# Value Function-Guided MPPI for Hierarchical Planning and Control in Off-Road Autonomous Navigation

Carnegie Mellon University Robotics Institute

Anoushka Alavilli, Raymond Song, Guanya Shi, Jeff Schneider

Robotics Institute, Carnegie Mellon University

LeCAR Learning and Control for Agile Robotics Lab

Auton Lab

## 1. Motivation & Introduction

**Off-road autonomy is challenging** due to steep slopes, ditches, and uneven terrain that can induce unsafe pitch and roll.

**Terrain cost is *controller*-dependent:** For e.g., a steep slope may be unsafe at high speed but traversable under an intelligent control strategy.

**MPPI typically relies on explicit terrain cost maps** that do not directly capture the capabilities of the *controller* executing the trajectory [1, 2]. Running MPPI with these cost maps is computationally expensive. **Reactive RL controllers capture execution-aware terrain effects** but lack the long-horizon reasoning of a planner.

**Our idea:** use the RL controller's **value function as an execution-aware MPPI cost**, incorporating *controller* feedback into *planning* while reducing reliance on expensive terrain cost maps.

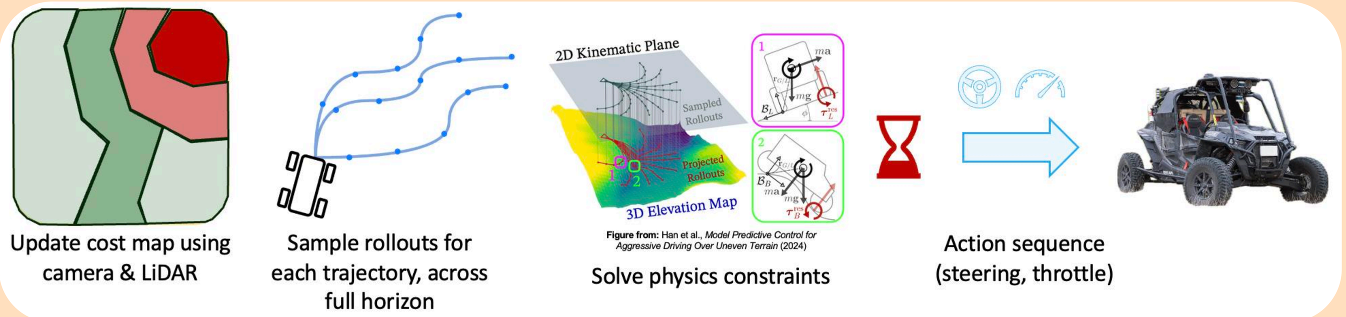


**Fig. 1** Traditional MPPI control often requires expensive computations. These create a bottleneck that limits the speed of the MPPI controller.

## 2. Methods

**1. Train a low-level RL controller** $\pi_\theta(s)$ to track waypoint sequences using local terrain geometry. Its critic $V_\phi(s)$ estimates the expected return, where the reward function encourages waypoint progress while penalizing damage, vertical jerk, and acceleration.

$$r(s,a) = \beta_1 \Delta \text{waypt dist} + \beta_2 \Delta \text{damage} + \beta_3 \text{clip}(\text{jerk}_z) + \beta_4 \text{clip}(\text{acceleration}_z)$$

**2. Use MPPI to sample candidate waypoint trajectories** and score each trajectory using both goal progress and the RL critic as an execution-aware residual cost.

$$c(w_i, o_i) = \lambda_{\text{goal}} \|w_i - g\|_2 + \lambda_{\text{res}} \alpha_i \left(-V_{\text{ppo}}(o_i)\right)$$

$$C(\tau) = \sum_{i=1}^{N} c(w_i, o_i)$$

**3. Online adaptation:** driving experience is used to continually fine-tune the critic so that the planning cost adapts to terrain encountered at deployment.

