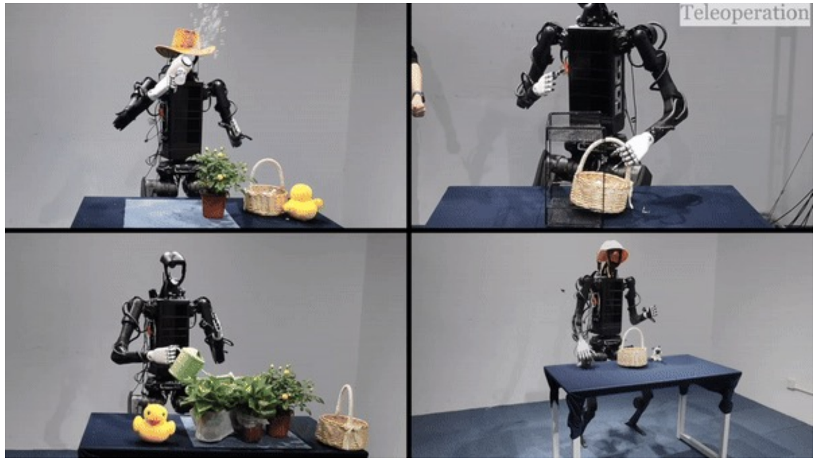
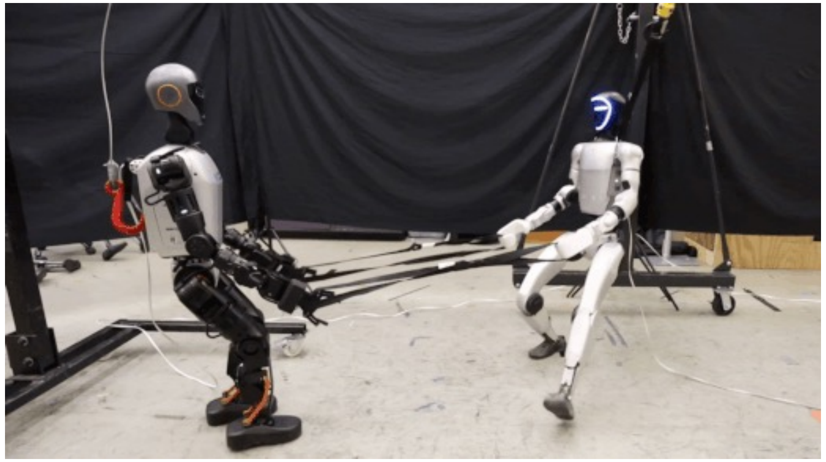
From *Sim2Real* 1.0 to 4.0 for Humanoid Whole-Body Control and Loco-Manipulation



