and red dotted curves represent the real and learned parameters respectively.

mounted on the two UAVs respectively. The design parameters of the directional antennas are selected as $n = 8$, and $d_a = \frac{\lambda}{10}$.

We first simulate the case when the GPS is available but the RSSI model is unknown. Gaussian noises are added to the GPS measurements. Estimation for UAV 1 is based on UKF with known maneuver ($v_1[k]$ and $r_1[k]$), while the estimation for UAV 2 is based on the integration of UKF and MPCM as described in Section III-A with unknown $v_2[k]$ and $r_2[k]$. Figs. 4(a) and 4(b) show the trajectories of UAV 1 and UAV 2 respectively in one realization with the simulation time $T = 45$ s and sampling period $\delta = 1$ s. To find the statistics of the estimation performance, 10 realizations with randomly generated trajectories are conducted. The mean estimation distance errors for the two UAVs are calculated over all realizations as $e_1 = 0.84$ m and $e_2 = 0.89$ m respectively. It can be seen from the simulations that the estimated trajectories for UAVs 1 and 2 are both close to their real trajectories, indicating that the proposed state estimation algorithm performs well in both known and unknown maneuver cases.

With the estimated states, we simulate the RL-based stochastic optimal control algorithm. To simulate the long-distance communication scenario, the minimum received signal strength is assumed to be 0, and in this case, the directional antennas' minimum gain ($G^{\min}_{t|dBi}$) can be calculated accordingly. Figs. 5(a) and 5(b) show the learned environment-specific antennas' maximum gain ($G^{\max}_{t|dBi}$) and the shift angle caused by the environment ($\theta_{t_{env}}$) respectively. Gaussian noises are added to the RSSI measurements. To avoid unnecessary divergence, we limit the maximum values of the two parameters. In particular, we assume

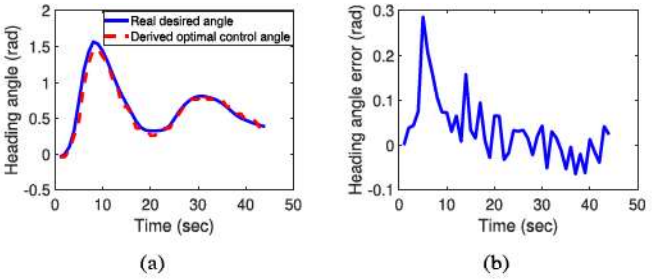



Fig. 6. (a) Obtained optimal heading angles with GPS signals and unknown RSSI model. The blue solid curve is the real optimal angles, and the red dotted curve is the obtained optimal angles. (b) Heading angle errors between the derived heading angles and the real optimal heading angles.

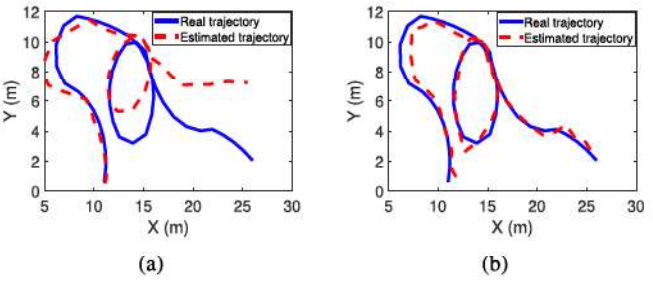



Fig. 7. (a) Trajectories of UAV 2 in (a) GPS-denied, and (b) GPS-available environments. The blue solid curves are the real trajectories, and the red dotted curves are the estimated trajectories.

