Geometry Adaptive Deep Q-Network for UAV-Based Emitter Localization in Cluttered RF Environments

Christopher Peters
Darwin Deason Institute for Cyber Security
Southern Methodist University
P.O. Box 750122
Dallas, TX 75275-0122
peterscl@mail.smu.edu

Michael Watts
Darwin Deason Institute for Cyber Security
Southern Methodist University
P.O. Box 750122
Dallas, TX 75275-0122
mcwatts@smu.edu

Eric C. Larson
Darwin Deason Institute for Cyber Security
Southern Methodist University
P.O. Box 750122
Dallas, TX 75275-0122
eclarson@smu.edu

Mitchell A. Thornton
Darwin Deason Institute for Cyber Security
Southern Methodist University
P.O. Box 750122
Dallas, TX 75275-0122
mitch@smu.edu

***Abstract*— This paper presents a prototype geometry-adaptive deep Q-network (DQN) for cooperative unmanned aerial vehicle localization of radio-frequency emitters in dense multipath environments. UAVs act as reconfigurable array elements and are repositioned by a reinforcement learning controller built on time-difference-of-arrival measurements, with rewards shaped by composite uncertainty reduction, signal strength, and trajectory efficiency. Monte Carlo simulations show the DQN reduces error by up to 60 % within fewer repositioning steps as compared to baseline cases. Digital twin evaluations in the Cyber Autonomy Range using Keysight EXata, that incorporates multipath, fading, and hardware inaccuracies, demonstrates rapid convergence in Rician fading and multipath conditions. These results demonstrate robust performance beyond idealized models and establish a foundation for scalable cooperative UAV-based emitter localization systems.**

TABLE OF CONTENTS

1. INTRODUCTION

Accurate localization of unknown-location radio frequency (RF) emitters is critical for aerospace and defense missions, including search-and-rescue, spectrum monitoring, and electromagnetic spectrum operations (EMSO). Static sensor arrays suffer from limited coverage and degraded accuracy in cluttered RF environments where multipath propagation and synchronization errors dominate performance [1]-[7]. Cooperative unmanned aerial vehicle (UAV) systems (Figure 1), offer an alternative by operating as mobile sensor arrays that can dynamically reconfigure their geometry to improve localization accuracy while maintaining acceptable operational standoff range [2], [4].

In this work, we design and evaluate a prototype cooperative multi-agent reinforcement learning (MARL) framework that enables a swarm of UAVs to autonomously locate RF emitters whose positions are unknown in dense multipath environments. Each UAV functions as a geometry-reconfigurable antenna array element, and a deep-Q network (DQN) [5] controller adapts the swarm geometry to improve time difference of arrival (TDOA) localization, achieving accuracy goals in very few consecutive measurements or "snapshots".

Building upon established time-based localization algorithms, such as multilateration, we exploit time difference of arrival (TDOA) measurements to estimate emitter position. Prior work established that the TDOA-based Location on a Conic Axis (LOCA) algorithm method [8], [9] can approach the Cramér-Rao lower bound (CRLB) under realistic bounded error conditions [1]. However, real deployments face multipath, atmospheric absorptions, and hardware inaccuracies that degrade performance, particularly in urban settings [3]. Adaptive repositioning mitigates these effects, and metrics such as received signal strength (RSS) convergence or centroid equidistant repositioning can be used to find improved geometries [4]. While effective in controlled settings, these objective-based approaches can plateau in suboptimal geometries and therefore often fail to generalize to complex environments. Reinforcement learning (RL) localization approaches such as DQNs [6] offer a policy-based alternative that can overcome these geometry limitations and intelligently guide the system towards improved geometries in degraded environments [7]-[12].

Moreover, adaptive array reconfiguration allows for improved localization performance in the elevation as well as in the azimuthal planes, improving the measurement fields-of-view and yielding higher emitter location accuracies. In our architecture, a two-phase control strategy guides the swarm into a LOCA-optimized formation. First, a 2-D azimuthal-plane localization stage reduces the circular error probable (CEP) area, then elevation is refined through 3-D localization by leveraging array geometry diversity to

979-8-3315-7360-7/26/$31.00 ©2026 IEEE

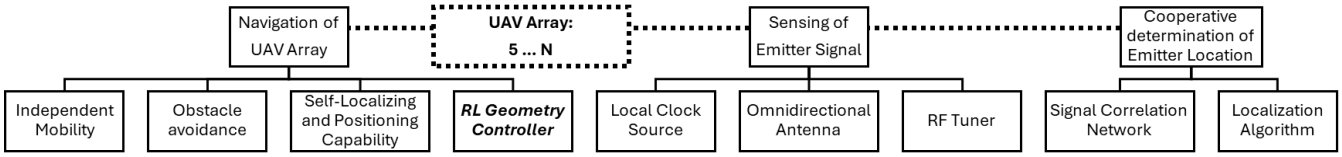


Figure 1. Representative block diagram for UAV-based emitter localization system.

improve spherical error probable (SEP) position and volume [19]. This framework, implemented with column lattice formations and radius-threshold corrections, achieves a significant reduction in localization uncertainty relative to an operator-specified 3-D SEP goal, thereby enhancing mission effectiveness in cluttered RF terrains.

A sequential extended Kalman filter (EKF) filters TDOA measurements at each timestep, refining current localization estimates and integrating prior estimates. Our training and verification process is configurable by number of drones in the UAV array swarm, allowing the DQN policy to scale to varying swarm sizes without hardware or software modifications. Using increasing swarm sizes, we are able to realize even lower localization error and earlier convergence [4], however, our current work evaluates a fixed swarm size.

A robust digital twin of the swarm and environment, including multipath, weather effects, and hardware tolerances such UAV on-board positioning and clock jitter, validates our performance. Using the Southern Methodist University (SMU) Cyber Autonomy Range (CAR) facility [13], the system performance is evaluated with realistic digital twins of the external environment and the localization system. Our results include many test scenarios with varying simulation parameters allowing our prototype to be exercised much more broadly than if a physical test range were used. Our digital twins are implemented at the physical layer with physics simulation engines for RF propagation, fading, scattering, kinematic dynamics with gravity field interactions, and weather effects using physical layer accurate protocols including 5G and IEEE 802.11 standards. The system twin includes dominant sensor hardware error sources such as UAV local clock inaccuracies and UAV positioning drift. The environmental twin is implemented using Keysight's EXata high fidelity urban environment within the CAR. With these realistic constraints, our functional evaluations confirm the system's low localization error and SEP under degraded RF conditions, demonstrating its viability for aerospace and other applications.

To our knowledge, this is the first RF emitter location system implemented within an autonomous UAV swarm exploiting dynamic spatial array diversity through customized reinforcement learning. Our composite DQN reward function balances current and future array geometry, current emitter location accuracy and uncertainty, and minimal UAV repositioning path length. This strategy provides rapid convergence to low localization error, typically within ten or fewer snapshots, each with a different array configuration.

