Information-Theoretic RL-Based Sensor Placement for Half-Plane AOA Localization

Seongyeol Park, Hanvit Kim, Jongchan Won, and Sunwoo Kim
Department of Electronics Engineering, Hanyang University, Seoul, South Korea
Email: {c18006, dante0813, wonjc71, remero}@hanyang.ac.kr

***Abstract*—This paper investigates the multi-sensor placement problem for angle-of-arrival (AOA)-based localization in a surveillance scenario over a half-plane region. In such settings, simple regular layouts often leave large areas with very high localization error. To address this issue, we formulate the placement problem as minimizing the spatially averaged position error bound (PEB) over the surveillance region. To solve this combinatorial optimization problem, we first employ a reinforcement learning (RL) policy with a PEB-based reward derived from the Cramér–Rao lower bound (CRLB). However, this purely data-driven approach does not explicitly capture the underlying sensing geometry and does not guarantee an optimal placement. Therefore, we propose a hybrid framework that initializes the layout using this RL policy and then refines it via a physics-aware greedy 1-swap search. Simulation results show that the proposed method achieves a lower spatially averaged PEB and reduces high-error lobes over the surveillance region compared with RL-based baselines.**

I. Introduction

Bearings-only localization using angle-of-arrival (AOA) measurements is widely used in radar, sonar, and electronic warfare (EW) systems [1], [2]. Because passive AOA sensors provide only bearing information and no direct range, localization accuracy is highly sensitive to the sensor–target geometry. In ideal symmetric layouts such as convex-hull geometries, sensors surround the target region and provide diverse viewing angles, which yields relatively uniform accuracy. In practical surveillance scenarios with unknown target positions, however, sensors must be deployed behind a front line and observe only an opposing half-plane [3], as illustrated in Fig. 1. In this half-plane configuration, distant targets are viewed from similar directions, which makes the Fisher information matrix ill-conditioned and leads to broad high-error bands in the Cramér–Rao lower bound (CRLB) over the surveillance region.

Therefore, classical geometry-based sensor placement rules are not directly applicable to asymmetric half-plane scenarios [4], [5]. To overcome these limitations, recent work has explored reinforcement learning (RL) to optimize sensor layouts based on performance-driven rewards in simulation environments [6], [7]. RL is well suited for such problems, as it can explore very large discrete placement spaces and handle irregular deployment constraints. However, standard model-free RL agents typically rely on stochastic exploration guided by scalar rewards, without explicitly exploiting the analytical sensor–target geometry or Fisher information structure. Consequently, RL-only designs may settle for suboptimal local layouts, leaving regions with high localization error.

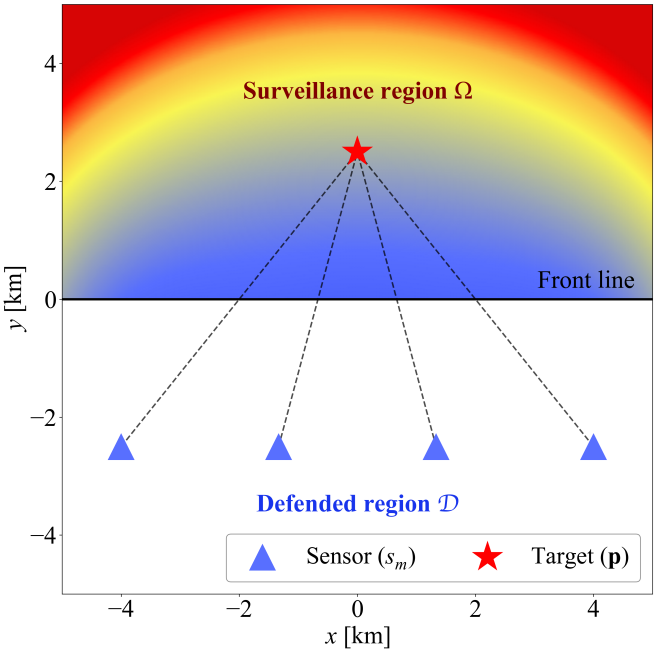


Fig. 1. Half-plane AOA surveillance geometry.

