# GenPlan: Generative Sequence Models as Adaptive Planners

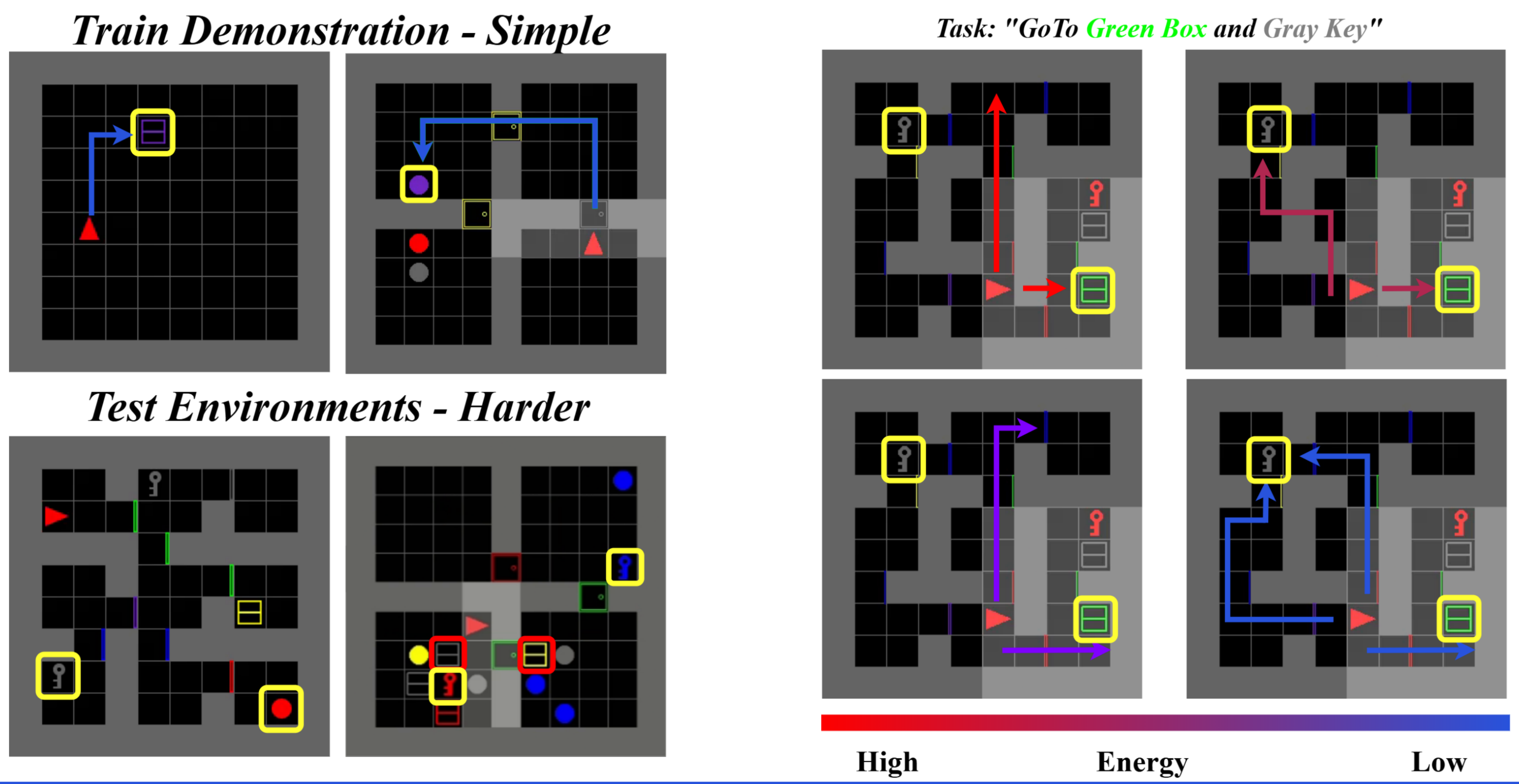

Akash Karthikeyan    Yash Vardhan Pant

Department of Electrical and Computer Engineering, University of Waterloo


## 1. Overview

We tackle the problem of adaptive planning, in multi-task missions. Autonomous agents must generalize to new tasks and environments. This challenge is addressed by GenPlan, which:

A. learns a stochastic policy, facilitating adaptability, to harder and unseen tasks at runtime.
B. is capable of abstract-reasoning and skill composition, all the while being sample-efficient.
C. Is capable of iterative refinement and generating unconditional long-horizon rollouts

