Presented during NeurIPS 2026, Atlanta, December 12, 2026

Policy-DRIFT: Dynamic Reward-Informed Flow Trajectory Steering

Atharva Mahajan[1,*], Abhijeet Vishwasrao[1,*], Yuning Wang[1], Ricardo Vinuesa[1]

[1]Department of Aerospace Engineering, University of Michigan, Ann Arbor
[*]Equal contribution

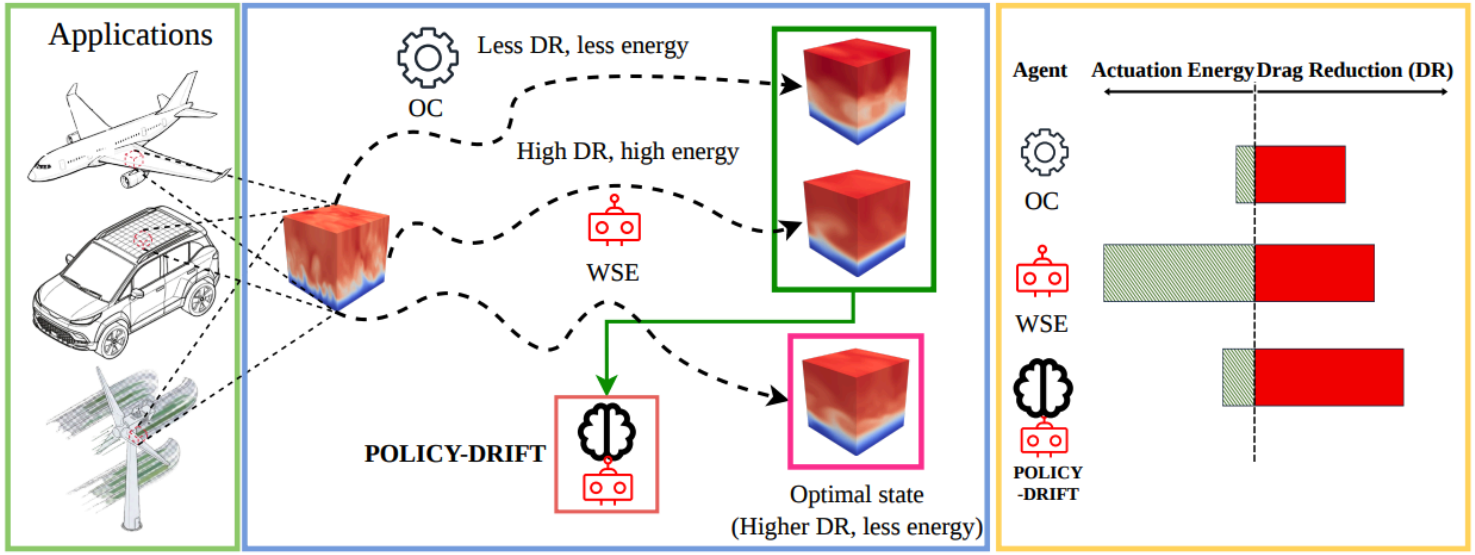
UNIVERSITY OF MICHIGAN

VINUESA LAB

Drag control with reinforcement learning

A scalar reward caps what a DRL controller can learn

- **Skin friction: about half of aircraft drag [1]**
  A few percent less drag
  already saves fuel and CO_2
- **DRL: the leading real-time controller [2]**
  Its ceiling is set by the reward,
  a scalar proxy of the physics
- **Reward gradients pass through the policy**
  High drag reduction is bought
  with high actuation energy



The drag-energy trade-off of known controllers, and the state we aim for

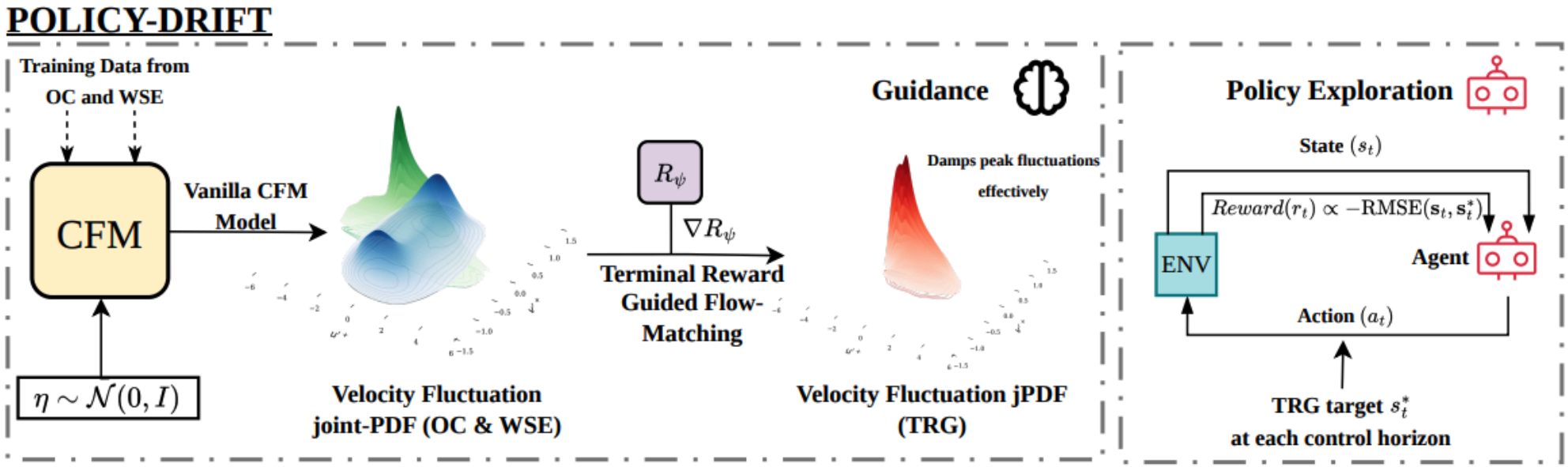
Our solution, Policy-DRIFT: the reward steers a generative model of the flow, not the policy