The contributions of this paper are threefold:

(1) Geometry-aware DQN controller: A DQN with variance-aware state descriptors, adaptive action scaling, and hybrid reward shaping (SEP+geometry) enables robust operation under multipath and jitter, allowing the swarm to achieve target accuracy within only a few snapshots.
(2) Comprehensive simulation evaluation: Monte Carlo trials show that the DQN achieves up to 75 % median SEP reduction in adverse conditions within 10 steps, outperforming baseline localization methods.
(3) High-fidelity digital twin validation: Using Keysight EXata in the Cyber Autonomy Range, we validate the controller in a high-fidelity environment with multipath, weather effects, hardware inaccuracies and other error sources. Results confirm effective convergence relative to simulation.

2. Background and Related Works

Throughout this section and the remainder of the discussion, reference terms and notation are defined in the Appendix in Table 3.

Cooperative UAVs for Emitter Localization

Emitter localization using unmanned aerial vehicles and systems has been studied extensively due to its applications in surveillance, spectrum monitoring, and search-and-rescue [1]-[2], [6]-[7], [10]-[12] . Past approaches usually rely on single or multiple UAVs converging on the RSS of an RF or other signal type [14]-[15], or multi-UAV multilateration or angle-of-arrival techniques [16]-[17], but accuracy is highly sensitive to ill-formed geometry that limits operational fields of view resulting in degraded TDOA techniques and measurements that are sensitive to environmental multipath effects [18].

UAVs equipped with sensing antennas can form a cooperative wireless sensor array through ad hoc networking. Unlike static arrays, UAV arrays have the ability to dynamically adjust their geometry to improve line-of-sight (LoS) coverage and reduce multipath errors [4].

Time-Based Localization Algorithms

For a sensor at position $s_i = (x_i, y_i, z_i)$ and emitter $\mathbf{E} = (x_e, y_e, z_e)$, the emitter signal time of arrival (TOA) is

$$t_i = \frac{\|\mathbf{E} - s_i\|}{c} + \varepsilon_{t,i}, \quad (1)$$

with c as propagation speed and ε_t as timing error.

The TDOA between sensors i and j is $\Delta t_{ij} = t_i - t_j$, that defines a hyperboloid of possible emitter locations. LOCA [8]-[9] is an RF emitter geolocation approach using distributed sensors [1], [4] where the TDOA of a signal among sensor triads form conic axes coincident with the emitter's location and can be applied in 2-D or 3-D cases. Computing the intersection of the conic axes finds the emitter location, as visualized in Figure 2. Sensor triads are chosen by any combination of three sensors in the array, and in 3-D the geolocation method can function with as few as four UAVs per swarm. The use of additional sensors leads to an increasing number of geolocation estimates; however, using swarms larger than eight UAVs provides diminishing improvement in terms of emitter location convergence rate [1], [4].

Aggregating across these sets of sensor triads forms a solution cloud,

$$\Lambda = \{\lambda_m\}_{m=1}^{M} \subset \mathbb{R}^3 \tag{2}$$

that we robustly filter based on the interquartile range (IQR) of the solution set to obtain an emitter location estimate $\hat{E}$ using the filtered results.

The quality of the localization estimate can be expressed in terms of spherical error probable (SEP_{50}) [19], where the radius defines a spatial region such that the true location, **E**, is present with probability 0.50:

$$SEP_{50} = q_{0.5}\{\|\lambda_m - \hat{E}\|\}. \tag{3}$$

Error Sources

As impacts to localization performance, we consider hardware inaccuracies such as UAV positional drift and clock jitter. Realistic error values are used by assuming that each UAV is equipped with on-board LiDAR for self-positioning correction and a chip-scale atomic clock (CSAC) as a local time source [1]. We account for multipath scattering and fading by using Rician and Rayleigh propagation models to incorporate atmospheric effects that attenuate the signal and contribute to path delays. Additionally, the EXata environment also includes various atmospheric weather models and all of these model aspects contribute to sensor and clock offsets that impact UAV localization operation [20]. These various error sources affect sensor-to-emitter range, are height-dependent, and vary according to the physical clutter of the surrounding environment with typical performance impacts as stochastically characterized in [21]. In consideration of these sources, we generated appropriate error models with parameters as summarized in Table 1. For multipath delay spread impacts [21], we assume a 50 % mix of line of sight (LoS) and non-LoS conditions [20].

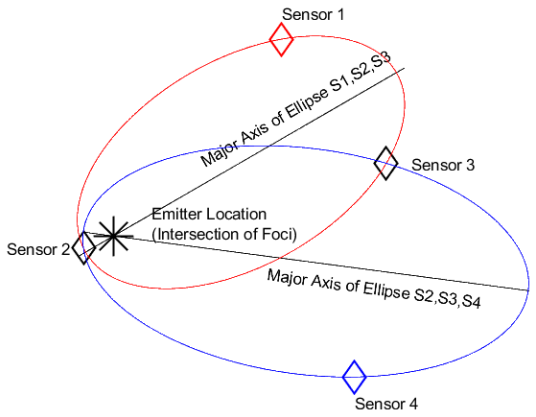


Figure 2. Intersection of two conic major axes at emitter location found by LOCA algorithm.

Measurement Fusion and Sequential Estimation

The cloud Λ from Equation (2) is combined with sensor jitter variance (known) and multipath variance (estimated) weights from Table 1 to yield a set (z_k, Σ_z), that represents the system measurement z_k and the measurement covariance $\Sigma_z = P_k$ [22]. This approach enables evaluation of a "variance-fused" estimate in terms of the sensor i covariance:

$$\begin{aligned} &\text{Sensor variances: } \sigma_i^2 = \sigma_{env}^2(\hat{E}, i) + \sigma_{hw}^2(i) \\ &\text{weights: } \omega_q \propto \left(\Sigma_i a_i(\hat{E}, \lambda_q)\sigma_i^2\right)^{-1} \\ &z_k = \frac{\Sigma_q \omega_q \lambda_q}{\Sigma_q \omega_q}, \Sigma_z = \frac{\Sigma_q \omega_q (\lambda_q - z_k)(\lambda_q - z_k)^T}{\Sigma_q \omega_q}. \end{aligned} \tag{4}$$

To track sequential emitter location estimates and stabilize the geometries against sensor jitter, we apply a position-only Kalman filter (KF) [23], gated by normalized innovation squared (NIS). A KF is a common statistical method to reduce measurement fluctuations and fuse noisy measurements. Our use of an extended KF (EKF) utilizes a nonlinear measurement model of the emitter state, as outlined below:

1) Define the process model as a random walk:
   - $x_{k+1} = x_k + w_k$, $w_k \sim \mathcal{N}(0, Q)$ with $Q = 10^{-2} I_3$
2) Define measurement model:
   - $z_k = x_k + v_k$, $v_k \sim \mathcal{N}(0, R_k)$ with $R_k = \Sigma_z$
3) Predict:
   - State: $x_{k|k-1} = x_{k-1|k-1}$,
   - Covariance: $P_{k|k-1} = P_{k-1|k-1} + Q$
4) Determine innovation, gate EKF with NIS:
   - Residual: $v_k = z_k - x_{k|k-1}$,
   - Innovation covariance: $S_k = P_{k|k-1} + R_k$

Table 1. Error parameters for localization model.

