Optimizing the Sky: Machine Learning-Based Aerial Network Planning for UAM

Hyeon Woo Jeon
Department of Electrical and Computer Engineering
Inha University
Republic of Korea
hw3652@inha.edu

InTaek Lee
6G Development Team
SK Telecom
Republic of Korea
intaek31.lee@sk.com

Duk Kyung Kim
Department of Electrical and Computer Engineering
Inha University
Republic of Korea
kdk@inha.ac.kr

***Abstract*— Urban Air Mobility (UAM) demands a reliable low-altitude communication fabric, yet planning aerial networks is intrinsically combinatorial because base-station (BS) activation and 3D beam orientation (swing/tilt/HPBW) jointly determine coverage and interference. We formulate this design as a high-dimensional optimization over real corridor data and present a two-stage BO–MARL framework with sequential multi-agent learning (MA-Sequential). First, Bayesian optimization (BO) explores BS on/off states and initializes beam angles to prune the search space. Then, a sequential MARL phase treats each active beam as an agent that updates its policy in turn, improving stability and sample efficiency relative to parallel multi-agent updates. A coverage-centric objective, augmented with distance and angular-overlap penalties, encourages wide coverage while curbing redundant infrastructure. Evaluations on an urban UAM corridor corroborate that the proposed pipeline delivers effective and efficient network plans, yielding robust convergence and improved coverage under diverse path conditions.**

Keywords*— *Urban Air Mobility (UAM), Aerial Network Planning, Multi-Agent Reinforcement Learning (MARL), Bayesian Optimization (BO)

I. INTRODUCTION

Urban Air Mobility (UAM) has emerged as a promising solution to mitigate severe traffic congestion in rapidly expanding metropolitan areas. UAM refers to a short-distance air transportation system that utilizes three-dimensional airspace through low-noise, eco-friendly electric vertical takeoff and landing (eVTOL) aircraft and supporting infrastructure. Recently, the concept has expanded to encompass Regional Air Mobility (RAM) under the broader framework of Advanced Air Mobility (AAM).

In South Korea, UAM has been identified as a strategic future industry, with active efforts underway through the K-UAM policy roadmap. This includes regulatory revisions, the establishment of a technical roadmap, and phased demonstration projects[1]. In particular, the K-UAM Grand Challenge (GC), a public-private joint demonstration project, has been launched to define operational standards for UAM services. The project spans from 2022 to 2024 and consists of two phases. The first phase focuses on testing vehicle stability and traffic management in open areas, while the second phase extends to semi-urban and eventually urban environments with higher population densities[2]. Unlike drones operating below 150 meters or commercial aircraft flying above 18 kilometers, UAM vehicles are expected to operate at altitudes ranging from 300 to 600 meters within designated air corridors, which function as structured aerial routes similar to roads for ground transportation. During the initial deployment phase (starting in 2025), services will operate with a limited number of vertiports and fixed corridors. As the technology matures, the growth phase (starting around 2030) will enable operations between multiple vertiports using remotely piloted aircraft, followed by the maturity phase (from 2035 onward) featuring fully autonomous operations and automated traffic control systems.

To support the deployment of UAM services, both national and international efforts are being made to establish dedicated aerial communication infrastructures. In South Korea, government-led demonstration projects are underway to evaluate technologies for aerial traffic management and wireless communication in low-altitude urban air corridors, involving collaboration between public institutions and private industry to lay the groundwork through staged testing and regulatory development. Concurrently, research and development into eVTOL aircraft and network integration, including terrestrial and satellite-based communication, is progressing.

International studies highlight that aerial networks differ significantly from terrestrial networks, particularly in radio propagation and interference patterns. Because aerial links have a high probability of line-of-sight (LoS), airborne UEs experience stronger inter-cell interference than terrestrial users, which can degrade performance for both aerial and ground users. In addition, most terrestrial base stations are down-tilted for ground coverage, so aerial UEs are often served via antenna sidelobes; this can cause attachment to more distant cells, frequent serving-cell changes, and reduced effectiveness of conventional handover procedures [3]. In addition, UAM services involve heterogeneous traffic demands: command and control (C2) data typically require low latency and moderate throughput, while application services such as real-time video and autonomous flight control demand much higher data rates and stricter delay constraints. These latency and reliability targets imply ultra-reliable and low-latency communication (URLLC)-like capabilities; aerial networks therefore require prioritized low-latency paths, resource reservation, and multi-connectivity to meet diverse QoS demands [4].

