2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Simulated Annealing for Multi-Robot Ergodic Information Acquisition Using Graph-Based Discretization

Benjamin Wong[1], Aaron Weber[1], Mohamed M. Safwat[1], Santosh Devasia[1], and Ashis G. Banerjee[2]

Abstract— **One of the goals of active information acquisition using multi-robot teams is to keep the relative uncertainty in each region at the same level to maintain identical acquisition quality (e.g., consistent target detection) in all the regions. To achieve this goal, ergodic coverage can be used to assign the number of samples according to the quality of observation, i.e., sampling noise levels. However, the noise levels are unknown to the robots. Although this noise can be estimated from samples, the estimates are unreliable at first and can generate fluctuating values. The main contribution of this paper is to use simulated annealing to generate the target sampling distribution, starting from uniform and gradually shifting to an estimated optimal distribution, by varying the coldness parameter of a Boltzmann distribution with the estimated sampling entropy as energy. Simulation results show a substantial improvement of both transient and asymptotic entropy compared to both uniform and direct-ergodic searches. Finally, a demonstration is performed with a TurtleBot swarm system to validate the physical applicability of the algorithm.**

I. Introduction

The robotics community has been interested in multi-robot systems for their versatility and efficiency in performing time consuming and repetitive tasks in a parallel manner. Many of these tasks belong to active information acquisition, including surveillance, inspection, environmental monitoring, and disaster response [1]–[4]. In these scenarios, a team of robots is tasked with collecting or gathering information by traveling between various sites, often to locate targets, such as defects in inspection operations, survivors in disaster responses, or sources of hazard during surveillance and environmental monitoring. By using a multi-robot system, resources can be distributed (allocated) among different regions to effectively perform time-critical tasks.

The key to effectively carrying out the tasks is then to characterize the *quality of information* in the different regions and allocate the robots accordingly. *Ergodic control* is well suited for this purpose which amounts to sampling from a statistical perspective. This is particularly true for a cost-effective multi-robot system, where a large team of robots are equipped with commodity sensors that provide uncertain measurements. These uncertain (noisy) measurements are expressed in terms of probability, where estimates can be improved by repeated measurements. Moreover, the confidence of the estimates is directly correlated to the measurement noise (e.g., the target is hiding in a visually cluttered environment as opposed to a target in plain sight), such that more samples are needed for a noisier target to reach the same level of confidence as a less noisy target. This provides an ideal number of total visitations to each target site (region) that the team of robots has to maintain as a whole. This aligns with the core objective of ergodic control, where a control law is devised such that the time-averaged visitation frequency is equal to the (spatial) target distribution [5], which can be obtained from information quality in the case of sampling.

In traditional ergodic control, the target distribution is assumed to be known, either from an oracle, prior experience, or experts' demonstrations. However, if we relax this assumption for realistic sampling tasks, the challenge becomes that information quality is initially unknown to the robot team. In fact, none of the parameters of the sampling distribution, including the sampling variance, are initially known to the robots. Hence, the optimal target distribution cannot be derived. While the *posterior* variance can be estimated, as in [6], the sample variance and process variance are still required. Alternatively, the sample variance can be estimated, along with the mean, based on the collected samples, as the robots start exploring [7]. However, such estimation is unreliable when the sample size is small and fluctuates as new samples are incorporated. This leads to a highly time-varying and unreliable target distribution that causes further misallocation of resources.

To avoid this problem, the target distribution should focus on collecting information on the variance at the beginning and gradually shift toward the optimal solution as the estimated variance becomes reliable. This is done by applying *simulated annealing* to the target distribution. Annealing is a process of heating up a material and slowly cooling it down to allow the molecules to reach the lowest energy state [8]. Simulated annealing applies this idea to stochastic optimization to encourage exploration by introducing substantial randomness (high entropy) at the beginning and gradually reducing the randomness until the optimal solution is reached. In this work, the robot team is treated as a particle swarm with the noise level of each region corresponding to the energy state. The high entropy state corresponds to the robots spreading out to all the regions regardless of the estimated noise level, and low entropy corresponds to all the robots focusing on the highest noise level region. This allows for a smooth interpolation between a uniform target distribution and a greedy distribution. Consequently, the robot team can collect information about the variance

[1]B. Wong, A. Weber, M. M. Safwat, and S. Devasia are with the Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA. bycw,aweber6,mohsaf,devasia@uw.edu

[2]A. G. Banerjee is with the Department of Industrial & Systems Engineering and Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA. ashisb@uw.edu

and smoothly transition to the optimal distribution. The main contributions of this work are summarized below.

- Formulating a multi-robot information acquisition problem where ergodic control provides an optimal solution
- Incorporating simulated annealing to handle unreliable information (noisy measurements) in the ergodic controller
- Demonstrating superior performance in terms of posterior entropy compared to direct and uniform ergodic methods

II. Related Works

A. Active Sampling

In a broad sense, this work falls in the domain of active planning, where agents plan trajectories that satisfy an objective and continuously adjust their trajectories as new information arrives. In this case, the objective is to obtain information through sampling. A popular choice of doing so is to find a trajectory that covers the entire space [9], [10]. This choice assumes that all the points in the space are equally important and lacks the flexibility to adapt to the quality of information. Alternatively, to consider information density, sensor placement through Voronoi partitioning has been used [11]–[13]. This is well-suited for a static placement problem, where the robots converge to the optimal position until a new event arises. This requires a known number of robots and assumes all robots to be active, and does not account for the need of redundancy. On the other hand, in ergodic control, all the agents provide coverage individually, which make the system more robust to sensor and actuation failures and varying numbers of active robots.

