

Decentralized Multi-Robot Exploration with Probabilistic Peer Intent and Multi-hop Plan Propagation

Saurbh Singh Jamwal, Nived Chebrolu, Shivaram Kalyanakrishnan
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
 saurbh, nived, shivaram@cse.iitb.ac.in

Abstract—Efficient coordination under limited communication remains a key challenge in decentralized multi-robot exploration. While centralized approaches benefit from global information sharing, they are often impractical in large-scale or communication-constrained environments. Existing Monte Carlo Tree Search (MCTS)-based approaches, such as Decentralized Monte Carlo Exploration (DMCE), enable decentralized planning by taking peer intent into account. This peer intent is obtained by communicating sequences of planned waypoints with robots within direct communication range. In this work, we extend this idea by introducing Probabilistic Peer Intent (PPI), which converts peer trajectories into a continuous spatial representation of predicted intent and incorporates it into local MCTS action evaluation. We additionally study the effects of sharing peer intent beyond direct communication range by propagating plans over multiple hops. Experiments across multiple simulated environments and team sizes show that PPI and Multi-hop propagation can each improve decentralized exploration, with their relative benefits depending on environment structure and team size. We also demonstrate the real-world deployment of our method on three robots operating in different environment types.

Index Terms—Multi-Robot Exploration, Decentralized Coordination, Monte Carlo Tree Search, Peer Intent, Multi-hop Communication

I. INTRODUCTION

Autonomous exploration with multi-robot systems is a fundamental capability for time-critical applications such as search and rescue [1], subterranean mapping [2], and planetary exploration [3]. Decentralized coordination is particularly attractive in these settings due to its robustness to single points of failure and its ability to operate under limited and intermittent communication. A central challenge is to achieve efficient map coverage while limiting redundant sensing and traversal. At the same time, robots must maintain sufficient awareness of the team’s exploration progress.

Recent approaches have moved beyond reactive frontier-based heuristics towards predictive planning methods that reason over future exploration actions [4], [5]. Decentralized Monte Carlo Tree Search (MCTS) [4] provides one such framework, allowing each robot to evaluate candidate trajectories through simulated sensing and evaluate information gain while incorporating communicated peer plans. Nevertheless, we identify two limitations when this framework is applied under restricted communication. First, representing peer intent

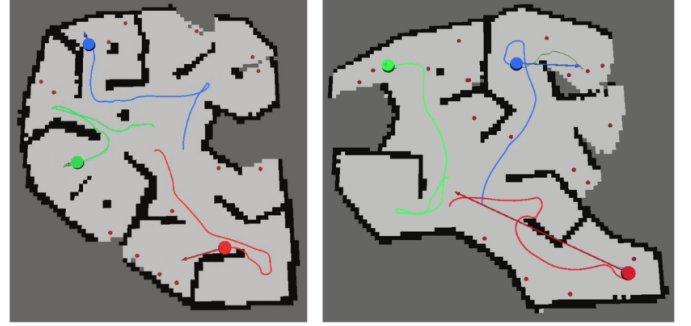


Fig. 1. **Multi-robot exploration in two real-world indoor arenas.** Online occupancy maps showing the regions explored during execution, together with the trajectories traced by the three robots.

only as sequences of planned waypoints does not explicitly capture how a peer’s intended motion should influence nearby candidate paths. As a result, robots may independently select similar high-value regions despite having access to peer plans. Second, direct plan exchange is constrained by the instantaneous communication graph, limiting planning awareness as robots disperse [4], [6].

We address these limitations through two complementary extensions to decentralized MCTS. **(i) Probabilistic Peer Intent (PPI)** converts other robots trajectories into a continuous spatial representation of predicted peer intent and incorporates this information into local action evaluation. **(ii) Multi-hop (MH) peer-plan propagation** relays trajectories from agents beyond the instantaneous one-hop neighborhood without requiring global connectivity or a central coordinator. This augments the planning information available to each robot. PPI and Multi-hop propagation influence coordination through two distinct mechanisms. PPI modifies local action evaluation, whereas Multi-hop propagation changes the peer information available for distributed planning.

We study their individual and combined effects across team sizes and environments ranging from open or branching spaces to cluttered and corridor-constrained tunnel environments. Our experiments show that both mechanisms can improve exploration individually, but their benefits are not uniformly additive. In constrained environments, increased peer-plan availability can also increase spatial intent exposure along

shared passages, negatively affecting the overall exploration goal. This highlights an important distinction between extending the reach of coordination information and determining how that information should influence local planning.

The primary contributions of this paper are:

- 1) We introduce Probabilistic Peer Intent (PPI), which converts communicated peer trajectories into a continuous spatial representation of predicted peer intent and incorporates it directly into the decentralized MCTS exploration objective.
- 2) We introduce Multi-hop (MH) plan propagation, extending peer plan availability beyond directly communicating robots under range-limited and time-varying connectivity.
- 3) We systematically evaluate PPI and MH across team sizes and environment structures, showing that both can improve exploration individually while their combination is topology dependent and not uniformly additive.
- 4) We validate the proposed framework on three robots in previously unknown indoor environments, demonstrating the physical deployment of the same decentralized planning, PPI, and Multi-hop mechanisms used in simulation.

The remainder of the paper is organized as follows. Section II reviews related work, Section III formulates the decentralized exploration problem, and Section IV presents PPI and Multi-hop propagation. Section V reports simulation and hardware results, followed by conclusion and future directions in Section VI.

II. RELATED WORK

Frontier-based exploration introduced by Yamauchi [7] directs robots toward boundaries between known and unknown space and remains a widely used exploration strategy. Sampling-based methods, such as Multi-RRT exploration by Umari et al. [8], instead sample trajectories through the environment rather than selecting only explicit frontier targets. In multi-robot settings, however, selecting exploration goals without coordination can lead to redundant coverage.