Error Parameter	Error Distribution
Self-positioning error (m)	$\varepsilon_U = \mathcal{N}(0, 0.02)$
Local clock inaccuracy (ns)	$\varepsilon_c = \mathcal{N}(0, \sigma_t + 10)$
Multipath delay spread (ns)	$\varepsilon_m = 0.5 * \mathcal{N}(23R_i^{0.26}, 5.5R_i^{0.35})$
Time of arrival error (ns)	$\varepsilon_{t,i} = \varepsilon_m + \varepsilon_c$
Scattering loss (dB)	$\varepsilon_P = \mathcal{N}(3, 5)$

- If: $v_k^T S_k^{-1} v_k \leq$ NIS gate: $\begin{cases} K_k = P_{k|k-1} S_k^{-1} \\ x_{k|k} = x_{k|k-1} + K_k v_k \\ P_{k|k} = (I - K_k) P_{k|k-1} \end{cases}$
- Else: $x_{k|k} = x_{k|k-1}, \qquad P_{k|k} = P_{k|k-1}$

5) Set $\bar{E}_k = x_{k|k}, \qquad P = P_{k|k}$

Reinforcement Learning for Geometry Control

While filtering provides robustness against stochastic errors, the overall geometry still plays a dominant role in reducing SEP. To dynamically adapt the array geometry, reinforcement learning has emerged as an alternative to repositioning approaches that only rely on real-time system measurements. Specifically, value-based learners such as Q-Learning and DQN [5], [14] have been used for trajectory planning. These methods optimize state-action value assignment through a neural network while encoding high-level feature combinations into the network. Reference [10] applied Q-learning for UAV-based illegal station localization in simplified environments, showing the potential of RL to outperform traditionally static, ground-based measurement approaches. RL approaches have been studied for many UAV trajectory planning [7], mapping [6], and localization tasks [12] and have been extended to cooperative arrays of UAV in the multi-agent Q-Learning (MQL) case [11].

Many approaches use the RSS of the emitter at the UAV and optimize the UAV trajectory towards physical convergence on the emitter by rewarding increased RSS values[6] [11]. This approach has been shown to outperform non-RL RSS-based methods [10]. However, this approach requires the environment to allow direct convergence, which may be infeasible in geography-constrained scenarios (for example, no-fly zones, hazardous areas, or obstacles). Additionally, RSS-only convergence can lead to degenerate array geometries that degrade TDOA performance and increase localization error [22]. Beyond the limitations of RSS-only methods, other MQL approaches incorporate multilateration for localization [11] and promote cooperative actions, such as trajectory decisions, in a shared state and action space to maximize the reward. These studies and similar non-RL approaches [15] form a framework promoting an optimized geometry for localization, but many methods use pre-determined information [16] and they would generally not be suited for dynamic environments.

In this work we extend prior studies by introducing a DQN-based controller for dynamically adaptive swarm geometry reconfiguration, integrating LOCA-based TDOA fusion metrics directly into the learning process. We compare our DQN method to a geometry-heuristic based (non-RL) approach [4] that continuously repositions the UAV array to geometries that improve SEP. This prior approach was able to demonstrate, at best, a 75 % reduction in SEP volume and 50 % reduction in localization error after 25 steps of repositioning, achieving an average localization error of 16 m. Using the heuristic approach as a baseline, we first evaluate whether applying the DQN RL method to the geometry controller can achieve greater performance improvement in fewer repositioning steps. Further, we validate system performance in a high-fidelity digital twin environment inclusive of a variety of error sources. We demonstrate RL-based geometry adaptation with physics-based EXata validation, bridging the gap between theoretical feasibility and operational viability.

EXata

EXata is a network modeling software application [24] designed to generate digital twins of real-world network topologies under a variety of conditions such as terrain and weather effects. These digital twins are capable of running real time, modeling the environment as data moves across the network stack with high fidelity, with the ability to respond to control messages near identically to a real network. Using this tool, one can simulate a variety of network protocols running across multiple devices and receive real time statistics on the behavior of devices and data in the network. The EXata engine is capable of interactions with external live applications for the analysis of this data. For this study, the primary value of EXata is its robust modeling of wireless network topology, such that multipath impacts to signal strength and path delay are superior to the error models we include in our simulation. Therefore, we use EXata for verification analysis through a digital twin of our system versus performing flight trials that are costly, time-consuming, and introduce risks to prototype hardware.

3. ARCHITECTURE

Each UAV functions as a mobile array element with omnidirectional sensing of emitter signals. Although we can use a variety of RF transmitter frequencies in our digital twin environment, the results described here use 5 GHz. At each measurement snapshot or step, UAVs acquire TDOA and RSS measurements, form LOCA-based solution clouds, and fuse results with variance-aware weighting. Intra-swarm communication is achieved with an ad hoc wireless network optimized to minimize transmit power and with extremely short message frames. A controller UAV is elected within the swarm that receives measurement data, performs computations, and issues repositioning commands. The localization estimates are performed in 2-D and 3-D, depending on the dominant geometry characteristics, and the controller determines a new geometry based on the metrics from the localization result. Because we develop a policy with a DQN, the approach runs in two phases: training within the digital twin environment, where true emitter location information is utilized to guide the improved array geometry, and evaluation, where the true emitter location is unknown, but a Kalman filter is used to refine the localization estimates. Our system performance is evaluated by the localization error and uncertainty (framed by SEP_{50}), the number of repositioning steps required to converge on the emitter, and the distances the UAVs travel across a localization episode. These metrics serve as operationally-relevant performance parameters providing a measure of system effectiveness in its ability to accurately, rapidly, and efficiently localize an unknown-location emitter.

Localization Estimation

After initial TDOA and RSS measurements, the 2-D or 3-D LOCA solution cloud Λ is fused with the sensor jitter and multipath variance to yield the solution set (z_k, Σ_z). As described in Section 2, we apply to the variance-fused measurements a position-only EKF [23] with a NIS gate of 16.3 (~95 % in 3-D). During training the EKF is disabled so agents learn directly from noisy estimates, while during evaluation the use of the EKF reduces jitter to accelerate convergence.

Deep Q-Network

The DQN controller UAV receives normalized geometry descriptors, per agent features, and compact statistics, utilizing two fully-connected layers with dropout regularization and an experience replay buffer [5]. Actions are 3-D displacements, scaled adaptively by current SEP_{50} radius to ensure stability across mission phases. The reward function combines localization improvements, geometry formation, and a path penalty to shape the policy towards a pseudo-optimal geometry.

At each repositioning step k, each UAV determines a state vector

$$\phi_k = \left[\frac{\hat{s}_i - \bar{x}_k}{\|\hat{s}_i - \bar{x}_k\|}, \widehat{RSS}_i, \widehat{\Delta t}_{i,j}, P, SEP_{50}\right] \tag{5}$$

including inter-UAV distances and relative emitter estimate offsets, RSS, TDOA measurements, SEP_{50}, and the state error covariance P. In our architecture, the RSS is provided only as context and is not included in the reward function. The agents are trained to improve geometry for emitter localization, not to maximize RSS as in prior power-seeking approaches, which can produce degenerate TDOA geometries. However, because RSS correlates with signal quality and multipath conditions, exposing it in the state allows the policy to use it as a contextual indicator when making geometry decisions.