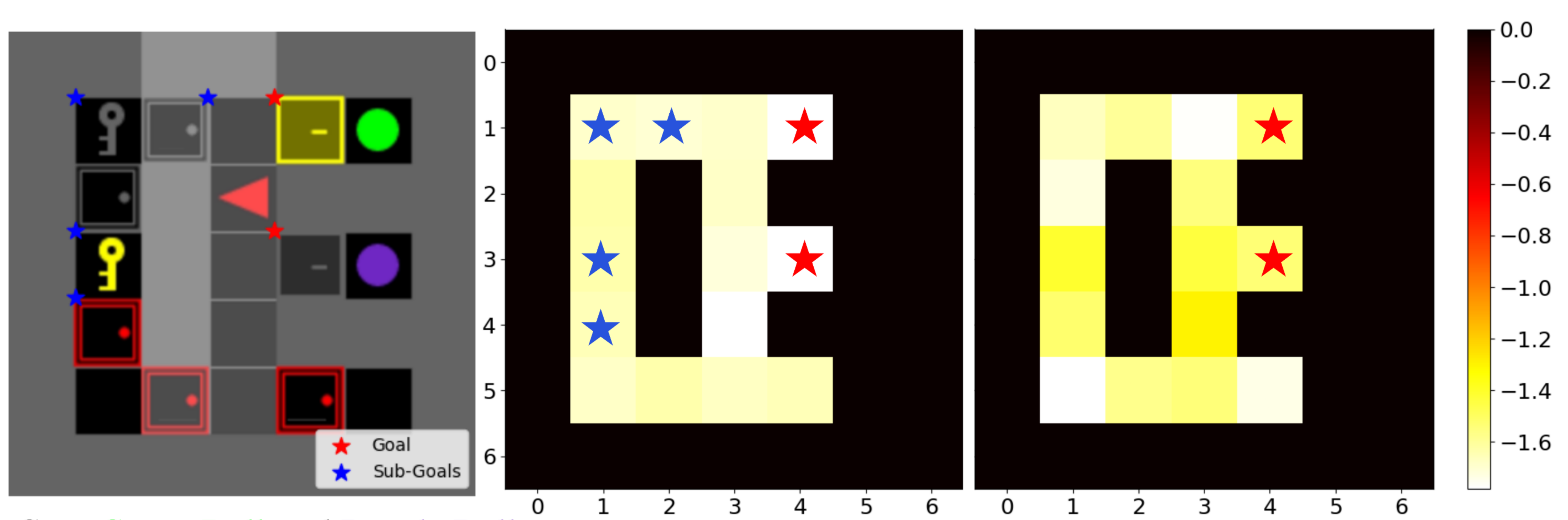



## 2. Problem setup

Given the set of demonstrations $\mathcal{T} = \{(s_0^i, a_0^i, s_1^i, \dots, s_{T^i}^i, a_{T^i}^i)\}_{i=1}^N$, we seek to learn an energy function. This energy function assigns lower energy to an optimal action sequence, while being subject to a lower bound on entropy β. We also jointly learn a goal and state distribution to facilitate unconditional generation. Thus, at inference, we can sample from the energy model by simulating a CTMC, similar to a discrete flow model.

$$\min_{\theta} \mathbb{E}_{\mathbf{a}^0 \sim p_0, \mathbf{o} \sim \mathcal{D}} \Bigg[ \underbrace{\sum_{k=1}^{H} -\log p_{1|t}(\mathbf{a}^1 | \mathbf{a}^0, \mathbf{o})}_{\text{Energy}} \Bigg], \ \text{s.t. } \mathbb{E}_{\mathbf{a}^0 \sim p_0, \mathbf{o} \sim \mathcal{D}} \Bigg[ \underbrace{\sum_{k=1}^{H} \mathcal{H}(p_{1|t}(\mathbf{a} | \mathbf{a}^0, \mathbf{o}))}_{\text{Entropy}} \Bigg] \geq \beta$$