To address this issue, we formulate the half-plane sensor placement problem as minimizing a position error bound (PEB)-based cost over the surveillance region. The spatially averaged PEB serves as an error metric for the region, since it directly reflects localization accuracy. We then propose a hybrid framework in which a PEB-based RL policy first generates an initial layout, and a physics-aware greedy 1-swap search refines the RL-based layout by explicitly minimizing the same cost. This design combines RL's global exploration with physics-based local refinement and yields a lower spatially averaged PEB and fewer high-error regions than RL-based baselines.

II. System Model

A. Half-Plane Geometry and Discretization

As illustrated in Fig. 2, we consider a two-dimensional xy-plane in which the front line coincides with $y = 0$. The defended region $\mathcal{D}$ and the surveillance region Ω are defined as

$$\begin{aligned}\mathcal{D} &= \{(x,y) \mid -W/2 \le x \le W/2,\ -D \le y \le 0\},\\ \Omega &= \{(x,y) \mid -W/2 \le x \le W/2,\ 0 \le y \le d\},\end{aligned} \tag{1}$$

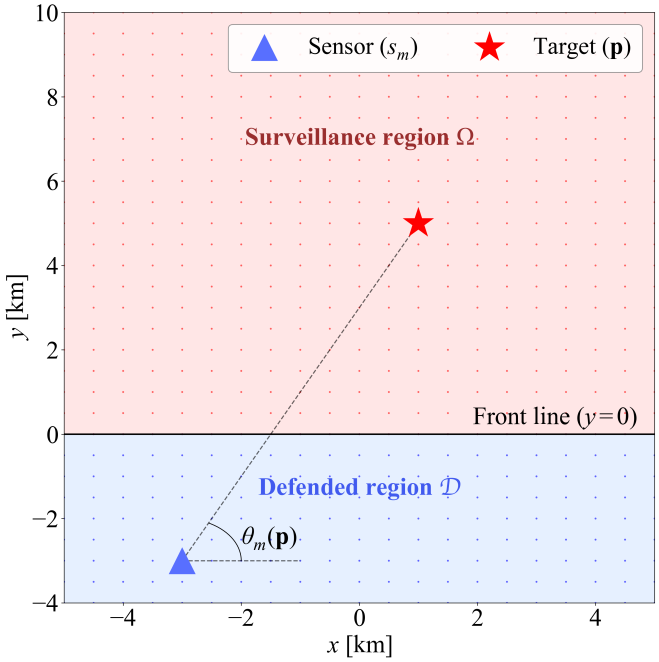


Fig. 2. Illustration of the half-plane sensor placement scenario with defended region $\mathcal{D}$ and surveillance region Ω.

where W, D, and d denote the front-line width, defended depth, and surveillance depth, respectively.

For numerical evaluation, both regions are discretized on a uniform grid with spacing $(\Delta x, \Delta y)$. For notational convenience, we reuse $\mathcal{D}$ and Ω to denote the sets of discrete grid points. Region $\mathcal{D}$ contains M candidate sensor locations $\{q_i\}_{i=1}^{M}$, and Ω contains L evaluation points $\{\mathbf{p}_\ell\}_{\ell=1}^{L}$. We denote by $\mathcal{S}_N$ the family of all N-sensor placements:

$$\mathcal{S}_N = \{S \subset \mathcal{D} : |S| = N\}, \tag{2}$$

where N is the number of sensors and $S = \{s_1, \ldots, s_N\}$ denotes their positions.

B. AOA Measurement Model

Let the m-th sensor be located at $s_m = (x_m, y_m) \in \mathcal{D}$ and the target position vector be denoted by $\mathbf{p} = [x\ y]^\top \in \Omega$. The bearing measured by sensor m is modeled as [2]

$$\theta_m(\mathbf{p}) = \tan^{-1}\left(\frac{y - y_m}{x - x_m}\right) + w_m, \tag{3}$$

where $\tan^{-1}(\cdot)$ denotes the arctangent, and $w_m \sim \mathcal{N}(0, \sigma_\theta^2)$ is zero-mean Gaussian noise, independent across sensors.

