

Toward Safe and Trustworthy Aerial Robotics Through Hybrid MPC–RL

Rudolf Reiter, Angel Romero, and Davide Scaramuzza

ICRA 2026 Workshop on AI-Driven Safe Aerial Robotics

Vienna, Austria · June 5, 2026

Robotics and Perception Group, University of Zurich



**University of
Zurich**^{UZH}



Motivation

MPC-RL Synthesis Landscape

Actor-Critic Model Predictive Control

Experimental Results

Open Challenges

Motivation

Why Hybrid MPC–RL for Aerial Robotics?

Safety-critical aerial autonomy demands controllers that are:

- High-performing
- Verifiable
- Robust

Two dominant paradigms:

- **Model Predictive Control**
- **Reinforcement Learning**

Key insight: Their weaknesses are each other's strengths \Rightarrow combine them



Agile drone racing at up to 21 m/s — an ideal benchmark for safe, high-performance control.

MPC-RL Synthesis Landscape

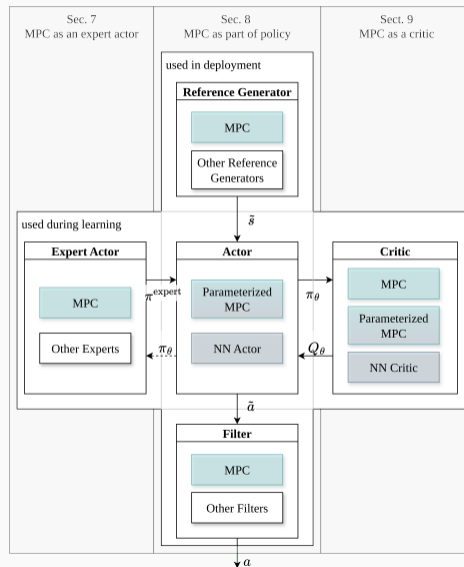
Taxonomy of MPC–RL Combinations

The survey classifies combinations along two axes:

Three main roles of MPC:

1. **Expert actor:** MPC provides demonstrations for RL
2. **Within the policy:**
 - Reference generator (pre-processing)
 - Actor (MPC as policy layer)
 - Safety filter (post-processing)
3. **As critic:** value function derived from MPC cost

Reiter et al., “Synthesis of MPC and RL,” *Annu. Rev. Control*, 2026.



Taxonomy of MPC–RL Combinations

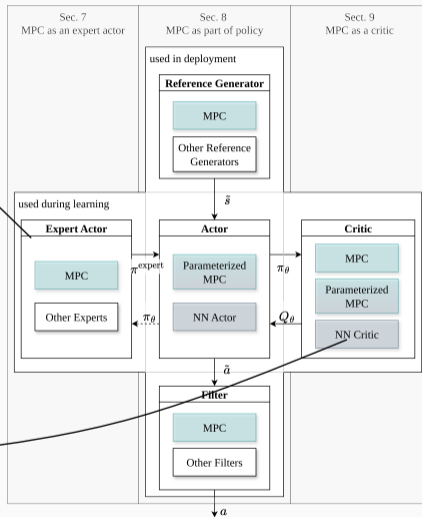
Deep Drone Acrobatics

Elia Kaufmann^{*†}, Antonio Loquercio^{*†}, René Ranftl[†], Matthias Müller[†], Vladen Koltun[†], Davide Scaramuzza[†]



A Painless Deterministic Policy Gradient Method for Learning-based MPC

Akhil S Anand, Dirk Reinhardt, Shambhuraj Sawant, Jan Tommy Gravdahl, Sebastien Gros