Each agent selects a bounded displacement vector:

$$a_i^k = \eta \hat{\delta}_i \in \mathbb{R}^3 \tag{6}$$

where $\hat{\delta}_i = \frac{\hat{s}_i - \bar{E}_k}{\|\hat{s}_i - \bar{E}_k\|}$ is a normalized UAV-emitter displacement direction, and η is an adaptive step size scaled to current SEP. During training, action exploration to the optimal geometry is further encouraged with an annealing ϵ-greedy policy [25]:

$$a_i^k = \begin{cases} a_{i,\text{greedy}}^k & \text{probability } 1 - \epsilon_k \\ a_{i,\text{random}}^k & \text{probability } \epsilon_k \end{cases}. \tag{7}$$

With probability $1 - \epsilon_k$, each agent executes the bounded displacement action a_i^k, and with probability ϵ_k, each agent executes a random bounded displacement. ϵ_k decays throughout the training process to guarantee exploration and escape poor local geometries while later converging to geometry-improving moves. In evaluation, we set $\epsilon_k = 0$ to ensure a pure greedy action selection.

The action selection is additionally guided by a 2-D-to-3-D phase strategy to optimize the LOCA geometry. Initial measurements utilize a 2-D localization step to reduce the circular error probable (CEP) area in an azimuthal plane and to promote, or encourage, an array geometry in a ring-like formation around the estimated emitter location which is optimal for the LOCA methodology. Once the initial geometry is formed, the elevation estimate of the emitter location is refined through 3-D localization by leveraging array geometry diversity though evenly-spaced column lattice formations to improve SEP position and volume.

The reward function combines localization improvement, geometry shape, and convergence efficiency:

$$r_k = \alpha \Delta SEP_{50} + \beta \Delta \|\mathbf{E} - \bar{E}_k\| + \gamma \hat{\delta}_i - \kappa \tag{8}$$

where α, β, γ are tuned weights, $\kappa > 0$ is a per-step penalty, and $\beta > 0$ only during model training. The geometry shape reward encourages the UAVs to form an equidistant shape lattice around the emitter, similar to the best-performing heuristic approach identified in [4]. The architecture stops iterating when SEP_{50} meets a user-specified goal, and smaller SEP probabilities can easily be accommodated through the change of a parameter. The RL controller algorithm is shown in Algorithm 1, using the parameters defined in Appendix Table 3.

Algorithm 1. Emitter localization with UAV geometry control

Initialize: UAV starts $\{s_i^0\}_{i=1}^N$; unknown emitter $\mathbf{E} \in \mathbb{R}^3$.
Set $k \leftarrow 0$, SEP goal ς_{SEP}, EKF $x_{0|0} = z_0$, $P_{0|0} = 10^6 I_3$
Repeat (per step k):
1: Compute noisy sensing: draw Δt_{ij}, RSS_i from **E**, then
$\hat{s}_i = s_i^k + \varepsilon_{U,i}$, $\widehat{\Delta t}_{i,j} = \Delta t_{i,j} + \varepsilon_{t,ij}$, $\widehat{RSS}_i = RSS_i + \varepsilon_{P,i}$
2: Perform 2-D or 3-D LOCA cloud estimation (2) and prune outliers outside 50 % interquartile region.
3: Determine cloud statistics (3).
4: Perform variance-aware fusion (4) to generate (z_k, Σ_z)
5: Perform position-only EKF:
i: *if* training, skip EKF: $\bar{E}_k = z_k$, $P = \Sigma_z$
ii: *else* evaluation, perform EKF: set $\bar{E}_k = x_{k|k}$, $P = P_{k|k}$
6: RL Control:
i: Generate state (5)
ii: Determine bounded action (6)-(7) and apply elevation diversity
iii: Compute reward (8)
iv: Update policy
7: Advance to next geometry ($s_i^{k+1} = s_i^k + a_i^k$) and continue
8: Stop: $SEP_{50} \leq \varsigma_{SEP}$ or plateau
Return: $\bar{E}_k$, covariance P, SEP_{50}, total steps, total distance traveled.

4. Simulation and Digital Twin Verification Approach

Simulation Environment

The system was simulated using the Actor-Environment Cycle Environment class of PettingZoo in Python which was developed to simulate multi-agent environments with an observation space defined by Box from Gym [26]. We conducted 1000 Monte Carlo runs over a 4000 m×4000 m×500 m urban volume populated with Rayleigh-distributed scatterers following ITU-R P.1411 recommendations [21]. Emitters broadcasting a 5 GHz signal were randomly placed within the environment. An eight-element UAV swarm is deployed with randomized initial positions.

UAVs are modeled as independently mobile systems with omnidirectional antennas and precise self-positioning capability. Clock accuracy is simulated as synchronized pre-mission by CSAC oscillators, and TDOA errors were drawn from Gaussian distributions consistent with CSAC-specified jitter parameters. RSS values are perturbed by log-normal shadowing and multipath fading. Measurement noise is injected at each step, with variance values selected to replicate realistic hardware tolerances. Multipath scattering and fading is modeled in accordance with ITU-R P.1411 [21]. Since these values are frequency and elevation-dependent and are stochastically specified, we consider a 5 GHz operating frequency with UAVs located at varying elevations giving primary signal contributions in a mix of LoS and non-LoS conditions such that we reduce the multipath effects by 50 %. The above errors follow the methodology in [4] and are detailed in Table 1.

RL Training Setup

The DQN was trained within our digital twin environment for 1000 episodes of random initial geometries with each episode consisting of up to 250 localization steps. Each episode terminated when the SEP radius fell below a mission-defined threshold (we use 5 m for the results described here) or the system performed 250 repositioning steps. Reward shaping followed the formulations described in Section 3. The resultant policy from the RL training was utilized in the same environment for an initial evaluation of the DQN.

EXata Digital Twin Environment Setup

The EXata network simulator is integrated with the SMU CAR [13] to create a digital twin of both the UAV swarm and the RF environment. The following elements are included within our digital twin:

(1) Propagation: ITU-R P.1411 multipath propagation and Rician fading with configurable *K*-factor, simulated urban environment.
(2) Weather effects: rain attenuation modeled at 0.5 dB/km at 5 GHz; atmospheric absorption by oxygen and water vapor following ITU-R P.676 [27].
(3) Hardware errors: UAV hover drift modeled as Gaussian-distributed errors ($\varepsilon_p = 2$ cm); clock jitter drawn from Gaussian distributions with $\varepsilon_t = 10$ ns.
(4) Kinematics: UAV motion followed gravity field dynamics with wind perturbations.

At each snapshot, UAVs collect TDOA and RSS measurements generated by EXata's physical-layer accurate RF models and store the measurements within an internal EXata system database. These measurements are parsed and passed to the DQN architecture that selects discrete displacement actions for each UAV based on the trained policy. Updated positions are then re-injected into EXata for the next iteration. This process allows for a hardware-in-the-loop ready framework for use in the future to evaluate how the same policy would be executed on real UAVs.

