UAV Swarm Tracking in Low-Altitude Economies: A Robust Online Learning-based Control Strategy

Minjie Tang*, Chenyuan Feng†, Xijun Wang⋆, Chenxi Liu‖, Yongming Huang¶,§, Tony Q. S. Quek‡,§
* Department of Communication Systems, EURECOM, France
† Department of Computer Science, University of Exeter, U.K.
⋆ School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China
‖ Beijing University of Posts and Telecommunications, Beijing, China
¶ School of Information Science and Engineering, Southeast University, Nanjing, China
§ Purple Mountain Laboratories, Nanjing, China
‡ Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore
Emails: Minjie.Tang@eurecom.fr, c.feng@exeter.ac.uk, wangxijun@mail.sysu.edu.cn, chenxi.liu@bupt.edu.cn, huangym@seu.edu.cn, tonyquek@sutd.edu.sg

***Abstract*—With the rapid development of low-altitude applications, ensuring robust control of UAV swarms under uncertain communication environments has become a critical challenge. This paper investigates online robust tracking control for a UAV swarm with one leader and multiple followers communicating over unreliable multiple-input multiple-output (MIMO) channels. A dynamic model is constructed to capture uncertainties arising from both inter-UAV coordination and wireless communication. The control problem is reformulated as a virtual ergodic optimal control task for an auxiliary system that explicitly accounts for these uncertainties, and the optimal solution is shown to guarantee robust stabilization. To address the curse of dimensionality, we derive a reduced-order formulation and develop an online learning framework based on stochastic approximation (SA) with almost sure convergence guarantees. Simulation results verify that the proposed method significantly enhances tracking stability, convergence speed, and computational efficiency compared with widely used baselines under stochastic channel conditions and coupled system–communication uncertainties. This work not only establishes theoretical foundations for robust UAV swarm control but also provides practical strategies for large-scale deployment.**

***Index Terms*—Robust tracking control, UAV swarm, MIMO fading channel, stochastic approximation, online learning.**

I. Introduction

With the rapid rise of the low-altitude economy, UAV swarms are expected to play a pivotal role in autonomous logistics, urban air mobility, and emergency response, where robust and intelligent control strategies are essential for ensuring safety, efficiency, and scalability. We consider a representative UAV swarm configuration consisting of a *leader UAV* and multiple *follower UAVs*, as illustrated in Fig. 1. The leader UAV collects real-time state information (e.g. positions, velocities, and angular velocities) from the followers and periodically generates tracking control signals, which are transmitted to the followers via a wireless communication network. Upon reception, each follower adjusts its state to maintain predefined trajectories. However, the wireless channels between the leader and followers are inherently unreliable, often subject to fading and additive noise. Moreover, UAV system dynamics may not be precisely known due to modeling inaccuracies and system complexity. These uncertainties can degrade tracking performance and compromise swarm coordination, posing significant challenges for robust control design.

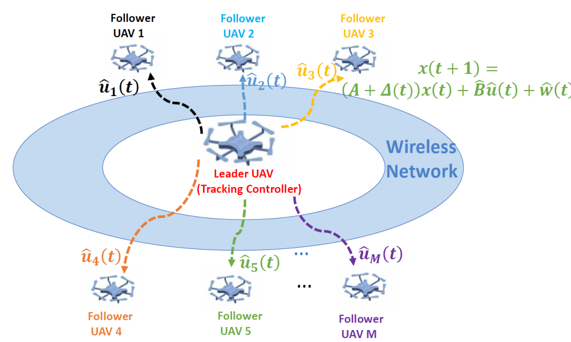


Fig. 1: Example of the UAV swarm tracking control.

Extensive research has focused on tracking control of UAV swarms, often under the assumption of static and reliable leader–follower communication [1]–[9]. Early works rely on frequency-domain analysis with Proportional-Integral-Derivative (PID) control [1], where empirically tuned parameters generate control signals based on tracking errors. Although simple to implement, PID controllers lack robustness and perform poorly in practice. More advanced approaches, including Linear-Quadratic-Tracker (LQT) and Model Predictive Control (MPC), optimize control policies by explicitly modeling UAV dynamics [2]. However, these methods are generally developed assuming perfect system knowledge and static communication,

This work was funded in part by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee MSCA-PF Grant EP/Z53433X/1, and in part by the National Natural Science Foundation of China under Grant 62301328. *Corresponding author: Chenyuan Feng*.

limiting their effectiveness under time-varying wireless channels and system uncertainties. To examine UAV control under unreliable communication, robust control under packet loss has been studied using simplified independent and identically distributed (i.i.d.) on–off models [3], while more realistic fading models such as Gaussian, Rayleigh, and Rician have been investigated in UAV swarm coordination [4] and event-triggered regulation [5]. However, these approaches often assume perfect channel state information (CSI) or system dynamics, which is rarely attainable in practice. In reality, UAV dynamics are affected by environmental variability and modeling errors, while CSI acquisition suffers from estimation inaccuracies, leading to degraded performance in swarm scenarios with both system and communication uncertainties. To further handle system-level uncertainties, disturbance-observer-based control (DOBC) mitigates external disturbances in adaptive finite-time control enhances responsiveness under dynamic conditions [6], and trajectory planning improves coordination in uncertain environments [7]. However, most existing methods assume ideal communication and neglect wireless channel impairments. In practice, fading, interference, and CSI estimation errors are stochastic and time-varying, and cannot be modeled as simple bounded disturbances. This makes conventional strategies inadequate for UAV swarms, highlighting the need for approaches that jointly address both system and channel uncertainties.

To address these challenges, we address robust tracking control for UAV swarms operating over general wireless MIMO fading channels. The key contributions of this work are summarized as follows: i) Unlike prior studies assuming perfect dynamics or simplified channels with ideal CSI, we formulate the UAV swarm control problem under random fading channels with both system and channel uncertainties, and recast it as a virtual ergodic optimal tracking problem enabling efficient closed-form robust controller design; ii) While robust control can be derived by solving the Bellman optimality equation [10], this is computationally prohibitive due to the continuous state spaces of CSI and UAV states. To address this, we develop a structured reduced-order optimality equation that enables efficient online learning of the robust policy. Using Lyapunov analysis [11], we further establish a sufficient condition guaranteeing robust tracking stability under both system and channel uncertainties; iii) To handle time-varying CSI and coupled system–channel uncertainties, we further develop an online learning algorithm based on the structured reduced-order optimality equation and stochastic approximation (SA) [12], with an almost sure convergence guarantee.

