

Predictive Lightweight MARL for Resilient Coverage in Sparse-Signaling Aerial Networks

Chuan-Chi Lai, *Member, IEEE*, and Ang-Hsun Tsai, *Member, IEEE*

Abstract—This letter proposes the Predictive Lightweight Multi-Agent Reinforcement Learning (PL-MARL) framework to ensure resilient coverage in bandwidth-constrained UAV swarms. To counter coordination collapse caused by sparse signaling and information aging, we introduce a Kinematic-Aware Inference Engine that proactively reconstructs neighbor trajectories via physical priors. This approach enables an efficient computation-for-communication trade-off, decoupling structural resilience from signaling frequency. Simulations confirm that PL-MARL maintains superior coverage and mission continuity under extreme signaling scarcity and node failure. Our results validate proactive inference as a scalable, low-latency solution for robust aerial coordination, effectively minimizing control overhead to preserve spectrum for payload services while ensuring resilience against interference.

Index Terms—Sparse signaling, Age of Information (AoI), Spatio-temporal inference, Resilient coverage, UAV swarms, Multi-agent reinforcement learning.

I. INTRODUCTION

IN the 6G era, *Unmanned Aerial Vehicle* (UAV) networks are pivotal for ubiquitous connectivity. Seamless coordination requires continuous state exchange, yet conventional *Multi-Agent Reinforcement Learning* (MARL) schemes rely on high-frequency synchronization. In bandwidth-constrained environments, this overhead exhausts spectral resources and aggravates co-channel interference, causing severe information aging [1], [2]. While *Digital Twin* (DT) frameworks [3] provide a stable training backplane, decentralized execution remains hindered by the difficulty of maintaining state awareness under sparse signaling.

Recent research explores swarm intelligence through graph-enhanced planning [4], DT-enhanced resource allocation [5], and secure communications [6]. While *Quantum-Assisted*

Frameworks [7] offer algorithmic potential, their specialized hardware dependencies are impractical for resource-constrained UAVs. More critically, state-of-the-art frameworks utilizing spatio-temporal attention for partial observability fundamentally rely on frequent signaling regimes; they lack the onboard generative capacity to synthesize unobserved states during prolonged silence, rendering them untenable in extreme *Flying Ad-hoc Networks* (FANETs). Specifically, our prior work, *Topology-Aware Graph MAPPO* (TAG-MAPPO) [8], achieves resilient reconfiguration via centralized graph aggregation but remains bound to continuous state synchronization for topological observability. Consequently, in signaling-sparse environments, these models face coordination collapse; lacking the capacity to synthesize unobserved states, they leave a critical gap between delayed feedback and the real-time requirements of agile aerial control. By contrast, this study shifts the paradigm from reactive synchronization to proactive, physics-guided state reconstruction, enabling robust coordination where prior methodologies fundamentally fail.

To bridge the aforementioned gap, this letter proposes the *Predictive Lightweight Multi-Agent Reinforcement Learning* (PL-MARL) framework. Our core contribution is a Kinematic-Aware Inference Engine that functions as a local generative unit. It proactively reconstructs neighbor trajectories via physical motion priors to decouple decentralized policy execution from communication frequency. Unlike black-box recurrent baselines hindered by computational bottlenecks, this mechanism enables sub-millisecond inference and subverts sequential processing limitations. Furthermore, we introduce a topology-aware graph attention mechanism that balances computational efficiency with mission reliability. By deliberately trading redundant topological cohesion for an expanded spatial footprint, PL-MARL mitigates the coverage collapse observed in traditional reactive models and ensures graceful performance degradation. Finally, empirical results validate that PL-MARL establishes a robust control backbone capable of sustaining mission continuity under the dual pressure of prolonged sparse signaling and abrupt node failure. This confirms its deployment readiness for latency-aware aerial operations, even when physical feedback loops are effectively severed.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a UAV swarm $\mathcal{N} = \{1, \dots, N\}$ performing 3D coverage, where each node i maintains state $\mathbf{s}_i(t) = [\mathbf{p}_i(t), \mathbf{v}_i(t)] \in \mathbb{R}^6$. As illustrated in Fig. 1, we implement a decoupled DT architecture. A ground server facilitates centralized training to mitigate non-stationarity. Once deployed,