Using EXata, we evaluated two scenarios: (i) a Rician-faded urban multipath environment, and (ii) the same environment under degraded weather conditions. These two scenarios used the same set of 200 randomized emitter/UAV initializations and UAV system on-board positioning and clock error conditions. Results were compared with DQN simulations utilizing the same initial geometries and comparable multipath conditions.

5. Results

Evaluation Metrics

Performance is evaluated by the number of steps required to achieve a target SEP_{50} radius, the error in the final localization estimate from the true target location, and the distance the system traveled to localize the emitter. Each of these performance metrics indicates the effectiveness of the system to accurately and efficiently localize an emitter. We analyze the mean and median performance of these metrics in addition to the performance distribution to understand the stability of the system in different scenarios.

Simulation Results

Using the policy developed in the RL training in Section 3, the DQN is evaluated in the same dense multipath simulation environment for a first measure of performance prior to integration with EXata. As shown in Figure 3, the starting SEP_{50} and localization error was ~80 m. The initial reduction in localization error is due to the 2-D ring formation phase, and further improvement is found through the geometry refinement steps after the 2-D-to-3-D localization transition. The DQN simulation realized an average 75 % reduction in SEP_{50} in 10 steps and ~50 % error reduction in 20 steps, which outperforms the average performance of the heuristic-based approach [4]. However, considering the large standard deviation for the system error, the performance bounds are weighted by some outliers provided in the distributions shown in Figure 4 and Figure 5. Here, SEP_{50} most often reached a minimum of ~3 m with ~5.4 m localization accuracy. In the DQN simulation, this accuracy required 15-25 repositioning steps where each UAV in the system explored around 200 m to find the optimal geometry.

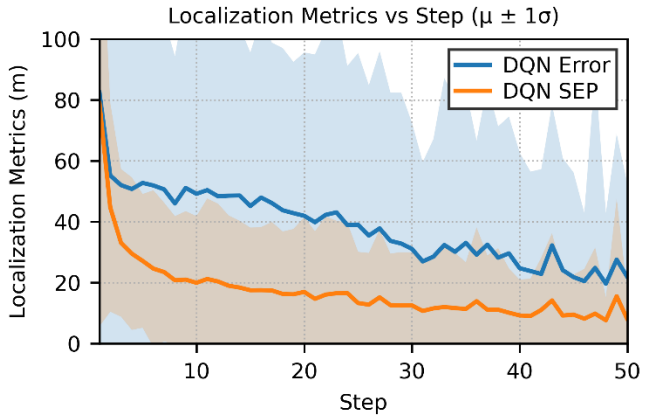


Figure 3. Localization error and SEP per repositioning step demonstrates convergence in dense multipath conditions.

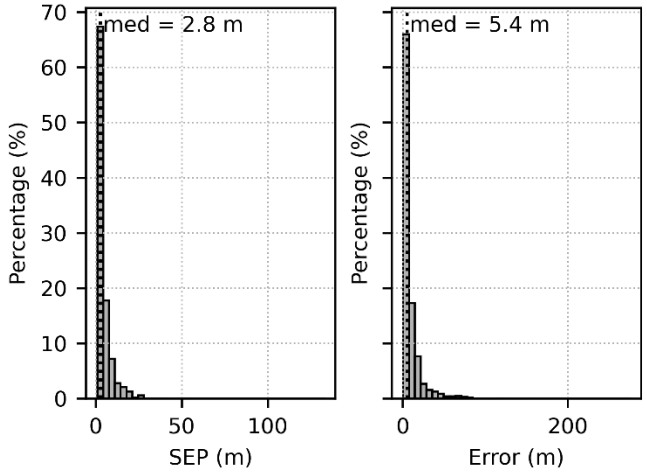


Figure 4. Localization metrics and median performance at convergence on emitter for 1000 DQN episodes, SEP radius (m) and localization error (m), show low median error and SEP for majority of evaluation episodes.

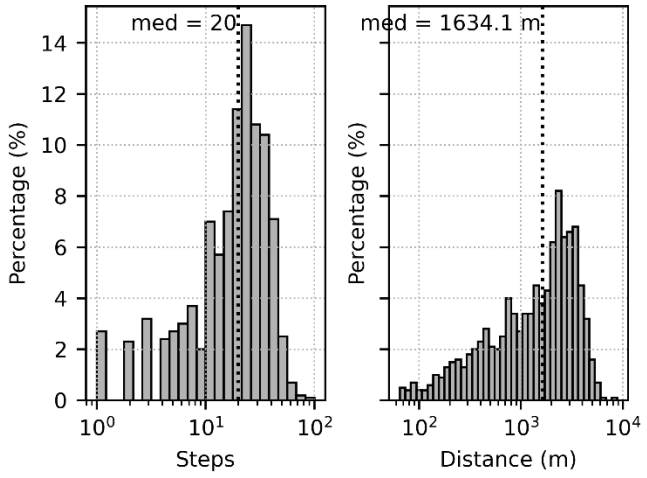


Figure 5. Repositioning metrics for 1000 DQN episodes show distance traveled by UAV system (m) and number of repositioning steps, with convergence in approximately 20 steps, traveling 200 m per UAV.

While these results demonstrate that the DQN achieves significant error reduction in a controlled setting, the initial simulation model has inherent limits in representing multipath and environmental dynamics. The approach relies on generalized error terms (Table 1) that inject noise and multipath contributors across trials, but these models do not fully represent all spatial conditions, especially as the UAVs converge to more optimal geometries and improve their LoS to the emitter. This limitation sometimes made the results appear worse than what would be expected in practice. The conservative modeling was still useful, as it allowed the policy to train under challenging error conditions and respond appropriately when exposed to a range of multipath scenarios. To assess performance under more realistic RF and environmental dynamics, we next evaluate the system using the EXata digital twin.

EXata Digital Twin Results

After evaluating the DQN RL policy in the dense multipath simulation environment, the DQN and policy were integrated into the EXata environment and analyzed under Rician fading with and without the presence of adverse weather conditions. The same initial geometries from the EXata simulations were also provided to the Python DQN simulation and evaluated for comparison. To better model the dominant LoS conditions of the Rician fading model in the Python DQN simulation, we reduced the Multipath Delay Spread factor in Table 1 by 50 %. This change in the environment also allowed us to evaluate our policy under different multipath conditions to ensure its suitability in other scenarios.

The primary evaluation metrics for these experiments are summarized in Table 2, showing the DQN with more stable performance in the EXata environment, requiring fewer repositioning steps to localize the emitter, and achieving an overall higher localization accuracy. Further, we see that the adverse weather conditions primarily impacted the initial system error, but once the geometry refinement began, the impacts were quickly overcome by the array geometry improvements indicating robustness of the approach in a variety of weather conditions.

Full metrics are further explored in Figure 6-Figure 9. For the EXata-based evaluations, each approach reached a minimum error in ten steps, while the Python DQN simulation reached approximately double the average error at the ten-step mark.

Table 2. Summary metrics for Python DQN Simulation (Sim) vs EXata Digital Twin study with Rician fading conditions and adverse weather.