Our formulation of obtaining information from discrete locations is essentially a problem of sequential experiment design [14]. A particular form that has been thoroughly studied is the multi-armed bandit (MAB) problem, where the agents maximize their rewards under unknown distributions [15]. Traditionally, MAB assumes that any arm can be sampled at any time. Recently, work has been done on applying the MAB problem constrained on a graph [16], similar to our problem formulation. While we can consider the information as a form of reward in our formulation, the biggest difference with the MAB problem is that the MAB rewards are fungible, where agents can collect solely from the state with the maximum amount of reward. On the other hand, in our formulation, sufficient information has to be collected from all the regions, i.e. missing information in one region cannot be substituted by more information from another region.

B. Ergodic Control

In ergodic control, a controller is designed such that the time-averaged trajectories of the agents equal the spatial target distribution. It is, therefore, often used in robotics to solve multi-agent resource allocation and coverage problems [17]–[22]. Early foundational works assumed the target distribution to be given [5], [20], [23]. Other application-oriented works considered ergodic control as a means to promote exploration by assigning a distribution around the optimal [24]–[26]. Recently, discrete ergodic control formulations have been investigated for robot exploration in topologically challenging environments [27]–[29]. In this work, we aim to extend such graph-based formulations to address a specific multi-robot allocation problem, where the target distribution is modeled as an objective for the ergodic controller. In other words, our ergodic controller is designed to yield an optimal strategy for allocating robots to different regions so as to acquire information in a Bayesian manner.

C. Simulated Annealing

This work utilizes simulated annealing to transition from exploration to exploitation. Simulated annealing originated from statistical mechanics and is widely applied to complex optimization problems such as the traveling salesman problem [30]. It is rooted in the study of ergodic systems, with the most common implementation being the Metropolis-Hasting (M-H) algorithm as a stochastic alternative to gradient descent [31, chapter 8]. It has been applied to robotic systems for path planning problems as a means to escape local minima [32]–[34]. In these works, the target distribution is implicitly defined under the acceptance-rejection step of M-H. In comparison, as an ergodic control framework, the (annealed) target distribution in our work is explicitly defined, which yields a method that has more desired characteristics, such as convergence rate.

III. Problem Statement

A. Problem Formulation

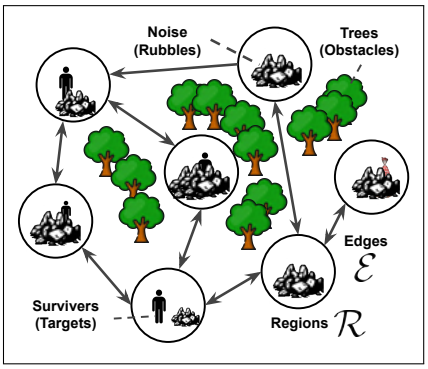


Fig. 1: Example information gathering task of locating survivors in a map $\mathcal{G}$ with regions, modeled as set of nodes $\mathcal{R}$, and edges $\mathcal{E}$ caused by blockage of trees. The regions contain varying amount of rubble, which causes differences in information quality i.e., noise levels σ_i^2, across the regions.

This article considers information gathering tasks that are scattered in a large area, with regions of interest separated by obstacles. A team of N robots has to estimate the states, x_i, internal to each region. An example task of finding survivors in disaster relief is shown in Fig. 1, where the states to be estimated are the survivors' locations in all the regions (in the case of scalar x_i, we assume there is exactly one survivor

per region[1]).

The regions are modeled as a set of nodes $\mathcal{R}$. The connections between pairs of regions are represented by an edge, and $\mathcal{E}$ is the set of all the edges. Then, the tuple of nodes and edges forms the graph $\mathcal{G} = (\mathcal{R}, \mathcal{E})$. Information gathering is considered an estimate of the scalar state x_i for each region $r_i \in \mathcal{R}$. Each robot can collect a single observation from the region at every time step, with the noisy observation model

$$z_i = x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2). \tag{1}$$

The variance σ_i^2 varies between the regions but not between the robots, i.e., some regions have lower information quality per observation than others. The varying variance can be due to a lack of landmarks for robot localization or a high visual noise in the region that conceals the target[2]. The state x_i can be estimated from the observations by

$$\bar{x}_i = \frac{1}{\nu_i} \sum_{j=1}^{\nu_i} z_j \approx x_i, \quad \text{Var}[\bar{x}_i] = \frac{\sigma_i^2}{\nu_i} \tag{2}$$

where ν_i is the number of observations acquired collectively by the robots. The standard error, i.e., the variance of the estimation $\text{Var}(\bar{x}_i)$, is a function of the number of observations ν_i and the region noise σ_i^2[3].

B. Optimal Distribution

To ensure no one region has more standard error than the others, the goal is to minimize the maximum posterior variance after any time horizon K, i.e. all regions are equally confident at all time,

$$\begin{aligned} \min_{\bar{\rho}} \quad & \max_i \left[\frac{\sigma_i^2}{KN\bar{\rho}_i} \right] \\ \text{s.t.} \quad & \mathbf{1}^T \bar{\rho} = 1 \\ & \bar{\rho} \succeq 0 \end{aligned} \tag{3}$$

where $\bar{\rho}_i$ is the relative visitation frequency of region i, and $\bar{\rho}$ is the target distribution, which is a vector with the i-th entry being $\bar{\rho}_i$; N is the number of robots. For a known sample variance σ^2, the optimal target distribution is

$$\bar{\rho}_i^* = \frac{1}{\sum_i \sigma_i^2} \sigma_i^2. \tag{4}$$

Then, by substituting $\bar{\rho}^*$ to $\bar{\rho}$ in (3), the variance of the estimation at any time step K is[4]

$$\text{Var}[\bar{x}_i] = \frac{\sigma_i^2}{KN\bar{\rho}_i} = \frac{\sum_i \sigma_i^2}{KN} \tag{5}$$

[1]Multiple survivors can be modeled as a multivariate Gaussian distribution, which is outside the scope of this work. Alternatively, it can be done by considering x_i as the number of survivors instead of the location of one survivor.