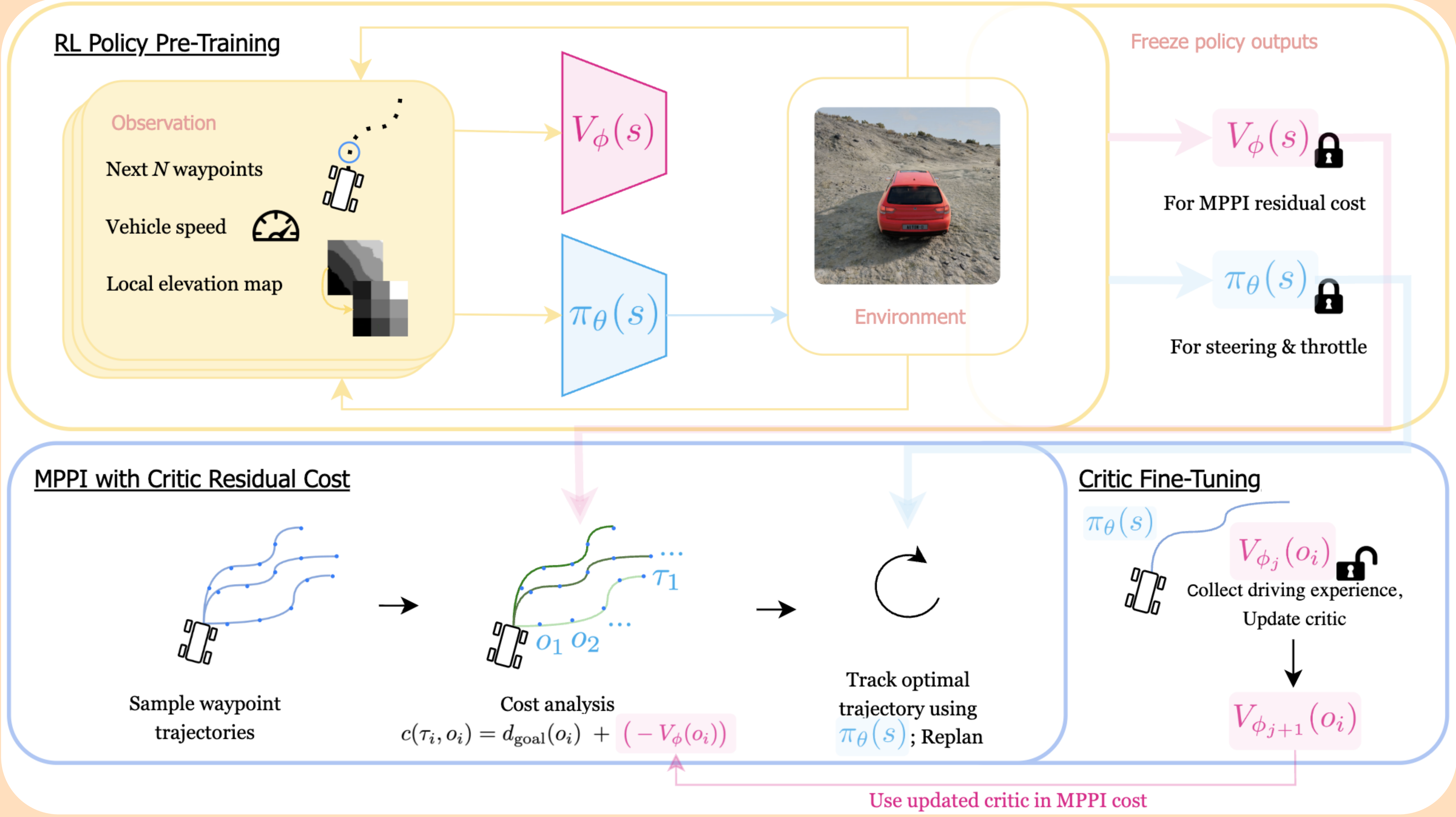


**Fig. 2.** Value-guided planning framework. A **pre-trained RL policy** provides an actor ($\pi_\theta(s)$) and value function ($V_\phi(s)$). **MPPI** evaluates sampled trajectories using goal progress and ($-V_\phi(s)$) as a **control-aware cost**. The selected trajectory is executed by the policy, with online experience optionally used to update the value function.