II. System Model

A. UAV Dynamic Model

A typical UAV swarm consists of a leader UAV and $M \in \mathbb{Z}_+$ geographically distributed followers, interconnected via an unreliable wireless network, as illustrated in Fig. 1. The leader and each follower UAV are equipped with N_t transmit and N_r receive antennas, respectively. For $m \in \{1, 2, \dots, M\}$, the dynamics of the m-th follower UAV are described by the following first-order coupled equations:

$$\mathbf{x}_m(t+1) = (\mathbf{A}_{mm} + \Delta_{mm}(t))\mathbf{x}_m(t) + \mathbf{B}_m \widehat{\mathbf{u}}_m(t) + \mathbf{w}_m(t) + \sum_{n \neq m} (\mathbf{A}_{mn} + \Delta_{mn}(t))\mathbf{x}_n(t), \tag{1}$$

where $\mathbf{x}_m(t) = [p_{m,x}(t), p_{m,y}(t), p_{m,z}(t), v_{m,x}(t), v_{m,y}(t), v_{m,z}(t), a_{m,x}(t), a_{m,y}(t), a_{m,z}(t)]^T \in \mathbb{R}^{9\times 1}$ denotes the state vector of the m-th follower UAV with initial condition $\mathbf{x}_m(0) = \mathbf{x}_0$, with $p_{m,n}(t)$, $v_{m,n}(t)$, and $a_{m,n}(t)$ representing the position, velocity, and acceleration along the n-th axis. The matrices $\mathbf{A}_{mm} \in \mathbb{R}^{9\times 9}$ and $\mathbf{B}_m \in \mathbb{R}^{9\times N_r}$ denote the internal transition and actuation matrices of the m-th UAV, while $\mathbf{A}_{mn} \in \mathbb{R}^{9\times 9}$ characterizes the state-transition coupling from UAV n to UAV m. The uncertainties $\Delta_{mm}(t) \in \mathbb{R}^{9\times 9}$ and $\Delta_{mn}(t) \in \mathbb{R}^{9\times 9}$ capture deviations in the internal dynamics of UAV m and the coupling from UAV n, respectively. The received control input is $\widehat{\mathbf{u}}_m(t) \in \mathbb{R}^{N_r\times 1}$, and $\mathbf{w}_m(t) \in \mathbb{R}^{9\times 1}$ denotes additive plant noise with covariance $\mathbf{W}_m \in \mathbb{S}^9_{++}$.

By aggregating the individual UAV dynamics, the swarm admits the global state-space model

$$\mathbf{x}(t+1) = (\mathbf{A} + \Delta(t))\mathbf{x}(t) + \widehat{\mathbf{B}}\widehat{\mathbf{u}}(t) + \widehat{\mathbf{w}}(t), \tag{2}$$

where $\mathbf{x}(t) = [\mathbf{x}_1^T(t), \dots, \mathbf{x}_M^T(t)]^T \in \mathbb{R}^{9M\times 1}$ is the global state vector, and $\widehat{\mathbf{u}}(t) = [\widehat{\mathbf{u}}_1^T(t), \dots, \widehat{\mathbf{u}}_M^T(t)]^T \in \mathbb{R}^{9N_r\times 1}$ is the global received control input. The global transition matrix is $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{M1} & \cdots & \mathbf{A}_{MM} \end{bmatrix} \in \mathbb{R}^{9M\times 9M}$, and the global actuation matrix is $\widehat{\mathbf{B}} = \mathrm{Diag}(\mathbf{B}_1, \dots, \mathbf{B}_M) \in \mathbb{R}^{9M\times N_r M}$. The transition uncertainty is $\Delta(t) = \begin{bmatrix} \Delta_{11}(t) & \cdots & \Delta_{1M}(t) \\ \vdots & \ddots & \vdots \\ \Delta_{M1}(t) & \cdots & \Delta_{MM}(t) \end{bmatrix} \in \mathbb{R}^{9M\times 9M}$. The global additive noise is modeled as $\widehat{\mathbf{w}}(t) \sim \mathcal{N}(\mathbf{0}_{9M\times 1}, \mathrm{Diag}(\mathbf{W}_1, \dots, \mathbf{W}_M))$.

For efficient robust controller design, we make the following assumption regarding the system uncertainty $\Delta(t)$.

Assumption 1 (System Uncertainties): There exist a positive definite matrix $\mathbf{F} \in \mathbb{S}^{9M}_{+}$ and a positive constant $\mu > 0$ such that $\mu^{-1}\mathbb{E}[\Delta^T(t)\Delta(t)] \leq \mathbf{F}$.

Note that μ and $\mathbf{F}$ in Assumption 1 are both essential: $\mathbf{F}$ embeds the structural distribution of uncertainties into the per-stage cost (10), while μ acts as a slack variable decoupling the bound from $\mathbf{F}$ and enabling flexible stability analysis (see Theorem 2). Together, $(\mu, \mathbf{F})$ ensure tractability and stability guarantees. Unlike many robust control works [13], [14] requiring uniformly bounded uncertainties, our assumption only requires average boundedness, thus being less restrictive.

B. Wireless Communication Model

The leader UAV observes the follower states $\mathbf{x}(t)$ using a depth-of-field (DOF) camera [15] and generates a tracking control signal $\mathbf{u}_m(t) \in \mathbb{R}^{N_t\times 1}$ for each follower UAV m. These signals are transmitted over an orthogonal frequency-division multiple access (OFDMA) MIMO network, where each follower is allocated a dedicated subcarrier to avoid interference. The received signal at the m-th follower, denoted $\widehat{\mathbf{u}}_m(t) \in \mathbb{R}^{N_r \times 1}$, is given by:

$$\widehat{\mathbf{u}}_m(t) = \delta_m(t)\mathbf{H}_m(t)\mathbf{u}_m(t) + \mathbf{v}_m(t), \quad 1 \le m \le M, \quad (3)$$

where $\delta_m(t) \in \{0, 1\}$ is an i.i.d. Bernoulli random variable across UAVs and timeslots with $\Pr(\delta_m(t) = 1) = p \in [0, 1]$ representing the probabilistic activation of the communication link between the leader and the m-th follower. The additive Gaussian noise is $\mathbf{v}_m(t) \sim \mathcal{N}(\mathbf{0}_{N_r \times 1}, \mathbf{I}_{N_r})$, and the MIMO fading gain $\mathbf{H}_m(t) \sim \mathcal{N}(\mathbf{0}_{N_r \times N_t}, \mathbf{I}_{N_r})$ models the random channel coefficients induced by multipath propagation, which is constant within each timeslot and i.i.d. across UAVs and timeslots [16].