This research was supported by the National Science and Technology Council, Taiwan, R.O.C., under Grant Nos. NSTC 114-2221-E-194-062-, NSTC 115-2221-E-194-042-MY2, and NSTC 115-2221-E-035-050-. This work was also partially supported by the Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. In addition, this work was sponsored by Feng Chia University under Grant 25H00812. (*Corresponding author: Chuan-Chi Lai.*)

Chuan-Chi Lai is with the Department of Communications Engineering, National Chung Cheng University, Minxiong Township, Chiayi County 621301, Taiwan, and also with the Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI), National Chung Cheng University, Minxiong Township, Chiayi County 621301, Taiwan (e-mail: chuan-clai@ccu.edu.tw).

Ang-Hsun Tsai is with the Department of Communications Engineering, Feng Chia University, Taichung 407102, Taiwan.

© 2026 IEEE. Personal use of this material is permitted. For any other uses, permission must be obtained from IEEE.

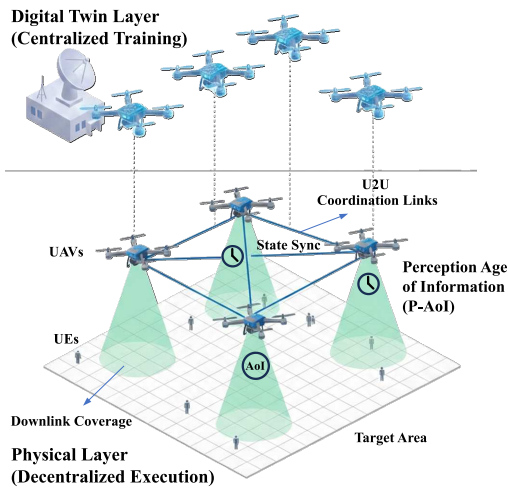


Fig. 1. Architecture of the terminal-edge cooperative DT framework for signaling-efficient aerial networks.

each UAV acts as an autonomous edge node to execute decentralized control. However, during sparse signaling, agents must rely on delayed historical information, leading to severe perception gaps within the communication range R_{comm} .

UAVs communicate over a shared channel where a link requires the *Signal-to-Interference-plus-Noise Ratio* (SINR) to exceed a threshold Γ . To mitigate interference in dense swarms, inter-UAV updates are restricted to a sparse interval $T_{\text{up}} \in \mathbb{Z}^+$. Node i broadcasts its true state $\mathbf{s}_i(t)$ only at $t = kT_{\text{up}}$, where $k \in \mathbb{Z}_{\geq 0}$ is the index of the signaling cycle. Information staleness is quantified by the *Perception Age of Information* (P-AoI):

$$\Delta_{ij}(t) = \begin{cases} 1, & t = kT_{\text{up}} \\ t - kT_{\text{up}}, & t \in (kT_{\text{up}}, (k+1)T_{\text{up}}) \end{cases} \quad (1)$$

A larger T_{up} explicitly induces higher perception uncertainty.

We formulate the coordination as a *Decentralized Partially Observable Markov Decision Process* (Dec-POMDP). At timestep t , node i selects a continuous action $\mathbf{a}_i(t)$ (i.e., acceleration) using its available local perception $\mathcal{O}_i(t) = \{\mathbf{s}_i(t), \{\mathbf{s}_j(kT_{\text{up}})\}_{j \in \mathcal{N}_i(t)}\}$, where $\mathcal{N}_i(t) = \{j \mid \|\mathbf{p}_i(t) - \mathbf{p}_j(t)\| \leq R_{\text{comm}}\}$ is the neighbor set, and $\mathbf{s}_j(kT_{\text{up}})$ represents the stale state received at the last synchronization instant.