Several works extend exploration to multi-robot systems through explicit coordination. Yu et al. [9] propose SMMR-Explore, a submap-based framework combined with a multi-target potential field method to distribute robots across exploration goals. More broadly, multi-robot exploration systems commonly coordinate through target assignment, utility functions, or local interactions [10]. While these mechanisms can reduce redundant exploration, they primarily coordinate the selection or allocation of exploration targets rather than explicitly reasoning over the future trajectories of other robots.

Planning-based approaches explicitly reason over future robot actions. Best et al. [4] introduced Decentralized Monte Carlo Tree Search (Dec-MCTS), where agents coordinate by exchanging distributions over candidate plans, primarily for active perception tasks. Bone et al. [5] extended this idea to multi-robot exploration through Decentralized Monte Carlo Exploration (DMCE), where each robot performs local MCTS

and replays communicated peer trajectories in simulation to account for their expected exploration. However, coordination remains limited to directly communicated trajectories, without explicitly representing spatial intent around the predicted paths. We address these limitations separately: PPI introduces a continuous spatial representation of available peer trajectories, while Multi-hop propagation extends peer-plan availability beyond direct neighbors.

Communication-constrained coordination has also been studied independently of exploration planning. MOCHA [6] uses gossip-based Multi-hop communication to relay information beyond direct contacts, but operates primarily at the communication layer rather than integrating propagated information into exploration planning. Bramblett and Bezzo [11] address communication loss through epistemic belief updates and task allocation, while other approaches reduce communication requirements through low-bandwidth position sharing [12] or wireless signal sensing [13]. In contrast, our Multi-hop mechanism propagates planned trajectories and makes non-local peer plans directly available to the decentralized exploration planner.

Recent work has also explored learning-based approaches for multi-robot coordination. MARVEL [14] uses multi-agent reinforcement learning to learn decentralized exploration policies, while Meng et al. [15] employ graph neural networks to generate coordinated multi-agent behaviors. These approaches learn coordination policies from data, whereas our work focuses on model-based online planning. Specifically, we augment decentralized MCTS with an explicit spatial representation of communicated peer intent, allowing its influence on planning decisions to be directly controlled and analyzed without policy training.

Overall, existing approaches address complementary aspects of multi-robot coordination, including exploration-goal allocation, future-plan reasoning, and communication-constrained operation. Building on decentralized MCTS, we study two distinct extensions: PPI provides a continuous spatial representation of communicated peer trajectories for local planning, while Multi-hop propagation extends the availability of peer plans beyond direct neighbors. This separation allows us to examine how the representation and reach of coordination information independently affect multi-robot exploration.

III. PROBLEM FORMULATION

A. Decentralized Multi-Robot Exploration as an MDP

We consider a team of N robots exploring an initially unknown planar environment represented by an occupancy-grid map. Each robot maintains a local map from its sensor observations and makes exploration decisions using locally available information. For robot i , we formulate the finite-horizon exploration problem as an MDP

$$\mathcal{M}^i = (\mathcal{S}^i, \mathcal{A}^i, T^i, R^i). \quad (1)$$

The state $s_k^i \in \mathcal{S}^i$ consists of the robot position $p_k^i = (x_k^i, y_k^i)$ and its current local occupancy map M_k^i . At the

exploration-planning level, robot orientation and motion dynamics are abstracted away, and actions specify feasible spatial waypoints. The displacement action space is continuous in direction, with each action specifying a fixed-length motion from the current position, while frontier locations provide additional goal-directed actions. An action is feasible if the resulting waypoint lies within the map and can be reached through known free space.

The transition function T^i describes the evolution from the current state to the state associated with the selected waypoint, together with the corresponding update of the robot's local map from new observations. A sequence of actions over a finite horizon H results in a trajectory

$$\tau^i = (p_0^i, p_1^i, \dots, p_H^i). \quad (2)$$

The reward R^i quantifies the exploration utility of a trajectory. The objective for each robot is to select a feasible trajectory that maximizes this reward,

$$\tau^{i*} = \arg \max_{\tau^i} R^i(\tau^i). \quad (3)$$

Since decision making is decentralized, each robot operates using its local map and information received from other robots. Under range-limited and intermittent communication, this information may be incomplete, which can lead to different robots exploring overlapping regions.

B. Assumptions

We assume that all robots are localized in a common reference frame and that the localization error remains sufficiently small for exchanged positions and trajectories to be interpreted consistently, as in prior decentralized exploration systems [12], [16]. All explorable regions are assumed to be reachable, while low-level collision avoidance is handled independently of the high-level exploration planner. Communication is opportunistic and range-limited, resulting in intermittent local connectivity without requiring a globally connected communication network [4], [6].

IV. METHODOLOGY

A. Baseline: Decentralized MCTS-Based Exploration

We build upon Decentralized Monte Carlo Exploration (DMCE) [5], where each robot independently performs MCTS to select exploration actions using its locally available map and peer information. Coordination is introduced by exchanging planned trajectories between directly communicating robots. During planning, received peer trajectories are sampled and replayed through simulated sensing, updating the predicted map to account for regions that peers are expected to explore before evaluating the local robot's actions.

We retain the underlying DMCE planning procedure and introduce two extensions to peer plan coordination, detailed in the following subsections: (i) **Probabilistic Peer Intent (PPI)**, which modifies how available peer plans influence local action evaluation; and (ii) **Multi-hop propagation (MH)**, which extends peer plan availability beyond direct communication

neighbors. In this work, we evaluate peer-intent representation and information reach separately and in combination.