OmniH2O (CoRL'24)
https://omni.human2humanoid.com/

ASAP (RSS'25)
https://agile.human2humanoid.com/



FALCON
https://lecar-lab.github.io/falcon-humanoid/

Guanya Shi

Assistant Professor, Robotics Institute, CMU

https://lecar-lab.github.io/

LeCAR
Learning and Control for Agile Robotics Lab

Carnegie Mellon University
Robotics Institute

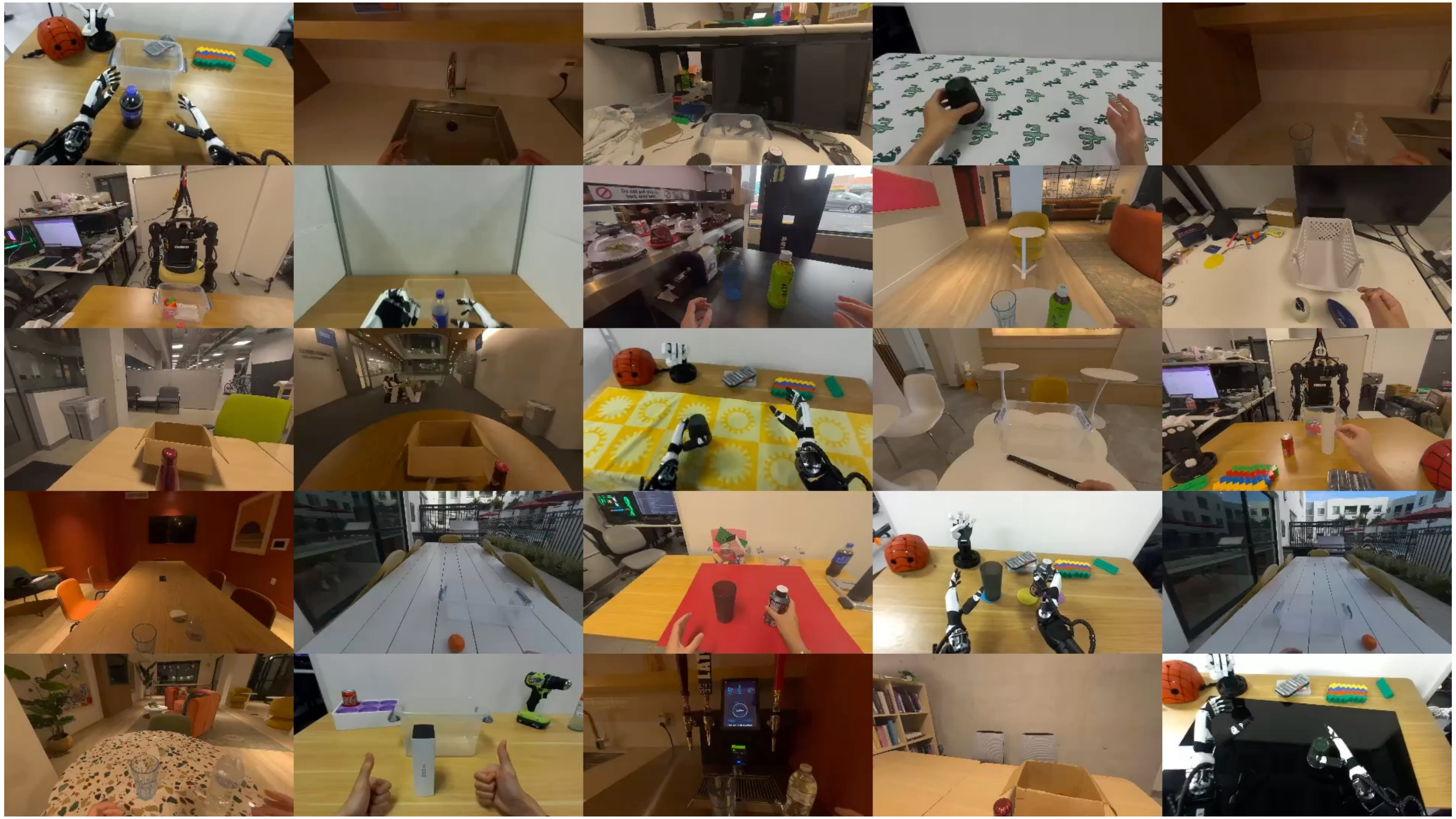
Teleoperation or Learning from Videos Seems Really Promising

- Basic recipe: behavior cloning from labeled actions
  - Action space: joint angle or end-effector pose
  - Low-level control is simple and accurate (PD or IK + PD)

Physical Intelligence $\pi 0.5$ (VLA)
Learning from teleoperation data

Tesla Optimus
Learning from mixed teleoperation & human video data

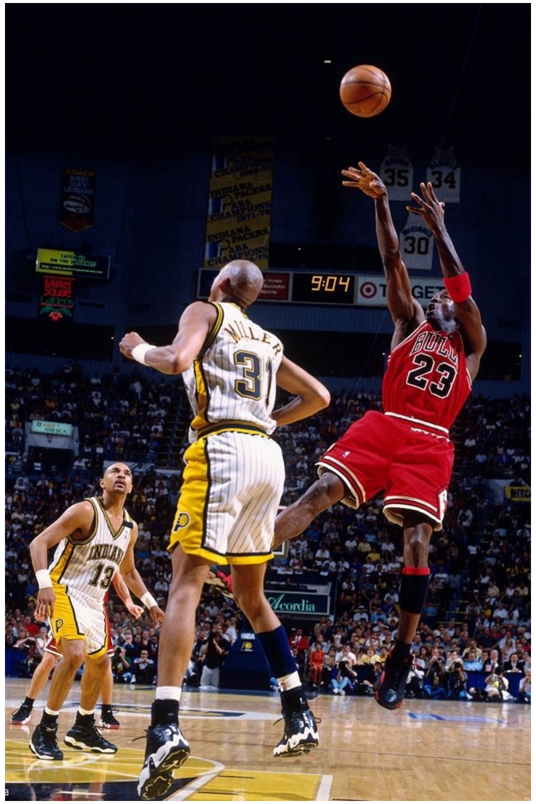
Teleoperation or Learning from Videos Seems Really Promising