To capture channel estimation errors, we assume that the leader UAV has access to an uncertain estimate $\mathbf{H}^e_m(t) \in \mathbb{R}^{N_r \times N_t}$ of the true channel gain $\mathbf{H}_m(t)$, modeled as [17]

$$\mathbf{H}^e_m(t) = \alpha \mathbf{H}_m(t) + \sqrt{1 - \alpha^2}\, \mathbf{E}_m, \quad (4)$$

where $\alpha \in [0, 1]$ is a reliability factor reflecting the quality of the estimate, and $\mathbf{E}_m \sim \mathcal{N}(\mathbf{0}_{N_r \times N_t}, \mathbf{I}_{N_r})$ is an auxiliary random matrix capturing the estimation uncertainty.

C. Robust Tracking Control Problem Formulation

Let the target state $\mathbf{r}(t) \in \mathbb{R}^{9M \times 1}$ evolve as

$$\mathbf{r}(t + 1) = \mathbf{G}\mathbf{r}(t), \quad (5)$$

with initial condition $\mathbf{r}(0) = \mathbf{r}_0$, where $\mathbf{G} \in \mathbb{R}^{9M \times 9M}$ denotes the target transition matrix. The objective of the leader UAV is to design a control signal $\mathbf{u}(t)$ such that the swarm state $\mathbf{x}(t)$ tracks the target trajectory $\mathbf{r}(t)$, as formally stated in Problem 1.

Problem 1 (Robust Tracking Control Problem): Design a control sequence $\pi = \{\mathbf{u}(0), \mathbf{u}(1), \dots\}$ such that the swarm state $\mathbf{x}(t)$, evolving according to (2), achieves mean-square tracking stability, i.e., $\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\|\mathbf{x}(t) - \mathbf{r}(t)\|^2] < \infty$ for all system uncertainties $\Delta(t)$ satisfying Assumption 1 and channel uncertainties modeled in (4).

III. Problem Reformulation and Robust Tracking Control Solution

A. Reformulation of Robust Tracking Control Problem

Recent studies [13] have shown the effectiveness of optimal control theory for robust control. Building on this foundation, we propose a robust tracking strategy that optimizes a control policy within the nominal plant dynamics, obtained by integrating the UAV dynamics (2), target dynamics (5), and wireless communication models (3) and (4), as follows:

$$\bar{\mathbf{x}}(t + 1) = (\bar{\mathbf{A}} + \bar{\Delta}_1(t))\bar{\mathbf{x}}(t) + (\bar{\mathbf{B}}(t) + \bar{\Delta}_2(t))\widehat{\mathbf{u}}(t) + \bar{\mathbf{w}}(t), \quad (6)$$

where the aggregated state is $\bar{\mathbf{x}}(t) = [\mathbf{x}^T(t), \mathbf{r}^T(t)]^T \in \mathbb{R}^{18M \times 1}$, with transition matrix $\bar{\mathbf{A}} = \mathrm{Diag}(\mathbf{A}, \mathbf{G}) \in \mathbb{R}^{18M \times 18M}$ and actuation matrix $\bar{\mathbf{B}}(t) = [(\widehat{\mathbf{H}}^e(t))^T \widehat{\mathbf{B}}^T, \mathbf{0}_{MN_t \times 9M}]^T \in \mathbb{R}^{18M \times MN_t}$.

The aggregated CSI is $\widehat{\mathbf{H}}^e(t) = \mathrm{Diag}(\delta_1(t)\mathbf{H}^e_1(t), \dots, \delta_M(t)\mathbf{H}^e_M(t)) \in \mathbb{R}^{MN_r \times MN_t}$. The transition uncertainty is $\bar{\Delta}_1(t) = \mathrm{Diag}(\Delta(t), \mathbf{0}_{9M}) \in \mathbb{R}^{18M \times 18M}$, while the channel uncertainty is $\bar{\Delta}_2(t) = [\mathrm{Diag}((\alpha - 1)\mathbf{H}^e_1(t) + \sqrt{1 - \alpha^2}\mathbf{E}_1(t), \dots, (\alpha - 1)\mathbf{H}^e_M(t) + \sqrt{1 - \alpha^2}\mathbf{E}_M(t))^T \widehat{\mathbf{B}}^T, \mathbf{0}_{MN_t \times 9M}]^T \in \mathbb{R}^{18M \times MN_t}$. The aggregated noise is $\bar{\mathbf{w}}(t) = [\widehat{\mathbf{w}}^T(t) + [\mathbf{v}^T_1(t), \dots, \mathbf{v}^T_M(t)]\widehat{\mathbf{B}}^T, \mathbf{0}_{1 \times 9M}]^T \in \mathbb{R}^{18M \times 1}$.

We apply a pseudoinverse decomposition to split $\bar{\Delta}_1(t)$ into a matched component $\bar{\mathbf{B}}(t)\bar{\mathbf{B}}(t)^\dagger \bar{\Delta}_1(t)$ and a mismatched component $(\mathbf{I}_{9M} - \bar{\mathbf{B}}(t)\bar{\mathbf{B}}(t)^\dagger)\bar{\Delta}_1(t)$.

$$\bar{\Delta}_1(t) = \bar{\mathbf{B}}(t)\bar{\mathbf{B}}(t)^\dagger \bar{\Delta}_1(t) + (\mathbf{I}_{9M} - \bar{\mathbf{B}}(t)\bar{\mathbf{B}}^\dagger(t))\bar{\Delta}_1(t), \quad (7)$$

where $(\cdot)^\dagger$ denotes the Moore–Penrose pseudoinverse.