[2]The observation model can be considered as an abstraction of a robot collecting observations continuously within one region for a fixed amount of time before transitioning, where z_i, σ_i^2 are the result of integrating all the observations collected within the time step.

[3]By the central limit theorem, this is also extendable to a non-Gaussian distribution with a sufficiently large ν.

[4]It can be observed that uniform values minimize the maximum value since to maintain a constant sum, any decrease of value in one region will cause an increase of equal amount in other regions, which increases the maximum value.

which is the same for all regions. Hence, there exists a need for a planning strategy such that the target $\bar{\rho}$ can be reached under graph connectivity constraints.

Moreover, the true sampling variance σ_i^2 is initially unknown to all the agents. The variance can also be estimated along side x_i as the robots start collecting samples, which is done by

$$\bar{\sigma}_i^2 = \frac{1}{\nu_i - 1} \sum_{j=1}^{\nu_i} (z_j - x_i)^2 \approx \sigma_i^2. \tag{6}$$

Similar to the mean $\bar{x}$, the estimated variance $\bar{\sigma}$ is unreliable at the beginning and is improved with increasing number of samples. As a result, $\bar{\rho}^*$ in (4) is unknown to the robots and the $\bar{\rho}$ generated by the estimated variance $\bar{\sigma}^2$ is unreliable and causes inefficient allocation. This motivates the research problem to develop a planning method that can utilize the estimated variance while accounting for the unreliable initial estimations. This can be considered as two subproblems in developing the planning method: (subproblem 1) to reach a given target distribution on a graph with a multi-robot system and (subproblem 2) to account for the unknown variance and unreliable estimates in the beginning.

IV. Methodology

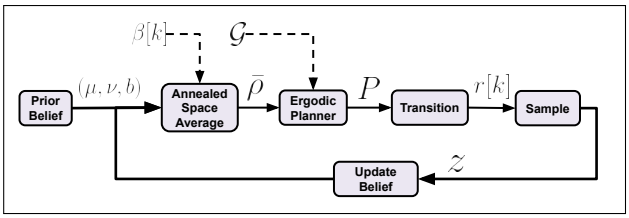


Fig. 2: Flowchart of the annealed ergodic information gathering algorithm.

The problem of achieving the target distribution is solved using *Rapidly Ergodic Markov Chain* (REMC), and the problem of unreliable initial estimated variance is solved using *simulated annealing*. The solution methods are explained in Sections IV-A and IV-B, respectively. The overall pipeline of the planning method is shown in Fig. 2.

A. Subproblem 1 Solution: Multi-Robot Ergodic Control

Ergodic control can solve the subproblem 1 to achieve the target distribution $\bar{\rho}$ for a given σ^2. In general, the goal for ergodic control is to synthesize a control law such that the dynamic system has a time average equal to the space average for almost all initial conditions. This is formulated in a graph space as

$$\lim_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} (F(r[k])) = \frac{1}{\mu(\mathcal{R})} \sum_{r_i \in \mathcal{R}} F(r_i) \mu(r_i) \tag{7}$$

where $r[\cdot]$ is the region trajectory; $\mathcal{R}$ is the set of regions; μ is a measure on the region set; and F is an arbitrary μ-measurable function. To measure visitation, F is defined as the indicator function

$$I(r_i) \triangleq \begin{bmatrix} \delta_{1,i} & \delta_{2,i} & \cdots & \delta_{n,i} \end{bmatrix}^T \tag{8}$$

where $\delta_{i,j}$ is the Kronecker delta, and

$$\hat{\rho} \triangleq \lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \left(I(r[k])\right), \quad \bar{\rho} \triangleq \frac{1}{\mu(\mathcal{R})} \sum_{r_i \in \mathcal{R}} I(r_i)\mu(r_i). \tag{9}$$

The ergodic objective for a single agent has been achieved by generating the trajectory using a *rapidly ergodic Markov chain (REMC)* [29]. The Markov chain is generated by randomly sampling the next region based only on the current region, i.e. $r[k+1] \sim \mathbb{P}(R \mid r[k])$, with R being the random variable of the possible region. In a finite graph space, the transition probability can be represented by a stochastic matrix P. The time average is then expressed in terms of the expected value as

$$\begin{aligned} \mathbb{E}\left[\hat{\rho}\right] &= \lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \left(I(\mathbb{E}[r[k]])\right) \\ &= \lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \left(P^k \rho[0]\right) \end{aligned} \tag{10}$$

with $\rho[0]$ as the initial distribution.

The transition matrix that guarantees ergodicity and also optimizes the convergence rate is found by the following convex program:

$$\begin{aligned} \arg\min_{P} \quad & \lambda_{\max}\left(\frac{1}{2}\left(\tilde{P} + \tilde{P}^T\right) - 2\bar{\rho}^{1/2}\bar{\rho}^{T/2}\right) \\ \text{s.t.} \quad & \mathbf{1}^T P = \mathbf{1}^T \quad \text{(Stochastic } P\text{)} \\ & P\bar{\rho} = \bar{\rho} \quad \text{(Target Distribution)} \\ & P_{i,j} \geq 0 \quad \text{(Stochastic } P\text{)} \\ & P_{i,j} = 0 \quad \text{if} \quad (j,i) \notin \mathcal{E} \quad \text{(Transitions in } \mathcal{G}\text{)} \\ & \tilde{P} = \text{diag}(\bar{\rho}^{-1/2}) P \text{diag}(\bar{\rho}^{1/2}). \end{aligned} \tag{11}$$