To enable such operations, a dedicated aerial communication infrastructure, referred to as the aerial network, must be established. Signal propagation in aerial links is generally favorable; however, this also increases the risk of long-range interference from neighboring base stations. To ensure reliable coverage within the three-dimensional corridor structure, base station antennas must be uptilted, and beam pattern designs must accommodate both vertical and horizontal dimensions. Moreover, the complexity of aerial network design is heightened by diverse flight routes, varying service requirements, and deployment constraints tied to

979-8-3315-7896-1/26/$31.00 ©2026 IEEE

existing ground infrastructure. These factors make traditional ground-based network design strategies difficult to apply, emphasizing the need for a dedicated aerial network design approach. Given that no aerial network has yet been deployed in practice, prior research and validation of specialized design methodologies are essential.

Various optimization and learning-based approaches have been explored to address the challenges of aerial network design in complex UAM corridor environments. In particular, BO improves cellular-network performance by tuning antenna uptilt, half-power beamwidth (HPBW), and transmission power. High-dimensional variants jointly optimize uptilt and HPBW, yielding notable SINR gains while preserving ground-user performance [5]. BO also balances aerial and terrestrial SINR through coordinated antenna and power control [6]. While these studies demonstrate the efficacy of BO in infrastructure parameter tuning, their applicability remains limited in real-world UAM scenarios characterized by dynamically varying flight paths, traffic demands, and channel conditions. When both base station activation states and beam orientation angles (swing/tilt) must be considered, the configuration space grows exponentially, making traditional search-based optimization methods inefficient and slow to converge.

To overcome these limitations, we propose a reinforcement learning (RL)-based framework suited for policy-driven control and sequential decision-making. Specifically, each base station is modeled as an independent agent operating under a multi-agent sequential (MA-Sequential) learning structure, where agents act in sequence without explicit coordination. Although conventional decentralized multi-agent RL approaches offer scalability and independence, they are susceptible to convergence instability and difficulties in learning binary on/off behaviors. To address this, we leverage BO to generate near-optimal initial configurations, which are then used to initialize the MA-Sequential training process. This hybrid approach enhances convergence stability and learning efficiency, as validated through simulations conducted under diverse aerial corridor conditions.

This paper makes the following contributions:

- Problem formulation: We cast UAM aerial-network planning over real urban corridors as a joint discrete–continuous optimization that couples BS activation with 3D beam orientation (swing/uptilt/HPBW) under coverage/QoS constraints.
- Hybrid BO–MARL framework: We introduce a two-stage pipeline where BO prunes the combinatorial space (on/off and initial angles) and seeds a sequential multi-agent DQN that fine-tunes continuous parameters, improving sample efficiency and convergence.
- Learning design: We propose an MA-Sequential update rule that uses the next agent's Q-value as the target. and specify a compact 27-bit state encoding with a five-action control set tailored to practical antenna actuation.
- Objective shaping: We design a coverage-centric objective augmented with distance and angular-overlap penalties plus a conditional efficiency term to reduce redundant BS usage while meeting coverage targets.
- Empirical gains: On a real urban corridor, the method achieves 100% coverage with 7/9 BSs, outperforming a 9-BS baseline (90.87%) and BO-only (95.43%).

The remainder of this paper is organized as follows. Section II formulates the UAM aerial network design problem based on real-world corridor data. Section III presents the proposed aerial network planning framework, including its architecture and optimization procedure. Section IV describes the simulation environment and analyzes the performance of the proposed method under various path conditions and configuration combinations. Finally, Section V concludes the paper and discusses potential directions for future research.