By leveraging the decomposition in (7), we introduce the following auxiliary system:

$$\bar{\mathbf{x}}(t + 1) = \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \bar{\mathbf{B}}_1(t)\bar{\mathbf{u}}_1(t) + \bar{\mathbf{B}}_2(t)\bar{\mathbf{u}}_2(t) + \bar{\mathbf{w}}(t), \quad (8)$$

where $\bar{\mathbf{B}}_1(t) = \bar{\mathbf{B}}(t)$ and $\bar{\mathbf{B}}_2(t) = (\mathbf{I}_{9M} - \bar{\mathbf{B}}(t)\bar{\mathbf{B}}^\dagger(t))$. The virtual control inputs $\bar{\mathbf{u}}_i(t) \in \mathbb{R}^{MN_t \times 1}$, $i \in \{1, 2\}$, serve distinct roles: $\bar{\mathbf{B}}_1(t)\bar{\mathbf{u}}_1(t)$ counteracts the matched uncertainties $\bar{\mathbf{B}}(t)\bar{\mathbf{B}}(t)^\dagger \bar{\Delta}_1(t)$, while $\bar{\mathbf{B}}_2(t)\bar{\mathbf{u}}_2(t)$ mitigates the mismatched uncertainties $(\mathbf{I}_{9M} - \bar{\mathbf{B}}(t)\bar{\mathbf{B}}^\dagger(t))\bar{\Delta}_1(t)$.

With the auxiliary system (8), the robust tracking control of the UAV swarm under the original dynamics (1) is reformulated as a virtual ergodic optimal control problem:

Problem 2 (Robust UAV Swarm Tracking Control via Ergodic Optimal Control): The robust UAV tracking control policy π^* for Problem 1 can be obtained by extracting $\bar{\pi}^*_1$ from the optimal control policy $\{\bar{\pi}^*_1, \bar{\pi}^*_2\}$ of the following ergodic optimal control formulation:

$$\min_{\{\bar{\pi}_1, \bar{\pi}_2\}} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\xi^t c(\bar{\mathbf{x}}(t), \bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t))]$$

$$\text{s.t.} \quad \text{system dynamics given in (8)}, \quad (9)$$

where the policies are defined as $\bar{\pi}_i = \{\bar{\mathbf{u}}_i(0), \bar{\mathbf{u}}_i(1), ...\}$. The per-stage cost function is expressed as:

$$\begin{aligned} c(\bar{\mathbf{x}}(t), \bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t)) =& \bar{\mathbf{x}}^T(t)\bar{\mathbf{Q}}\bar{\mathbf{x}}(t) + \bar{\mathbf{x}}^T(t)(\zeta^2\bar{\mathbf{F}} + \beta^2\bar{\mathbf{I}})\bar{\mathbf{x}}(t) \\ &+ \bar{\mathbf{u}}^T_1(t)\mathbf{R}_1\bar{\mathbf{u}}_1(t) + \bar{\mathbf{u}}^T_2(t)\mathbf{R}_2\bar{\mathbf{u}}_2(t), \end{aligned} \quad (10)$$

where $\zeta, \beta \in \mathbb{R}$ are design parameters, and the weighting matrices are given by $\bar{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & -\mathbf{Q} \\ -\mathbf{Q} & \mathbf{Q} \end{bmatrix} \in \mathbb{S}^{18M}_{++}$, $\bar{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & -\mathbf{F} \\ -\mathbf{F} & \mathbf{F} \end{bmatrix} \in \mathbb{S}^{18M}_{++}$, $\bar{\mathbf{I}} = \begin{bmatrix} \mathbf{I}_{9M} & -\mathbf{I}_{9M} \\ -\mathbf{I}_{9M} & \mathbf{I}_{9M} \end{bmatrix} \in \mathbb{S}^{18M}_{++}$, with $\mathbf{Q} \in \mathbb{S}^{9M}_{+}$, $\mathbf{F} \in \mathbb{S}^{9M}_{+}$, and $\mathbf{R}_i \in \mathbb{S}^{MN_t}_{+}$. $\xi \in (0, 1)$ is a factor to penalize the open-pole induced by the target dynamics $\mathbf{G}$.

Unlike conventional optimal control formulations, the cost function $c(\bar{\mathbf{x}}(t), \bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t))$ in Problem 2 contains additional penalties $\bar{\mathbf{x}}^T(t)(\zeta^2\bar{\mathbf{F}} + \beta^2\bar{\mathbf{I}}_S)\bar{\mathbf{x}}(t)$ and $\bar{\mathbf{u}}^T_2(t)\mathbf{R}_2\bar{\mathbf{u}}_2(t)$, explicitly introduced to compensate for the time-varying uncertainties $\bar{\Delta}_i(t)$ in the system dynamics (6).

B. Robust Tracking Control Solution

Solving Problem 2 entails the Bellman optimality equation [10], formalized below.

$$\begin{aligned} \rho + V(\mathbf{S}(t)) = \min_{\bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t)} &[c(\bar{\mathbf{x}}(t), \bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t)) \\ &+ \xi \mathbb{E}[V(\mathbf{S}(t + 1)) \mid \mathbf{S}(t), \bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t)]], \end{aligned} \quad (11)$$

where $V(\mathbf{S}(t)) \in \mathbb{R}$ is the value function, and $\mathbf{S}(t) = \{\bar{\mathbf{x}}(t), \delta_1(t)\mathbf{H}_1^e(t), \ldots, \delta_M(t)\mathbf{H}_M^e(t)\}$ denotes the aggregated system state, incorporating the plant state $\mathbf{x}(t)$, target state $\mathbf{r}(t)$, and CSI. The optimal control sequence $\{\bar{\mathbf{u}}_1^*(t), \bar{\mathbf{u}}_2^*(t)\}$ is obtained by minimizing the right-hand side (R.H.S.) of (11), which yields the optimal solution to Problem 2. The parameter $\rho > 0$ is a positive constant.

Conventionally, solving Problem 2 requires iterative algorithms such as value iteration or Q-learning [18] to compute the Bellman equation (11). However, these methods suffer from the *curse of dimensionality* due to the continuous state space $\mathbf{S}(t)$. To address this, we exploit the statistical independence between the plant state $\mathbf{x}(t)$, target state $\mathbf{r}(t)$, and CSI $\{\delta_i(t)\mathbf{H}_i^e(t)\}$ to derive a structured reduced-order Bellman equation, formalized in the following theorem.