Let $C(t) \in [0, 1]$ be the instantaneous spatial coverage rate. We maximize the expected cumulative coverage after a topological perturbation at t_f under policy π , subject to the following constraints for all $i \in \mathcal{N}$:

$$\max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=t_f}^T \gamma^{t-t_f} C(t) \right] \quad (2a)$$

$$\text{s.t. } 1/T_{\text{up}} \leq \Omega, \quad (2b)$$

$$\mathbb{P}(\text{SINR}_{ij}(t) \geq \Gamma) \geq 1 - \epsilon, \quad \forall j \in \mathcal{N}_i(t) \quad (2c)$$

$$\|\mathbf{p}_i(t) - \mathbf{p}_j(t)\| \geq d_{\text{safe}}, \quad \forall j \neq i \quad (2d)$$

$$\|\mathbf{v}_i(t)\| \leq V_{\text{max}}, \quad \|\mathbf{a}_i(t)\| \leq A_{\text{max}} \quad (2e)$$

where $\gamma \in [0, 1)$ is the discount factor, T is the horizon, Ω bounds the signaling frequency, ϵ is the outage threshold,

d_{safe} is the minimum safe distance, and $V_{\text{max}}, A_{\text{max}}$ denote kinematic limits. Sensing and communication ranges are fixed to isolate signaling staleness. Instead of explicitly penalizing (2c), connectivity is implicitly maintained because disconnections naturally degrade $C(t)$. Constraints (2d) and (2e) act as soft penalties.

III. PREDICTIVE LIGHTWEIGHT MULTI-AGENT REINFORCEMENT LEARNING FRAMEWORK

To resolve coordination degradation from stale observations, we propose the PL-MARL framework. Built upon a Digital Twin (DT)-assisted *Centralized Training with Decentralized Execution* (CTDE) paradigm, it explicitly integrates physical kinematic constraints to overcome standard black-box limitations. As illustrated in Fig. 2, the architecture synergistically combines the Kinematic-Aware Inference Engine with proactive kinematic feature fusion, operating as a unified onboard generative engine to enable robust decentralized control with minimal computational overhead.

A. Kinematic-Aware Inference Engine

During periods of sparse signaling, the delayed perception of neighbor states severely degrades multi-agent coordination. To maintain continuous spatial awareness, node i maintains a local reconstructed trajectory buffer $\bar{\mathcal{H}}_{ij}(t-1) = \{\hat{\mathbf{s}}_{ij}(t-\tau)\}_{\tau=1}^L$, which stores the L most recent states inferred by node i regarding neighbor j . Instead of recurrent sequential unrolling, the L temporal states are concatenated into a single flattened vector and processed by a *Multi-Layer Perceptron* (MLP) in one highly efficient forward pass (hyperparameters detailed in Table I). The generative predictor $\phi(\cdot)$ operates autoregressively utilizing this flattened buffer and the accumulated P-AoI $\Delta_{ij}(t)$ to estimate the current neighbor position:

$$\hat{\mathbf{s}}_{ij}(t) = \phi(\bar{\mathcal{H}}_{ij}(t-1), \Delta_{ij}(t); \theta_{\text{inf}}), \quad (3)$$

where θ_{inf} denotes the learnable parameters of the inference engine. Once generated, the newly inferred state $\hat{\mathbf{s}}_{ij}(t)$ is appended to the buffer, replacing the stale component in the decision-making loop. The augmented policy input is thus refined from $\mathcal{O}_i(t)$ to $\hat{\mathcal{O}}_i(t) = \{\mathbf{s}_i(t), \{\hat{\mathbf{s}}_{ij}(t)\}_{j \in \mathcal{N}_i(t)}\}$, ensuring uninterrupted topological reasoning even under extreme signaling scarcity.

To ensure physical plausibility, we formulate trajectory reconstruction as a geometry-constrained estimation problem. Beyond minimizing coordinate reconstruction error, we enforce spatial consistency by constraining the predicted relative displacement $\Delta \hat{\mathbf{p}}_{ij}(t)$ to preserve local topological structure. This yields the objective $\mathcal{L}_{\text{inf}} = \|\mathbf{s}_j^{\text{true}}(t) - \hat{\mathbf{s}}_{ij}(t)\|^2 + \lambda \|\text{PE}(\Delta \mathbf{p}_{ij}^{\text{true}}(t)) - \text{PE}(\Delta \hat{\mathbf{p}}_{ij}(t))\|^2$. The first term ensures trajectory fidelity, while the second term functions as a kinematic-aware regularizer grounding inference in coverage geometry. Here, λ is a weighting coefficient, $\text{PE}(\cdot)$ denotes the positional encoding function (detailed in Section III-B), and $\Delta \mathbf{p}_{ij}^{\text{true}}(t)$ represents the true relative displacement at time t .