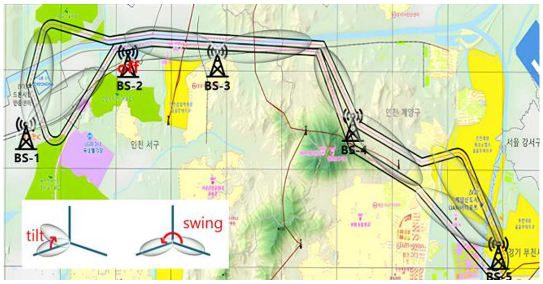


Fig. 1. UAM corridor and Base station deployment

II. System Model

In this section, the aerial network planning problem is modeled based on a real-world UAM flight scenario. Fig. 1 illustrates an example of the flight corridor and the corresponding base station deployment on a map-based layout. The flight path spans several tens of kilometers across an urban area, and the UAM vehicle is assumed to follow this predefined route at a fixed altitude of 300 meters. The path is discretized into points at intervals of up to 100 meters, and the Reference Signal Received Power (RSRP) is measured at each point to evaluate coverage performance. Each base station is deployed at a fixed location and equipped with two independent antenna beams. For each beam, configurable parameters include power state (On/Off), uptilt angle, swing angle, and beamwidth, which enable flexible coverage configurations. This flexibility significantly enlarges the search space per base station, making efficient parameter control and optimization a key challenge addressed in this study.

The wireless environment parameters used in the simulation are based on the antenna element pattern for macro base stations, with beamwidth set to 60° in the vertical domain and 30° in the horizontal domain [7]. The path loss model adopts the UMa-AV LoS scenario specified by 3GPP [8]. The channel frequency is set to 866 MHz, with a total of 25 resource blocks (RBs) allocated within a 5 MHz bandwidth. The RSRP threshold is defined as −83 dBm. The aerial corridor must satisfy predefined traffic demand requirements, which necessitates sufficient coverage from surrounding base stations. Consequently, the problem is formulated as a high-dimensional combinatorial optimization task that jointly considers base station selection, antenna beam configuration, and coverage constraints.

III. Proposed Aerial Network Planning

This section presents a two-stage framework for aerial network optimization that combines Bayesian Optimization (BO) and Multi-Agent Reinforcement Learning (MARL).

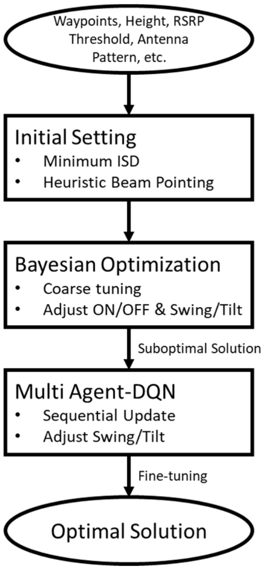


Fig. 2. Overall framework of the proposed BO-MARL-based aerial network optimization

The proposed framework begins by defining an initial configuration based on real-world flight routes, candidate base station (BS) locations, antenna specifications, and system coverage requirements. Using this input, an initial beam configuration is heuristically constructed by considering the minimum Inter-Site Distance (ISD) between base stations and setting antenna pointing directions accordingly. This initial setting serves as the starting point for BO exploration. In the first stage, BO performs a global search over the configuration space, which includes the on/off status of each base station and the initial swing and tilt angles of each antenna beam. BO approximates the objective function using a Gaussian Process (GP)-based surrogate model and selects promising configurations by maximizing an acquisition function. Through this process, BO refines the on/off states to improve resource utilization and reduce redundancy, and the resulting binary decisions are carried forward to the MARL phase.

In the second stage, MARL is applied to optimize the continuous parameters, namely the swing and tilt angles of the activated beams determined in the BO stage. Each antenna beam is modeled as an independent agent, and the learning process follows a sequential update structure (MA-Sequential), where agents are updated one at a time rather than simultaneously. This sequential decision-making approach improves convergence stability and mitigates inefficient exploration, which is often observed in conventional parallel MARL. By combining BO for global exploration of binary variables with MARL for fine-tuning continuous parameters, the proposed framework achieves both search efficiency and robust convergence. The overall architecture and data flow are visually summarized in Fig. 2.