Theorem *1 (Equivalent Reduced-Order Optimality Equation):* If an optimal solution to Problem 2 exists, it can be equivalently obtained by solving the reduced-order optimality equation:

$$\widehat{\rho} + \widehat{V}(\bar{\mathbf{x}}(t)) = \mathbb{E}_{\widehat{\mathbf{H}}^e(t)}\Big[\min_{\bar{\mathbf{u}}_1(t),\bar{\mathbf{u}}_2(t)}\big[c(\bar{\mathbf{x}}(t), \bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t)) + \xi\mathbb{E}\big[\widehat{V}(\bar{\mathbf{x}}(t+1)) \mid \bar{\mathbf{x}}(t), \widehat{\mathbf{H}}^e(t), \bar{\mathbf{u}}_1(t), \bar{\mathbf{u}}_2(t)\big]\big]\Big], \quad (12)$$

where $\widehat{V}(\bar{\mathbf{x}}(t)) = \bar{\mathbf{x}}^T(t)\mathbf{P}\bar{\mathbf{x}}(t)$ is the structured reduced-order value function with kernel $\mathbf{P} \in \mathbb{S}_{++}^{18M}$. The scalar bias $\widehat{\rho} = \rho = \mathrm{Tr}(\xi\mathbf{P}_{1:9M}\mathbf{W} + \xi\widehat{\mathbf{B}}^T\mathbf{P}_{1:9M}\widehat{\mathbf{B}})$. For $i \in \{1, 2\}$, the optimal solution to Problem 2 and (12) is $\bar{\mathbf{u}}_i^*(t) = \mathbf{K}_i(\mathbf{P}, t)\bar{\mathbf{x}}(t)$, where $\mathbf{K}_i(\mathbf{P}, t) \in \mathbb{R}^{MN_t \times 18M}$ is the corresponding feedback gain, given by

$$\begin{aligned}\mathbf{K}_1(\mathbf{P}, t) = &-(\mathbf{R}_1 + \xi\bar{\mathbf{B}}_1^T(t)\mathbf{P}\bar{\mathbf{B}}_1(t) - \xi^2\bar{\mathbf{B}}_1^T(t)\mathbf{P}\bar{\mathbf{B}}_2(t)\\ &\times \mathbf{N}_2^{-1}(t)\bar{\mathbf{B}}_2^T(t)\mathbf{P}\bar{\mathbf{B}}_1(t))^{-1}(\xi\bar{\mathbf{B}}_1^T(t)\mathbf{P}\bar{\mathbf{A}} - \xi^2\bar{\mathbf{B}}_1^T(t)\mathbf{P}\bar{\mathbf{B}}_2(t)\\ &\times \mathbf{N}_2^{-1}(t)\bar{\mathbf{B}}_2^T(t)\mathbf{P}\bar{\mathbf{A}}), \quad (13)\\ \mathbf{K}_2(\mathbf{P}, t) = &-(\mathbf{R}_2 + \xi\bar{\mathbf{B}}_2^T(t)\mathbf{P}\bar{\mathbf{B}}_2(t) - \xi^2\bar{\mathbf{B}}_2^T(t)\mathbf{P}\bar{\mathbf{B}}_1(t)\\ &\times \mathbf{N}_1^{-1}(t)\bar{\mathbf{B}}_1^T(t)\mathbf{P}\bar{\mathbf{B}}_2(t))^{-1}(\xi\bar{\mathbf{B}}_2^T(t)\mathbf{P}\bar{\mathbf{A}} - \xi^2\bar{\mathbf{B}}_2^T(t)\mathbf{P}\bar{\mathbf{B}}_1(t)\\ &\times \mathbf{N}_1^{-1}(t)\mathbf{P}\bar{\mathbf{B}}_1^T(t)\bar{\mathbf{A}}) \quad (14)\end{aligned}$$

and $\mathbf{N}_i(t) = \mathbf{R}_i + \xi\bar{\mathbf{B}}_i^T(t)\mathbf{P}\bar{\mathbf{B}}_i(t)$.

Proof: The proof follows a similar approach as in [19], based on backward induction for the finite-horizon total cost problem associated with Problem 1, and then taking the limit to obtain the structured form. Detailed steps are omitted due to space constraints. ∎

Unlike the Bellman equation (11), which requires estimating the full value function $V(\mathbf{S}(t))$ over an uncountable state space, the reduced-order equation (12) involves only the structured function $\widehat{V}(\bar{\mathbf{x}}(t))$, fully characterized by a single matrix $\mathbf{P}$. This condensation of learning complexity into one parameter greatly alleviates the *curse of dimensionality* and improves computational efficiency.

C. Robust Stability Analysis

Leveraging Lyapunov stability analysis [11], we establish sufficient conditions for the robust tracking stability of the UAV swarm using the optimal solution from Theorem 2, as stated below.

Theorem *2 (Sufficient Conditions for Robust Tracking Stability of the UAV Swarm):* Denote $\mathbf{K}_i^*(t) = \mathbf{K}_i(\mathbf{P}, t)$. For $i \in \{1, 2\}$, if the optimal solution $\bar{\mathbf{u}}_i^*(t) = \mathbf{K}_i^*(t)\bar{\mathbf{x}}(t)$ in Problem 2 exists, then $\bar{\mathbf{u}}_1^*(t) = \mathbf{K}_1^*(t)\bar{\mathbf{x}}(t)$ is also a solution to Problem 1, provided that the following conditions hold:

- **Condition on μ:**
$$\mu^{-1}\mathbf{I}_{9M} - \mathbf{P} > \mathbf{0}_{9M}. \quad (15)$$
- **Condition on ζ:**
$$\zeta > \sqrt{2}. \quad (16)$$
- **Condition on β:**
$$\begin{aligned}&\mathbb{E}\Big[3(\bar{\mathbf{A}} + \bar{\mathbf{B}}_1(t)\mathbf{K}_1^*(t))^T(\mathbf{P} - \mu\mathbf{I}_{9M})^{-1}(\bar{\mathbf{A}} + \bar{\mathbf{B}}_1(t)\mathbf{K}_1^*(t)) + (2-\alpha)(\mathbf{K}_1^*(t))^T\mathbf{K}_1^*(t)\Big]\\ &\le \mathbb{E}\Big[(\bar{\mathbf{A}} + \bar{\mathbf{B}}_1(t)\mathbf{K}_1^*(t) + \bar{\mathbf{B}}_2(t)\mathbf{K}_2^*(t))^T\mathbf{P}(\bar{\mathbf{A}} + \bar{\mathbf{B}}_1(t)\mathbf{K}_1^*(t) + \bar{\mathbf{B}}_2(t)\mathbf{K}_2^*(t)) + (\mathbf{K}_1^*(t))^T\mathbf{R}_1\mathbf{K}_1^*(t)\\ &\quad + (\mathbf{K}_2^*(t))^T\mathbf{R}_2\mathbf{K}_2^*(t)\Big] + \bar{\mathbf{Q}} + \beta^2\bar{\mathbf{I}}. \quad (17)\end{aligned}$$

Proof: The proof proceeds by constructing a Lyapunov function $L(\bar{\mathbf{x}}(t)) = \bar{\mathbf{x}}^T(t)\mathbf{P}\bar{\mathbf{x}}(t)$ and establishing negative drift under the stated conditions. Detailed derivations are omitted due to space constraints. ∎

According to Theorem 2, swarm tracking stability is ensured by appropriately choosing the parameters $\{\mu, \beta, \zeta\}$: a small μ reduces the effect of uncertainties, while sufficiently large β and ζ enhance the robustness margin.