Define the relative coordinates between the target and sensor m as $\Delta x_m = x - x_m$ and $\Delta y_m = y - y_m$, and the squared range as $r_m^2 = \Delta x_m^2 + \Delta y_m^2$. The gradient of $\theta_m(\mathbf{p})$ with respect to the position vector $\mathbf{p}$ is given by [5]

$$\mathbf{g}_m(\mathbf{p}) = \frac{1}{r_m^2}\begin{bmatrix} -\Delta y_m \\ \Delta x_m \end{bmatrix}, \tag{4}$$

where $\mathbf{g}_m(\mathbf{p}) \in \mathbb{R}^2$ encodes the local sensitivity of the bearing measurement at sensor m with respect to the target position $\mathbf{p}$.

C. Regularized CRLB and Spatial Cost

Assuming independent measurements across sensors and the Gaussian noise model in Section II-B, the Fisher information matrix (FIM) at location $\mathbf{p} \in \Omega$ under placement S is expressed as

$$\mathbf{J}(\mathbf{p}; S) = \frac{1}{\sigma_\theta^2}\sum_{m=1}^{N} \mathbf{g}_m(\mathbf{p})\, \mathbf{g}_m^\top(\mathbf{p}). \tag{5}$$

For any unbiased estimator $\hat{\mathbf{p}}$, the error covariance satisfies $\mathrm{Cov}(\hat{\mathbf{p}} - \mathbf{p}) \succeq \mathbf{J}^{-1}(\mathbf{p}; S)$ [8].

In the half-plane geometry, $\mathbf{J}(\mathbf{p}; S)$ can become ill-conditioned when the target lies on or near the sensors' baseline, leading to very large error variances (blind zones). To ensure numerical stability during optimization and to avoid singularities, we apply diagonal loading and define the regularized CRLB covariance matrix as

$$\mathbf{C}(\mathbf{p}; S) = \left(\mathbf{J}(\mathbf{p}; S) + \delta \mathbf{I}\right)^{-1}, \tag{6}$$

where $\delta > 0$ is a small regularization parameter and $\mathbf{I}$ is the 2×2 identity matrix.

We define the position error bound (PEB) at point $\mathbf{p}$ as

$$\mathrm{PEB}(\mathbf{p}; S) = \sqrt{\mathrm{tr}\big(\mathbf{C}(\mathbf{p}; S)\big)}, \tag{7}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace. The PEB provides a lower bound on the position root-mean-square error (RMSE) at $\mathbf{p}$, since $\mathrm{tr}(\mathbf{C}(\mathbf{p}; S))$ is the sum of the position variances. This regularization keeps the bound well defined in nearly singular geometries, while still closely approximating the unregularized CRLB-based PEB in well-conditioned regions [9].

We use the following PEB-based cost function over the surveillance region:

$$J_{\mathrm{mean}}(S) = \frac{1}{L}\sum_{\mathbf{p}\in\Omega} \mathrm{PEB}^2(\mathbf{p}; S). \tag{8}$$

This quantity is a lower-bound surrogate for the mean-square error (MSE) of localization under a uniform target distribution over Ω, and heavily penalizes regions with very large errors, thereby suppressing blind zones [9], [10].

For performance reporting, we also define the spatially averaged PEB as

$$\overline{\mathrm{PEB}}(S) = \frac{1}{L}\sum_{\mathbf{p}\in\Omega} \mathrm{PEB}(\mathbf{p}; S), \tag{9}$$

representing the mean error bound.

III. Proposed Hybrid Framework

A. Problem Formulation

Given the candidate set $\mathcal{D}$ and sensor count N, we seek a placement S^* that minimizes $J_{\mathrm{mean}}(S)$:

$$S^*(N) = \arg\min_{S\in\mathcal{S}_N} J_{\mathrm{mean}}(S), \tag{10}$$

where $\mathcal{S}_N$ denotes the set of all valid N-sensor subsets. Since $J_{\mathrm{mean}}(S)$ is non-convex and the combinatorial search space $\binom{M}{N}$ is huge, exhaustive search is infeasible; instead, we adopt a hybrid approach that combines RL-based global exploration with physics-aware greedy refinement.

