SFU
SIMON FRASER UNIVERSITY

ICRA 2026 VIENNA

Implicit Maximum Likelihood Estimation for Real-time Generative Model Predictive Control

Grayson Lee, Minh Bui, Shuzi Zhou, Yankai Li, Mo Chen, Ke Li

Department of Computing Science, Simon Fraser University

Generative models have seen a lot of success in a wide range of areas such as images, video, and language.

"duck walking in magical forest"

Dall-E 2 by OpenAI

"cute cate riding a pineapple while wearing a race helmet"

Sora by OpenAI

SFU
SIMON FRASER UNIVERSITY

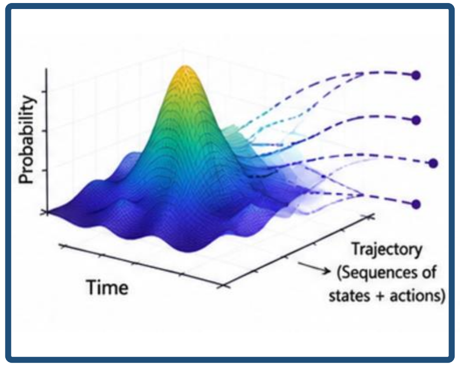
ICRA 2026 VIENNA

Applying Generative Models to Robotics

Difficult Tasks

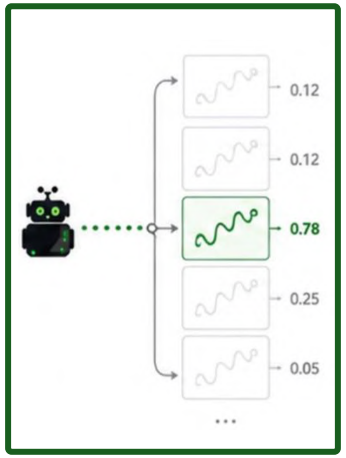
- Tasks like social navigation are inherently **uncertain** and hard to predict
- The **goals** and **rationale** of the humans are unknown
- Humans balance **implicit** and **explicit** social rules which are hard to define

Generative Model



- Learns a probability distribution over entire **trajectories** (sequences of states + actions) from data

Model Predictive Control



Sample many possible futures with the generative model

Evaluate and **refine** each using costs & constraints

Execute the first action of the **best trajectory** and repeat

Two Key Challenges

Multi-Modal Trajectories

Many valid paths to the same goal

High Precision Required

Plan precisely in complex dynamic environments

SFU

SIMON FRASER UNIVERSITY

ICRA 2026 VIENNA

Diffusion Based Methods

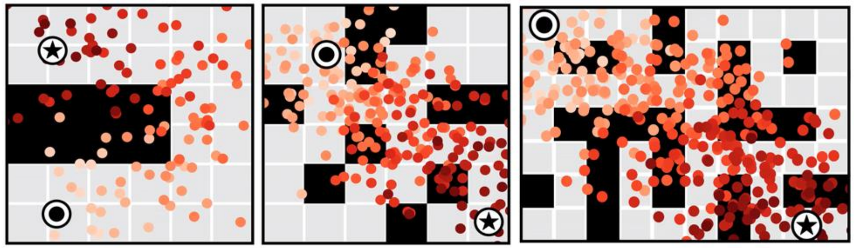
Why Diffusion Works

- Models complex **multi-modal** trajectory distributions
- Generates **diverse** and valid trajectories
- Enables **long-horizon,** goal-conditioned planning

Iterative Denoising Process

Trajectories are generated through many **sequential denoising steps.**

Planning with Diffusion [1]:



Time / Denoising Steps

Challenge for Robotics

- Requires many **denoising steps** (50+ typically)
- High inference **latency (< 10 Hz)**
- Difficult for **real-time** predictive control

Key Challenges

Slow Iterative Sampling
Many sequential refinement steps

High Computational Cost
High compute and optimized for GPU

Limited Real-Time Use
Hard to integrate into real-time MPC

[1]: Janner, Michael, et al. "*Planning with diffusion for flexible behavior synthesis*." Proceedings of the 39th International Conference on Machine Learning (ICML). 2022.

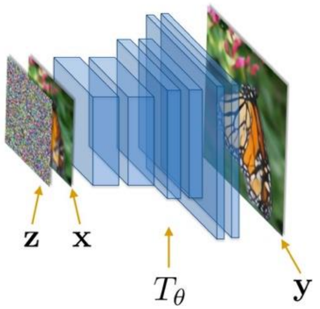
Implicit Maximum Likelihood Estimation (IMLE)

IMLE is a **single-shot** generative model that directly maps noise to samples, enabling fast and efficient trajectory generation.

1. From Noise to Sample in One Step

IMLE directly maps noise to high-quality samples in a **single** forward pass.

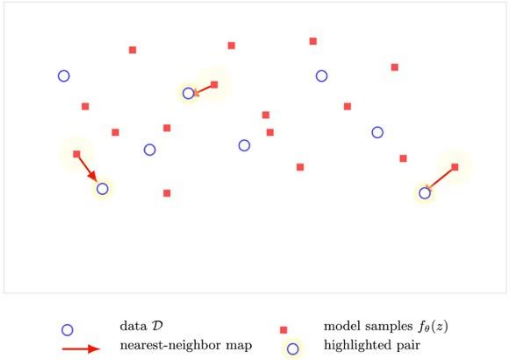
$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$
$$\mathbf{y} = T_\theta(\mathbf{x}, \mathbf{z})$$



Single-shot generation
No iterative denoising required

2. IMLE Objective

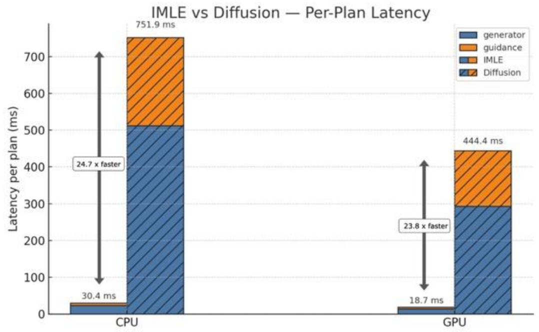
IMLE learns the distribution by **matching and minimizing the distance** of each real sample to a generated sample.