A. Bayesian Optimization

Bayesian Optimization is a sample-efficient global optimization technique designed for black-box functions that are expensive to evaluate and have unknown internal structures. BO employs a surrogate model-typically a GP or Random Forest-to approximate the true objective function. Based on this surrogate model, an acquisition function evaluates the expected improvement at potential input configurations by balancing exploration and exploitation. The configuration that maximizes the acquisition function is selected as the next candidate and evaluated through the actual process. The result of this evaluation is then used to update the surrogate model with new data, refining its prediction accuracy. Fig. 3 illustrates this iterative BO workflow, where the surrogate model, acquisition function, and evaluation process operate in a closed loop to progressively search for the optimal configuration.

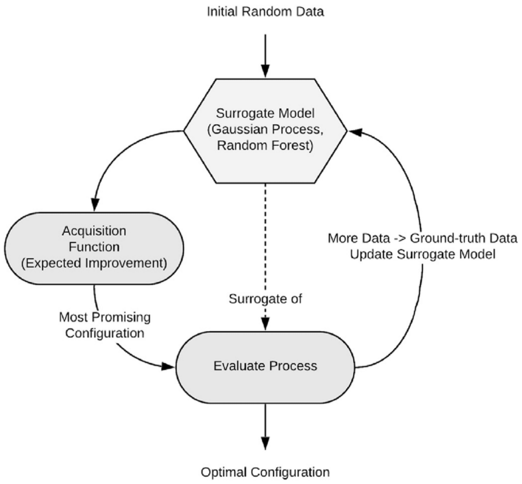

Initial Random Data
Surrogate Model (Gaussian Process, Random Forest)
Acquisition Function (Expected Improvement)
More Data -> Ground-truth Data Update Surrogate Model
Surrogate of
Most Promising Configuration
Evaluate Process
Optimal Configuration

Fig. 3. Conceptual diagram of Bayesian Optimization[9]

In this study, BO is applied to derive a near-optimal initial configuration for aerial network planning. The objective function is defined based on the overall network coverage performance, and the surrogate model is constructed using the RSRP values measured at each observation point. The input variables include the on/off states of base stations, as well as the uptilt and swing angles of their antenna beams. RSRP is calculated for each observation point based on the geometric relationship, including distance, azimuth angle (ϕ), and elevation angle (θ), together with the antenna gain. The antenna gain is determined by the relative difference between the observation point direction (ϕ , θ) and the configured beam orientation defined by the swing and tilt angles, so that these control variables are explicitly incorporated into the RSRP computation. During the BO process, an observation point is considered covered if the maximum RSRP among all base stations exceeds a predefined threshold ($Threshold_{max}$), and total network coverage is evaluated accordingly.

The loss function is designed to not only maximize coverage but also reduce structural inefficiencies in the network configuration. Specifically, it incorporates the total coverage ratio ($cover$), a distance penalty (p_{dis})for overly proximate base stations, and an angular penalty (p_{angle}) to prevent excessive overlap between beams from the same base station. These components are combined into a base loss function, as defined in Equation (1).

$$Loss = \ - w_{cov} \cdot cover + p_{dis} + p_{angle} \tag{1}$$

Coveage weight (w_{cov}) ensures that the coverage ratio, which is bounded between 0 and 1, contributes at a scale comparable to the penalty terms. The distance penalty is defined in Equation (2). Here, d_{ij} denotes the distance between base stations i and j, and $\boldsymbol{\mathcal{D}}$ is the set of unique BS pairs $(i < j)$ whose distance is below the threshold d_{th}. The threshold parameter d_{th} serves as a minimum distance criterion to discourage overly dense BS deployments that could lead to excessive interference.

$$p_{dis} = \sum_{(i,j)\in\boldsymbol{\mathcal{D}}} \frac{d_{th}}{d_{ij}}, \qquad \boldsymbol{\mathcal{D}} = \{(i,j)|i < j, d_{ij} < d_{th}\} \tag{2}$$

The angular penalty is defined in Equation (3), where $\Delta\theta_k$ denotes the angular separation between two beams of the same base station, and $\boldsymbol{\Theta}$ is the set of beam pairs with separation smaller than the angular threshold θ_{th}. The parameter θ_{th} prevents excessive overlap between beams, while the normalization constant p_{max} limits the maximum scale of this penalty to maintain balance with the coverage term. N_{BS} denotes the total number of candidate base stations considered in the optimization.