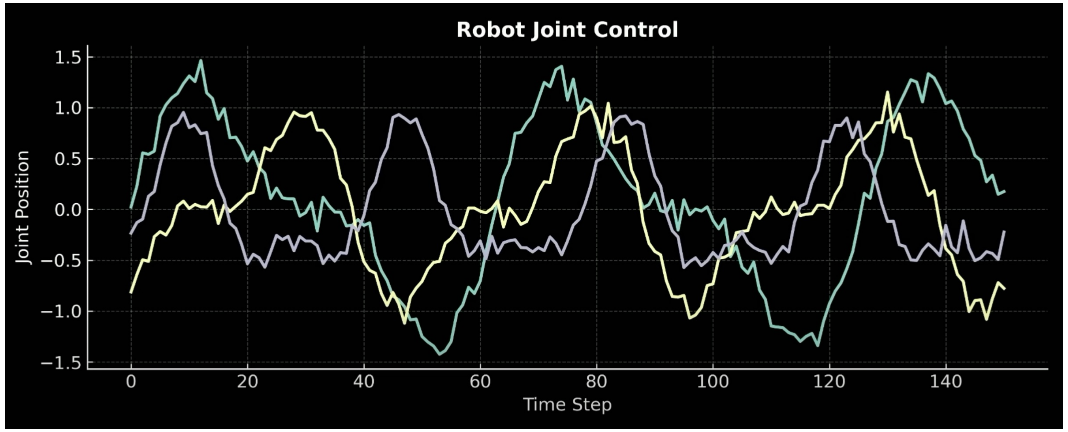
Humanoid Policy ~ Human Policy
(human data and humanoid data co-training)
https://human-as-robot.github.io/

… How About Tasks Involving Whole-Body Agility?

- For those tasks, *impossible or extremely hard* to:
  - Teleoperate (if you can, you already solve the problem)
  - Get labeled action (imagine ask MJ: “I wanna learn fadeaway jumper. Could you tell me your joint trajectories?”)
  - Use simple low-level controllers for tracking



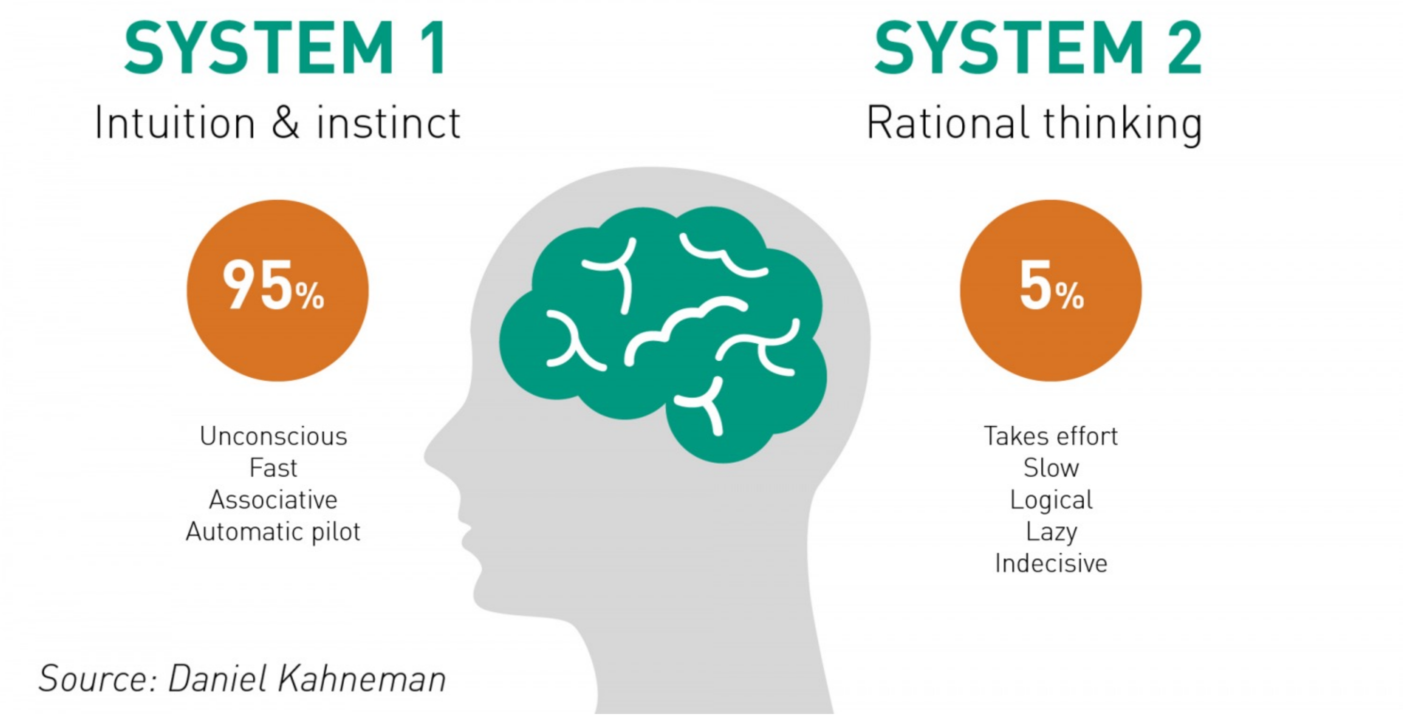
≠



… How About Tasks Involving Whole-Body Agility?