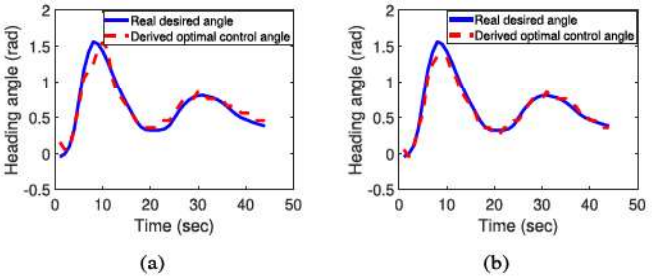


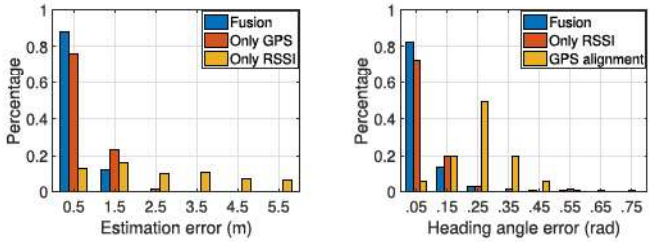

Fig. 8. Obtained optimal heading angles in (a) GPS-denied, and (b) GPS-available environments. The blue solid and red dotted curves are the real optimal heading angles and derived heading angels respectively.

the directional antenna's maximum gain is no more than the maximum gain given in the data sheet, and the environment-specific shift angle is no more than 20 degrees. As shown in the figures, the learned parameters are very close to their true values, which indicates the effectiveness of the learning algorithm. Figs. 6(a) and 6(b) show the derived optimal heading angles of the local directional antenna and the heading angle errors between the derived and real optimal heading angles in one realization. The small angle errors indicate the good performance of the proposed RL-based stochastic optimal control algorithm.

With the learned RSSI model, we simulate the proposed stochastic optimal control algorithms in both GPS-denied and GPS-available environments. Note that in the GPS-denied environment, RSSI is the only measurement signal, while in the GPS-available case, both GPS and RSSI signals are fused. Figs. 7 and 8 show the performance of state estimation and optimal control

TABLE I
ESTIMATION PERFORMANCE

| | UKF-based | | | GPS signals |
|---|---|---|---|---|
| | GPS | RSSI | Fusion | |
| Mean distance error (m) | 0.89 | 5.13 | 0.56 | 1.25 |




(a) (b)

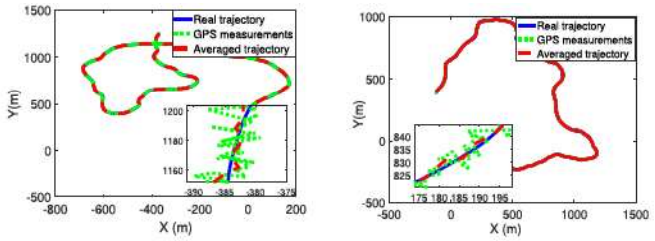
Fig. 9. Barplots of (a) Estimation errors, and (b) Heading angle errors.

TABLE II
CONTROL PERFORMANCE

| | RL-based | | GPS alignment-based |
|---|---|---|---|
| | RSSI | Fusion | |
| Mean RSSI (dBm) | -31 | -29 | -47 |
| Mean angle error (rad) | 0.09 | 0.07 | 0.3 |

algorithms respectively, in both GPS-denied and GPS-available environments. We have simulated 10 realizations with randomly generated UAV trajectories, and calculated the mean estimation errors, RSSI signals, and heading angle errors over all 10 realizations. The system state estimation performance is shown in Table I and Fig. 9(a), where "GPS," "RSSI," and "Fusion" represent the UKF-based state estimation using only GPS, only RSSI, and both GPS and RSSI signals respectively. The column "GPS signals" shows the raw GPS measurements. The optimal control performance is shown in Table II and Fig. 9(b), where "RSSI" and "Fusion" represent the control algorithms based on only RSSI, and both RSSI and GPS respectively. It can be seen from the tables and plots that: 1) the estimated system states and derived heading angles are very close to their real states and optimal heading angles in both GPS-denied and GPS-available cases, indicating that the proposed algorithms work well in both GPS-available and GPS-denied environments; 2) the estimation errors and heading angle errors in the GPS-available case are much smaller than that in the GPS-denied case, indicating that the fusion of the GPS and RSSI promises a better performance.

To provide comparative studies, we also simulate the GPS alignment-based directional antenna control algorithm developed in [11]. In this algorithm, each directional antenna points towards the GPS location of the other UAV to align the directional antennas, and RSSI is not used as a measurement signal nor value function. The control performance of this GPS alignment-based algorithm is also shown in Table II. The barplots of the controlled heading angle errors are shown in Fig. 9(b). It can be seen from the tables and plots that the optimal control algorithm developed in our paper performs much better than the GPS alignment-based algorithm, with larger RSSI signals and less heading angle errors.