## 3. Results

| Metric | MPPI | Ours | Ours Online |
|---|---|---|---|
| *Task Performance* | | | |
| Success (sum) | 35 (55%) | **55 (86%)** | 52 (81%) |
| Failure | 29 (45%) | 9 (14%) | 12 (19%) |
| Stuck | 28 | 7 | 6 |
| Timeout | 1 | 2 | 6 |
| Dist. Traveled (m) | 5495 | 9252 | $1.29 \times 10^5$ |
| Displacement (m) | 5157 | 6966 | 6813 |
| *Z-axis Motion* ($\times 10^2$) | | | |
| $z$-jerk p10 | 0.17 | 0.14 | **0.12** |
| $z$-jerk p90 | 6.60 | 2.75 | **2.60** |
| $z$-jerk mean | 2.9 ± 5.9 | 1.27 ± 1.4 | **1.15 ± 1.3** |
| $z$-accel p10 | 1.2 | 0.6 | **0.4** |
| $z$-accel p90 | 32 | 12 | **11** |
| $z$-accel mean | 14 ± 17 | 5.4 ± 5.9 | **4.8 ± 5.4** |
| *Orientation (rad.)* ($\times 10^2$) | | | |
| Pitch p10 | 2.0 | **1.0** | **1.0** |
| Pitch p90 | 27 | **20** | 21 |
| Pitch mean | 13 ± 9.9 | **8.0 ± 7.6** | **8.0 ± 7.9** |
| Roll p10 | 2.0 | **1.0** | **1.0** |
| Roll p90 | 26 | 19 | **18** |
| Roll mean | 12 ± 9.1 | **8.0 ± 7.9** | 9.0 ± 7.0 |