## 3. Energy Landscape

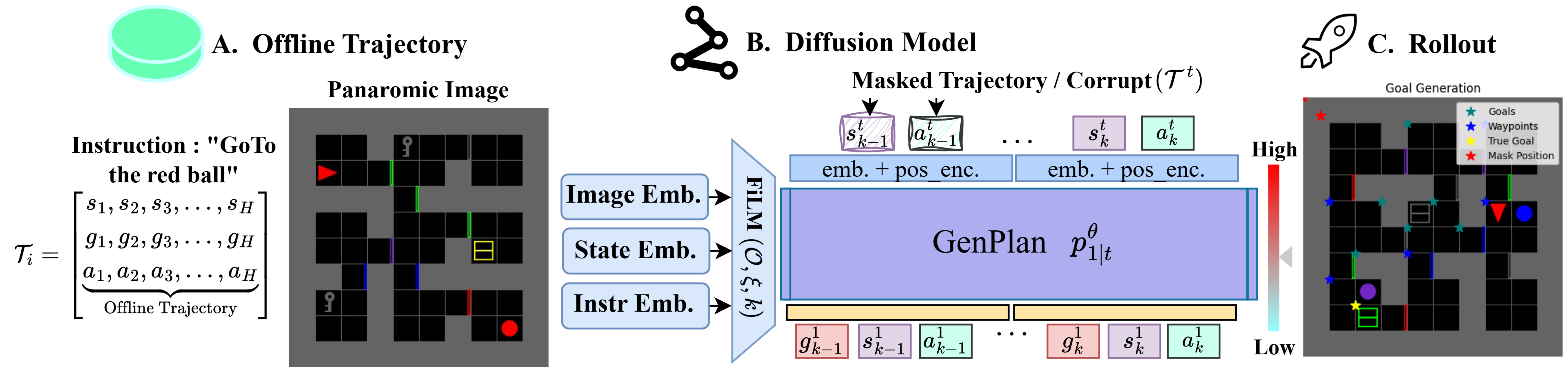



*Goto Green Ball and Purple Ball*    **GenPlan**    **LEAP (Chen *et al.* 2023)**

GenPlan, learns intermediate sub-goals and task hierarchy, we observe that it implicitly assigns minimum energy values to subgoals *(pick-up key, open doors)* required for the task.

## 4. Method

The GenPlan, trained on offline data **(A)**, learns to jointly model action, goal, and state distributions. In **(B)**, the joint denoising model takes in a corrupted trajectory $\tau^0$ and predicts the clean trajectory $\tau^1$ **(C)** Demonstrates the joint inference of goals and actions by simulating the reverse CTMC. Thus reframing planning as iterative denoising through discrete flow models.




## 5. Simulation Studies

We evaluate GenPlan on grid-world environments for trajectory planning, instruction completion and adaption tasks in the following aspects. 1) Generalization to unseen environments and tasks 2) Adaptation to harder tasks, note that during training we only have demonstrations from simpler tasks/ constraints.