$$p_{angle} = \sum_{k\in\boldsymbol{\Theta}} \frac{p_{max}}{N_{BS}}, \qquad \boldsymbol{\Theta} = \{k|\Delta\theta_k < \theta_{th}\} \tag{3}$$

Furthermore, to balance coverage with infrastructure efficiency, a conditional loss function is applied as shown in Equation (4). If coverage is below the target threshold $(Target_{cover})$, the optimizer prioritizes coverage improvement when the target is not met. otherwise, the number of inactive base stations (off_{BS}) is maximized. off_{BS} denotes the number of inactive base stations among all candidates, taking integer values from 0 to N_{BS}. In this study, withN_{BS} = 9, the range of off_{BS} is there fore [0,9]. The coefficient λ_{off} controls the contribution of this efficiency term, ensuring that resource saving is encouraged without dominating coverage optimization.

$$Loss_{RSRP} = \begin{cases} Loss & cover < Target_{cover} \\ Loss - \lambda_{off} \cdot off_{BS} & otherwise \end{cases} \tag{4}$$

B. Multi-Agent DQN

The learning architecture proposed in this study defines each antenna beam of a base station as an individual agent, resulting in two agents per base station and a total of 18 agents in the system. To effectively represent each agent's state while minimizing dimensionality, The state of each agent is represented as a 27-bit multi-hot vector encoding the swing and tilt angles of its antenna beam. Swing is encoded with 12 bits, consisting of a 2-bit quadrant index (0-360° divided into four quadrants) and a 10-bit single-hot vector representing the tens digit within each quadrant (0-90°). Tilt is encoded with 15 bits, using a 10-bit single-hot vector for the tens digit (0-90° in steps of 10°) and a 5-bit single-hot vector for the units digit (0–9 in steps of 2°). Concatenating these yields the 27-bit state vector.

The action space is defined based on the adjustment parameters that each base station can apply. While a joint action approach could be considered, since Swing and Tilt can be adjusted independently, this leads to an exponential increase in the number of possible actions, making it impractical. Instead, each of Swing and Tilt includes two actions for increasing or decreasing the angle, along with a Stay action that keeps the current value unchanged. As a result, a total of five discrete actions are defined. Table 1 summarizes the actions associated with each action index.

TABLE I. ACTIONS

INDEX	0	1	2	3	4
ACTION	Swing-10°	Swing+10°	Tilt-2°	Tilt+2°	STAY

Each agent receives the current state from the environment and selects an action accordingly. Based on the chosen action, the next state and resulting coverage are calculated. The reward is computed as the difference between the updated coverage and the previous coverage. That is, if the agent's action leads to improved coverage, it receives a positive reward; if the coverage decreases, a negative reward is assigned.

In Equation (5), $\bar{C}_n$ represents the coverage after agent n has taken an action, $\boldsymbol{S}$ denotes the set of all feasible states under the given constraints, and s refers to the current state. The reward is calculated as the coverage difference between the current and previous states when the resulting state remains valid, as shown below:

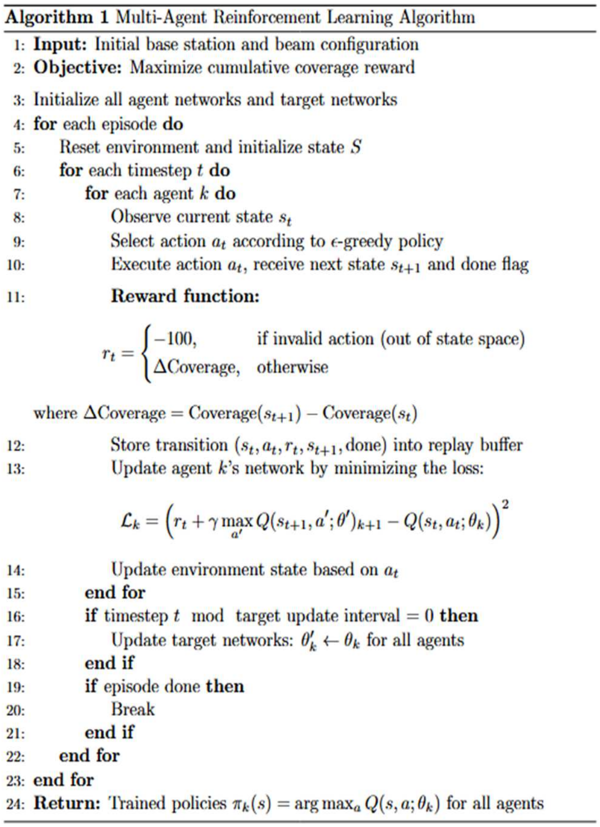
$$R_n = \bar{C}_n - \bar{C}_{n-1}, \qquad s \in \boldsymbol{S} \tag{5}$$