Taxonomy of MPC–RL Combinations



DTC: Deep Tracking Control

FABIAN JENELTEN,^{1*} JUNZHE HE,¹ FARBOD FARSHIDIAN,² AND MARCO HUTTER¹

¹ Fabian Jenelten, Junzhe He, Farbod Farshidian, Marco Hutter



Contents lists available at ScienceDirect

Automatica

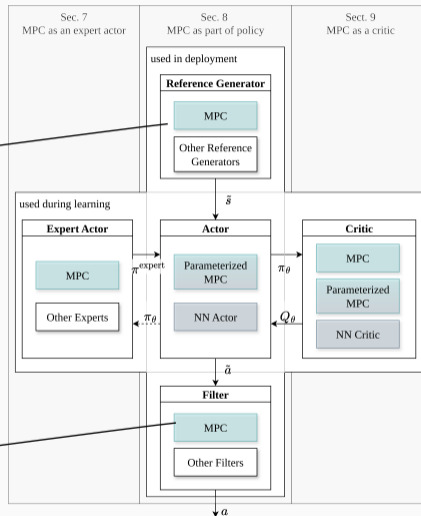
journal homepage: www.elsevier.com/locate/automatica



A predictive safety filter for learning-based control of constrained nonlinear dynamical systems[☆]

Kim Peter Wabersich^{*}, Melanie N. Zeilinger

Institute for Dynamic Systems and Control, ETH Zurich, Zurich, Switzerland



MPC as the Actor: Two Learning Paradigms - Examples

IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, VOL. 34, NO. 1, JANUARY 2026

395

AC4MPC: Actor-Critic Reinforcement Learning for Guiding Model Predictive Control

Rudolf Reiter[✉], Andrea Ghezzi[✉], Katrin Baumgärtner[✉], Jasper Hoffmann, Robert D. McAllister[✉], and Moritz Diehl[✉], *Member, IEEE*



Actor-Critic Model Predictive Control: Differentiable Optimization meets Reinforcement Learning for Agile Flight

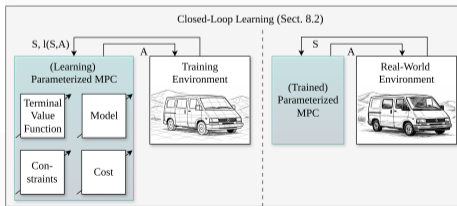
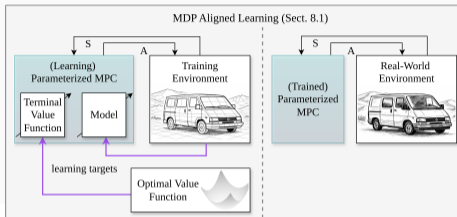
Angel Romero, Elie Aljalbout, Yunlong Song, Davide Scaramuzza



MPC within the Deployed Policy (Sect. 8)

Architecture During Learning

Architecture During Deployment



Implementation Design Choices

What is the actual state?

- MPC may be warm-started → solver state carries history
- Markov property may be violated

Where to add exploration noise?

- On the MPC parameters (cost matrices)?
- On the actions (after MPC)?

Which RL algorithm?

- On-policy: PPO (stable, sample-heavy)
- Off-policy: SAC (sample-efficient, complex)

How to combine NN and MPC?

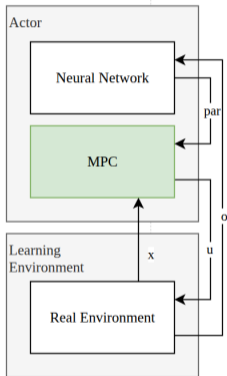
- Learned model (integrated)
- Reference guidance (hierarchical)
- Parallel correction
- Algorithmic (warm-start, sampling)

How to train with the MPC?

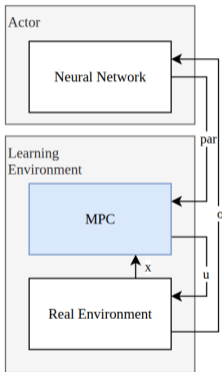
- **Differentiable MPC:** backpropagate gradients through the solver
- **MPC as environment:** treat MPC as black box, learn via RL

Differentiate Through MPC?

MPC as Part of Actor



MPC as Part of Environment



A Hierarchical Approach for Strategic Motion Planning in Autonomous Racing

Rudolf Reiter¹, Jasper Hoffmann², Joschka Boedecker² and Moritz Diehl^{1,3}



Actor-Critic Model Predictive Control: Differentiable Optimization meets Reinforcement Learning for Agile Flight



Angel Romero, Elie Ajalbout, Yunlong Song, Davide Scaramuzza



Actor-Critic Model Predictive Control

AC-MPC: Architecture

Core idea: Add differentiable MPC as actor output layer.