IV. Online Robust Tracking Controller Design

Note that Theorem 2 gives a fixed-point equation w.r.t. the structured kernel $\mathbf{P}$:

$$\bar{\mathbf{x}}^T(t)\mathbf{P}\bar{\mathbf{x}}(t) = \bar{\mathbf{x}}^T(t)\big(\mathbb{E}[g(\mathbf{P}, \widehat{\mathbf{H}}^e(t))]\big), \quad (18)$$

where the function $g(\mathbf{P}, \widehat{\mathbf{H}}^e(t))$ is defined as:

$$g(\mathbf{P}, \widehat{\mathbf{H}}^e(t)) = \bar{\mathbf{Q}} + \zeta^2\bar{\mathbf{F}} + \beta^2\bar{\mathbf{I}} - \xi^2 \times \begin{bmatrix}\bar{\mathbf{B}}_1^T(t) & \mathbf{P}\bar{\mathbf{A}}\\ \bar{\mathbf{B}}_2^T(t) & \mathbf{P}\bar{\mathbf{A}}\end{bmatrix}^T \begin{bmatrix}\mathcal{M}_{11}(t) & \mathcal{M}_{12}(t)\\ \mathcal{M}_{21}(t) & \mathcal{M}_{22}(t)\end{bmatrix} \begin{bmatrix}\bar{\mathbf{B}}_1^T(t) & \mathbf{P}\bar{\mathbf{A}}\\ \bar{\mathbf{B}}_2^T(t) & \mathbf{P}\bar{\mathbf{A}}\end{bmatrix}. \quad (19)$$

The matrices $\mathcal{M}_{ij}(t)$ are given by $\mathcal{M}_{11}(t) = \mathbf{R}_1 + \xi\bar{\mathbf{B}}_1^T(t)\mathbf{P}\bar{\mathbf{B}}_1(t)$, $\mathcal{M}_{12}(t) = \xi\bar{\mathbf{B}}_1^T(t)\mathbf{P}\bar{\mathbf{B}}_2(t)$, $\mathcal{M}_{21}(t) = \mathcal{M}_{12}^T(t)$, and $\mathcal{M}_{22}(t) = \mathbf{R}_2 + \xi\bar{\mathbf{B}}_2^T(t)\mathbf{P}\bar{\mathbf{B}}_2(t)$.

Since (18) defines a fixed-point equation in the unknown $\mathbf{P}$, SA theory can be used to iteratively learn $\mathbf{P}$. The learned kernel then enables efficient evaluation of the reduced-order value function $\widehat{V}(\bar{\mathbf{x}}(t))$ and the optimal control $\bar{\mathbf{u}}_i^*(t)$ from Theorem 2, which are subsequently applied to achieve robust control of the UAV swarm dynamics (6).

To standardize the formulation, we express (18) as $f(\mathbf{P}) = \mathbf{0}_{18M}$, where:

$$f(\mathbf{P}) = \mathbb{E}[g(\mathbf{P}, \widehat{\mathbf{H}}^e(t))] - \mathbf{P}. \quad (20)$$

Solving for $\mathbf{P}$ in the robust UAV swarm control problem requires computing the root of $f(\mathbf{P})$. We achieve this via the SA algorithm outlined in Algorithm 1. Specifically, at the $(t+1)$-th timeslot, the learned structured kernel $\mathbf{P}(t) \in \mathbb{S}_{++}^{18M}$ is updated using:

$$\mathbf{P}(t+1) = \mathbf{P}(t) + \nu(t)\big(g(\mathbf{P}(t), \widehat{\mathbf{H}}^e(t)) - \mathbf{P}(t)\big), \quad (21)$$

where $\{\nu(t)\}_{t\in\mathbb{Z}}$ represents the step size sequence satisfying $\sum_{t=0}^{\infty}\nu(t) = \infty$ and $\sum_{t=0}^{\infty}\nu^2(t) < \infty$. The function $g(\mathbf{P}(t), \widehat{\mathbf{H}}^e(t))$ serves as an unbiased estimator of $\mathbb{E}[g(\mathbf{P}(t), \widehat{\mathbf{H}}^e(t))]$ in the fixed-point equation (18).

We summarize the convergence results of Algorithm 1 in the following theorem.

Theorem *3 (Almost Sure Convergence of the Proposed Online Robust Tracking Control Algorithm for the UAV Swarm):* If the solution to Problem 2 exists, then: the tracking control solution $\mathbf{u}(t)$ in Step 3 of Algorithm 1 converges to the optimal control solution $\bar{\mathbf{u}}_1^*(t)$ of Problem 2, as stated in Theorem 1, almost surely, i.e., $\Pr(\lim_{t\to\infty}\mathbf{u}(t) = \bar{\mathbf{u}}_1^*(t)) = 1$.