Metric	EXata Rician	EXata Weather	Sim.
Mean final SEP (m)	4.1	4.1	3.9
Mean initial error (m)	18.5	22.7	76
Mean % reduction in error	81 %	85 %	87 %
Mean step count	2.8	2.8	9.4
Mean system total distance traveled (m)	1544	1588	660

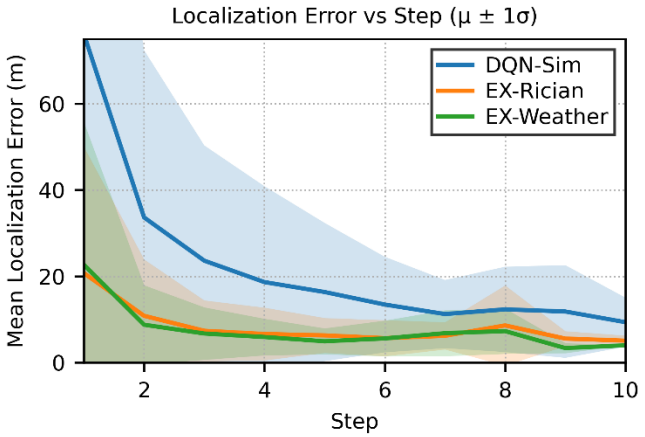


Figure 6. Comparison of average and standard deviation of localization error in the first ten reconfiguration steps among the Python simulations and the EXata digital twin trials shows convergence on emitter in approximately ten steps in Rician channel conditions.

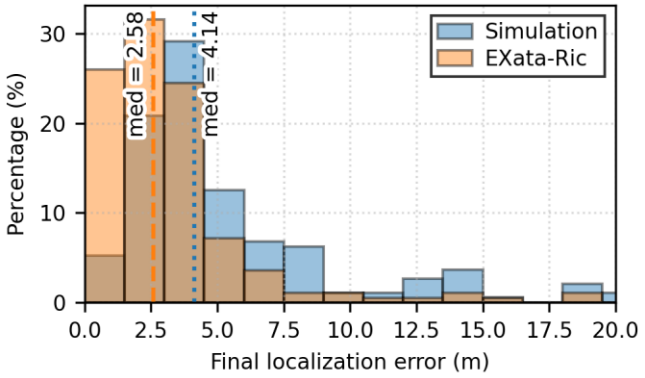


Figure 7. Histogram and median comparison of final localization error between Python simulations and EXata digital twin evaluations (<5 % of distribution > 20 m) shows low median error for majority of trials.

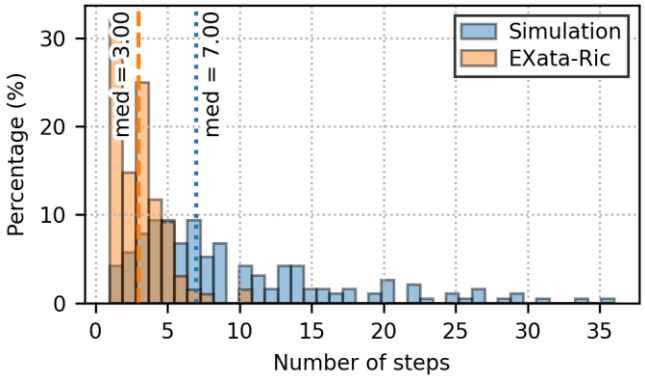


Figure 8. Histogram and median comparison of repositioning steps taken per episode between the Python simulations and EXata digital twin models shows majority of trials completing in under ten steps in both evaluation architectures.

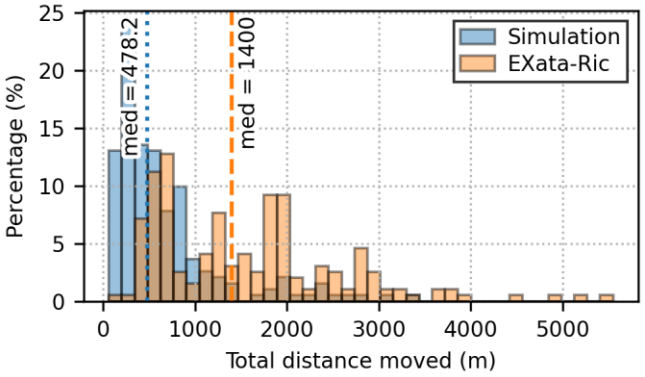


Figure 9. Histogram and median comparison of distance traveled by UAV system between simulation and EXata shows distance traveled was much lower in Python simulation for similar convergence metrics.

To consider the general performance of the system, in Figure 7-Figure 9 we overlay histograms of the primary metrics between the DQN Python simulation with the DQN EXata Rician scenario. In Figure 7, the error response shows a higher distribution of low-valued errors for the EXata trials and a higher spread for the Python DQN simulations, although both had comparable median error in the 2.6 m - 4.1 m range. In Figure 8, the typical performance in EXata showed convergence in as few as three repositioning steps, whereas in the Python simulation the system explored more geometries and converged in closer to seven steps. While the system explored over fewer steps in the EXata trials, in Figure 9 we observe the overall distance traveled in the EXata trials for the system is greater than that of the Python simulations.

Figure 10 shows a 2-D visualization of one representative EXata trial. Two agents make a large x-y excursion from their initial positions, which is an expected behavior that is seen consistently in both our DQN simulation and EXata aligned with our two-phase controller. The policy's SEP radius displacement vector guidance moves these agents mostly tangentially to correct the error in the azimuth plane with nominal z-axis change (not shown), and some correction back toward the ring geometry as they improve SEP and ultimately converge on a low localization error.

The EXata trials show an overall improved performance over the simulations, attributed to the robustness of EXata's multipath model versus the more stringent multipath error constraints given to the system in the simulation environment. In the simulation runs, we assume a mix of LoS and non-LoS conditions (with dominant LoS to approximate a Rician condition) but we provide no LoS advantage to the system when the UAVs converge on the emitter. When the UAVs are very close to, or even above the emitter, the system can experience pure LoS scenarios that increase the likelihood of an accurate measurement by reducing the error strictly to the clock and positioning hardware errors. In a more robust simulation environment like that provided by the EXata digital twin, when this convergence occurs in the first