(a) (b)

Fig. 10. Trajectories of (a) UAV 1, and (b) Trajectories of UAV 2. The blue solid curves are real trajectories, red dotted curves are estimated trajectories, and green dots are GPS measurements.

TABLE III
PERFORMANCE OF ONLINE INTENTION ESTIMATION

| Random variables | $v_2[T_i^2]$ | | $\frac{1}{r_2[T_i^2]}$ | | $\tau_2[T_i^2]$ |
|---|---|---|---|---|---|
| Parameters to be estimated | $\mu_v$ | $\sigma_v^2$ | $\mu_2$ | $\sigma_2^2$ | $\lambda_2$ |
| Estimated value | 12.7 | 6.3 | $10^{-4}$ | $10^{-3}$ | 2.07 |
| Real value | 12.5 | 2.1 | 0 | $10^{-4}$ | 2 |

Finally we simulate the remote UAV uncertain intention estimation algorithm. The total simulation time in this part is set as $T = 10$ minutes, with the sampling period $\delta = 1$ s. The system states are estimated from the GPS measurements by adopting the moving average method. Figs. 10(a) and 10(b) show trajectories of the two UAVs respectively. The performance of the uncertain intention estimation algorithm is shown in Table III. Note that $v_2[T_i^2]$, $\frac{1}{r_2[T_i^2]}$, and $\tau_2[T_i^2]$ follow uniform, Gaussian, and Possion distributions respectively. As such, the parameters to be estimated in their pdfs are: $\mu_v$ and $\sigma_v$ for $v_2[T_i^2]$, $\mu_2$ and $\sigma_2$ for $\frac{1}{r_2[T_i^2]}$, and $\lambda_2$ for $\tau_2[T_i^2]$ respectively. It can be seen from the table that the estimated means of $v_2[T_i^2]$, $\frac{1}{r_2[T_i^2]}$, and $\tau_2[T_i^2]$ match with their real mean values perfectly, indicating the effectiveness of the proposed estimation algorithm. The estimated variance of $v_2[T_i^2]$ and $1r_2[T_i^2]$ show small biases to their real values, caused by Gaussian GPS noises.

## VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we developed a RL-based online directional antenna control solution for the ACDA system to establish a robust long-distance air-to-air communication channel using pure directional antennas. In particular, to capture the uncertain intentions of UAVs executing surveillance-like missions for better tracking, we adopted a UAV ST RMM. With this nonlinear random switching mobility model, a new state estimation algorithm that integrates MPCM and UKF was developed. To account for an unstable GPS environment, we developed a new RL-based stochastic optimal control solution, which features a learning of communication RSSI models to provide an additional measurement that compensates GPS signals. This solution also features an integration of RL and MPCM to learn the environment-specific RSSI model and to provide online optimal control solutions. With the learned RSSI model, the optimal solutions in both GPS-available and GPS-denied environments

were developed. The learning and uncertainty-exploited decision framework is generally applicable to distributed decision-making of nonlinear multi-agent systems that are governed by random intentions in an uncertain environment. In the future work, we will further investigate the ACDA system in 3-D UAV RMMs. In addition, we will expand the two-UAV aerial communication link study to multi-UAV communication networks equipped with multi-sector directional antennas. We will also pursue implementation of the entire ACDA solution in hardware platforms.

## REFERENCES


[1] M. Liu, Y. Wan, S. Li, and F. L. Lewis, "A learning and uncertainty-exploited directional antenna control solution for robust aerial networking," in *Proc. IEEE Vehicle Technol. Conf.*, Honolulu, HI, 2019.

[2] S. Winkler, S. Zeadally, and K. Evans, "Privacy and civilian drone use: The need for further regulation," *IEEE Secur. Privacy*, vol. 16, no. 5, pp. 72–80, Sep./Oct. 2018.

[3] O. S. Oubbati, N. Chaib, A. Lakas, P. Lorenz, and A. Rachedi, "Uav-assisted supporting services connectivity in urban VANETs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3944–3951, Apr. 2019.

[4] S. Hu, "On ergodic capacity and optimal number of tiers in UAV-assisted communication systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2814–2824, Mar. 2019.