If an agent's action results in a state outside the feasible set $\boldsymbol{S}$, the episode is terminated early and a penalty of -λ_{pen} is applied to discourage invalid state transitions.

Conventional multi-agent reinforcement learning (MARL) frameworks typically adopt a simultaneous execution scheme, where all agents observe their states and perform actions at the same time. However, in scenarios like this study, where the number of agents is large and each agent has a high-dimensional action space, such synchronous updates can lead to unstable convergence and inefficient exploration.

To address this issue, this study adopts a sequential multi-agent update structure, referred to as MA-Sequential. In this framework, each agent receives the current state from the environment, selects an action, and updates its policy based on the resulting reward. Unlike conventional MARL, where all agents act simultaneously, the MA-Sequential structure updates agents one at a time within a single timestep. The process begins with the first agent and proceeds sequentially to the last. Although each agent makes decisions independently, it does so in a state that reflects the updates made by preceding agents. This allows each agent not only to respond to changes in the environment, but also to implicitly account for the influence of other agents' actions during learning.

The loss function in the sequential multi-agent (MA) framework is formulated differently from that of the conventional MA-DQN. In this study, the loss is computed using the Mean Squared Error (MSE) approach. For a given agent k, let the Q-value be represented as $Q(s,a)_k$. The target value is calculated based on the reward $\mathrm{r_t}$ at the current timestep t and the maximum Q-value of the next state s_{t+1}, obtained from the target network. In conventional MA-DQN, the target network is updated independently within each agent. The loss function used in the standard MA-DQN framework is defined as shown in Equation (6).



Algorithm 1 Multi-Agent Reinforcement Learning Algorithm

```
Input: Initial base station and beam configuration
Objective: Maximize cumulative coverage reward
Initialize all agent networks and target networks
for each episode do
    Reset environment and initialize state $S$
    for each timestep $t$ do
        for each agent $k$ do
            Observe current state $s_t$
            Select action $a_t$ according to $\epsilon$-greedy policy
            Execute action $a_t$, receive next state $s_{t+1}$ and done flag
            Reward function:
                $r_t = \begin{cases} -100, & \text{if invalid action (out of state space)} \\ \Delta\text{Coverage}, & \text{otherwise} \end{cases}$
    where $\Delta\text{Coverage} = \text{Coverage}(s_{t+1}) - \text{Coverage}(s_t)$
            Store transition $(s_t, a_t, r_t, s_{t+1}, \text{done})$ into replay buffer
            Update agent $k$'s network by minimizing the loss:
                $\mathcal{L}_k = \left(r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta')_{k+1} - Q(s_t, a_t; \theta_k)\right)^2$
            Update environment state based on $a_t$
        end for
        if timestep $t$ mod target update interval $= 0$ then
            Update target networks: $\theta'_k \leftarrow \theta_k$ for all agents
        end if
        if episode done then
            Break
        end if
    end for
end for
Return: Trained policies $\pi_k(s) = \arg\max_a Q(s, a; \theta_k)$ for all agents
```