Remark: Ground-truth trajectory information ($\mathbf{s}_j^{\text{true}}$) is utilized solely during centralized training within the DT environment as a supervised signal. During decentralized execution, global

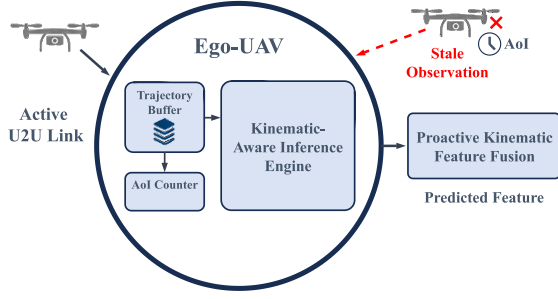


Fig. 2. Architecture of the onboard kinematic-aware inference engine. The framework reconstructs physical states $\hat{\mathbf{s}}_{ij}$ from sparse signaling updates to compensate for perception aging, subsequently performing topology-aware feature fusion for decentralized decision-making.

states are unavailable. UAVs rely exclusively on local observations and pre-trained inference, ensuring compliance with real-world, sparse-signaling operational constraints.

B. Proactive Kinematic Feature Fusion

Upon reconstructing the local belief states $\hat{\mathbf{s}}_{ij}(t)$, the unified engine performs topology-aware aggregation. Standard feature-based attention mechanisms often neglect spatial proximity, which limits their effectiveness in coverage optimization. To address this, we augment neighbor embeddings with explicit positional information:

$$\hat{\mathbf{h}}_{ij}(t) = \text{Encoder}(\hat{\mathbf{s}}_{ij}(t)) + \text{PE}(\Delta \mathbf{p}_{ij}(t)), \quad (4)$$

where $\text{Encoder}(\cdot)$ extracts the latent semantics and $\Delta \mathbf{p}_{ij}(t) = \hat{\mathbf{s}}_{ij}(t) - \mathbf{s}_i(t)$ is the relative displacement. $\text{PE}(\cdot)$ acts as a learnable projection that maps relative physical displacements into the latent space. This positional encoding ensures that the proactive attention coefficients $e_{ij}(t) = \text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i(t) \parallel \mathbf{W}\hat{\mathbf{h}}_{ij}(t)])$ are kinematically informed by inter-node geometry, where $\mathbf{W}$ is a shared weight matrix, $\mathbf{a}$ is the attention vector, $\parallel$ denotes concatenation, and $\mathbf{h}_i(t)$ is the ego-state embedding. To capture the asymmetric importance, these coefficients are normalized via a softmax operation to obtain $\alpha_{ij}(t) = \exp(e_{ij}(t)) / \sum_{k \in \mathcal{N}_i(t)} \exp(e_{ik}(t))$. Consequently, the aggregated message is computed inline as $\mathbf{m}_i(t) = \sigma(\sum_{j \in \mathcal{N}_i(t)} \alpha_{ij}(t) \mathbf{W}\hat{\mathbf{h}}_{ij}(t))$ to fuse both latent semantic features and explicit spatial priors, where σ denotes the activation function (e.g., ELU). This architecture allows the decentralized controller to implicitly prioritize neighbors based on their actual physical impact on network coverage, significantly enhancing the coordination precision in dynamic aerial environments.

C. Training and Lightweight Efficiency

Algorithm 1 summarizes the predict-then-coordinate flow. We train the framework via centralized MAPPO within an edge DT. To satisfy Section II constraints and enhance resilience against node or link failures, we employ an adaptive global reward $R(t)$:

$$R(t) = \omega_1 C(t) + \omega_2(f) P(t) - \omega_3(f) D(t) - \omega_4 \Psi_{\text{safe}}(t), \quad (5)$$

where ω_1 to ω_4 are scaling weights, $C(t)$ is the coverage rate, $P(t)$ is the potential reward measured by the proximity between the network center and service centroid, and $D(t)$