[5] Y. Wan and S. Fu, "Communicating in remote areas or disaster situations using unmanned aerial vehicles," *Homeland Secur. Today Mag.*, pp. 32–35, 2015.

[6] K. Li, R. C. Voicu, S. S. Kanhere, W. Ni, and E. Tovar, "Energy efficient legitimate wireless surveillance of UAV communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2283–2293, Mar. 2019.

[7] S. Hayat, E. Yanmaz, and R. Muzaffar, "Survey on unmanned aerial vehicle networks for civil applications: A communications viewpoint," *IEEE Commun. Surv. Tut.*, vol. 18, no. 4, pp. 2624–2661, Fourthquarter 2016.

[8] B.-N. Cheng, A. Coyle, S. McGarry, I. Pedan, L. Veytser, and J. Wheeler, "Characterizing routing with radio-to-router information in a heterogeneous airborne network," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4183–4195, Aug. 2013.

[9] A. I. Alshbatat and L. Dong, "Performance analysis of mobile ad hoc unmanned aerial vehicle communication networks with directional antennas," *Int. J. Aerosp. Eng.*, vol. 2010, 2010.

[10] S. Li *et al.*, "The design and implementation of aerial communication using directional antennas: Learning control in unknown communication environments," *IET Control Theory Appl.*, 2019.

[11] J. Chen *et al.*, "Long-range and broadband aerial communication using directional antennas (acda): Design and implementation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10 793–10 805, Jul. 2017.

[12] Y. Gu, M. Zhou, S. Fu, and Y. Wan, "Airborne wifi networks through directional antennae: An experimental study," in *Proc. IEEE Wireless Commun. Netw. Conf.*, New Orleans, LA, 2015, pp. 1314–1319.

[13] J. Xie, F. AI-Emrani, Y. Gu, Y. Wan, and S. Fu, "UAV-carried long-distance wi-fi communication infrastructure," in *Proc. AIAA Infotech@ Aerosp.*, San Diego, CA, 2016.

[14] E. Yanmaz, R. Kuschnig, and C. Bettstetter, "Achieving air-ground communications in 802.11 networks with three-dimensional aerial mobility," in *Proc. IEEE INFOCOM*, Turin, Italy, 2013, pp. 120–124.

[15] C. Danilov *et al.*, "Experiment and field demonstration of a 802.11-based ground-UAV mobile ad-hoc network," in *Proc. IEEE Mil. Commun. Conf.*, Boston, MA, 2009, pp. 1–7.

[16] K. Wakita, J. Huang, P. Di, K. Sekiyama, and T. Fukuda, "Human-walking-intention-based motion control of an omnidirectional-type cane robot," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 1, pp. 285–296, Feb. 2013.

[17] Y. Li and S. S. Ge, "Human-robot collaboration based on motion intention estimation," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 3, pp. 1007–1014, Jun. 2014.

[18] H. Modares, I. Ranatunga, F. L. Lewis, and D. O. Popa, "Optimized assistive human–robot interaction using reinforcement learning," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 655–667, Mar. 2016.

[19] T. Takeda, Y. Hirata, and K. Kosuge, "Dance step estimation method based on HMM for dance partner robot," *IEEE Trans. Ind. Electron.*, vol. 54, no. 2, pp. 699–706, Apr. 2007.

[20] S. Lefèvre, C. Laugier, and J. Ibañez-Guzmán, "Exploiting map information for driver intention estimation at road intersections," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Baden-Baden, Germany, 2011, pp. 583–588.

[21] J. Xie, Y. Wan, J. H. Kim, S. Fu, and K. Namuduri, "A survey and analysis of mobility models for airborne networks," *IEEE Commun. Surv. Tut.*, vol. 16, no. 3, pp. 1221–1238, Third Quarter 2014.

[22] M. Liu, Y. Wan, and F. L. Lewis, "Analysis of the random direction mobility model with a sense-and-avoid protocol," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Singapore, Singapore, 2017, pp. 1–6.

[23] M. Liu and Y. Wan, "Analysis of random mobility model with sense and avoid protocols for UAV traffic management," in *Proc. AIAA Inf. Syst.-AIAA Infotech@ Aerospace*, Kissimmee, FL, 2018.