B. Probabilistic Peer Intent (PPI)

DMCE incorporates a communicated peer trajectory by replaying the peer's predicted motion and sensing into the simulated map, thereby accounting for regions that the peer is expected to explore. However, this mechanism does not explicitly quantify the spatial interaction between a local candidate trajectory and the peer's predicted path. We therefore introduce Probabilistic Peer Intent (PPI), which transforms each available peer trajectory into a continuous spatial field. As illustrated in Fig. 2(a), trajectories passing close to predicted peer motion receive greater exposure, while spatially separated trajectories receive weaker influence.

For each peer j , we retain its most recent valid trajectory and discard plans older than a freshness threshold T_c . Consecutive waypoints are linearly interpolated with spacing Δ_p to provide spatial support independent of the original waypoint spacing. Let $\mathbf{q}_{k,l}^j \in \mathbb{R}^2$ denote the l -th interpolated support point between future waypoints k and $k+1$ of peer j . Its influence at location $\mathbf{x}$ is

$$P_{k,l}^j(\mathbf{x}) = \exp(-\lambda_s k) \exp\left(-\frac{\|\mathbf{x} - \mathbf{q}_{k,l}^j\|^2}{2\sigma^2}\right), \quad (4)$$

where σ controls the spatial extent of peer intent and λ_s discounts positions farther along the communicated trajectory. The PPI field available to robot i is then

$$P_i(\mathbf{x}) = \max_{j,k,l} P_{k,l}^j(\mathbf{x}), \quad P_i(\mathbf{x}) \in [0, 1]. \quad (5)$$

The Gaussian kernel provides smoothly decreasing influence around predicted paths, while future-step decay gives greater importance to near-term peer motion. Max aggregation retains the strongest predicted intent at each location without increasing the field magnitude due to dense interpolation or overlapping peer trajectories. We use $\lambda_s = 0.02$, $T_c = 5.0$ s, and $\Delta_p = 0.10$ m throughout the experiments.

C. PPI-based Reward Integration

The underlying DMCE planner evaluates each MCTS rollout according to the information expected to be acquired through simulated sensing. Let M_0^i and M_k^i denote the simulated occupancy maps at the beginning of a rollout and after step k , respectively. The explored-map value is

$$V(M_k^i) = 1 - H_{\text{rel}}(M_k^i), \quad (6)$$

where $H_{\text{rel}}(\cdot)$ is the relative map entropy. The resulting information gain is

$$\text{IG}(s_k^i) = V(M_k^i) - V(M_0^i). \quad (7)$$

PPI augments this rollout evaluation by penalizing candidate trajectories that traverse regions of high predicted peer intent. For a candidate trajectory τ^i , cumulative peer-intent exposure is defined as

$$C_{\text{PPI}}(\tau^i) = \sum_{k=1}^H P_i(\mathbf{p}_k^i), \quad (8)$$

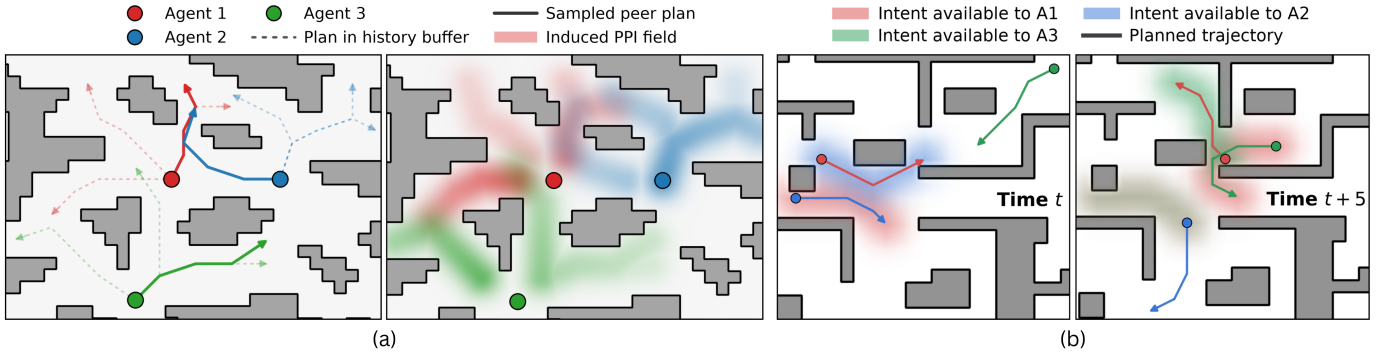


Fig. 2. **Overview of the proposed coordination mechanisms.** (a) Probabilistic Peer Intent (PPI), where communicated peer trajectories are transformed into spatial intent fields used during local planning. (b) Multi-hop (MH) plan propagation, which extends peer-plan information beyond direct communication neighbors by relaying fresh peer plans while preserving their source identity and generation time.

where $\mathbf{p}_k^i$ is the robot position at rollout step k . The resulting rollout objective becomes

$$R(\tau^i) = \text{IG}(s_H^i) - \lambda_{\text{PPI}} C_{\text{PPI}}(\tau^i), \quad (9)$$

where λ_{PPI} balances expected information gain against exposure to predicted peer intent.

Only the local robot's candidate trajectory is evaluated using PPI. Communicated peer trajectories continue to serve their original role in DMCE by predicting future map observations during rollout simulation, while the PPI penalty evaluates how the local trajectory relates to the resulting peer-intent field. Thus, PPI complements the original peer-plan reasoning by modifying the rollout objective rather than the underlying map prediction process.