- ❑ Most dexterity and agility (especially whole-body) come from **system 1**
- ❑ (I suspect) *Most human teleoperation involves little system 1*

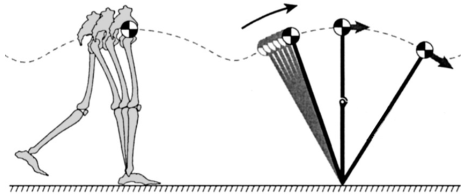
- ❑ How to learn **system 1** agility and dexterity?
    - We need a “model/simulator” and **sim2real** learning!



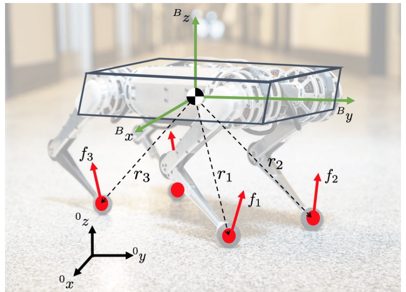
Source: Daniel Kahneman

Sim2Real 1.0: Simplified Model + Online Reasoning

- The control community has been doing sim2real for many decades!



Inverted pendulum model



Single rigid body model

Simulator: Reduced-ordered models

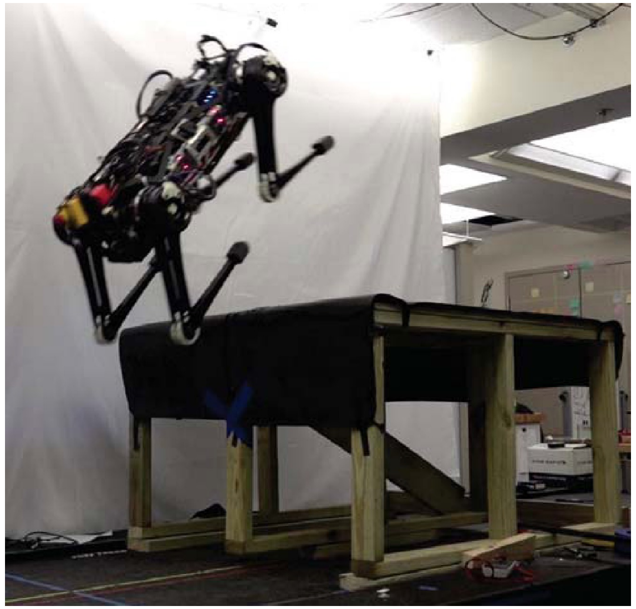
$$\min_{\boldsymbol{x}} \sum_{k=0}^{N-1} l(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) + l_N(\boldsymbol{x}_N, \boldsymbol{\theta})$$

subject to

$$\boldsymbol{x}_0 = \hat{\boldsymbol{x}}$$
$$\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \ \forall\ k \in \{0, ..., N-1\}$$
$$\boldsymbol{g}(\boldsymbol{x}_k, \boldsymbol{u}_k, \boldsymbol{\theta}) \leq \boldsymbol{0} \ \forall\ k \in \{0, ..., N-1\}$$
$$\boldsymbol{g}_N(\boldsymbol{x}_N, \boldsymbol{\theta}) \leq \boldsymbol{0},$$

Online model predictive control

a_t



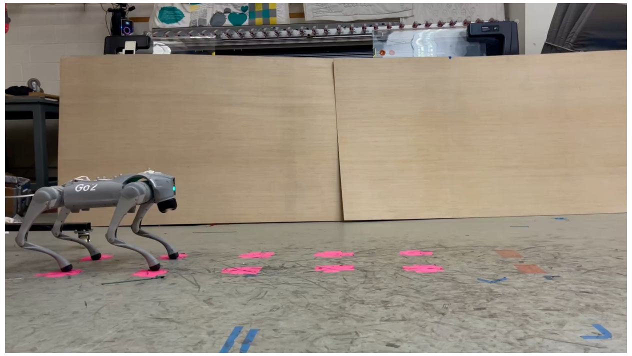
[Nguyen et al., ICRA'19]