[24] Y. Wan, K. Namuduri, Y. Zhou, and S. Fu, "A smooth-turn mobility model for airborne networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3359–3370, Sep. 2013.

[25] J. Xie, Y. Wan, K. Namuduri, S. Fu, G. L. Peterson, and J. F. Raquet, "Estimation and validation of the 3D smooth-turn mobility model for airborne networks," in *Proc. IEEE Mil. Commun. Conf.*, San Diego, CA, 2013, pp. 556–561.

[26] J. Yan, Y. Wan, S. Fu, X. J., S. Li, and K. Lu, "Rssi-based decentralized control for robust long-distance aerial networks using directional antennas," *IET Control Theory Appl.*, vol. 11, no. 11, pp. 1838–1847, 2016.

[27] M. K. Haider and E. W. Knightly, "Mobility resilience and overhead constrained adaptation in directional 60 ghz wlans: Protocol design and system implementation," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Paderborn, Germany, 2016.

[28] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, Third Quarter 2009.

[29] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Syst.*, vol. 32, no. 6, pp. 76–105, Dec. 2012.

[30] J. Xie, Y. Wan, K. Mills, J. J. Filliben, and F. L. Lewis, "A scalable sampling method to high-dimensional uncertainties for optimal and reinforcement learning-based controls," *IEEE Control Syst. Lett.*, vol. 1, no. 1, pp. 98–103, Jul. 2017.

[31] J. Xie, Y. Wan, and F. L. Lewis, "Strategic air traffic flow management under uncertainties using scalable sampling-based dynamic programming and Q-learning approaches," in *Proc. IEEE 11th Asian Control Conf.*, Gold Coast, QLD, Australia, 2017, pp. 1116–1121.

[32] A. Fotouhi *et al.*, "Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Commun. Surv. Tut.*, 2019.

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

[34] B. Galkin, J. Kibilda, and L. A. DaSilva, "Coverage analysis for low-altitude UAV networks in urban environments," in *Proc. IEEE Global Commun. Conf.*, Singapore, 2017, pp. 1–6.

[35] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *Proc. IEEE Global Commun. Conf.*, San Diego, CA, 2015, pp. 1–6.

[36] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. Tata McGraw-Hill Education, 2002.

[37] T. Li, Y. Wan, M. Liu, and F. L. Lewis, "Estimation of random mobility models using the expectation-maximization method," in *Proc. IEEE 14th Int. Conf. Control Autom.*, Anchorage, AK, 2018, pp. 641–646.

[38] T. S. Rappaport *et al.*, *Wireless Communications: Principles and Practice*. PTR New Jersey, 1996.

[39] L. Josefsson and P. Persson, *Conformal Array Antenna Theory and Design*. Hoboken, NJ, USA: John Wiley & Sons, 2006.

[40] "Ubiquity nanostation loco m5," in https://www.ubnt.com/airmax/nanostationm/

[41] J. D. Kraus, *Antennas*. New York, NY, USA: McGraw-Hill Education, 1988.

[42] Y. Zhou *et al.*, "Multivariate probabilistic collocation method for effective uncertainty evaluation with application to air traffic flow management," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 44, no. 10, pp. 1347–1363, Oct. 2014.

[43] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
[44] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter," *Kalman Filtering and Neural Networks*, pp. 221–280, 2001.
[45] A. Gil, J. Segura, and N. M. Temme, *Numerical Methods for Special Functions*. Siam, 2007.
[46] W. Y. Chung and E. E. L. Lau, "Enhanced RSSI-based real-time user location tracking system for indoor and outdoor environments," in *Proc. Int. Conf. Convergence Inf. Technol.*, Gyeongju, South Korea, 2007, pp. 1213–1218.
[47] Z. Fang, Z. Zhao, D. Geng, Y. Xuan, L. Du, and X. Cui, "RSSI variability characterization and calibration method in wireless sensor network," in *Proc. IEEE Int. Conf. Inf. Autom.*, Harbin, China, 2010, pp. 1532–1537.
[48] A. P. Subramanian, H. Lundgren, T. Salonidis, and D. Towsley, "Topology control protocol using sectorized antennas in dense 802.11 wireless networks," in *Proc. 17th IEEE Int. Conf. Netw. Protocols*. Princeton, NJ, 2009, pp. 1–10.
[49] M. Liu, Y. Wan, and F. L. Lewis, "Adaptive optimal decision in multi-agent random switching systems," *IEEE Control Syst. Lett.*, vol. 4, no. 2, pp. 265–270, Apr. 2019.
[50] J. Xie, Y. Wan, B. Wang, S. Fu, K. Lu, and J. H. Kim, "A comprehensive 3-dimensional random mobility modeling framework for airborne networks," *IEEE Access*, vol. 6, pp. 22 849–22 862, Mar. 2018.
[51] N. E. Nahi, *Estimation Theory and Applications*. New York, NY, USA: Wiley, 1969.