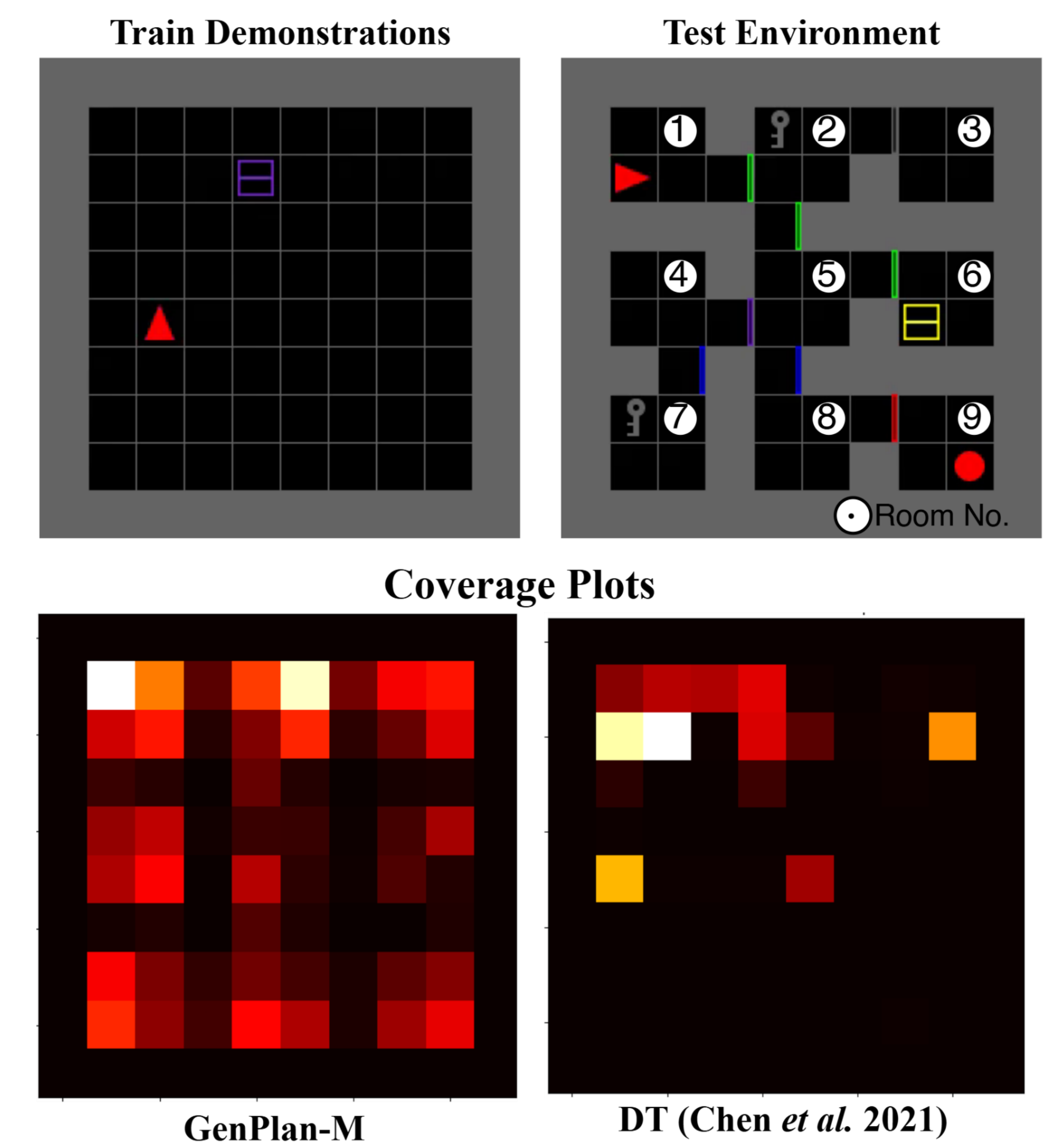
| Env. | Uncond. Rollouts | | | Cond. Rollouts | |
|---|---|---|---|---|---|
| | GP-U | GP-M | LEAP⊖GC | LEAP | DT |
| **Traj. Planning (TP)** | | | | | |
| MazeS4G1 | 52.4% | **62%** | 44% | 49.2% | 46.8% |
| MazeS7G2 | **21.2%** | 19.6% | 3.6% | 4% | 13.6% |
| **Instr. Completion (IC)** | | | | | |
| BlockUn | 13.2% | **16%** | 0% | 0.8% | 0% |
| KeyCorS3R3 | 11.6% | **17.6%** | 0% | 0.4% | 3.6% |

| | Uncond. Rollouts | | | Cond. Rollouts | |
|---|---|---|---|---|---|
| **Environment** | **GP-U** | **GP-M** | **LEAP⊖GC** | **LEAP** | **DT** |
| **Adaptive Planning (AP)** | | | | | |
| MazeS4N3G1 | 56% | **62%** | 44.8% | 48% | 24% |
| MazeS4G2 | 28.8% | **34.8%** | 14% | 18.4% | 3.6% |

Quantitative evaluation on MiniGrid Tasks. Success rates of the models across different environments are presented.

| Sampling / Objective | | GoToObjMazeS4G1 | KeyCorridorS3R3 | DoorsOrder |
|---|---|---|---|---|
| **Energy Models** | Random | 29.2% | 2.8% | 26% |
| | CEM | 38.9% | 6.4% | 28.4% |
| **DDPM** | Diffusion BC | 25.2% | 0.8% | 29.2% |
| | Energy (Gradient) | 2% | 0% | 0% |
| **GenPlan** | Energy+DFM | **62%** | **17.6%** | **35.2%** |

Comparison with other generative baselines on MiniGrid Tasks. We compare various sampling techniques and generative objectives on discrete planning tasks.




State Coverage. State visit frequency is evaluated across 10 unseen maze layouts with varying goal positions (Rooms 1-9), fixing the start position.

## Conclusion

We study the problem of learning to plan from demonstrations, particularly for unseen tasks and environments. We propose GenPlan, an energy-DFM-based planner that learns annealed energy landscapes and uses DFM sampling to iteratively denoise plans. Through simulation studies, we demonstrate how joint energy-based denoising improves performance in complex and long-horizon tasks.