Proof: The proof is established by showing the convergence of $\mathbf{P}(t)$ in Algorithm 1 to $\mathbf{P}$ in Theorem 1, which amounts to demonstrating that $\mathbf{P}(t)$ asymptotically follows the trajectory of a stable ordinary differential equation (ODE). Detailed steps are omitted due to space constraints. ■

Algorithm 1 Online Robust Tracking Control for the UAV Swarm over MIMO Fading Channels

---

Initialization: We initialize the UAV state as $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}_{9M\times 1}, \mathbf{I}_{9M})$ and the target state as $\mathbf{r}_0 \in \mathbb{R}^{9M\times 1}$. The structured kernel is initialized by $\mathbf{P}(0) \in \mathbb{S}_{++}^{18M}$, which yields the initial reduced-order value function $\widehat{V}_0(\bar{\mathbf{x}}) = \bar{\mathbf{x}}^T\mathbf{P}(0)\bar{\mathbf{x}}$ for all $\bar{\mathbf{x}} \in \mathbb{R}^{18M\times 1}$. The initial robust control is given by $\mathbf{u}(0) = \mathbf{K}_1(\mathbf{P}(0), 0)\bar{\mathbf{x}}(0)$, where $\mathbf{K}_1(\mathbf{P}(0), 0)$ is defined in (13). The control signal is structured as $\mathbf{u}(0) = [\mathbf{u}_1^T(0), \dots, \mathbf{u}_M^T(0)]^T \in \mathbb{R}^{18M\times 1}$, with each $\mathbf{u}_m(0) \in \mathbb{R}^{N_t\times 1}$ corresponding to the robust control signal for the m-th follower UAV.

For $t = 1, 2, \dots$**:**

- **Step 1 (Update of the Structured Kernel):**
  Update the structured kernel according to (21):
  $$\mathbf{P}(t) \leftarrow \mathbf{P}(t-1) + \nu(t-1)\big(g(\mathbf{P}(t-1), \widehat{\mathbf{H}}^e(t-1)) - \mathbf{P}(t-1)\big).$$
- **Step 2 (Update of the Reduced-Order Value Function):**
  Update the value function using the current structured kernel:
  $$\widehat{V}_t(\bar{\mathbf{x}}) \leftarrow \bar{\mathbf{x}}^T\mathbf{P}(t)\bar{\mathbf{x}}, \quad \forall \bar{\mathbf{x}} \in \mathbb{R}^{18M\times 1}.$$
- **Step 3 (Update of the Robust Control Solution):**
  Compute the tracking control solution:
  $$\mathbf{u}(t) \leftarrow \mathbf{K}_1(\mathbf{P}(t), t)\bar{\mathbf{x}}(t),$$
  where $\mathbf{K}_1(\mathbf{P}(t), t)$ is given by (13).

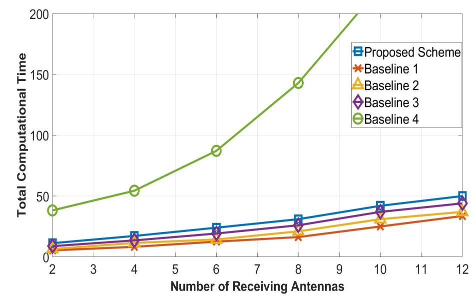
End For

V. Numerical Results

A. Experiment Setup

To evaluate the proposed online robust tracking control scheme for UAV swarms, we compare its performance against the following baseline methods over random MIMO fading channels with system and channel uncertainties:

- **Baseline 1** *(PID-based UAV Swarm Control):* The control input $\mathbf{u}(t)$ is generated via PID control with the gain matrices pre-tuned offline using pole placement based on the nominal system parameters $\{\mathbf{A}, \widehat{\mathbf{B}}\}$.
- **Baseline 2** *(Robust Control via Optimal Control Under Matching Condition and Static Channels):* $\mathbf{u}(t) = \mathbf{K}\bar{\mathbf{x}}(t)$, where the constant gain matrix $\mathbf{K} \in \mathbb{R}^{MN_r\times 18M}$ is computed by solving an optimal control problem formulated based on the nominal system parameters.
- **Baseline 3** *(Robust Control via Optimal Control Under Mismatching Condition and Static Channels):* $\mathbf{u}(t) = \mathbf{K}\bar{\mathbf{x}}(t)$, where the constant gain matrix $\mathbf{K} \in \mathbb{R}^{MN_t\times 18M}$ is obtained by solving an optimal control problem based on the nominal system parameters.
- **Baseline 4** *(Robust Control via Brute-Force Value Iteration over Bellman Equation):* $\mathbf{u}(t) = \mathbf{K}(t)\bar{\mathbf{x}}(t)$, where the time-varying gain $\mathbf{K}(t) \in \mathbb{R}^{MN_t\times 18M}$ is computed by solving Problem 2 through brute-force value iteration over the Bellman equation (11).



ing

Fig. 2: Total CPU computational time versus number of receiving antennas.

We consider a UAV swarm with $M = 6$ followers. Each UAV has an internal transition matrix $\mathbf{A}_{mm} \in \mathbb{R}^{9\times 9}$ and actuation matrix $\mathbf{B}_m \in \mathbb{R}^{9\times 6}$, randomly generated from $\mathcal{N}(0, 1)$. The coupling is defined as $\mathbf{A}_{mn} = 0.1\mathbf{A}_{mm}$ if $n = (m \bmod M) + 1$, and $\mathbf{A}_{mn} = \mathbf{0}_9$ otherwise. Plant noise follows $\mathbf{w}_m(t) \sim \mathcal{N}(\mathbf{0}_9, 10^{-5}\mathbf{I}_9)$. Dynamics uncertainties are modeled as $\Delta(t) = \sin(t+1)\Phi(t)$, where $\Phi(t)$ has i.i.d. entries $\sim U[0, 1]$. System parameters are $N_t = N_r = 6$, $\xi = 0.8$, and activation probability $p = 0.4$. Weighting matrices are $\mathbf{Q} = \mathbf{I}_{9M}$, $\mathbf{F} = 15\mathbf{I}_{9M}$, $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{I}_{6M}$, with $\beta = 7$ and $\zeta = 4$. The target transition matrix is $\mathbf{G} = 1.02\mathbf{I}_{9M}$ if $t \bmod 8 < 4$, and $0.95\mathbf{I}_{9M}$ otherwise. Initial states are $\mathbf{r}_0 = 5\mathbf{I}_{9M\times 1}$ and $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}_{9M\times 1}, \mathbf{I}_{9M})$. $M = 6$.

B. Performance Comparison and Analysis

1) CPU Computational Time Analysis: Fig. 2 shows the CPU computation time over 10^4 iterations for different numbers of receiving antennas N_r. The proposed scheme requires far less time than Baseline 4, which relies on brute-force value iteration and suffers from the curse of dimensionality. In contrast, our method learns only the structured kernel of the reduced-order value function, greatly improving efficiency. The computation time is slightly higher than Baselines 1–3, as those methods adopt simpler strategies that ignore system or channel uncertainties.