B. RL-Based Layout Initialization

We formulate the sensor placement problem as an N-step Markov decision process (MDP). At step t ($t = 0, \dots, N-1$), the state ξ_t encodes the current sensor set $S_t \subset \mathcal{D}$ with $S_0 = \emptyset$. The action a_t selects a new location $q_{j_t} \in \mathcal{D} \setminus S_t$, and the next sensor set is updated as

$$S_{t+1} = S_t \cup \{q_{j_t}\}. \tag{11}$$

After N steps, the episode terminates with the complete final placement S_N.

To match the objective in (10), we use a terminal reward

$$R(S_N) = -J_{\text{mean}}(S_N), \tag{12}$$

and set intermediate rewards to zero. We parameterize a stochastic softmax policy $\pi_\theta(a_t \mid \xi_t)$ over the remaining candidate sites and train it using a policy-gradient method (REINFORCE with a value baseline) to maximize the expected return $J_{\text{RL}}(\theta) = \mathbb{E}_{\pi_\theta}[R(S_N)]$. This formulation allows the agent to exploit global sensor–target geometric dependencies that are often missed by purely local greedy heuristics. At test time, we perform a greedy rollout by selecting the action with maximum probability under π_θ at each step, yielding the initial layout S_{RL}, which serves as $S^{(0)}$ for the refinement stage.

C. Physics-Aware Refinement: Greedy Search

Starting from $S^{(0)} = S_{\text{RL}}$, we refine the initial layout using a greedy 1-swap search to further reduce J_{mean}. The 1-swap neighborhood of a placement S is defined as

$$\mathcal{N}(S) = \{(S \setminus \{s_i\}) \cup \{q_j\} \mid s_i \in S,\ q_j \in \mathcal{D} \setminus S\}, \tag{13}$$

which preserves exactly N sensors within the defended region. At iteration k ($k = 0, 1, \dots$), we identify

$$\tilde{S} = \arg \min_{S' \in \mathcal{N}(S^{(k)})} J_{\text{mean}}(S'), \tag{14}$$

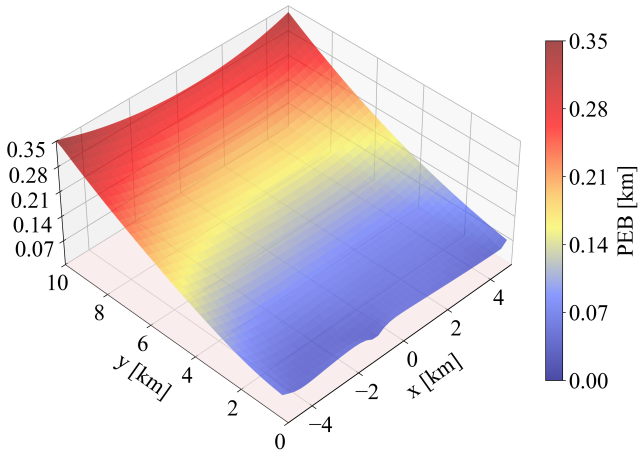
and update the layout according to

$$S^{(k+1)} = \begin{cases} \tilde{S}, & \text{if } J_{\text{mean}}(\tilde{S}) < J_{\text{mean}}(S^{(k)}), \\ S^{(k)}, & \text{otherwise.} \end{cases} \tag{15}$$

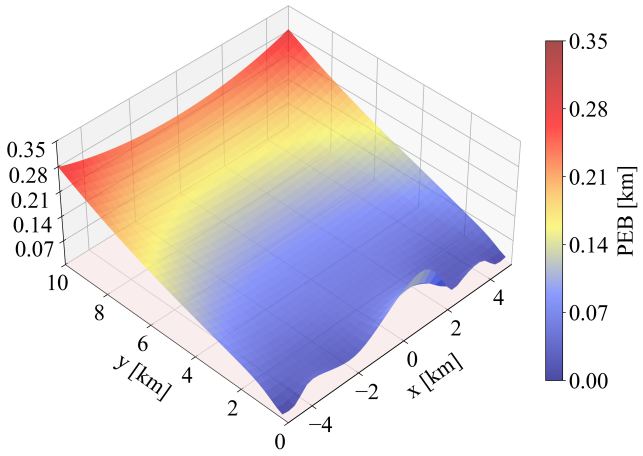
The search stops when $S^{(k+1)} = S^{(k)}$, at which point $S^{(k)}$ is a 1-swap local minimum of J_{mean}. Since J_{mean} is derived directly from the CRLB and the AOA measurement model, each 1-swap explicitly evaluates the Fisher information structure, and thus, this refinement acts as a physics-aware local search.