As ergodicity is a property derived from ensemble systems, (7) is extensible to multi-robot system. When all the robots follow the same Markov chain, (9) is modified to account for the trajectories of all the robots as

$$\begin{aligned} \mathbb{E}[\hat{\rho}] &= \frac{1}{N} \sum_{a=1}^{N} \mathbb{E}[\hat{\rho}_a] \\ &= \frac{1}{N} \sum_{a=1}^{N} \left(\lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} P^k \rho_a[0] \right) \\ &= \lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \left(P^k \frac{1}{N} \sum_{a=1}^{N} \rho_a[0] \right) \\ &= \lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} \left(P^k \rho[0]\right). \end{aligned} \tag{12}$$

Here, ρ_a denotes the distribution for agent a, and ρ is averaged over all the robots. Consequently, the more robots are in the team, the less variance the Markov chain would have. That is, the true trajectory follows more closely to the expected value

$$\frac{1}{N} \sum_{a=1}^{N} I(r_a[k]) \to P^k \rho[0] \quad \text{as} \quad N \to \infty. \tag{13}$$

B. Subproblem 2 Solution: Annealing for Space Average

To account for the second subproblem of unreliable initial variance, the target distribution is designed to focus on collecting information on the variance at the beginning and gradually shift toward the oracle solution as the variance becomes reliable. This is achieved by simulated annealing, where the robots starts at the most random configuration, i.e. uniform random, and shift toward the optimal solution by varying a temperature parameter.

To achieve this, the *Gibbs measure*, also known as the *Boltzmann distribution in finite space*, is chosen as the measure μ in (7), with

$$\mu(r_i) = \exp(-\beta E(r_i)) \tag{14}$$

where $E(r_i)$ is the energy associated with the state r_i, and β is the *coldness* (or *inverse temperature*) parameter. Since the region set is disjointed, the normalizing factor is

$$\mu(\mathcal{R}) = \sum_{i=1}^{n} \exp(-\beta E(r_i)). \tag{15}$$

In particular, we define the energy to be the negative (differential) entropy of the sample distribution as

$$E(r_i) = -\ln(2\pi e \bar{\sigma}_i^2). \tag{16}$$

The space average in (9) is then

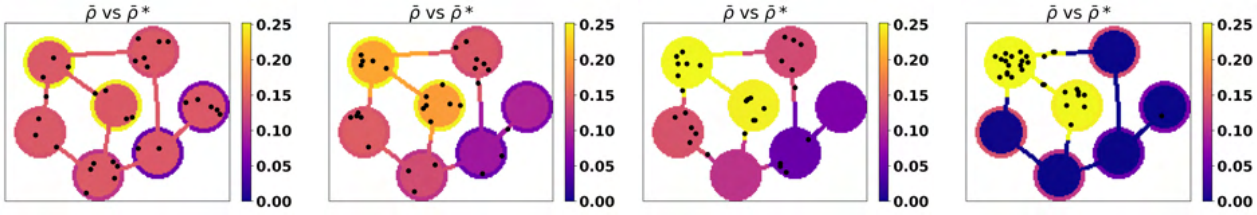
$$\begin{aligned} \bar{\rho}(\beta) &= \frac{1}{Z(\beta)} \sum_{i=1}^{n} I(r_i) \exp(\beta \ln(\bar{\sigma}_i^2)) \\ Z(\beta) &= \sum_{i} \exp(\beta \ln(\bar{\sigma}_i^2)). \end{aligned} \tag{17}$$

An additional advantage of using the Gibbs measure is that the target distribution $\bar{\rho}$ is guaranteed to be reachable by an ergodic Markov chain, as all the entries are strictly positive for any real-valued energy level $E(r_i)$. More importantly, the coldness parameter β provides a smooth control on the uniformness of the target distribution, with $\beta = 0$ generating a uniform distribution regardless of the energy level; and $\beta \to \infty$ generating a delta function at the minimum energy state.

In our setup, the target distribution will assign more samples to regions with higher sampling entropy as β increases. Specifically, when $\beta = 1$, the target distribution is

$$\begin{aligned} \bar{\rho}(1) &= \frac{1}{Z(1)} \sum_{i=1}^{n} I(r_i) \exp(\ln(\bar{\sigma}_i^2)) \\ &= \frac{1}{Z(1)} \sum_{i=1}^{n} I(r_i) \bar{\sigma}_i^2. \end{aligned} \tag{18}$$

If $\bar{\sigma}^2 = \sigma^2$, then this is the estimation of the optimal sample size in (4). An example of the effect of the value of β is shown in Fig. 3. As a result, if β varies gradually from 0 to 1, i.e., annealed, the goal of transitioning from uniform to



(a) $\beta = 0$ (b) $\beta = 0.5$ (c) $\beta = 1$ (d) $\beta = 10$

Fig. 3: Example distribution of 30 robots (black markers) with various constant coldness parameter β after a sufficient number of steps to reach equilibrium ($K = 100$). The color of the region border represents the relative variance σ^2 and the inner color represents the target distribution $\bar{\rho}(\beta)$ generated using the ground truth variance. (a) The robots are uniformly spread out regardless of the variance; (b) the distribution of robots is in between uniform and optimal; (c) the distribution of robots is proportional to the variance, which is optimal for information gathering; (d) the robots are concentrated in the two regions with the highest variance, which causes severe under-sampling in the rest of the regions.

optimal is achieved. In this work, annealing is done by the first-order step response as[5]

$$\beta(k) = 1 - \exp(-\alpha k), \tag{19}$$

with k being the time step and α the cooling rate. The choice of α is discussed in Section VII.