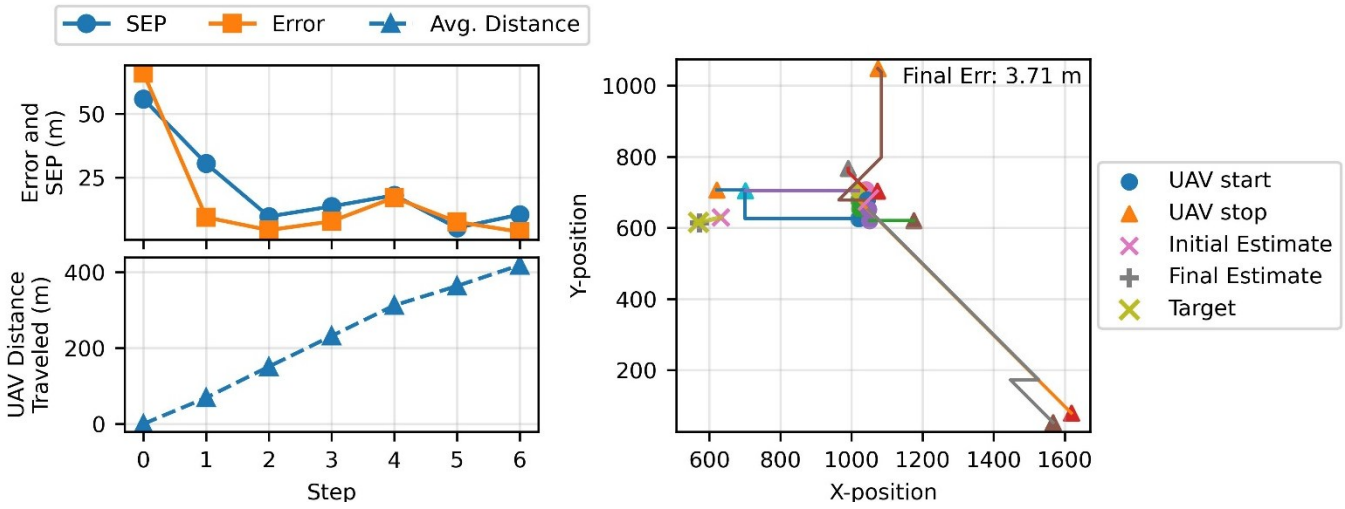


Figure 10. Visualization of EXata trial shows path strategy for localization convergence.

two or three repositioning steps, the system is more likely to realize a low-noise and low-scattering TDOA measurement and can rapidly converge on the emitter location. The trajectory shown in Figure 10 follows an immediate 50 m drop in error after the first repositioning step. This indicates that the UAV system started with an undesirable geometry, then moved towards an improved geometry nearly immediately while continuing to iterate and refine that geometry for subsequent snapshots. In the Python simulations, the environmental model always assumed multipath error, even in pure LoS cases, reducing simulation realism in comparison to the EXata digital twin evaluations. Therefore, the improved performance observed with EXata is expected and is representative of a realistic operating environment with the presence of more variable dynamics.

6. Conclusion and Summary

We presented a DQN-based cooperative UAV localization framework validated in both Python-based simulations and digital twin evaluations. In the initial simulations, the DQN achieved significant error reduction, and demonstrated further robustness in the more realistic EXata trials. With each intelligent repositioning step, the system improves its emitter localization accuracy and is able to converge on the emitter location estimate, typically with very low error and low uncertainty in very few snapshots. In all trials, RSS contributed as a contextual feature in the observation vector but did not dominate agent behavior, confirming that convergence was driven by SEP reduction rather than power maximization. With the performance validation in EXata, the system achieves low localization error in fewer than ten snapshots, even under multipath and adverse weather conditions. These results advance RL-based cooperative localization toward real-world deployment in aerospace and other applications. Although we focus on a passive RF emitter localization application here, DQN-based dynamic array repositioning can be applied to many other applications that depend upon autonomous UAS swarming.

Future work will pursue policy shaping for scalability to *N*-UAVs. In addition, the evaluated hardware will extend to directional antennas, with a straightforward incorporation of realistic antenna patterns into EXata. Further, we will expand the emitter dynamics to both nonstationary emitters and multiple emitter localization in contested RF environments.

Appendix

Table 3. Algorithm terms and parameters.

Term	Parameter
$s_i = (x_i, y_i, z_i)$	UAV location
$\widehat{E} = (\hat{x}_e, \hat{y}_e, \hat{z}_e)$	Emitter location estimate
Δt_{ij}	Time difference of arrival of emitter signal to sensors
c	Signal propagation speed
d_i	Euclidean distance of sensor to system origin
$\Lambda = \{\lambda_m\}_{m=1}^{M}$	Cloud of LOCA solutions λ_m
$SEP_{50} = q_{0.5}\{\|\lambda_m - \widehat{E}\|\}$	Spherical error probable of LOCA solution to emitter location estimate
$\sigma_{t,i}^2 = \frac{3R_i^2}{4\pi^2 f T_s B^2 r_0^2 SNR_0}$	Sensor range dependent variance
f	Operating frequency
R_i	Emitter-sensor range
T_s	Sensor integration time
B	Sensor bandwidth
r_0	Minimum sensor operational range
SNR_0	Sensor minimum signal to noise ratio
P	State error covariance
z_k	Weighted-fused LOCA position
Σ_z	Measurement covariance of z_k
ω_q	Variance-aware cloud weights
w_k, v_k	EKF process and measurement noise

REFERENCES

[1] C. Peters and M. A. Thornton, "Cooperative UAS Geolocation of Emitters with Multi-Sensor-Bounded Timing and Localization Error," Proc. IEEE Aerospace Conf., Big Sky, MT, USA, 2023, pp. 1–13, doi: 10.1109/AERO55745.2023.10116023.

[2] W. Wang, et al., "Optimal Configuration and Path Planning for UAV Swarms Using a Novel Localization Approach," Applied Sciences, vol. 8, no. 6, Art. 973, 2018, doi: 10.3390/app8060973.

[3] S. Aditya, et al., "A Survey on the Impact of Multipath on Wideband Time-of-Arrival-Based Localization," IEEE Signal Processing Magazine, vol. 35, no. 4, pp. 59–89, 2018, doi: 10.1109/MSP.2018.2818159.

[4] C. Peters and M. A. Thornton, "Reducing Emitter Localization Error in Urban Environments with Geometry Adaptive UAS Arrays," Proc. IEEE Systems Conf. (SysCon), Montréal, QC, Canada, 2025.

[5] V. Mnih, K. Kavukcuoglu, D. Silver. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015).

[6] A. Guerra, F. Guidi, D. Dardari, and P. M. Djurić, "Reinforcement Learning for Joint Detection and Mapping Using Dynamic UAV Networks," IEEE Trans. Aerospace Electron. Syst., vol. 60, no. 3, pp. 2586–2601, Jun. 2024, doi: 10.1109/TAES.2023.3300813.

[7] Y. Ding, Z. Yang, Q.-V. Pham, Y. Hu, Z. Zhang, and M. Shikh-Bahaei, "Distributed Machine Learning for UAV Swarms: Computing, Sensing, and Semantics," IEEE Internet Things J., vol. 11, no. 5, pp. 7447–7473, Mar. 2024, doi: 10.1109/JIOT.2023.3341307.

[8] R. O. Schmidt, "A New Approach to Geometry of Range Difference Location," in IEEE Transactions on Aerospace and Electronic Systems, vol. AES-8, no. 6, pp. 821-835, Nov. 1972.

[9] R. O. Schmidt, "Least Squares Range Difference Location," IEEE Transactions on Aerospace and Electronic Systems, vol. 32, no. 1, pp. 234-242, 1996.

[10] S. Wu, "Illegal radio station localization with UAV-based Q-learning," China Communications, vol. 15, no. 12, pp. 122–131, Dec. 2018, doi: 10.12676/j.cc.2018.12.010.

[11] Y. J. Chen, D. K. Chang, and C. Zhang, "Autonomous Tracking Using a Swarm of UAVs: A Constrained Multi-Agent Reinforcement Learning Approach," IEEE Trans. Veh. Technol., vol. 69, no. 11, pp. 13702–13717, Nov. 2020, doi: 10.1109/TVT.2020.3023733.