Two components:

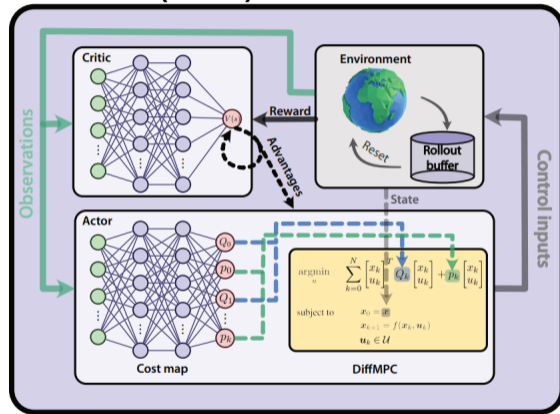
1. **Neural cost map:** $o_t \rightarrow (Q_k, R_k)$
2. **Differentiable MPC:** solves QP with quadrotor dynamics \rightarrow optimal action

Policy gradients propagate *through* the MPC \Rightarrow fully end-to-end RL training.

First RL-MPC variant with differentiable MPC verified in high-performance real-world robotics.

Romero et al., *IEEE T-RO*, 2025.

AC-MPC (Ours)



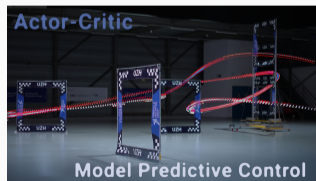
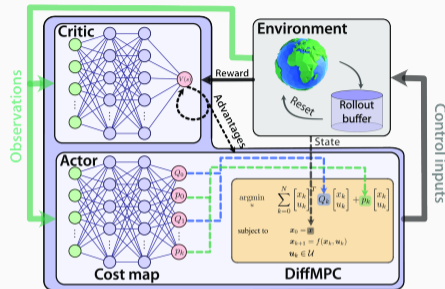
AC-MPC: Why It Works

Structural benefits:

- Physics-based prior \rightarrow feasible actions from start
- MPC layer provides guarantees

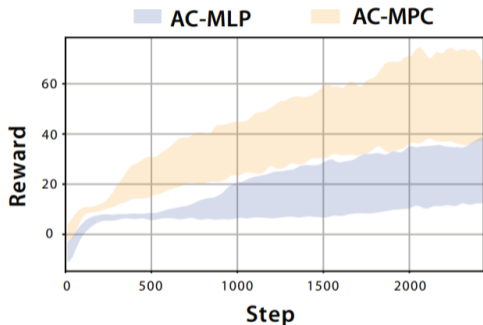
Model Predictive Value Expansion (MPVE):

- Reuse MPC rollouts as extra training signal for critic
- Better sample efficiency, same compute



Experimental Results

Training Efficiency: AC-MPC vs. AC-MLP



AC-MPC learns competitive racing policies significantly faster than an MLP actor.

Key observations:

- MPC prior accelerates early training
- Matches final RL performance
- Structured exploration: physics-based actions from the start

Why faster?

- MPC encodes dynamics as inductive bias
- MPVE provides additional critic training signal
- RL fine-tunes the cost, not the full policy

Comparative Properties

Property	MPC	RL	AC-MPC
Sample efficiency	+	-	+
Task performance	o	+	+
OOD robustness	+	-	+
Constraint handling	+	-	o
Online adaptability	+	-	+
Interpretability	+	-	o
Engineering effort	-	o	o

MPC-RL synthesis consistently retains the best of both worlds.

Open Challenges

1. Constraint Handling

- Caveat: AC-MPC only considers *input* constraints
- State constraints are essential for safety

2. World Models

- Create many new opportunities *and* challenges in this setting
- Local minima, large model size, inference time

3. GPU-Accelerated MPC Solvers

- Scaling differentiable MPC to GPU-parallel RL training pipelines

Exciting open problems — great opportunities for this community!

Thank you!

Rudolf Reiter, Angel Romero, Davide Scaramuzza
Robotics and Perception Group, University of Zurich



Survey: Reiter et al.,
Annu. Rev. Control, 2026



AC-MPC: Romero et al.,
IEEE T-RO, 2025