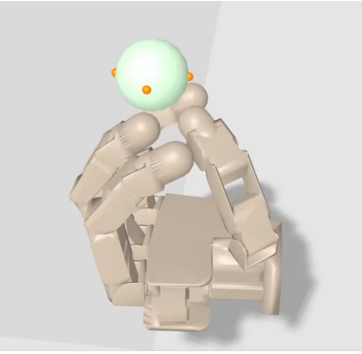
- What is fascinating (but also foolish): no "pretraining", 100% rely on very fast (>100Hz) online reasoning

Sim2Real 1.5: Full-Order Simulators + Online Reasoning

- We can do full-order MPC now using advanced sampling-based methods (e.g., DIAL-MPC)
- However, slow and require state estimation

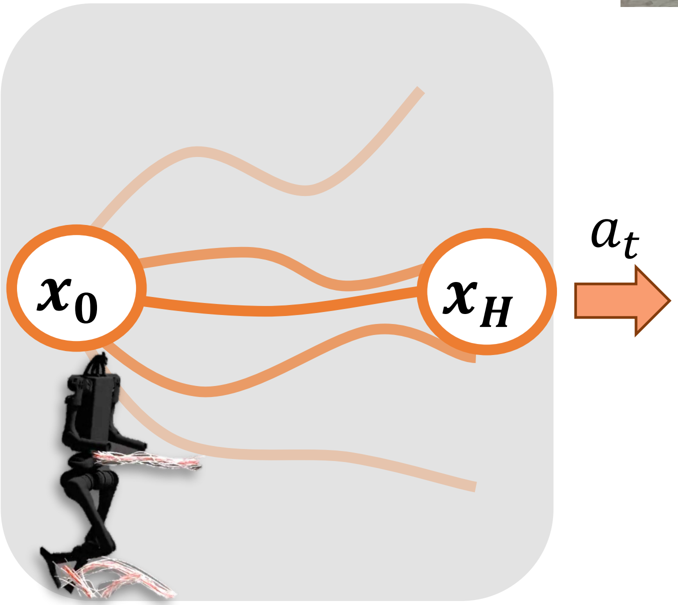
Theorem (informal) [Pan et al., NeurIPS'24][Xue et al., ICRA'25]
As $N \to \infty$, $U^+ = U + \Sigma \cdot \nabla \log p_\Sigma(U)$ where $p_\Sigma = p_0 * \mathcal{N}(0, \Sigma)$



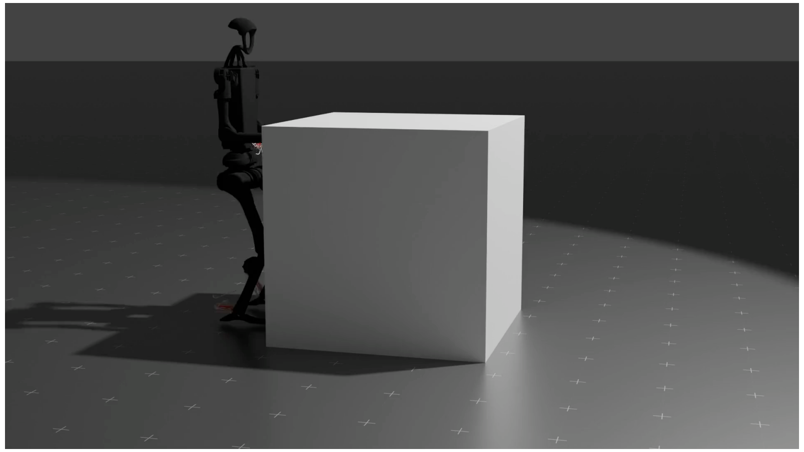


NVIDIA ISAAC
GENESIS
MuJoCo

Physical-based Simulators



Massively parallel
sampling-based optimization



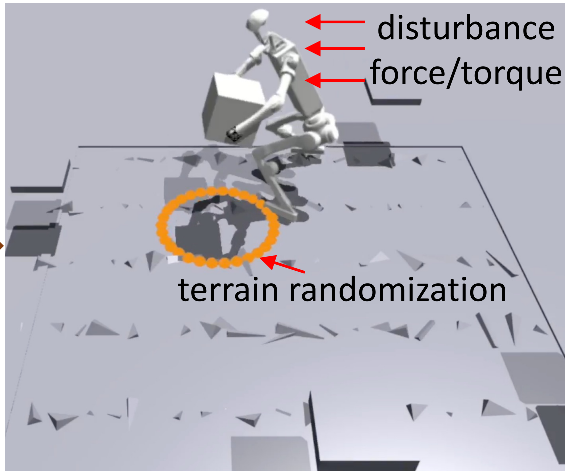
DIAL-MPC
https://lecar-lab.github.io/dial-mpc/
[Xue* and Pan* et al., ICRA'25]
Best Paper Award Finalist