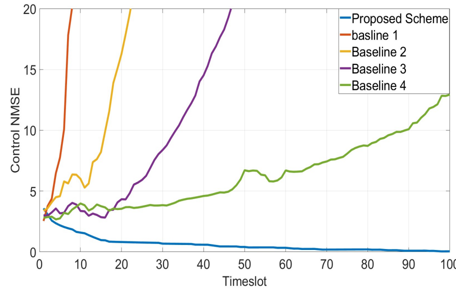


Fig. 3: NMSE between applied and optimal solution versus timeslot.

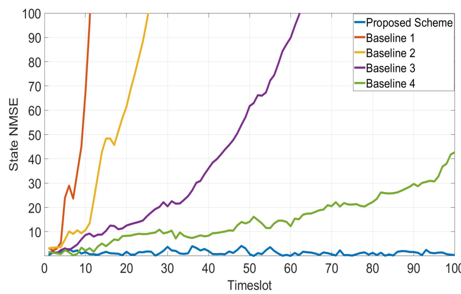


Fig. 4: NMSE between real-Time UAV state and target state versus timeslot.

2) Convergence Analysis: Fig. 3 shows the normalized mean square error (NMSE) between the applied control input and the optimal solution to Problem 2. The proposed scheme steadily converges as its structured kernel approximates the true reduced-order kernel, whereas the baselines diverge over time. Baseline 4 suffers from the curse of dimensionality, while Baselines 1–3 ignore random channel variations and time-varying uncertainties, leading to suboptimal performance.

3) Tracking Stability Analysis: Fig. 4 shows the NMSE between the UAV state and the target state over time. The proposed scheme achieves asymptotic tracking, while the baselines display growing deviations. This highlights the effectiveness of our method in ensuring robust swarm stabilization under system and channel uncertainties, whereas the baselines fail to account for such variations and lead to tracking instability.

VI. Conclusions

In this work, we propose an online robust tracking control framework for UAV swarms operating over MIMO fading channels subject to both system dynamics and channel uncertainties. The control task is reformulated as a virtual ergodic optimal control problem, from which we derive reduced-order optimality equations to mitigate the curse of dimensionality and facilitate efficient real-time learning of robust control policies. Theoretical analysis establishes rigorous guarantees, demonstrating both the robust stability of the closed-loop system and the almost sure convergence of the SA-based learning algorithm. Simulation results further verify the effectiveness of the proposed approach, showing notable improvements over baseline methods in terms of stability, convergence speed, and computational efficiency, thereby offering a practical and reliable solution for UAV swarm control under uncertain wireless environments. Beyond its theoretical contributions, this study offers practical insights for enhancing resilient UAV swarm coordination in the emerging low-altitude economy, with promising applications in large-scale Internet of Things deployments and AI-driven sensor networks.

References

[1] I. Lopez-Sanchez and J. Moreno-Valenzuela, "PID control of quadrotor UAVs: A survey," *Annu. Rev. Control*, vol. 56, p. 100900, 2023.

[2] D. Saccani, L. Cecchin, and L. Fagiano, "Multitrajectory model predictive control for safe UAV navigation in an unknown environment," *IEEE Trans. Control Syst. Technol.*, pp. 1–16, 2022.

[3] S. M. Lee, M. Shin, and H. Son, "Robust predictor-based control for multirotor UAV with various time delays," *IEEE Trans. Ind. Electron.*, vol. 70, no. 8, pp. 8151–8162, 2023.

[4] I. C. Lett., "Integrated communication and control for intelligent formation management of UAV swarms: A deep reinforcement learning approach," *IEEE Wireless Commun. Lett.*, 2025.

[5] Z. Cao, S. Wang, X. Ge, C. Peng, and H. Zhao, "A distributed self-triggered control scheme for multi-unmanned aerial vehicle containment control under Markov switching topologies and channel fading," *IEEE Trans. Autom. Sci. Eng.*, 2025.

[6] Y. Liu, H. Zhang, J. Chen, and M. Fu, "Distributed adaptive finite-time compensation control for UAV swarm with uncertain disturbances," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7432–7443, 2022.

[7] J. Zhang, P. Zhang, and J. Yan, "Distributed adaptive finite-time compensation control for UAV swarm with uncertain disturbances," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 2, pp. 829–841, 2020.

[8] C. Feng, H. H. Yang, D. Hu, Z. Zhao, T. Q. S. Quek, and G. Min, "Mobility-aware cluster federated learning in hierarchical wireless networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8441–8458, 2022.

[9] M. Tang, C. Feng, G. Min, and T. Q. S. Quek, "H-infinity tracking for intelligent edge-controlled systems over fading channels in ai-ran," *IEEE Trans. Veh. Techno.*, pp. 1–6, 2025.

[10] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.

[11] Y. Cui, V. K. Lau, R. Wang, H. Huang, and S. Zhang, "A survey on delay-aware resource control for wireless systems—large deviation theory, stochastic Lyapunov drift, and distributed stochastic learning," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1677–1701, 2012.

[12] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM J. Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.

[13] J. Na, J. Zhao, G. Gao, and Z. Li, "Output-feedback robust control of uncertain systems via online data-driven learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 6, pp. 2650–2662, 2020.

[14] Y. Yang, Z. Guo, H. Xiong, D.-W. Ding, Y. Yin, and D. C. Wunsch, "Data-driven robust control of discrete-time uncertain linear systems via off-policy reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3735–3747, 2019.

[15] G. Gallego, J. E. Lund, E. Mueggler, H. Rebecq, T. Delbruck, and D. Scaramuzza, "Event-based, 6-DOF camera tracking from photometric depth maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2402–2412, 2017.

[16] Y. Xi, A. Burr, J. Wei, and D. Grace, "A general upper bound to evaluate packet error rate over quasi-static fading channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1373–1377, 2011.

[17] L. Zhang, L. Song, M. Ma, and B. Jiao, "On the minimum differential feedback for time-correlated MIMO Rayleigh block-fading channels," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 411–420, 2012.

[18] R. S. Sutton, A. G. Barto *et al.*, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998.

[19] M. Tang, S. Cai, and V. K. Lau, "Online system identification and control for linear systems with multiagent controllers over wireless interference channels," *IEEE Trans. Autom. Control*, vol. 68, no. 10, pp. 6020–6035, 2022.