[12] M. Shurrab, R. Mizouni, S. Singh, and H. Otrok, "Reinforcement Learning Framework for UAV-Based Target Localization Applications," Internet of Things, vol. 23, p. 100867, 2023, doi: 10.1016/j.iot.2023.100867.

[13] D. L. Young, M. Bigham, M. Bradbury, E. Larson, and M. Thornton, "SMU-DDI Cyber Autonomy Range (CAR): Incorporation of Resiliency, Reliability, and Cyber Security in Autonomous Systems," in 2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, Oct 2022, pp. 1–5.

[14] X. Chen, C. Fu, and J. Huang, "A Deep Q-Network for Robotic Odor/Gas Source Localization: Modeling, Measurement and Comparative Study," Measurement, vol. 183, p. 109725, 2021, doi: 10.1016/j.measurement.2021.109725.

[15] X. Cheng, F. Shu, Y. Li, Z. Zhuang, D. Wu, and J. Wang, "Optimal Measurement of Drone Swarm in RSS-Based Passive Localization with Region Constraints," IEEE Open J. Veh. Technol., vol. 4, pp. 1–11, 2023, doi: 10.1109/OJVT.2022.3213866.

[16] Y. Dong, F. Li, C. Ma, C. He, and Z. J. Wang, "UAV-Based Dynamic Object Tracking with Radio Map," Proc. IEEE ICASSP, Seoul, South Korea, 2024, pp. 9166–9170.

[17] C. Yan, L. Fu, J. Zhang, and J. Wang, "A Comprehensive Survey on UAV Communication Channel Modeling," IEEE Access, vol. 7, pp. 107769–107792, 2019, doi: 10.1109/ACCESS.2019.2933173.

[18] D. W. Matolak and R. Sun, "Air–Ground Channel Characterization for Unmanned Aircraft Systems: The Near-Urban Environment," Proc. IEEE MILCOM, Tampa, FL, USA, 2015, pp. 1656–1660, doi: 10.1109/MILCOM.2015.7357682.

[19] W. E. Hoover and U. S., "Algorithms for confidence circles and ellipses," United States National Ocean Service Office of Charting and Geodetic Services, Tech. Rep., 1984, NOAA technical report NOS 107 C&GS 3. [Online]. https://repository.library.noaa.gov/view/noaa/23141.

[20] W. Khawaja, I. Guvenc, D. W. Matolak, U. -C. Fiebig and N. Schneckenburger, "A Survey of Air-to-Ground Propagation Channel Modeling for Unmanned Aerial Vehicles," in IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2361-2391, thirdquarter 2019, doi: 10.1109/COMST.2019.2915069.

[21] Radiocommunication Sector of International Telecommunication Union, "Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to

100 GHz", Rec. ITU-R P. 1411–12 ITU Recommendation, Aug. 2023.

[22] Y. T. Chan, K. C. Ho, "A simple and efficient estimator for hyperbolic location," Signal Processing, IEEE Transactions on, vol. 42, no. 8, pp. 1905-1915, 1994.

[23] D. Kim, J. Ha; K. You. "Adaptive extended Kalman filter based geolocation using TDOA/FDOA". Int. J. Control Autom. 2011, 4, 49–58.

[24] Keysight, "Network Digital Twin Ecosystem — EXata Product Description," Technical Overview, document no. 3122-1404, Keysight Technologies, Feb. 2024. [Online]. https://www.keysight.com/us/en/assets/3122-1404/technical-overviews/Network-Digital-Twin-Ecosystem.pdf.

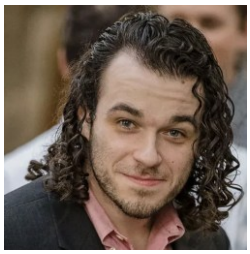
[25] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[26] Farama Foundation, "PettingZoo,", "Gym". [Online]: https://farama.org/ .

[27] Radiocommunication Sector of International Telecommunication Union. Recommendation ITU-R P.676-10: Attenuation by atmospheric gases 2013.

BIOGRAPHY

Christopher Peters *is currently pursuing his PhD with Southern Methodist University, and he is a Lead Systems Engineer with CACI where he develops RF architectures for airborne and maritime systems. Prior to CACI, he worked with L3Harris and Raytheon Technologies leading teams in manufacturing and testing of AESA radar systems and developing next-generation technologies for RF sensor systems. He received his M.S. in systems engineering from Johns Hopkins University, his M.S. in physics from the University of Mississippi, and his B.S. in physics and B.S. in mathematics from the University of Mississippi.*



Michael Watts *is currently pursuing his PhD with Southern Methodist University with a focus in Artificial Intelligence. Prior to Southern Methodist University, he worked with IBM working as an AI Consultant specializing in Generative AI system design and implementation in the telecom and energy sectors. He received his M.S. in Computer Science from Southern Methodist University and his B.S. in Computer Science from Southern Methodist University.*

Eric C. Larson *is an associate professor of computer science at Southern Methodist University. He is also the Interim Director of the Institute for Computational Biosciences and the Bobby B. Lyle Endowed Professor in Engineering Innovation. His research explores the interdisciplinary relationship of machine learning and signal/image processing with the fields of security, health, education, human-machine teaming, and ubiquitous computing—where he has secured over $12 million dollars in federal and corporate funding, including NSF, NIH, IES, DOE, ONR, USAFA, and others. He is a fellow of the Hunt Institute for Engineering Humanity, member of the Darwin Deason Institute for Cybersecurity, member of the SMU AT&T Center for Virtualization, and Member of the SMU Academy of Distinguished Teachers. Dr. Larson has published one textbook and disseminated his research in over 100 peer-reviewed conference and journal papers, garnering more than 6,500 citations and 9 best paper awards nominations. He received his Ph.D. from the University of Washington where he was an Intel Science and Technology fellow. At UW, he was co-advised by MacArthur Genius Fellow Shwetak Patel and IEEE Fellow Les Atlas. He also has an MS in Image and Signal Processing from Oklahoma State University, where he was advised by Damon Chandler.*

Mitchell A. (Mitch) Thornton *is currently the Cecil H. Green Chair of Engineering and Professor in the Department of Electrical and Computer Engineering at Southern Methodist University in Dallas, Texas. He also serves as the Executive Director of the Darwin Deason Institute for Cyber Security, a research unit at SMU. His prior industrial experience includes employment at Amoco Research Center, E-Systems, Inc (now L3Harris Communications, Inc.), and the Cyrix Corporation where he held a variety of engineering positions. His research areas include cyber security, applications of data science, signal processing, sensor-based systems in security, autonomous systems, and quantum informatics. He is an author or co-author of over 300 technical articles and five books and he is a named inventor on more than 20 patents and patents pending. He has performed sponsored research for several different government agencies and industrial organizations. He is a licensed professional engineer and he holds a general radiotelephone operator license with radar endorsement from the U.S. Federal Communications Commission. He received the PhD in computer engineering from SMU, MS in computer science from SMU, MS in electrical engineering from the University of Texas at Arlington, and BS in electrical engineering from Oklahoma State University.*