Sim2Real 2.0: Sim2Real Reinforcement Learning (RL)

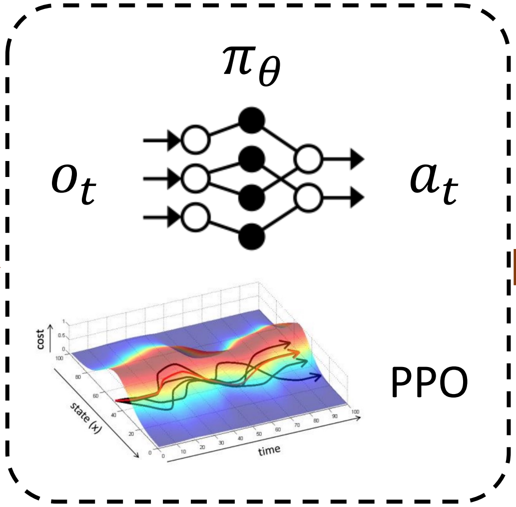
- ❑ Massively parallel policy gradient method (PPO) is such a strong policy optimizer
- ❑ No need for state estimation! Observation o_t is all you need

NVIDIA
ISAAC
GENESIS
MuJoCo

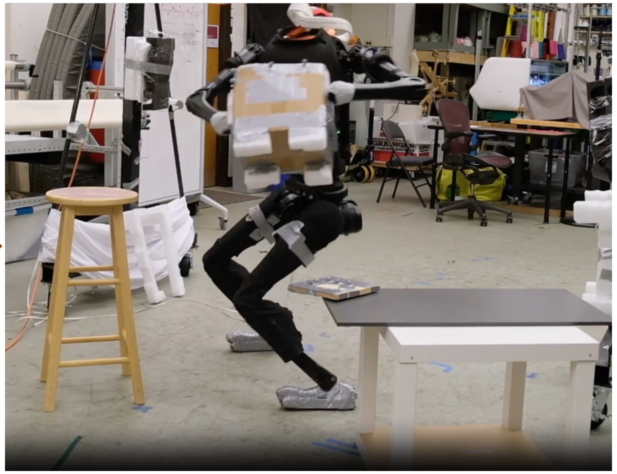
Physics-based Simulators



Massively Parallel
Training Environments



Policy Optimization
using RL



Real-world
Deployment

H2O: Human-to-Humanoid Whole-Body Control

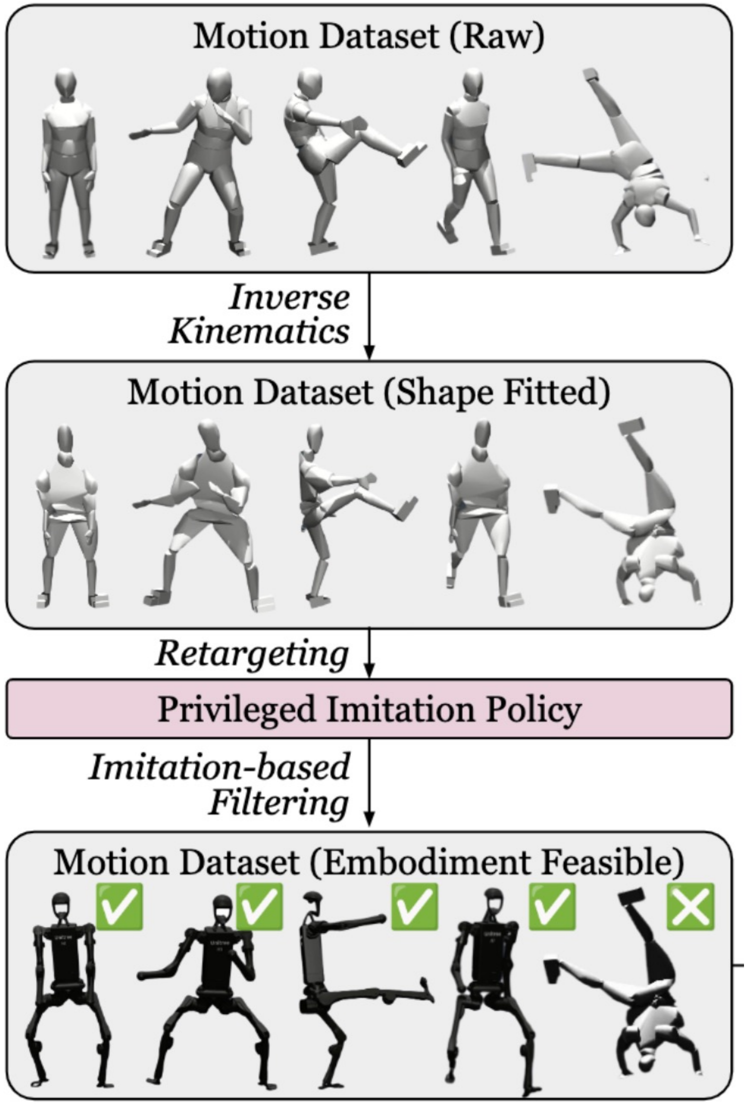
- **Goal**: Build an interface between whole-body human and humanoid motions
- Such an interface supports human whole-body teleoperation, imitation learning, integrating with VLMs, …
- **Key idea of H2O**: Sim2Real 2.0 from large-scale retargeted human motion dataset