**Table 1**

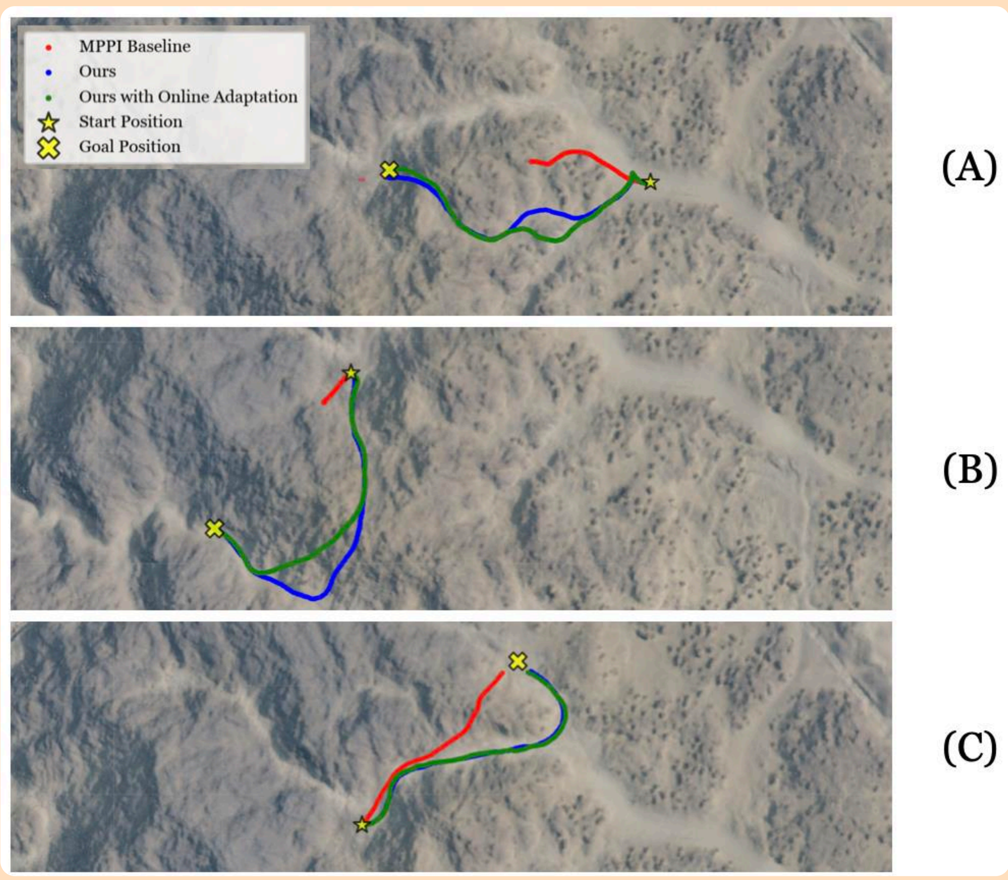


**Fig. 3**

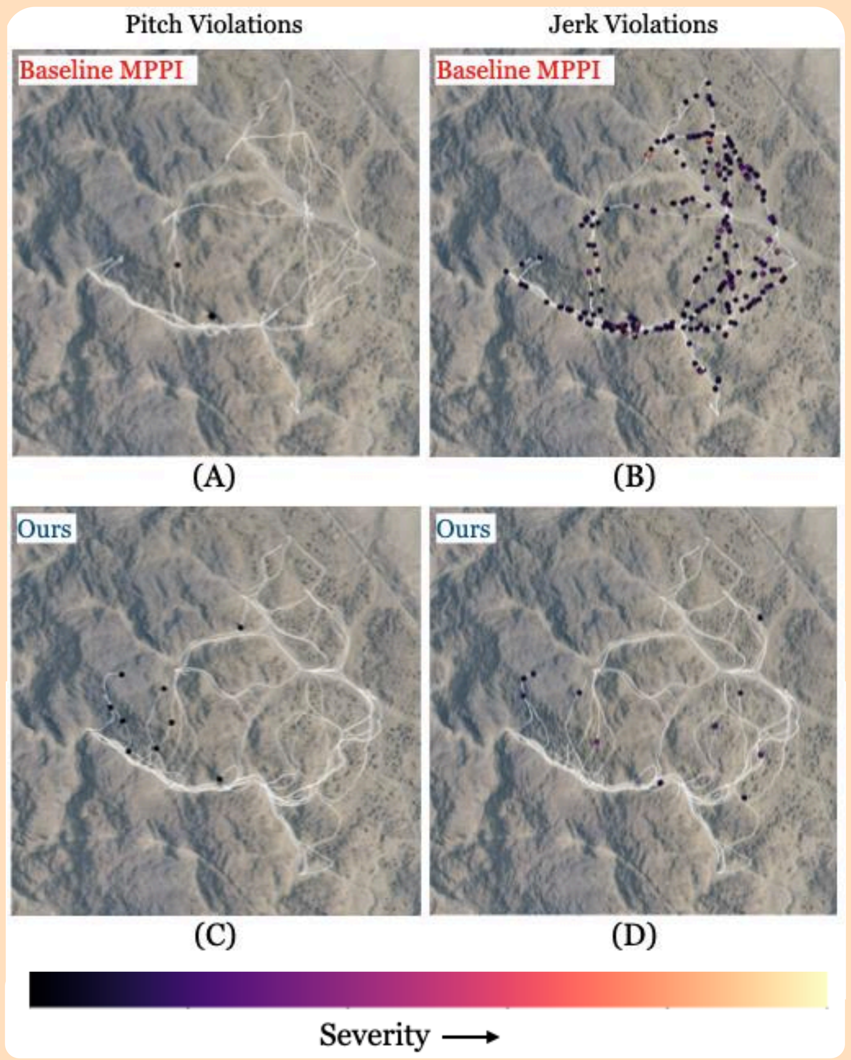


**Fig. 4**

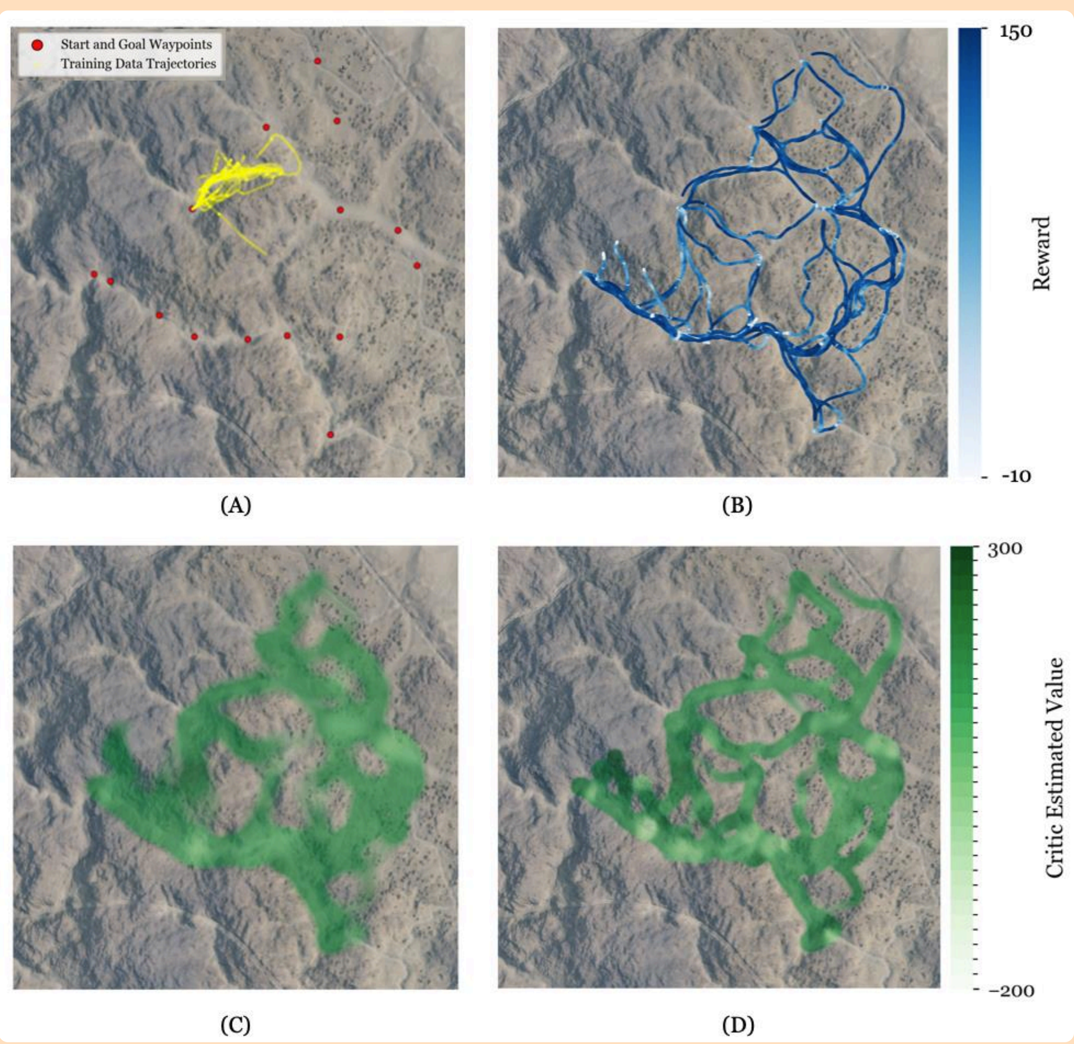


**Fig. 5**

- **Table 1:** Comparison of navigation performance across 64 randomly sampled start-goal pairs (unseen during training).
- **Fig 2:** Qualitative illustration of our algorithm's ability to follow existing trails, which provide safer traversal than off-trail terrain.
- **Fig 3:** Comparison of pitch and vertical jerk violations across all 64 evaluation trajectories. Pitch violations exceed 0.4 rad (22.9 deg.). Vertical jerk violations exceed 10e−2 m/s³.
- **Fig 4:** (A) Training data coverage, (B) trajectories traversed by ours with online adaptation, (C) and (D) value estimates of the traversed terrain before and after online adaptation.

## 4. References & Acknowledgments

[1] Tyler Han, Alex Liu, Anqi Li, Alexander Spitzer, Guanya Shi, and Byron Boots. Model predictive control for aggressive driving over uneven terrain. In Robotics: Science and Systems (RSS), Delft, Netherlands, 2024.

[2] Samuel Triest, Mateo Guaman Castro, Parv Maheshwari, Matthew Sivaprakasam, Wenshan Wang, and Sebastian Scherer. Learning risk-aware costmaps via inverse reinforcement learning for off-road navigation. In 2023 IEEE International Conference on Robotics and Automation (ICRA), pages 924–930, 2023.

This work was supported by the U.S. Army Research Office and the U.S. Army Futures Command under Contract No. W519TC-23-C-0030 and NSF #2512805 .