Algorithm 1: Kinematic-Aware Decentralized Coordination

Input: Time step t , Signaling interval T_{up} , Buffer $\bar{\mathcal{H}}_{ij}$, Neighbor index $\mathcal{N}_i(t)$, Ego state $\mathbf{s}_i(t)$
Output: Coordination action $\mathbf{a}_i(t)$

```

1 foreach neighbor  $j \in \mathcal{N}_i(t)$  do
2   if  $t \pmod{T_{\text{up}}} == 0$  then
3     Receive true state  $\mathbf{s}_j(t)$  via signaling update;
4     Update buffer  $\bar{\mathcal{H}}_{ij}$  and reset P-AoI  $\Delta_{ij}(t) \leftarrow 1$ ;
5      $\hat{\mathbf{s}}_{ij}(t) \leftarrow \mathbf{s}_j(t)$ ;
6   else
7     Increment P-AoI  $\Delta_{ij}(t) \leftarrow \Delta_{ij}(t) + 1$ ;
8     // Reconstruct physical state via inference engine
9      $\hat{\mathbf{s}}_{ij}(t) \leftarrow \phi(\bar{\mathcal{H}}_{ij}(t-1), \Delta_{ij}(t); \theta_{\text{inf}})$ ;
10  end
11  // Inject positional encoding into feature embedding
12   $\Delta \mathbf{p}_{ij}(t) \leftarrow \hat{\mathbf{s}}_{ij}(t) - \mathbf{s}_i(t)$ ;
13   $\hat{\mathbf{h}}_{ij}(t) \leftarrow \text{Encoder}(\hat{\mathbf{s}}_{ij}(t)) + \text{PE}(\Delta \mathbf{p}_{ij}(t))$ ;
14 end
15 Construct spatio-temporal graph  $\mathcal{G}(t)$  using  $\{\hat{\mathbf{h}}_{ij}(t)\}$ ;
16 Compute attention weights  $\alpha_{ij}(t)$  and message  $\mathbf{m}_i(t)$ ;
17 Generate action  $\mathbf{a}_i(t) \sim \pi(\mathbf{s}_i(t), \mathbf{m}_i(t))$ ;
18 return  $\mathbf{a}_i(t)$ 

```

indicates mobility cost. Crucially, $\Psi_{\text{safe}}(t)$ acts as a repulsive penalty for spatial distance violations below the collision threshold d_{safe} . Instead of imposing an explicit penalty for link degradation, we intentionally omit the SINR constraint from the reward formulation to foster emergent collaborative intelligence. Since kinematic-driven feature synthesis relies strictly on neighbors within the communication range R_{comm} , topological disconnections naturally degrade shared observation quality. This compels the model to implicitly discover the correlation between maintaining graph attention connectivity and maximizing the global coverage reward $C(t)$. Consequently, the swarm autonomously self-organizes the optimal trade-off between spatial dispersion for coverage and topological cohesion for feature synthesis, avoiding rigid inductive biases. We adaptively scale the spatial weights $\omega_2(f)$ and $\omega_3(f)$ based on the topological failure state f . Upon detecting a perturbation, the mechanism increases ω_2 to incentivize inward topology contraction, while decreasing ω_3 to facilitate large-scale network reconfiguration.

The computational overhead is primarily dominated by the generative inference and attention mechanisms. Let $M \leq N$ represent the average neighborhood size within R_{comm} . To ensure lightweight efficiency, we employ a *Single-Head Attention* (SHA) mechanism. Given a trajectory buffer of length L and a hidden dimension D , the local complexity per node is $\mathcal{O}(M(LD + D^2))$, which exhibits linear scaling with respect to M rather than the total swarm size N . Although onboard inference introduces marginal processing overhead, it substantially reduces invocations of the energy-intensive radio frequency front-end. Since radio frequency transmission typically dominates the UAV power budget, this computation-for-communication trade-off proves highly efficient for aerial networks under sparse signaling constraints.

IV. SIMULATION RESULTS AND ANALYSIS

A. Simulation Setup and Metrics

We evaluate a 1 km^2 3D coverage scenario with $N = 4$ UAVs and 240 Gauss-Markov ground users (max speed $v_u = 0.5 \text{ m/s}$). Due to space constraints, we focus on the Crowded Urban environment as a rigorous stress test against severe topological perturbations, though preliminary results confirm consistent PL-MARL gains in Suburban and Rural settings. To prevent overfitting to specific macro-mobility patterns, training employs a mixed-mobility curriculum (Random Waypoint, Gauss-Markov, and Reference Point Group Mobility). For realistic onboard processing, we implement a multi-rate architecture. Environment physics and the inference engine operate at a 1 s resolution for continuous tracking, while the decentralized policy updates actions every $\tau = 3 \text{ s}$ to conserve computation. Episodes span 200 s ($T = 200$ steps). Metrics are evaluated over 100 independent episodic runs, with shaded regions in all continuous plots denoting 95% confidence intervals. Furthermore, performance gains over baselines are strictly validated via the non-parametric Mann-Whitney U test ($p < 0.01$).