[Learning Human-to-Humanoid Real-Time Whole-Body Teleoperation, He* & Luo* et al., IROS'24 (**Oral**)]
[OmniH2O: Universal and Dexterous Human-to-Humanoid Whole-Body Teleoperation and Learning, He* & Luo* & He*, et al., CoRL'24]

H2O: Human-to-Humanoid Whole-Body Control

Step 1: Create a large-scale humanoid-feasible motion dataset

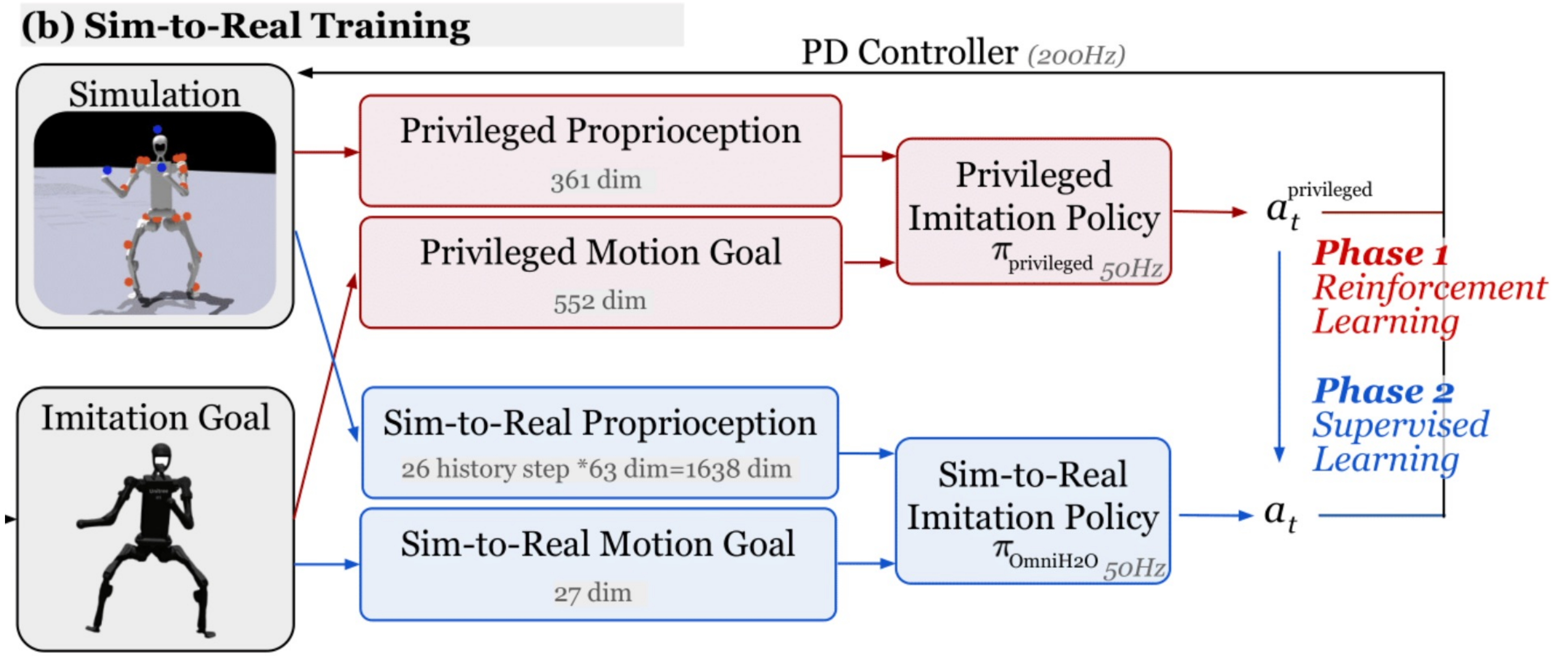
- ❑ >10K human motions from AMASS (ICCV'19)!
- ❑ Shape fitting using inverse kinematics
- ❑ Key: physics-based retargeting
  - Learn a privileged tracking policy to track all motions using RL
  - This policy knows all robot states
  - Generate *humanoid-feasible* motions and filter out impossible motions