D. Multi-hop Information Propagation

The original DMCE framework exchanges planned trajectories only between robots within direct communication range, limiting peer-plan availability to one-hop neighbors. Our Multi-hop (MH) extension allows each robot to forward both its own trajectory and previously received peer plans, enabling information to propagate across successive communication encounters as

$$\pi_j \rightarrow r_1 \rightarrow \dots \rightarrow r_m \rightarrow r_i, \quad (10)$$

where π_j is generated by robot r_j , and each relay preserves the original source identity and generation timestamp.

For each source robot, only the most recent received plan is retained. Because relayed plans keep their original timestamps, stale information is naturally discarded using the same timeout T_c defined in Section IV-B. Figure 2(b) illustrates this process: a robot can carry a previously received peer plan into a new communication neighborhood and relay it to a robot that never directly communicated with the original source. MH therefore extends peer-plan availability without changing the underlying MCTS planning or reward formulation.

V. RESULTS AND DISCUSSION

A. Experimental Setup

We evaluate our approach in simulation across four environments: *Open*, *Forest*, *Urban*, and *Tunnel*. *Open* is evaluated with a team size of $N = 5$, while the remaining environments use $N = \{5, 8, 12\}$. As visualized in Fig. 3, these environments range from open spaces to more branched and corridor-like layouts. We evaluate both the effects of PPI and Multi-hop propagation (MH) strategies individually as well as jointly. $\lambda_{\text{PPI}} = 1$ and $\sigma = 1$, with an 8 m sensor range for mapping and 15 m line-of-sight communication range are used throughout the experiments. Our primary evaluation metrics are coverage time (T90 and T95) and lifetime overlap, i.e., the fraction of the total explored area that is redundantly mapped by multiple robots (i.e., at least by two robots). Methods that have a lower time-to-coverage and lifetime overlap are considered better. Results are reported as the mean and standard deviation across 10 independent runs across methods.

B. Performance Analysis

In this section, we compare our proposed extensions (PPI and MH map propagation) against exploration methods including uncoordinated, sampling-based, decentralized, globally connected, and learned approaches. **MCTS (Uncoord.)** performs independent Monte Carlo tree search without peer-plan coordination, **RRT** [8] provides a sampling-based baseline, and **MMPF** [9] uses multi-robot potential fields. **DMCE** [5] is our primary decentralized baseline and operates under the same 15 m line-of-sight communication constraint as our extensions, while **DMCE-Global** provides a reference with globally available peer plans. We compare against **MARVEL** [14] as a learned multi-robot exploration baseline.

Table I compares the methods using $N = 5$ robots. Relative to DMCE, our proposed extensions exhibit environment-dependent trade-offs between coverage time and redundant exploration. PPI substantially reduces overlap in open environments and improves coverage time in forest environments. We observe that the MH strategy provides a similar improvement in the forest environment. By combining both PPI and MH, we

TABLE I

COMPARISON WITH OTHER METHODS ACROSS ENVIRONMENTS USING 5 ROBOTS, AVERAGED OVER 10 INDEPENDENT RUNS. LOWER IS BETTER. RUNS ARE LIMITED TO 450 s; “—” INDICATES THAT THE CORRESPONDING COVERAGE THRESHOLD WAS NOT REACHED IN ALL 10 RUNS.

Method	Open			Forest			Urban			Tunnel		
	T90 ↓	T95 ↓	Overlap ↓	T90 ↓	T95 ↓	Overlap ↓	T90 ↓	T95 ↓	Overlap ↓	T90 ↓	T95 ↓	Overlap ↓
MCTS (Uncoord.)	150.9	177.0	49.46	94.9	111.4	60.41	162.0	178.6	58.64	97.7	107.5	56.03
RRT	183.1	231.2	20.88	87.3	130.1	40.50	—	—	44.48	317.9	353.5	35.38
MMPF	—	—	48.14	107.2	123.4	54.32	—	—	69.58	—	—	70.11
DMCE	132.1	150.9	32.59	92.0	117.0	49.95	158.9	169.4	53.99	85.5	100.4	50.77
DMCE-Global	89.7	107.8	35.65	69.4	81.3	51.85	132.3	157.7	49.45	70.5	79.6	56.75
MARVEL	228.5	288.8	9.42	134.3	154.6	20.37	—	—	33.23	—	—	20.57
DMCE +PPI	132.9	161.3	26.22	73.3	94.5	47.64	170.3	179.5	53.42	95.7	106.4	48.92
DMCE +MH	121.1	154.3	38.74	71.2	93.2	48.58	154.1	170.7	58.51	95.7	103.2	49.74
DMCE +PPI +MH	139.1	167.4	22.62	79.0	95.1	44.08	159.0	177.0	51.26	109.7	117.4	49.72

TABLE II

COMPONENT ANALYSIS OF PPI AND MULTI-HOP PROPAGATION ACROSS TEAM SIZES. LOWER IS BETTER.

Robots	Method	Forest		Urban		Tunnel	
		T95 ↓	Overlap ↓	T95 ↓	Overlap ↓	T95 ↓	Overlap ↓
5	DMCE	117.0±30.6	49.95±6.75	169.4±19.9	53.99±6.53	100.4±16.1	50.77±4.96
	+PPI	94.5±18.8	47.64±8.30	179.5±16.8	53.42±8.41	106.4±11.3	48.92±7.23
	+MH	93.2±17.6	48.58±7.47	170.7±13.7	58.51±5.76	103.2±13.4	49.74±8.87
	+PPI+MH	95.1±11.7	44.08±8.06	177.0±16.9	51.26±10.80	117.4±10.5	49.72±6.59
8	DMCE	63.8±9.2	62.39±7.08	154.9±20.8	70.26±5.65	91.7±5.2	60.46±4.86
	+PPI	64.9±7.1	55.91±4.33	180.0±12.3	73.67±4.08	97.3±8.7	64.05±2.86
	+MH	67.7±8.1	64.93±3.44	153.0±18.2	73.35±2.82	94.6±7.7	65.97±5.34
	+PPI+MH	63.9±7.1	57.11±6.26	177.8±24.5	73.42±3.19	104.8±7.6	62.48±4.53
12	DMCE	59.5±4.2	74.10±3.13	171.6±15.3	79.73±3.53	106.4±9.5	74.52±4.09
	+PPI	55.0±3.9	69.10±2.68	168.3±16.4	83.15±3.62	99.4±9.3	73.19±3.32
	+MH	55.4±4.6	73.16±2.14	166.0±18.6	83.69±2.71	89.6±8.3	74.31±3.38
	+PPI+MH	62.0±4.6	70.42±3.67	183.8±16.7	82.90±1.45	108.7±12.2	75.42±3.33