To evaluate stringent bandwidth constraints, we vary the signaling interval $T_{\text{up}} \in \{5, 10, \dots, 50\}$ steps, where $T_{\text{up}} = 50$ imposes a severe 50 s silence. Despite large theoretical displacements, swarm connectivity persists as agents proactively bound their mobility to match ground dynamics. A buffer horizon $L = 10$ provides the inference engine with high-order motion priors, enabling smooth state fusion during intermittent updates. Furthermore, to foster topological resilience, training is executed for 1000 episodes; after the 500th episode, each episode incorporates a 0.2 probability of a single random node failure at $t = 100$. Detailed parameters are listed in Table I.

We compare PL-MARL against three benchmarks: 1) **TAG-MAPPO (Ablated Baseline)** [8], which relies solely on stale observations without trajectory inference; 2) **GRU-based MARL**, which substitutes the inference engine with a standard *Gated Recurrent Unit* (GRU) to evaluate the merit of physical priors; and 3) **Oracle Bound**, the theoretical coverage ceiling derived via sequential greedy grid-search under perfect global knowledge.

B. Generalization and Resilience to Signaling Scarcity

We evaluate the Crowded Urban scenario in Fig. 3. Fig. 3(a) shows training convergence under stochastic policy exploration. Here, continuous action exploration introduces high-entropy trajectory jitter, constraining the average coverage to approximately 0.4 . Despite severe dynamics, PL-MARL converges stably toward the theoretical Oracle Bound. While the GRU-based MARL baseline outperforms the ablated TAG-MAPPO by leveraging temporal features, it suffers from predictive degradation without explicit physical priors.

Fig. 3(b) demonstrates operational resilience under deterministic decision-making with extreme signaling scarcity ($T_{\text{up}} = 50$ steps). By removing exploratory variance, the swarm successfully exploits learned kinematic priors to stabilize at a higher operational coverage exceeding 0.8 , isolating true robustness from training noise. Although the GRU baseline achieves a competitive mean coverage, its wide confidence

TABLE I
SIMULATION AND TRAINING PARAMETERS

Parameter	Value
Network Area	$1,000 \text{ m} \times 1,000 \text{ m}$
Number of UAVs (N)	4
Max. UAV Speed (V_{max})	20 m/s
Max. UAV Acceleration (A_{max})	5 m/s^2
Max. Ground User Speed (v_u)	0.5 m/s
Comm. / Sensing Radii (R_{comm} / R_s)	$350 \text{ m} / 100 \text{ m}$
Default Signaling Interval (T_{up})	15 steps
Action Decision Interval (τ)	3 steps
Buffer Horizon (L)	10 steps
Adaptive Reward Weights	
Coverage Weight (ω_1)	50.0 / 50.0
Potential Weight (ω_2)	0.5 / 2.0
Mobility Cost (ω_3)	$2 \times 10^{-5} / 1 \times 10^{-5}$
Repulsive Penalty Weight (ω_4)	1.0 / 1.0
Inference Engine MLP Layers / Hidden Dims.	
Fusion Node Embedding Dimension	32
Learning Rate (Actor / Critic)	$1 \times 10^{-4} / 5 \times 10^{-4}$
Total Steps per Episode (T)	200
Discount Factor / Attention Heads	0.99 / 1
PPO Batch Size / Entropy Schedule	800 / $0.05 \rightarrow 0.02$
PPO Internal Epochs	5
Total Training Episodes	1,000
LR Annealing Episodes / Min. Ratio	800 / 0.1

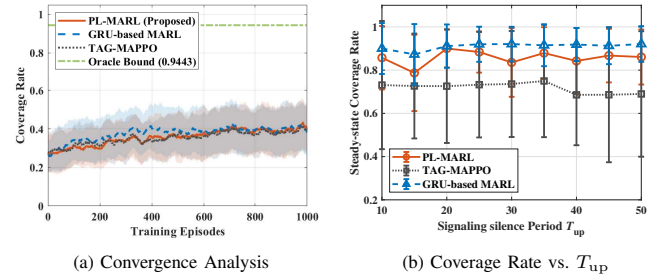


Fig. 3. Performance evaluation in the Crowded Urban scenario. (a) Convergence behavior compared against the Oracle Bound. (b) Steady-state coverage under deterministic decision policies across varying T_{up} .

intervals reveal substantial volatility and predictive drift inherent in pure data-driven recurrent architectures. Conversely, PL-MARL tightly bounds predictive uncertainty via kinematic-aware priors, ensuring highly stable spatial reconfiguration under bandwidth-constrained regimes.