V. Annealed Ergodic Algorithm

The complete algorithm for applying the annealed ergodic information gathering is shown in Algorithm 1. Since ergodic control can run persistently, no time horizon has to be specified, and the only parameter required is the cooling rate α. The main algorithm is composed of two stages: a) sampling (lines 4 - 10), b) planning (lines 11 - 19). Using the equations in (2) and (6) to estimate the mean and variance requires the entire set of observations to be stored. Instead, it is updated sequentially in a Bayesian manner using the normal-inverse gamma parameterization (ν, μ, b) [36, Section 6], where μ is the mean, ν is the number of samples, and b is an auxiliary variable to recover the variance, which is done using the equation in line 12. Once the most recent estimated variance σ^2 is recovered from the parameters, the target distribution $\bar{\rho}$ is calculated using the current temperature $\beta[k]$. The transition matrix P is then obtained using the REMC algorithm. The robots randomly transition to the next region individually according to the transition matrix.

VI. Experiments

In the section, the annealing algorithm is compared against uniform coverage

$$\bar{\rho}_{\text{uniform}} = [1/n, 1/n, \cdots] \tag{20}$$

and direct ergodic coverage

$$\bar{\rho}_{\text{direct}}(k) = \frac{1}{Z(1)} \sum_{i=1}^{n} I(r_i) \bar{\sigma}_i^2[k] \tag{21}$$

with different numbers of robots.

[5]Different annealing schedule can be used, such as tanh, with similar result. Here first-order step response is chosen with its similarity to Newton's law of cooling [35, eq.(1.22)].

Algorithm 1 Annealed Ergodic Information Gathering

1: **Parameter:** Cooling Rate: α
2: **Initialize** $\nu \leftarrow \text{ones}(n), b \leftarrow \text{ones}(n)$, $\mu \leftarrow \text{zeros}(n), k = 0$
3: **while** inspecting **do**
4: **for** each robot a **do**
5: $z = \text{Sample}(r(a))$ ▷ Sample from region of a
6: ▷ Update law for normal inverse-gamma dist.
7: $b[r(a)] \leftarrow b[r(a)] + \frac{1}{2} \frac{n[r(a)]}{n[r(a)]+1} (z - \mu[r(a)])$
8: $\mu[r(a)] \leftarrow \frac{n[r(a)]\mu[r(a)]+z}{n[r(a)]+1}$
9: $\nu[r(a)] \leftarrow \nu[r(a)] + 1$
10: **end for**
11: **for** $r_i \in \mathcal{R}$ **do**
12: $\sigma^2[i] \leftarrow 2b[i] \frac{\nu[i]+1}{(\nu[i])^2}$
13: **end for**
14: $\beta = 1 - \exp(-\alpha k)$
15: $\bar{\rho} \leftarrow \exp(\beta \ln(\sigma^2))$ ▷ Entry-wise exponential
16: $\bar{\rho} \leftarrow \bar{\rho} / \sum_i \bar{\rho}_i$
17: $P \leftarrow \text{REMC}(\bar{\rho})$
18: **for** each robot a **do**
19: $r[a] \leftarrow P(r[a])$
20: **end for**
21: $k \leftarrow k + 1$
22: **end while**

A. Region Graph

1) Methodology: Performance comparisons are done on a representative region graph shown in Fig. 1. The observation mean and noise, σ^2, are both uniformly randomly sampled from $(-10, 10]$ and $(0, 200]$, respectively. Both the parameters are sampled just once at the beginning and then used throughout the trials. Additionally, the noise is multiplied by the number of robots N to realize a fixed cost, wherein one can get a large team of robots with bad sensors or a small team of robots with good sensors. All the robots start from the same region (region 1) to emulate the condition that the team is activated at the same time from a base station. The annealing rate α is chosen as 0.025. 100 trials

are conducted for each method. The robots are modeled as collision-free point masses that synchronously transition to their planned regions at each time step k. Parameter estimation (of μ, ν, b) and Markov chain optimization are performed in a centralized manner. The resultant stochastic matrix P is published to the robots and planning is carried out by the robots independently.

2) Results: Fig. 4 shows that the annealing method outperforms the direct method during the transient phase, and the uniform method asymptotically with respect to the true posterior entropy. The true posterior entropy is obtained from the true sample variance σ^2 and is maximized over regions, i.e., the region with the worst entropy is chosen at each time step k, where

$$h(k) = \max_i \left[\ln \left(\frac{\sigma_i^2}{\nu_i(k)} \right) \right]. \tag{22}$$

It is seen that the direct method has a true posterior entropy higher than both uniform and annealing methods from $k = 0$ to $k = 200$. This is caused by the previously mentioned problem of variance estimation error, which leads the robots to initially misallocate the samples until the variance is more refined. Consequently, the direct method also has more variability over the trials because of its high dependence on the quality of the initial observations. Conversely, while the uniform method shows better initial performance, its long-term performance is plateaued by the regions with high observation noise, since all the regions are assigned the same number of samples with the maximum entropy dominated by the noisiest region. The annealing method shows advantages over both the uniform and direct methods. Initially, when the coldness parameter $\beta = 0$, it closely follows the performance of the uniform method. As more samples are collected and $\beta \to 1$, the robots shift toward the optimal distribution.

More insight can be gained from the bottom row of Fig. 4, which plots the posterior entropy estimated by the robots. Here, the oracle σ^2 is replaced by the estimated variance at each time step $\bar{\sigma}^2(k)$ to yield

$$\bar{h}(k) = \max_i \left[\ln \left(\frac{\bar{\sigma}_i^2(k)}{\nu_i(k)} \right) \right]. \tag{23}$$

Uniform and annealed methods show similar estimated posterior entropy to the true posterior entropy, except they are nosier at the beginning when the sample size is small. However, direct entropy shows a significantly smaller estimated entropy than the true entropy. In other words, with the direct ergodic method, the robots believe they have more information than they really have, which directly causes the aforementioned misallocation of resources. Therefore, this shows that starting with uniform search is more advantageous as it avoids the problem of overconfidence.