TABLE I Simulation and training parameters.

Parameter	Description	Value
W, D	Defended region size	10×4 km
d	Surveillance depth	10 km
$\Delta x, \Delta y$	Grid spacing in $\mathcal{D}$ and Ω	0.5 km
σ_θ	AOA noise std.	1°
δ	Regularization param.	10^{-6}
N (Figs. 3, 4)	Sensor counts (field / sweep)	6 / $4 \le N \le 8$
RL params	Episodes / early-stopping patience	$10^3/200$



(a) RL-based baseline



(b) Proposed method

Fig. 3. Comparison of PEB maps for $N = 6$ sensors.

IV. Simulation Results

A. Simulation Setup

In our experiments, both the defended region $\mathcal{D}$ and the surveillance region Ω are discretized on a uniform grid with spacing 0.5 km. A policy-gradient RL agent minimizes J_{mean} to obtain the initial layout S_{RL}, which is subsequently improved by the physics-aware greedy refinement. The detailed simulation settings, including geometric dimensions and training hyperparameters, are summarized in Table I.

B. PEB Field Analysis

Fig. 3 visualizes the PEB fields, where the color gradient indicates the PEB magnitude over the surveillance region. The baseline configuration in Fig. 3(a) exhibits high estimation errors, especially for targets far from the front line. In contrast, the proposed method in Fig. 3(b) effectively mitigates these error peaks, yielding a more uniform low-error distribution. The corresponding spatial averages, $\overline{\text{PEB}}$ and J_{mean}, serve as our performance metrics and are summarized in Table II.

TABLE II Average performance over 100 MC runs.

Metric	Baseline	Proposed	Improv.	Ref.
$\overline{\text{PEB}}$ [km]	0.145	0.124	14.55%	Eq. (9)
J_{mean} [km^2]	0.0258	0.0188	26.98%	Eq. (8)

Table II summarizes these gains over 100 Monte Carlo (MC) runs. The proposed method reduces $\overline{\text{PEB}}$ and J_{mean} by approximately 15% and 27%, respectively. Moreover, when we examine the PEB at each surveillance grid point, the refined placement achieves a smaller local PEB at about 92% of the points, with the remaining $\approx$ 8% of non-improved points confined to a narrow strip just above the front line.

Since each sensor's Fisher information contribution decays approximately as $1/r^2$ with the sensor–target range r, the refinement tends to pull the sensors toward the front line. Consequently, a thin band near the front line appears where the bearings from different sensors become nearly parallel, making the FIM nearly singular and the PEB in this band slightly higher than in the baseline. However, this localized increase in PEB is minor compared to the substantial reduction achieved over the rest of the surveillance region. In half-plane surveillance scenarios, the primary objective is to reduce large error lobes over the surveillance area, so it is more beneficial for the spatial average cost J_{mean} to suppress these wide-area errors than to preserve marginal gains in this narrow strip.

C. Impact of Sensor Count

To investigate the robustness of the proposed framework, we vary the number of sensors N from 4 to 8. For each N, we generate an RL-based baseline, apply the same 1-swap refinement, and then compute $\overline{\text{PEB}}$ over Ω. The resulting averaged values for both methods are shown in Fig. 4. As expected, increasing N improves accuracy for both methods due to the increased information gain. However, the proposed hybrid framework achieves a lower $\overline{\text{PEB}}$ than the RL-based baseline for all tested N. The performance gap remains noticeable even as N increases and the RL baseline improves, indicating that the physics-aware refinement provides a consistent gain on top of the learned policy.