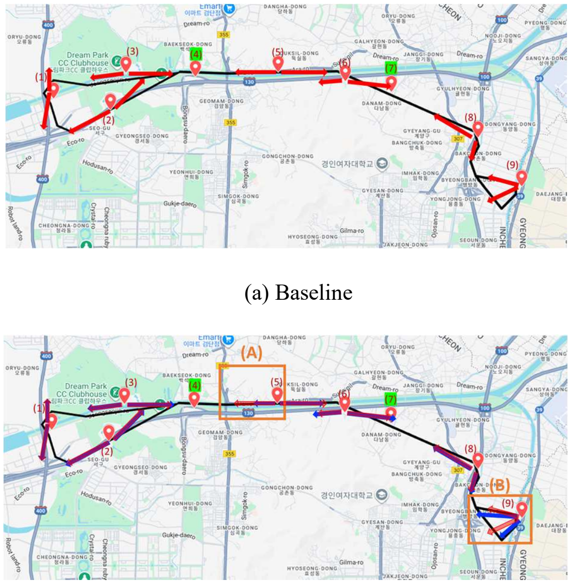
$$Loss = \left(r_t + \gamma max_{a'} Q(s_{t+1}, a'; \theta^-)_k - Q(s_t, a_t; \theta)_k\right)^2 \quad (6)$$

In contrast, the sequential multi-agent update framework differs from the conventional MA-DQN in that each agent performs its action one after another. As a result, the Q-value of the next state used in the target value corresponds to the Q-value predicted by the next agent, not the same agent. Therefore, the loss function in the proposed approach is formulated as shown in Equation (7).

$$Loss = \left(r_t + \gamma max_{a'} Q(s_{t+1}, a'; \theta^-)_{k+1} - (s_t, a_t; \theta)_k\right)^2 \quad (7)$$

In MA-Sequential, the environment state evolves step-by-step as agents act sequentially within the same timestep. The state s_{t+1} reached after agent k takes an action is observed and acted upon by agent $k+1$. Therefore, using the Q-value of the same agent to compute the target value can introduce a mismatch with the actual decision-making order, causing the update to diverge from the sequential structure of the framework. By instead using the Q-value predicted by the next agent, the proposed method ensures that the target value accurately reflects the state transitions that occur during sequential decision-making, leading to more coherent and stable learning.

Algorithm 1 summarizes the proposed reinforcement learning-based aerial network optimization framework. The algorithm operates on an episodic basis, where the environment is reset at the beginning of each episode. For every timestep, each agent is sequentially updated. Each agent observes its current state and selects an action according to the ε-greedy policy. Based on the executed action, the environment returns the next state and a reward, which reflects the change in coverage or applies a penalty for invalid actions. Transitions are stored in a replay buffer, and the agent's Q-network is updated by minimizing the loss function using the next agent's Q-value. This sequential update structure differs from the conventional MA-DQN by enabling each agent to learn with consideration of the prior agents' actions, enhancing learning stability and convergence. Target networks are periodically updated, and learning proceeds until the episode terminates. Through this process, each agent independently learns an optimal policy, ultimately aiming to maximize communication coverage in the UAM aerial corridor environment.



(a) Baseline

(b) BO & MA-DQN

Fig. 4. Simulation Results

IV. Performance Evaluation

The detailed simulation parameters are summarized in Table II below.

TABLE II. SIMULATION PARAMETERS

Coverage weight (w_{cov})	100
Distance threshold (d_{th})	1000m
Angular threshold (θ_{th})	90°
Max angle penalty (p_{max})	30
Off-BS weight (λ_{off})	10
Penalty coefficient (λ_{pen})	100
Discount factor (γ)	0.95
Exploration rate (ϵ)	1.0
Exploration decay (ϵ_{decay})	0.9975
Learning rate (α)	0.001
Hidden layer sizes	[1024, 256, 64]

The simulation results are visually summarized in Fig. 4, where (a) illustrates the baseline configuration, and (b) compares the BO-based configuration with the final result of the proposed MA-DQN learning. In (a), all nine candidate base stations (BSs) are activated and uniformly distributed along the flight path to ensure consistent coverage. However, in certain curved segments, beam directions were misaligned with the actual trajectory, resulting in coverage gaps despite full BS deployment.

In (b), both the results of the BO-based configuration (blue arrows) and the MA-DQN learning outcome (red arrows) are visualized. During the BO stage, several base stations that could be effectively substituted by neighboring ones were deactivated. Most agents adjusted their swing and tilt angles to better align with the path's geometry.