The region-wise behavior of the robots distribution can be seen in Fig. 5, which shows the space average and time average of one example trial from the annealed and direct methods. It can be seen that the annealed method generates smooth trajectories from uniform distribution to the optimal distribution for the time average to follow, while

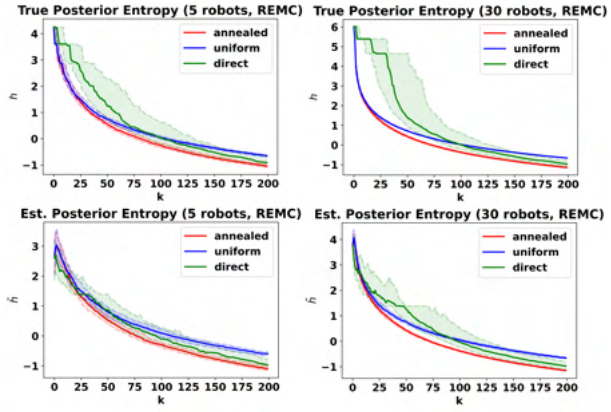


Fig. 4: Comparison of maximum posterior entropy between uniform, direct ergodic, and annealed ergodic methods over 100 trials. The solid line represents the median and the shaded region bounded by dashed lines represents the inter quartile region. The top row shows the true entropy obtained from an external oracle, where annealing is consistently performing better both transiently and asymptotically. The bottom row shows the estimated entropy from the internal belief of the robots, where the direct ergodic method has a problem of overestimating its information, as compared to the true entropy.

the direct method has a fluctuating space average caused by the initial noisy variance estimation with a sudden shift to the optimal distribution as the sample size increases. This causes a highly time-varying distribution that is difficult for the ergodic control to track and leads to the extreme values seen in the time average.

B. Erdős–Rényi Random Graph

To show more generalized results, additional experiments are carried out on Erdős–Rényi graphs, where the directed edges are generated independently with probability p. A total of 10 nodes with an edge probability of $p = \frac{2\ln(n)}{n}$ are considered. The rest of the setup is identical to Section VI-A. The results for 30 robots in Fig. 6 show similar behavior to that for the region graph, where the direct method performs the worst initially; the uniform method performs the worst near the end; and our annealed method performs the best throughout the trials. For the 5 robots case, direct method performs worse, where the true entropy remains high throughout the trials. This is potentially caused by the higher number of regions, which reduces the chance of robots exploring the regions wrongly perceived as having low variance, thereby requiring more time to recover from traversal decision errors.

C. Physical Demonstration

The simulation result is applied to a swarm system with an OptiTrack motion capture rig to validate the feasibility of executing the exploration on physical platforms. A total of five TurtleBot Burger robots are used with twelve Flex13 tracking cameras on the motion capture system. A 4-region

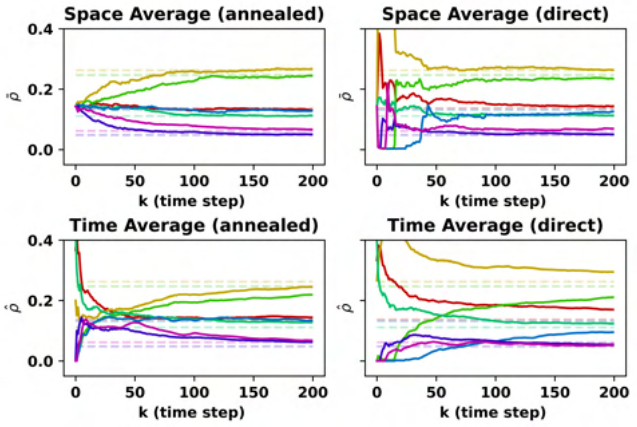


Fig. 5: Example space average $\bar{\rho}$ and time average $\hat{\rho}$ of each region shown in different colors, with the optimal distribution shown by dashed lines. The left and right columns show the results using annealing and direct ergodic traversal, respectively. Annealing yields a smooth transition from uniform to the optimal solution that is tracked by the time average; however, the direct ergodic method produces a fluctuating space average that causes an extreme time average.

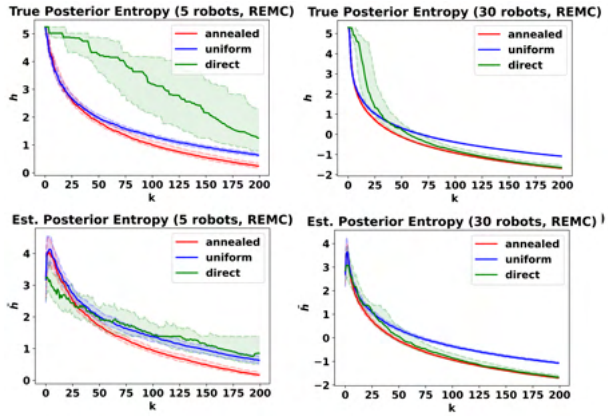


Fig. 6: Performance comparisons, with settings identical to Fig. 4, on directed Erdős–Rényi random graphs with 10 regions. The direct method shows an even worse performance potentially because of the increased number of regions and directed edges.

graph is projected to the ground. The cooling rate α is adjusted to 0.1. Snapshots of the system at 3 different time steps are shown in Fig. 7. The result shows that the planning pipeline can be executed in a physical platform with a suitable low-level collision avoidance technique.

VII. Discussion

An open question for the annealing process is the optimality of the cooling rate α. In this work, we choose α empirically according to the simulation time horizon. Conceptually, the optimality of the cooling rate is correlated to the convergence rate of the ergodic planner, e.g. the ergodic controller should be able to reach uniform distribution before the annealing completely cools down. It is also correlated to the convergence of the noise estimation. Some observation models may require more samples to reach accurate noise estimation. The cooling rate then has to be slower so that the robots spend more time in the information collection phase.