V. Conclusion

This paper addressed the multi-sensor placement problem for AOA-based localization in a half-plane surveillance scenario. We showed that an RL policy trained with a PEB-based reward can still leave regions with large localization errors. To overcome this limitation, we proposed a two-stage hybrid framework that minimizes a PEB-based cost by combining a PEB-driven RL policy with physics-aware greedy 1-swap refinement. By leveraging RL for global exploration and CRLB-based local search for geometry-aware adjustment, the method effectively optimizes the sensor placement and mitigates blind zones. Simulation results show that this physics-aware strategy consistently reduces the PEB-based cost compared with RL-based baselines and yields more robust localization performance across various sensor counts.

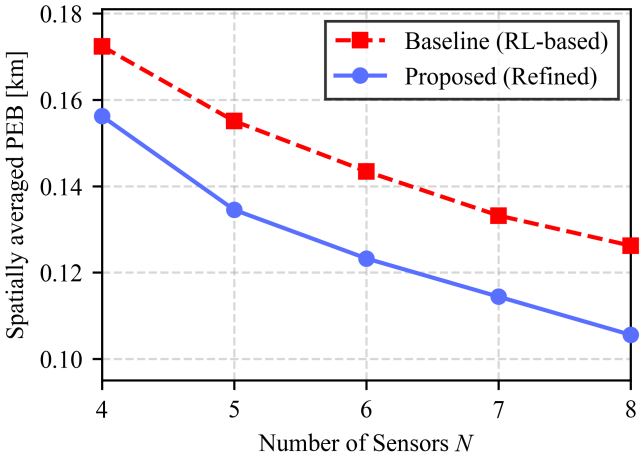


Fig. 4. Average PEB versus sensor count N (averaged over 100 MC runs).

Future work will extend the current grid-based formulation to continuous-space optimization and develop adaptive placement strategies that account for front-line width, surveillance depth, and obstacle-induced non-line-of-sight (NLOS) conditions. We will also investigate how these optimized static layouts can serve as robust baselines for dynamic tactical environments, including scenarios with maneuvering targets or time-varying sensor availability. Ultimately, this analysis will provide practical guidelines for resource allocation and deployment under realistic operational constraints.

Acknowledgment

This work was supported by Korea Research Institute for Defense Technology (KRIT)-Grant funded by Defense Acquisition Program Administration (DAPA) (KRIT-CT-24-004).

References

[1] S. C. Nardone, A. G. Lindgren, and K. F. Gong, "Fundamental properties and performance of conventional bearings-only target motion analysis," *IEEE Trans. Autom. Control*, vol. 29, no. 9, pp. 775–787, 1984.

[2] K. Doğançay, "Bearings-only target localization using total least squares," *Signal Process.*, vol. 85, no. 8, pp. 1695–1710, 2005.

[3] S. Kumar, T. H. Lai, and A. Arora, "Barrier coverage with wireless sensors," in *Proc. ACM MobiCom*, 2005, pp. 284–298.

[4] S. Xu and K. Doğançay, "Optimal sensor placement for 3-D angle-of-arrival target localization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1196–1211, Jun. 2017.

[5] A. N. Bishop, B. Fidan, B. D. O. Anderson, K. Doğançay, and P. N. Pathirana, "Optimality analysis of sensor–target localization geometries," *Automatica*, vol. 46, no. 3, pp. 479–492, 2010.

[6] J. A. Grant, A. Boukouvalas, R.-R. Griffiths, D. S. Leslie, S. Vakili, and E. Muñoz de Cote, "Adaptive sensor placement for continuous spaces," in *Proc. ICML*, 2019, pp. 2385–2393.

[7] R. Li *et al.*, "Flow field reconstruction with sensor placement policy learning," in *Proc. NeurIPS*, 2025.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall, 1993.

[9] M. Hamdollahzadeh, R. Amiri, and F. Behnia, "Optimal sensor placement for multi-source AOA localisation with distance-dependent noise model," *IET Radar, Sonar & Navigation*, vol. 13, no. 6, pp. 881–891, 2019.

[10] S. Xu, "Optimal Sensor Placement for Target Localization Using Hybrid RSS, AOA and TOA Measurements," *IEEE Commun. Lett.*, vol. 24, no. 9, pp. 1966–1970, Sep. 2020.