observe the lowest overlap among the DMCE variants in open, forest, and urban environments, however, the coverage time does not improve consistently. We note that DMCE-Global (which has access to all robot plans at all times) consistently performs well across all environments, illustrating the benefit of unrestricted peer-plan availability. In the next sections, we examine the individual contributions, scaling behavior, and interaction of PPI and MH.

C. Analysis of PPI and Multi-hop Strategies

Table II isolates the effects of PPI and Multi-hop propagation across team sizes. In Forest, both strategies consistently improve exploration time. For $N = 5$, PPI and MH reduce T95 from 117.0 s to 94.5 s and 93.2 s, respectively, while PPI+MH achieves the lowest overlap. Similar individual gains persist with $N = 12$, which indicates the benefits of both spatial intent reasoning and extended peer-plan availability as the team disperses in different directions.

The effects are more topology-dependent in Urban and Tunnel. At $N = 12$, uncoordinated MCTS is highly competitive in coverage time, potentially because of higher robot density relative to the available branches and passages. This can induce spatial separation even without explicit coordination. In such settings, environment geometry itself can partially distribute robots across distinct routes, reducing the immediate advantage of explicit peer-plan reasoning. Among coordinated methods, both PPI and MH improve T95 over DMCE, with MH providing the largest reduction in Tunnel (106.4 to

89.6 s), also outperforming DMCE-Global at high coverage in Fig. 3. This suggests that selectively extending peer-plan availability can be particularly useful once robots disperse and direct communication becomes sparse. The local map knowledge results in Fig. 3 further show that exploration time and distributed map knowledge need not improve together. Importantly, PPI+MH can underperform compared to either component individually, particularly in constrained environments. We examine this non-additive interaction in Sec. V-D.

D. Interaction Between PPI and Multi-hop Propagation

PPI and MH can each improve exploration, but their benefits are not always additive. At $N = 12$ in Tunnel, PPI and MH reduce T_{95} from 106.4 s to 99.4 s and 89.6 s, respectively, whereas PPI+MH increases it to 108.7 s. Similar non-additive behavior is observed across the evaluated environments.

To diagnose this interaction, we tested successful MCTS iterations to measure fresh peer plans and PPI exposure along simulated trajectories. In Tunnel, MH increases the mean number of fresh peer plans from 2.30 to 3.92, mean PPI exposure from 0.320 to 0.595, and the fraction of samples with non-zero exposure from 34.8% to 51.8%. Neither reducing the PPI penalty nor restricting PPI construction to directly received plans recovered the combined configuration. This indicates that the degradation is not explained solely by penalty magnitude or by applying PPI to relayed plans.

These results point towards a limitation of purely spatial intent reasoning. As illustrated in Fig. 4, Euclidean proximity