IMLE ensures each real trajectory has a generated sample nearby

3. Faster Inference, Lower Latency

By avoiding iterative sampling, IMLE achieves significantly **lower latency.**



IMLE is 20–25× faster
Ideal for real-time robotic control

Key Benefits of IMLE

Single-Shot Generation

Direct mapping from noise to sample in one forward pass

Mode Coverage

Each real data point has a generated sample

Low Latency

20–25× faster than diffusion methods

Real-Time Ready

Enables fast and responsive predictive control

[2] Li, Ke, et al. "Diverse image synthesis from semantic layouts via conditional IMLE." ICCV 2019, pp. 4220–4229.

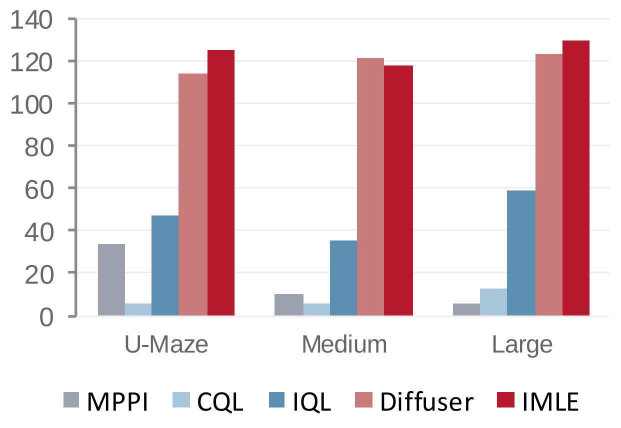
SFU
SIMON FRASER UNIVERSITY

ICRA 2026 VIENNA

Results: Real-Time Performance with Competitive Planning Quality

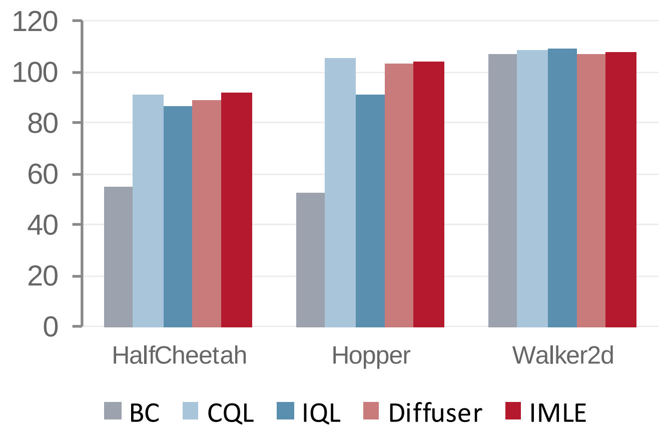
IMLE achieves state-of-the-art planning and control performance while being up to **119× faster** than diffusion-based methods.

1. Planning Performance (Maze2D)



IMLE **matches Diffuser** and outperforms RL baselines by a large margin in sparse-reward planning.

2. Control Performance (MuJoCo)



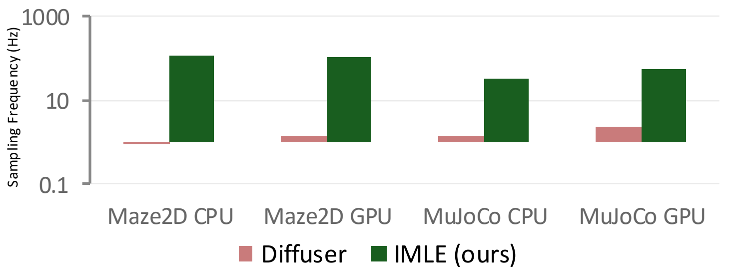
In continuous control, IMLE is **competitive** with strong RL and Diffuser baselines across diverse tasks.

3. Speed: IMLE vs Diffusion

Up to 119× faster than diffusion

Maze2D (CPU)	Maze2D (GPU)
119× faster	**74×** faster

MuJoCo (CPU)	MuJoCo (GPU)
25× faster	**24×** faster



Key Takeaways

State-of-the-Art Performance

Matches or outperforms Diffuser across tasks.

Dramatically Faster

Up to 119× faster sampling frequency than diffusion.

Real-Time Enabled

High-frequency (>10 Hz) planning for responsive robotic control.

Efficient & Practical

Single-shot generation with low latency on CPU and GPU.

Setup: AMD EPYC 9655 Zen 5 (CPU) · NVIDIA H100 2/7 MIG, 20 GB (GPU) · D4RL benchmark · 150 random seeds

SFU
SIMON FRASER UNIVERSITY

ICRA 2026 VIENNA

Real-Time Social Navigation

Single-shot IMLE enables responsive real-time navigation in crowded environments.

SFU
SIMON FRASER UNIVERSITY

ICRA 2026 VIENNA

Conclusion

1 Single-shot trajectory generation

IMLE generates diverse, high-quality trajectories in a single forward pass.

2 Competitive planning quality

Achieves comparable performance to state-of-the-art methods.

3 Real-time deployment at >50 Hz

Enables responsive closed-loop navigation in novel environments.

More Results & Real-World Videos:
https://gmpc-imle.github.io/