C. Zero-shot Scalability Analysis

To evaluate zero-shot scalability, the inference engine trained exclusively at $N = 4$ is directly deployed up to $N = 20$ without retraining. We proportionally scale task clusters alongside the swarm size, maintaining a constant density of 60 users per cluster to isolate coordination efficacy.

Fig. 4(a) illustrates the computational demand. The GRU-based MARL exhibits polynomial growth in inference latency, reaching 3.0 ms at $N = 20$ due to sequential recurrent unrolling and dense matrix multiplications as the aggregated observation dimension expands. Although a 3.0 ms latency fits within the 3 s control cycle, this non-linear scaling causes severe energy depletion and processor load on resource-constrained micro aerial vehicles. Conversely, PL-MARL maintains a near-constant latency of approximately 0.5 ms by combining a highly parallelizable feedforward architecture with a lightweight SHA mechanism. This decoupling

of computational complexity from swarm size makes our framework uniquely suited for battery-limited operations.

Fig. 4(b) demonstrates the steady-state coverage rate. While the GRU baseline achieves marginally higher coverage, it incurs substantial computational overhead. PL-MARL provides a superior trade-off, maintaining competitive coverage and minimal latency across all scales. Furthermore, it significantly outperforms TAG-MAPPO, confirming that purely reactive models fail to scale in dynamic environments. This validates PL-MARL as an efficient solution that successfully decouples coordination performance from the scaling bottlenecks inherent in recurrent architectures.

D. Resilience Analysis

Fig. 5 examines autonomous resilience under signaling scarcity. In Fig. 5(a), simulating a random node failure at $t = 100$ reveals that PL-MARL maintains graceful coverage degradation. As interpreted directly from the transient curves in Fig. 5(a), the graceful degradation process can be characterized by the initial coverage trough and time-to-stabilization. Specifically, TAG-MAPPO suffers a severe trough (dropping near 0.65) and fails to establish a stable plateau. The GRU baseline delays stabilization by 15 steps. Conversely, PL-MARL actively restricts its trough to approximately 0.81 and rapidly stabilizes to a new steady-state plateau within 10 steps, maximally exploiting the remaining physical capacity. By proactively managing the transient shock and stabilization latency, PL-MARL consistently bridges the gap between static baselines and fragile, high-performing heuristic approaches.

The underlying topological behaviors are exposed in Fig. 5(b), which quantifies integrity via algebraic connectivity (λ_2). Three distinct regimes emerge: TAG-MAPPO exhibits an overly redundant topology ($\lambda_2 \approx 6.5$), forcing excessive cohesion that restricts spatial exploration. Conversely, GRU-based MARL adopts an extremely sparse, almost disjointed policy ($\lambda_2 \approx 2.2$). While this extreme sparsity inadvertently avoids massive structural reconfigurations during node failure, it operates critically near the threshold of complete network partition, making the system vulnerable to environmental variations. PL-MARL maintains an optimal topological balance ($\lambda_2 \approx 3.0$). By deliberately trading unnecessary cohesion for an expanded spatial footprint, it maximizes coverage efficiency without sacrificing network coherence. This confirms that our inference engine successfully decouples structural resilience from signaling frequency, enabling robust, task-driven spatial reconfiguration rather than passive hovering or fragile sparsity.