In the subsequent MA-DQN learning stage, fine-tuning was conducted based on the configuration derived from BO. While swing angles remained mostly unchanged, tilt angles were updated in 11 agents. These adjustments resolved the remaining uncovered regions along the path, ultimately achieving complete coverage. In particular, region (A) required some antennas to cover relatively distant areas due to BSs deactivated during the BO stage. To address this, the MA-DQN phase compensated by lowering the tilt angles to widen the coverage range. In region (B), where the flight path formed a loop, swing angles were adjusted to realign the beam directions with the curved trajectory. These local refinements demonstrate that the proposed framework can effectively adapt to diverse path geometries.

Table III presents a summary of the number of base stations used and the corresponding coverage performance for each configuration. The baseline configuration achieved 90.87% coverage using nine base stations, while the BO Only configuration improved coverage to 95.43% with only seven base stations. The proposed MA-DQN method further enhanced the coverage to 100% using the same seven base stations. Therefore, the simulation results confirm that the proposed combination of BO-based search and MA-Sequential learning can achieve stable communication coverage with fewer base stations than conventional methods. This demonstrates the effectiveness of the approach in balancing resource efficiency and network performance in aerial network planning.

TABLE III. SIMULATION RESULTS

Technique	Number of Base Stations	Coverage
Baseline	9	90.87%
BO Only	7	95.43%
Proposed MA-DQN	7	100.00%

V. Conclusions

This paper addressed the problem of aerial network planning for UAM by considering a real-world urban flight corridor. To solve this problem, we proposed a hybrid framework that combines Bayesian Optimization with a sequential multi-agent reinforcement learning method (MA-Sequential DQN). Each base station was modeled as an independent agent with two beams, and the configuration-including antenna swing and tilt angles as well as power state-was sequentially optimized. Simulation results show ed that the proposed method achieved full coverage along the entire route while utilizing only 7 out of 9 candidate base stations. These results demonstrate that the proposed learning-based design approach is both efficient and practical for ensuring reliable aerial communication coverage.

Acknowledgment

This research was in part financially supported by the Ministry of Trade, Industry, and Energy (MOTIE), Korea, under the "Global Industrial Technology Cooperation Center program" supervised by the Korea Institute for Advancement of Technology (KIAT).

References

[1] S. Noh, J. Cho, M.-J. Ahn, and D. An, *A Study on Developing an Urban Air Mobility (UAM) Safety Assessment System*, Research Report RR-24-09, Korea Transport Institute, 2024.

[2] Ministry of Land, Infrastructure and Transport (MOLIT), *The First Step for Demonstrating K-UAM Working with 46 Companies (K-UAM Grand Challenge)*, Press Release, Seoul, Korea, Feb. 2023.

[3] 3GPP TR 36.777 V15.0.0, "Study on enhanced LTE support for aerial vehicles," Jan. 2019.

[4] 3GPP TS 22.125 V16.2.0, "Service requirements for the 5G system; Stage 1," Mar. 2022.

[5] Mohamed Benzaghta, Giovanni Geraci, David López-Pérez, Alvaro Valcarce, "Cellular network design for UAV corridors via data-driven high-dimensional Bayesian optimization," arXiv preprint, arXiv:2504.05176, Apr. 2025.

[6] Mohamed Benzaghta, Giovanni Geraci, David López-Pérez, Alvaro Valcarce, "Designing cellular networks for UAV corridors via Bayesian optimization," arXiv preprint, arXiv:2308.05052, Aug. 2023.

[7] 3GPP TR 37.885 V15.3.0, "Study on enhancement of 3D channel model for the frequency range from 6 GHz to 100 GHz," Jun. 2019.

[8] 3GPP TR 38.901 V16.1.0, "Study on channel model for frequencies from 0.5 to 100 GHz," Jan. 2020.

[9] Hud Wahab, Vivek Jain, Alexander Scott Tyrrell, Michael Alan Seas, Lars Kotthoff, Patrick Alfred Johnson, "Machine-learning-assisted fabrication: Bayesian optimization of laser-induced graphene patterning using in-situ Raman analysis," Carbon, vol. 167, pp. 609–619, 2020.