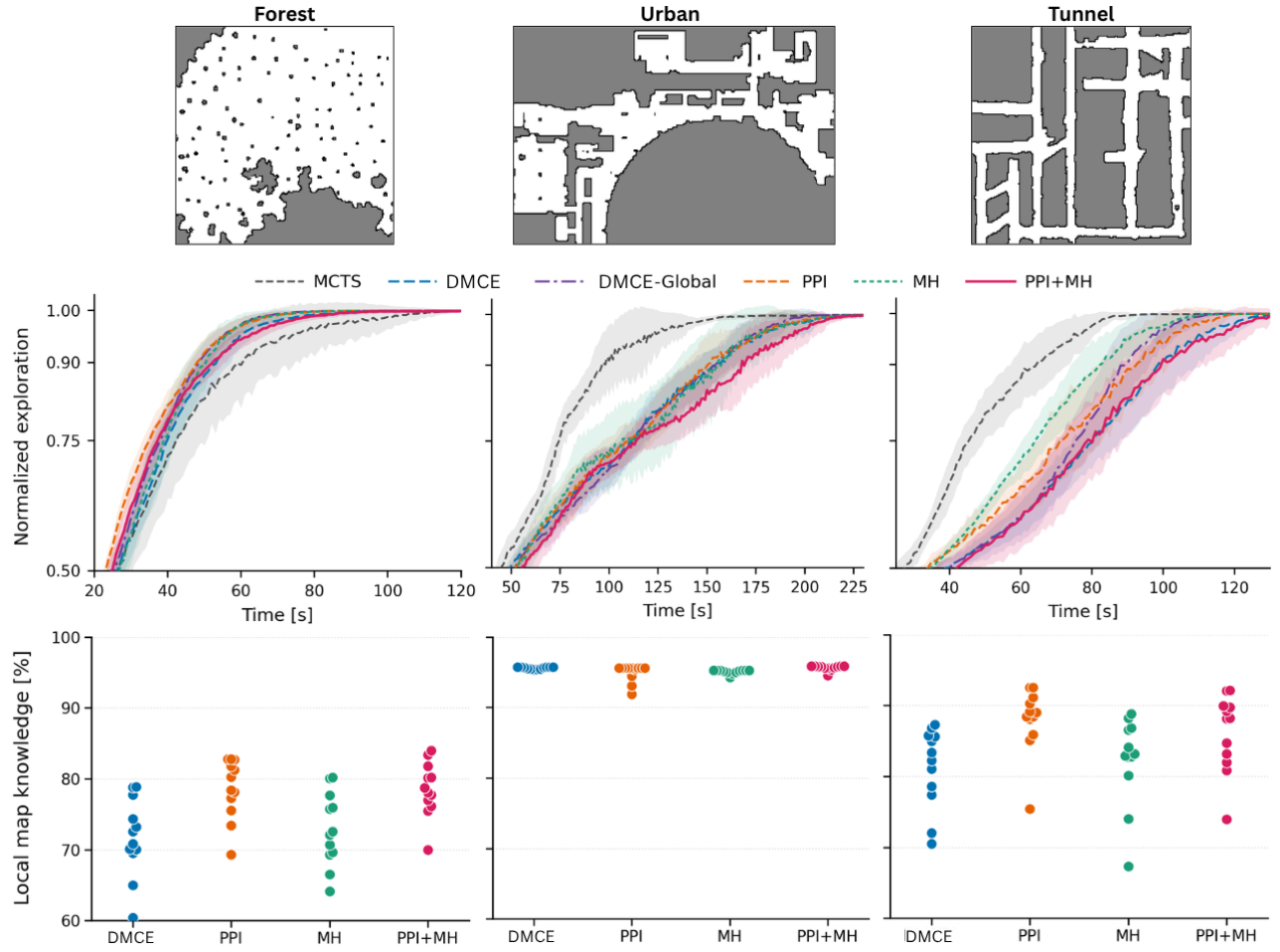


Fig. 3. Exploration behavior and map knowledge distribution. Experiments performed with 12 robots in (a) Forest, (b) Urban, and (c) Tunnel. **Top:** environment layouts. **Middle:** normalized coverage over time, with shaded regions indicating variation across 10 independent runs. **Bottom:** local map knowledge at T95 for the DMCE variants. Each particle represents the average knowledge of one within-run agent computed across the 10 runs. Agents are ranked from least- to best-informed within each run before averaging.

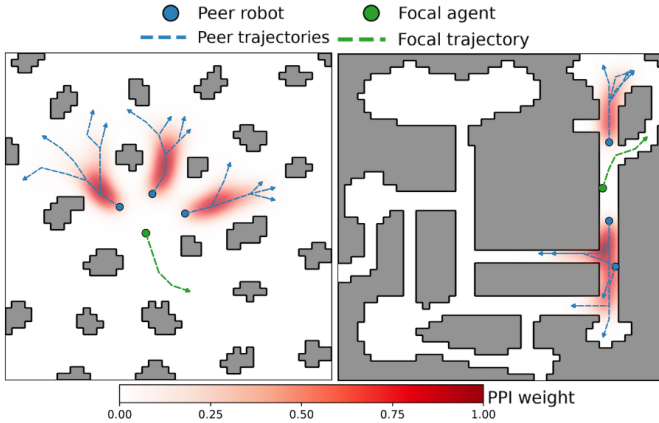


Fig. 4. **Conceptual illustration of environment-dependent peer intent.** The same spatial peer-intent representation can have different implications in open and topologically constrained environments.

alone cannot distinguish redundant exploration from necessary shared transit through bottleneck areas. When MH expands

peer-plan availability, this ambiguity can increase PPI exposure in shared corridors even when robots pursue different regions. Thus, PPI determines how peer information influences local planning, while MH determines how far that information propagates. Their interaction is topology dependent, which is a potential direction for future work.

E. Analysis of Communication and Distributed Information

Multi-hop propagation extends peer-plan availability beyond the immediate communication neighborhood at the cost of additional trajectory exchange. Table III reports the serialized ROS payload rate for $N = 5$. PPI+MH requires 230.66-342.37 kB/s compared with 46.61-78.11 kB/s for DMCE. Importantly, 26.15-54.15% of the source plans processed by PPI+MH are received through relays, confirming that a substantial fraction of planning information originates beyond direct neighbors.

We also examine how exploration progress is reflected in the local occupancy maps of individual robots. Since Multi-hop propagation exchanges trajectories rather than occupancy grids, greater plan availability does not necessarily imply

TABLE III

TRAJECTORY COMMUNICATION FOR $N = 5$ ROBOTS. PAYLOAD DENOTES MEAN SERIALIZED PAYLOAD RATE AND RELAY DENOTES THE FRACTION OF SOURCE PLANS RECEIVED THROUGH RELAYS.

Environment	DMCE	PPI+MH Payload [kB/s]↓	Relay [%]↑
Open	57.74	342.37	26.15
Forest	46.77	230.66	40.18
Urban	46.61	242.56	54.15
Tunnel	78.11	262.18	37.96

TABLE IV

LOCAL MAP KNOWLEDGE AT T95 FOR $N = 5$. HIGHER MEAN K. AND WORST K., AND LOWER GAP, ARE PREFERABLE.

Env.	Method	Mean K.↑ [%]	Worst K.↑ [%]	Gap↓ [%]
Open	DMCE	78.06	61.98	24.84
	PPI+MH	81.32	71.97	14.07
Forest	DMCE	72.65	64.10	13.65
	PPI+MH	69.33	56.15	21.25
Urban	DMCE	90.77	76.30	18.66
	PPI+MH	93.26	87.86	7.61
Tunnel	DMCE	79.18	67.24	18.77
	PPI+MH	82.43	69.71	21.37

uniform local map knowledge. At the T95 milestone, Table IV reports mean team knowledge, least-informed robot knowledge, and the best-to-worst knowledge gap. In Open and Urban, PPI+MH increases the least-informed robot knowledge by 9.99 and 11.56 percentage points while reducing the corresponding gaps by 10.77 and 11.05 points. Tunnel shows smaller gains, while Forest exhibits lower and less uniform local knowledge despite faster coverage. These results suggest that propagated planning information and occupancy-map information have different effects.