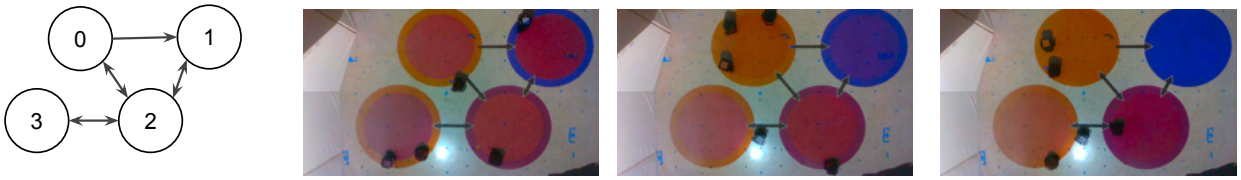
In this work, the target distribution is time-varying due to annealing and variance σ^2 updating. This is not included in the proof for the optimality of REMC. Conceptually, since the variance estimation in (6) converges as ν converges, and the annealing converges to the estimated variance, a convergence proof can be established. This is potentially related to the notion of *weak ergodicity*, which is defined for a time-varying Markov chain and is used to develop the convergence proof in simulated annealing [31].

One main challenge of the practicality of this framework is the assumption that all agents can communicate with the command/control center at all time steps. The Markov chain-based ergodic planner can be used in a fully decentralized manner without any communication for a static target distribution. However, due to the information gathering task, the target distribution is constantly being updated, which requires a new Markov chain to be generated and broadcast to the robots. In our future work, we will investigate a fully decentralized framework using percolation theory [37]. With the controller guaranteeing (weak) ergodicity, percolation theory can be applied [38] and, ideally, a critical transition point can be established for information to be fully shared in the robot network without requiring a command center.

For simplicity, the target information in this work is scalar and static. However, in many robotics applications, such as target tracking, the information will be multi-dimensional and dynamic. The core idea of "lower the quality of information, more robots should be allocated" should still be valid. Some extra considerations have to be accounted for, such as choosing a proper scalarization of variance (entropy, A-optimality, etc.) and establishing asymptotic confidence so that weak ergodicity holds true.

VIII. Conclusion

This paper addresses the problem of multi-robot information acquisition using ergodic control, by formulating the problem on a region graph-based discretization of the environment so as to allocate the acquisition effort according to the quality of the observed (sampled) information in the different regions. Accordingly, we introduce a multi-agent extension of a previously developed rapidly ergodic Markov chain planner that employs simulated annealing to generate the (space-averaged) target distribution of the agents. The estimated entropy is applied to the Boltzmann distribution such that annealing gradually varies the coldness parameter from 0 to 1, to control the space-averaged distribution from a uniform distribution to an optimal allocation. Substantial performance improvements over uniform and direct ergodic search methods are shown for a varying number of robots, both on a representative region graph and random directed graphs. The performance benefits are also validated on a ground robot swarm system.



(a) Graph structure (b) $k = 3 \quad (\beta = 0.26)$ (c) $k = 13 \quad (\beta = 0.73)$ (d) $k = 29 \quad (\beta = 0.94)$

Fig. 7: Experimental validation (see accompanying video). (a) shows the graph structure of the system used for physical demonstration. (b)-(d) shows the distribution of the TurtleBot Burger robots at different representative timestamps during the annealing process. The fill colors in the circular regions match the border colors as β increases, showing that the system has correctly identified the optimal target distribution.

REFERENCES

[1] A. Dutta, S. Roy, O. P. Kreidl, and L. Bölöni, "Multi-robot information gathering for precision agriculture: Current state, scope, and challenges," *IEEE Access*, vol. 9, pp. 161 416–161 430, 2021.

[2] B. Wong, W. Marquette, N. Bykov, T. M. Paine, and A. G. Banerjee, "Human-assisted robotic detection of foreign object debris inside confined spaces of marine vessels using probabilistic mapping," *Robot. Auton. Syst.*, vol. 161, p. 104349, 2023.

[3] S. M. T. Islam and X. Hu, "Towards decentralized importance-based multi-UAS path planning for wildfire monitoring," in *Annu. Syst. Syst. Eng. Conf.*, 2022, pp. 67–72.

[4] X.-Y. Dai, J.-Y. Wang, and Q.-H. Meng, "An infotaxis-based odor source searching strategy for a mobile robot equipped with a TDLAS gas sensor," in *Chinese Cont. Conf.*, 2019, pp. 4492–4497.

[5] G. Mathew and I. Mezić, "Metrics for ergodicity and design of ergodic dynamics for multi-agent systems," *Physica. D*, vol. 240, no. 4, pp. 432–442, 2011.

[6] K. B. Naveed, D. Agrawal, C. Vermillion, and D. Panagou, "Eclares: Energy-aware clarity-driven ergodic search," in *IEEE Int. Conf. Robot. Autom.*, 2024, pp. 14 326–14 332.

[7] J. Dokoupil, M. Papež, and P. Václavek, "Comparison of Kalman filters formulated as the statistics of the Normal-inverse-Wishart distribution," in *IEEE Conf. Decision Cont.*, 2015, pp. 5008–5013.

[8] F. J. Humphreys, G. S. Rohrer, and A. D. Rollett, *Recrystallization and related annealing phenomena*, third edition ed. Amsterdam: Elsevier, 2017.

[9] C.-H. Chung, K.-C. Wang, K.-T. Liu, Y.-T. Wu, C.-C. Lin, and C.-Y. Chang, "Path planning algorithm for robotic lawnmower using RTK-GPS localization," in *Int. Symp. Community-centric Syst.*, 2020, pp. 1–4.