**Mushuang Liu** received the B.S. degree in electrical engineering from the University of Electronic Science and Technology of China, Chendu, China in 2016. She is now working toward the Ph.D. degree in the department of electrical engineering in the University of Texas at Arlington. Her research interests include distributed decisions for multi-agent systems, uncertain systems, multi-player games, graphical games, reinforcement learning, and their applications to UAV traffic management and UAV networking.

**Yan Wan** is currently an Associate Professor with the Electrical Engineering Department at the University of Texas, Arlington. She received the Ph.D. degree in electrical engineering from Washington State University in 2008 and then did postdoctoral training at the University of California, Santa Barbara. Her research interests lie in the modeling and control of large-scale dynamical networks, cyber-physical system, stochastic networks, learning control, networking, uncertainty analysis, algebraic graph theory, and their applications to UAV networking, UAV traffic management, epidemic spread, complex information networks, and air traffic management. Dr. Wans research has led to over 170 publications and successful technology transfer outcomes. She has received prestigious awards, including the NSF CAREER Award, RTCA William E. Jackson Award, U.S. Ignite and GENI demonstration awards, IEEE WCNC and ICCA Best Paper Award, and Tech Titan of the Future-University Level Award.

**Songwei Li** received the B.S. degree from Nanchang Hangkong University, Jiangxi, China, in 2010, the M.S. degree from the University of North Texas in 2016, and the Ph.D. degree from the University of Texas at Arlington in 2019. His research interests lie in the path planning and tracking, wireless sensor networks, SLAM, and target detection and tracking based on Radar.

**Frank L. Lewis** received the bachelor's degree in physics/EE and the MSEE at Rice University, the M.S. degree in aeronautical engineering from Univ. W. Florida, and the Ph.D. degree from Ga. Tech. He works in feedback control, intelligent systems, cooperative control systems, and nonlinear systems. He is currently working as PE Texas, U.K., Chartered Engineer, UTA Distinguished Scholar Professor, UTA Distinguished Teaching Professor, and Moncrief-O'Donnell Chair at the University of Texas at Arlington Research Institute. He is Author of seven U.S. patents, numerous journal special issues, 420 journal papers, and 20 books. He received the Fulbright Research Award, NSF Research Initiation Grant, ASEE Terman Award, Int. Neural Network Soc. Gabor Award, U.K. Inst Measurement & Control Honeywell Field Engineering Medal, IEEE Computational Intelligence Society Neural Networks Pioneer Award, AIAA Intelligent Systems Award. Received Outstanding Service Award from Dallas IEEE Section, selected as Engineer of the year by Ft. Worth IEEE Section. He was listed in Ft. Worth Business Press Top 200 Leaders in Manufacturing. Texas Regents Outstanding Teaching Award 2013. He is a member of the National Academy of Inventors, fellow IEEE, fellow IFAC, fellow AAAS, fellow U.K. Institute of Measurement Control. He is also a Founding Member of the Board of Governors of the Mediterranean Control Association.

**Shengli Fu** received the B.S. and M.S. degrees in telecommunication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 1994 and 1997, respectively, the M.S. degree in computer engineering from Wright State University, Dayton, OH, in 2002, and the Ph.D. degree in electrical engineering from the University of Delaware, Newark, DE, in 2005. He is currently a Professor and the Chair in the Department of Electrical Engineering, University of North Texas, Denton, TX. His research interests include coding and information theory, wireless communications and sensor networks, aerial communication, and UAS networks.