A learned flow model proposes a realistic flow with less drag and less energy, and a small controller learns to steer toward it

[1] R. Vinuesa et al., The role of artificial intelligence in achieving the sustainable development goals, Nature Communications 11:233 (2020)
[2] L. Guastoni et al., Deep reinforcement learning for turbulent drag reduction in channel flows, The European Physical Journal E 46:27 (2023)

Drag control by steering a flow model

The reward guides a generative flow model, never the policy gradient [1]



1. Flow model

Learns $p(\mathbf{u}_1 \mid \mathbf{u}_0)$ from three control regimes

2. Terminal reward guidance

Nudge, then a flow-model step

$$\tilde{u}_s^+ = \tilde{u}_s + \gamma\, \delta s\, \nabla R_\psi(\tilde{u}_s, s)$$

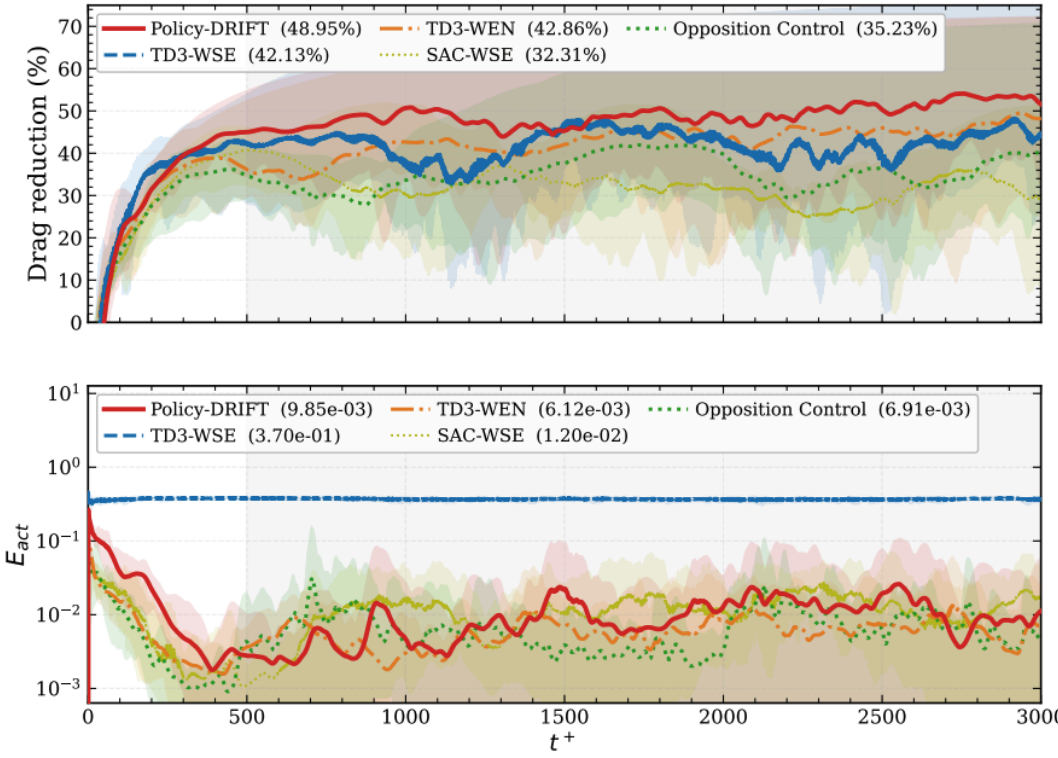
3. Tracking policy

TD3 tracks the guided target

[1] A. Mahajan & A. Vishwasrao et al., Policy-DRIFT: dynamic reward-informed flow trajectory steering, Advances in Neural Information Processing Systems, in press (2026)

Drag control: closed-loop results

48.95 % drag reduction at ~37× less actuation energy than TD3-WSE



Method	DR (%)	E_{act} ($\times 10^{-1}$)	$\bar{R}$
Policy-DRIFT	**48.95**	0.0990	**0.480**
TD3-WEN	42.86	**0.0609**	0.423
TD3-WSE [1]	42.13	3.692	0.052
SAC-WSE	32.31	0.1193	0.311
Opposition control [2]	35.23	0.0685	0.345