F. Ablation on Parameter Sensitivity

We evaluate the sensitivity of PPI to penalty weight λ_{PPI} and spatial spread σ in forest environment with $N = 5$. Within each experiment set, all methods use the same 10 random seeds, ensuring seed-matched comparisons. Separate experiment sets, including the main evaluation and parameter sweeps, use independently generated seed sets. Tables V and VI report the coverage time and overlap.

The effect of λ_{PPI} is non-monotonic. Increasing the peer-intent penalty does not consistently improve either coverage or overlap. $\lambda_{PPI} = 2$ achieves the lowest mean T95 and overlap, whereas intermediate values can perform worse. This indicates that the influence of PPI depends on its interaction with the underlying exploration objective rather than a simple avoidance behavior. We retain $\lambda_{PPI} = 1$ as a common setting throughout the main experiments rather than selecting a separate value for each environment.

Performance is similarly stable over a broad range of spatial scales, with no single σ simultaneously minimizing T90, T95, and overlap. The best mean T95 occurs at $\sigma = 0.5$, the lowest overlap at $\sigma = 0.8$, and the best T90 at $\sigma = 3.0$. We therefore use the intermediate setting $\sigma = 1$ to avoid environment specific parameters.

TABLE V

EFFECT OF λ_{PPI} WITH $\sigma = 1$ IN FOREST FOR $N = 5$. LOWER IS BETTER.

λ_{PPI}	T90 [s]↓	T95 [s]↓	Overlap [%]↓
0.1	73.8 ± 10.5	104.9 ± 33.3	46.67 ± 9.19
0.3	77.3 ± 7.8	112.9 ± 21.8	44.60 ± 8.99
0.5	86.3 ± 9.8	126.5 ± 26.6	45.84 ± 9.03
1.0	77.2 ± 9.7	103.9 ± 15.4	48.51 ± 7.13
2.0	80.2 ± 14.2	97.4 ± 15.3	43.69 ± 12.45

TABLE VI

EFFECT OF σ WITH $\lambda_{PPI} = 1$ IN FOREST FOR $N = 5$. LOWER IS BETTER.

σ	T90 [s]↓	T95 [s]↓	Overlap [%]↓
0.5	76.3 ± 14.9	96.9 ± 18.4	46.96 ± 7.98
0.8	81.3 ± 14.7	104.1 ± 25.8	42.97 ± 5.95
1.0	78.1 ± 10.0	97.2 ± 17.3	44.65 ± 11.15
1.5	81.0 ± 18.2	110.1 ± 25.8	44.47 ± 7.19
2.0	75.9 ± 9.3	101.0 ± 11.5	47.25 ± 6.78
3.0	71.5 ± 8.8	107.6 ± 31.4	46.83 ± 6.23

G. Hardware Implementation

We validate the proposed framework on a team of three TurtleBot3 robots operating in previously unknown environments. Each robot uses its onboard LiDAR measurements to compute the occupancy maps (restricted to 1.5 m). The plans for each robot were generated online using the same decentralized planning framework as in the simulation setup.

The exploration waypoint selected by DMCE+PPI+MH was executed using the standard ROS navigation stack, Dynamic Window Approach (DWA) was used for local trajectory execution and obstacle avoidance. Note that the underlying decentralized planner, PPI construction, and Multi-hop propagation remained unchanged from the simulation.

We evaluated the system in multiple indoor environments. Figure 1 shows online occupancy-map snapshots from two real-world arenas, together with the trajectories traced by the three robots during decentralized exploration. Figure 5 provides additional hardware-validation scenarios, pairing representative views of the physical testbed with the corresponding occupancy maps and executed trajectories. Across all settings, the robots successfully mapped the environment from onboard sensing, exchanged peer plans online, and executed decentralized exploration, demonstrating transfer of the proposed framework from simulation to physical robots.

VI. SUMMARY AND OUTLOOK

We introduced Probabilistic Peer Intent (PPI) and Multi-hop (MH) peer-plan propagation as complementary extensions to decentralized MCTS for multi-robot exploration. PPI provides a continuous spatial representation of available peer plans, while Multi-hop propagation extends their reach beyond direct communication neighbors. Experiments across environmental structures and team sizes show that both mechanisms can improve exploration individually, while their direct composition is dependent on the environment and scale. Physical experiments with three TurtleBot3 robots demonstrate the feasibility of the framework. The main limitations and corresponding directions for future work are:

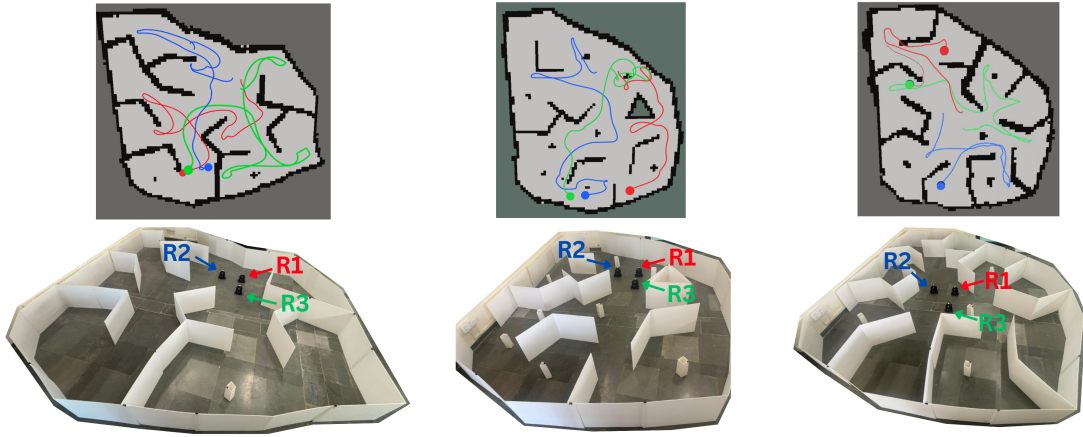


Fig. 5. **Hardware validation.** The proposed framework is deployed on three TurtleBot3 robots in a cluttered indoor environment using onboard LiDAR with an effective sensing range of 1.5 m and online occupancy mapping. The bottom row shows representative views of the physical testbed with the three robots and their initial positions indicated as R1, R2 and R3, while the top row shows the corresponding occupancy maps and executed trajectories obtained during the exploration.

- 1) **Spatial intent representation:** PPI cannot distinguish redundant exploration from exploring shared transit areas that are needed. Reducing the PPI penalty or excluding relayed plans from the PPI field did not resolve this behavior. This motivates the need for topological or goal-level intent.
- 2) **Communication scalability:** Multi-hop propagation increases peer-plan awareness at the cost of additional communication overhead. Hop-aware relay selection, trajectory compression, or communication-budgeted propagation could reduce this overhead while retaining useful non-local information.
- 3) **Physical evaluation:** Our hardware validation is limited to three homogeneous TurtleBot3 robots. Larger robot teams including heterogeneous platforms would be necessary for evaluating scalability and robustness.

REFERENCES

- [1] J. P. Queralta, J. Taipalmaa, B. C. Pullinen, V. K. Sarker, T. N. Gia, H. Tenhunen, M. Gabbouj, J. Raitoharju, and T. Westerlund, "Collaborative Multi-Robot Search and Rescue: Planning, Coordination, Perception, and Active Vision," *IEEE Access*, vol. 8, pp. 191 617–191 643, 2020.
- [2] H. Azpúrúa, M. Saboia, G. M. Freitas, L. Clark, A.-A. Agha-Mohammadi, G. Pessin, M. F. M. Campos, and D. G. Macharet, "A Survey on the Autonomous Exploration of Confined Subterranean Spaces: Perspectives from Real-World and Industrial Robotic Deployments," *Robotics and Autonomous Systems*, vol. 160, p. 104304, 2023.
- [3] S. Swinton, D. Mitchell, J. Blanche, E. McGookin, and D. Flynn, "The Evolution of Autonomous Systems for Planetary Cave Exploration: A Review," *Journal of Field Robotics*, 2026.
- [4] G. Best, O. M. Cliff, T. Patten, R. R. Mettu, and R. Fitch, "DecMCTS: Decentralized planning for multi-robot active perception," *The International Journal of Robotics Research*, vol. 38, no. 2-3, pp. 316–337, 2019.
- [5] S. Bone, L. Bartolomei, F. Kennel-Maushart, and M. Chli, "Decentralised multi-robot exploration using monte carlo tree search," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 7354–7361.
- [6] F. Cladera, Z. Ravichandran, I. D. Miller, M. A. Hsieh, C. J. Taylor, and V. Kumar, "Enabling large-scale heterogeneous collaboration with opportunistic communications," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 2610–2616.
- [7] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, 1997, pp. 146–151.
- [8] H. Umari and S. Mukhopadhyay, "Autonomous robotic exploration based on multiple rapidly-exploring randomized trees," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 1396–1402.
- [9] J. Yu, J. Tong, Y. Xu, Z. Xu, H. Dong, T. Yang, and Y. Wang, "SMMR-explore: Submap-based multi-robot exploration system with multi-robot multi-target potential field exploration method," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 8779–8785.
- [10] C. Wang, C. Yu, X. Xu, Y. Gao, X. Yang, W. Tang, and S. Yu, "Multi-robot system for cooperative exploration in unknown environments: A survey," *arXiv preprint arXiv:2503.07278*, 2025.
- [11] L. Bramblett and N. Bezzo, "Epistemic planning for multi-robot systems in communication-restricted environments," *Frontiers in Robotics and AI*, vol. 10, p. 1149439, 2023.
- [12] J. Bayer and J. Faigl, "Decentralized multi-robot exploration under low-bandwidth communications," *Autonomous Robots*, vol. 50, no. 1, p. 7, 2026.
- [13] N. Jadhav, M. Behari, R. J. Wood, and S. Gil, "WiSER-X: Wireless signals-based efficient decentralized multi-robot exploration without explicit information exchange," *arXiv preprint arXiv:2412.19876*, 2024.
- [14] J. Chiun, S. Zhang, Y. Wang, Y. Cao, and G. Sartoretti, "MARVEL: Multi-Agent Reinforcement Learning for constrained field-of-View multi-robot Exploration in Large-Scale Environments," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 11 392–11 398.
- [15] Y. Meng, N. Majcherczyk, W. Liu, S. Kiesel, C. Fan, and F. Pecora, "Reliable and efficient multi-agent coordination via graph neural network variational autoencoders," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 12 965–12 971.
- [16] L. Bartolomei, L. Teixeira, and M. Chli, "Fast Multi-UAV Decentralized Exploration of Forests," *IEEE Robotics and Automation Letters*, vol. 8, no. 9, pp. 5576–5583, 2023.