V. CONCLUSION

This letter proposes the PL-MARL framework for resilient coverage in bandwidth-constrained aerial networks. By integrating a Kinematic-Aware Inference Engine with topology-aware graph attention, we decouple policy execution from communication frequency. Experimental results validate that onboard predictive reasoning mitigates information aging, enabling task-driven spatial reconfiguration over passive, redundant cohesion. Unlike recurrent baselines restricted by polynomial computational scaling and structural fragility, PL-MARL achieves an optimal balance among inference ef-

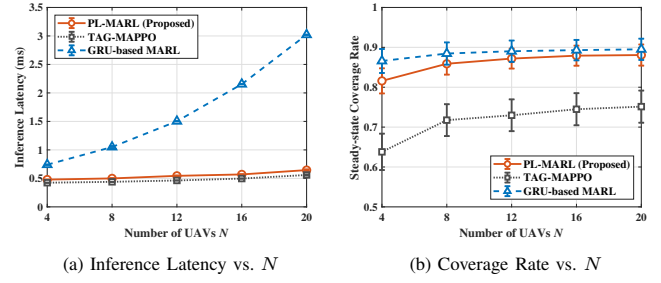


Fig. 4. Zero-shot scalability evaluation of the proposed PL-MARL and baseline models. (a) Inference latency versus the number of UAVs N . (b) Steady-state coverage rate versus the number of UAVs N .

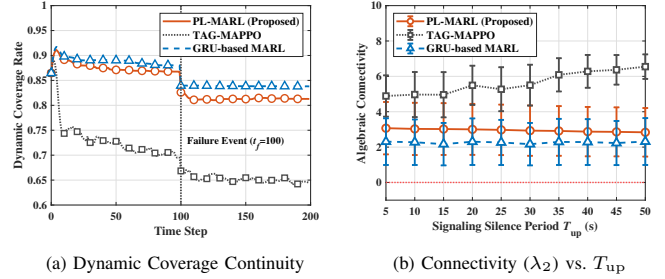


Fig. 5. Resilience evaluation in the Crowded Urban scenario. (a) Dynamic coverage continuity following node failure at $t = 100$. (b) Algebraic connectivity versus signaling silence T_{up} .

iciency, energy conservation, and mission reliability. The framework demonstrates a near-constant sub-millisecond latency profile, graceful performance degradation, and robust scalability, proving that predictive physical priors are essential for resource-constrained operations. Consequently, PL-MARL offers a deployment-ready solution for battery-limited and interference-prone aerial missions. By proactively minimizing the signaling footprint, it translates communication constraints into a strategic advantage, ensuring cyber-physical resilience while preserving critical spectrum for user services.

REFERENCES

- [1] Q.-V. Pham, R. Ruby, F. Fang, D. C. Nguyen, Z. Yang, M. Le, Z. Ding, and W.-J. Hwang, "Aerial computing: A new computing paradigm, applications, and challenges," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8339–8363, Jun. 2022.
- [2] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.
- [3] L. Zhou, S. Leng, Y. Liu, Z. Xiong, and T. Q. S. Quek, "Digital twins for low-altitude UAV networks—cooperation and learning," *IEEE Trans. Mobile Comput.*, vol. 25, no. 4, pp. 4839–4856, Apr. 2026.
- [4] Y. Du, N. Qi, X. Li, M. Xiao, A.-A. A. Boulogeorgos, T. A. Tsiftsis, and Q. Wu, "Distributed multi-UAV trajectory planning for downlink transmission: A GNN-enhanced DRL approach," *IEEE Wireless Commun. Lett.*, vol. 13, no. 12, pp. 3578–3582, Dec. 2024.
- [5] J. Luo, Z. Fei, X. Wang, L. Zhao, B. Li, and Y. Zhou, "GNN-based resource allocation for digital twin-enhanced multi-UAV radar networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 11, pp. 3137–3141, Nov. 2024.
- [6] X. Tang, K. Zhao, C. Shen, Q. Du, Y. Wang, D. Niyato, and Z. Han, "Deep graph reinforcement learning for UAV-enabled multi-user secure communications," *IEEE Trans. Mobile Comput.*, vol. 24, no. 9, pp. 8780–8793, Sep. 2025.
- [7] H. Zhang, M. Emu, and O. A. Dobre, "Quantum takes flight: Two-stage resilient topology optimization for UAV networks," in *IEEE International Conference on Communications (ICC)*, Glasgow, Scotland, UK, 2026.
- [8] C.-C. Lai, "Resilient topology-aware coordination for dynamic 3D UAV networks under node failure," 2026. [Online]. Available: <https://arxiv.org/abs/2602.10029>