Mean over six held-out initial conditions, plot band min to max
WSE: reward DR, WEN: reward DR − E_{act}, $\bar{R}$ = DR − E_{act}

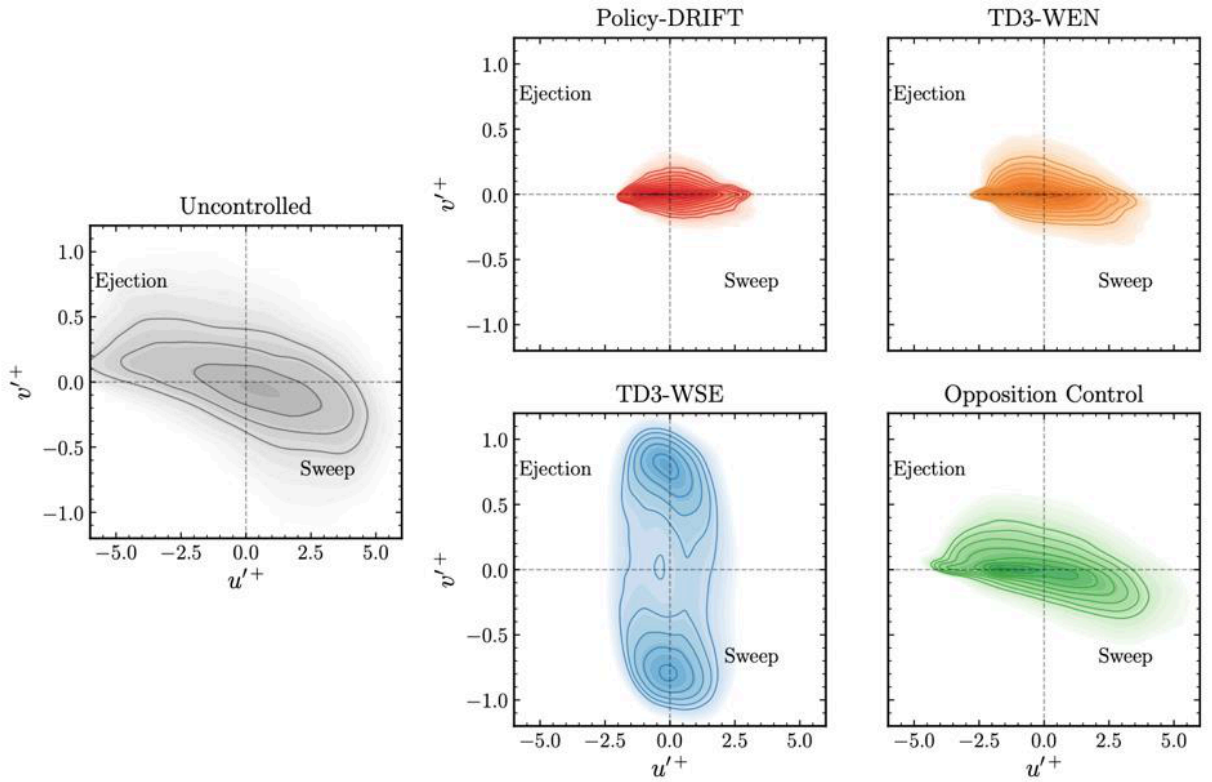
- Beats DRL with the same objective (TD3-WEN) by 6.1 percentage points
- Beats all baselines from $t^+ \approx 200$ on
- Close to the theoretical upper bound: optimal control with full DNS access reaches > 50 % drag reduction [3]

[1] L. Guastoni et al., Deep reinforcement learning for turbulent drag reduction in channel flows, The European Physical Journal E 46:27 (2023)
[2] H. Choi, P. Moin, J. Kim, Active turbulence control for drag reduction in wall-bounded flows, Journal of Fluid Mechanics 262:75 (1994)
[3] T. R. Bewley, P. Moin, R. Temam, DNS-based predictive control of turbulence: an optimal benchmark for feedback algorithms, Journal of Fluid Mechanics 447:179 (2001)

Drag control: physics and outlook

Policy-DRIFT damps ejections and sweeps without over-actuating

- **As compact as TD3-WEN, as damped as TD3-WSE**
  TD3-WSE over-actuates (bimodal v'^{+}), TD3-WEN damps less
- **The reward acts on the flow model, not the policy**
  A new objective retrains only R_ψ, and the deployed policy never queries the flow model
- **Next: higher Re and complex geometries**
  The same framework extends toward turbulent wings [2] and experimental data



Joint PDF of (u'^{+}, v'^{+}) at $y^{+} = 15$ [1]

[1] A. Mahajan & A. Vishwasrao et al., Policy-DRIFT: dynamic reward-informed flow trajectory steering, Advances in Neural Information Processing Systems, in press (2026)
[2] Y. Wang et al., Physics-guided surrogate learning enables zero-shot control of turbulent wings, arXiv:2604.09434 (2026)