[10] N. Shah, U. Dey, and K. Nishimiya, "End-to-end framework for robot lawnmower coverage path planning using cellular decomposition," *arXiv:2506.06028*, 2025.

[11] J. Li, R.-C. Wang, H.-P. Huang, and L.-J. Sun, "Voronoi based area coverage optimization for directional sensor networks," in *Int. Symp. Electronic Commerce Security*, vol. 1, 2009, pp. 488–493.

[12] S. Doodman, M. A. Mostafavi, R. Sengupta, and A. Afghantoloee, "A novel Voronoi-driven optimization approach for point-based sensor network deployment," *IEEE Access*, vol. 13, 2025.

[13] P. Manna, S. Majumder, N. H. Singh, and R. Bose, "A context-aware framework for sensor placement using PSO and Voronoi in dynamic scenarios," in *Int. Conf. Intell. Syst. Advanced Comput. Commun.*, 2025, pp. 1002–1006.

[14] H. Robbins, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, vol. 58, no. 5, pp. 527–535, 1952.

[15] A. Slivkins, "Introduction to multi-armed bandits," *Found. Trends Mach. Learn.*, vol. 12, no. 1–2, p. 1–286, 2019.

[16] T. Zhang, K. Johansson, and N. Li, "Multi-armed bandit learning on a graph," in *Annu. Conf. Inf. Sci. Syst.*, 2023, pp. 1–6.

[17] H. Salman, E. Ayvali, and H. Choset, "Multi-agent ergodic coverage with obstacle avoidance," in *Int. Conf. Autom. Plan. Schedul.*, vol. 27, 2017, pp. 242–249.

[18] S. Ivić, B. Crnković, H. Arbabi, S. Loire, P. Clary, and I. Mezić, "Search strategy in a complex and dynamic environment: The MH370 case," *Sci. Rep.*, vol. 10, no. 1, pp. 19 640–19 640, 2020.

[19] S. Patel, S. Hariharan, P. Dhulipala, M. C. Lin, D. Manocha, H. Xu, and M. Otte, "Multi-agent ergodic coverage in urban environments," in *IEEE Int. Conf. Robot. Autom.*, 2021, pp. 8764–8771.

[20] S. Ivić, A. Sikirica, and B. Crnković, "Constrained multi-agent ergodic area surveying control based on finite element approximation of the potential field," *Eng. Appl. Artif. Intell.*, vol. 116, p. 105441, 2022.

[21] C. Lerch, D. Dong, and I. Abraham, "Safety-critical ergodic exploration in cluttered environments via control barrier functions," in *IEEE Int. Conf. Robot. Autom.*, 2023, pp. 10 205–10 211.

[22] A. Xu, B. Vundurthy, G. Gutow, I. Abraham, J. Schneider, and H. Choset, "Measure preserving flows for ergodic search in convoluted environments," *arXiv preprint arXiv:2409.09164*, 2024.

[23] E. Ayvali, H. Salman, and H. Choset, "Ergodic coverage in constrained environments using stochastic trajectory optimization," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2017, pp. 5204–5210.

[24] I. Abraham, A. Prabhakar, and T. D. Murphey, "An ergodic measure for active learning from equilibrium," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 3, pp. 917–931, 2021.

[25] E. Pignat, J. Silvério, and S. Calinon, "Learning from demonstration using products of experts: Applications to manipulation and task prioritization," *Int. J. Robot. Res.*, vol. 41, no. 2, pp. 163–188, 2022.

[26] S. Shetty, J. Silverio, and S. Calinon, "Ergodic exploration using tensor train: Applications in insertion tasks," *IEEE Trans. Robot.*, vol. 38, no. 2, pp. 906–921, 2022.

[27] B. Crnković, S. Ivić, and M. Zovko, "Fast algorithm for centralized multi-agent maze exploration," *arXiv:2310.02121*, 2023.

[28] B. Shirose, A. Johnson, B. Vundurthy, H. Choset, and M. Travers, "GESCE: Graph-based ergodic search in cluttered environments," in *IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2024, pp. 7611–7616.

[29] B. Wong, R. H. Lee, T. M. Paine, S. Devasia, and A. G. Banerjee, "Rapidly converging time-discounted ergodicity on graphs for active inspection of confined spaces," *arXiv:2503.10853*, 2025.

[30] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Sci.*, vol. 220, no. 4598, pp. 671–680, 1983.

[31] P. Bremaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, ser. Texts in Applied Mathematics. Netherlands: Springer Nature, 2013, vol. 31.

[32] L. Claussmann, M. Revilloud, and S. Glaser, "Simulated annealing-optimized trajectory planning within non-collision nominal intervals for highway autonomous driving," in *IEEE Int. Conf. Robot. Autom.*, 2019, pp. 5922–5928.

[33] W. Shi, Z. He, W. Tang, W. Liu, and Z. Ma, "Path planning of multi-robot systems with Boolean specifications based on simulated annealing," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6091–6098, 2022.

[34] Z. Wen, X. Liu, G. Lu, and J. Liu, "Rapid autonomous exploration of large-scale environments for ground robots based on region partitioning," in *IEEE Int. Conf. Robot. Autom.*, 2025, pp. 13 067–13 073.

[35] J. H. Lienhard, *A Heat Transfer Textbook: Fourth Edition*. Dover Publications, 2013.

[36] K. P. Murphy, "Conjugate Bayesian analysis of the Gaussian distribution," 2007. [Online]. Available: https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf

[37] C. Kim, "Percolation theory." [Online]. Available: https://web.mit.edu/ceder/publications/Percolation.pdf

[38] R. Ghosh, "Ergodic control of multi-agent systems under communication constraints," 2025, DOI: 10.13140/RG.2.2.33017.38245.