H2O: Human-to-Humanoid Whole-Body Control

Step 2: Sim2Real RL training

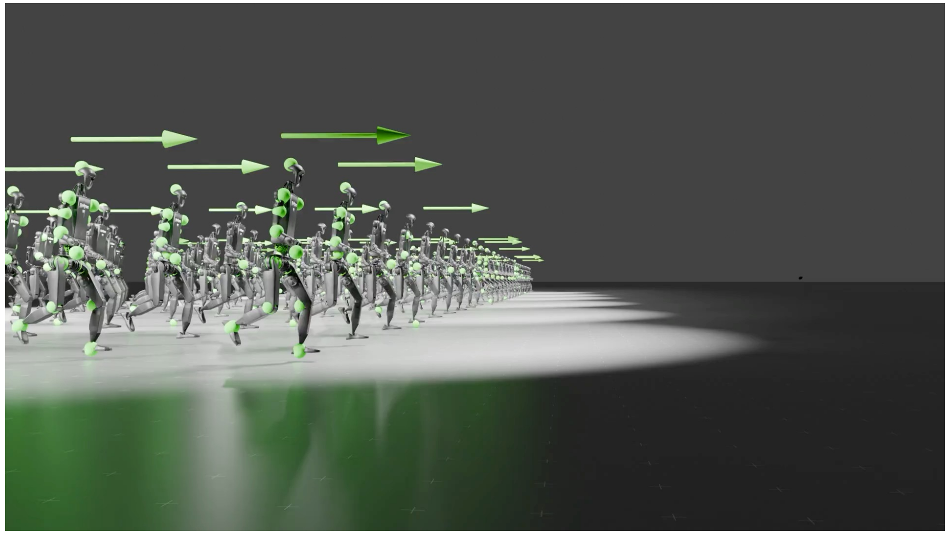
- ❑ Distill the privileged tracking policy to a deployable student policy in sim
  - The student policy only knows observations available in real
  - Key points as the motion goal (one head + two hands) for student policy
  - Domain randomization (DR) for robustness



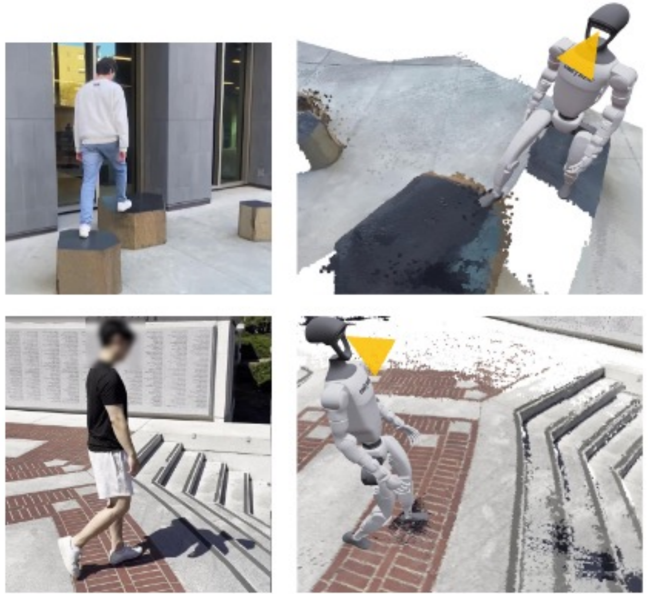
H2O: Human-to-Humanoid Whole-Body Control

- ❑ The H2O pipeline is highly extendable
  - *Step 1*: motion retargeting & learning a “tracking” policy in sim
  - *Step 2*: learn a “student” policy that can be deployed in real

- ❑ Motion source in step 1 is flexible: MoCap (AMASS), videos, …
- ❑ The student policy in step 2 is very flexible: Track different key points, vision-based, …



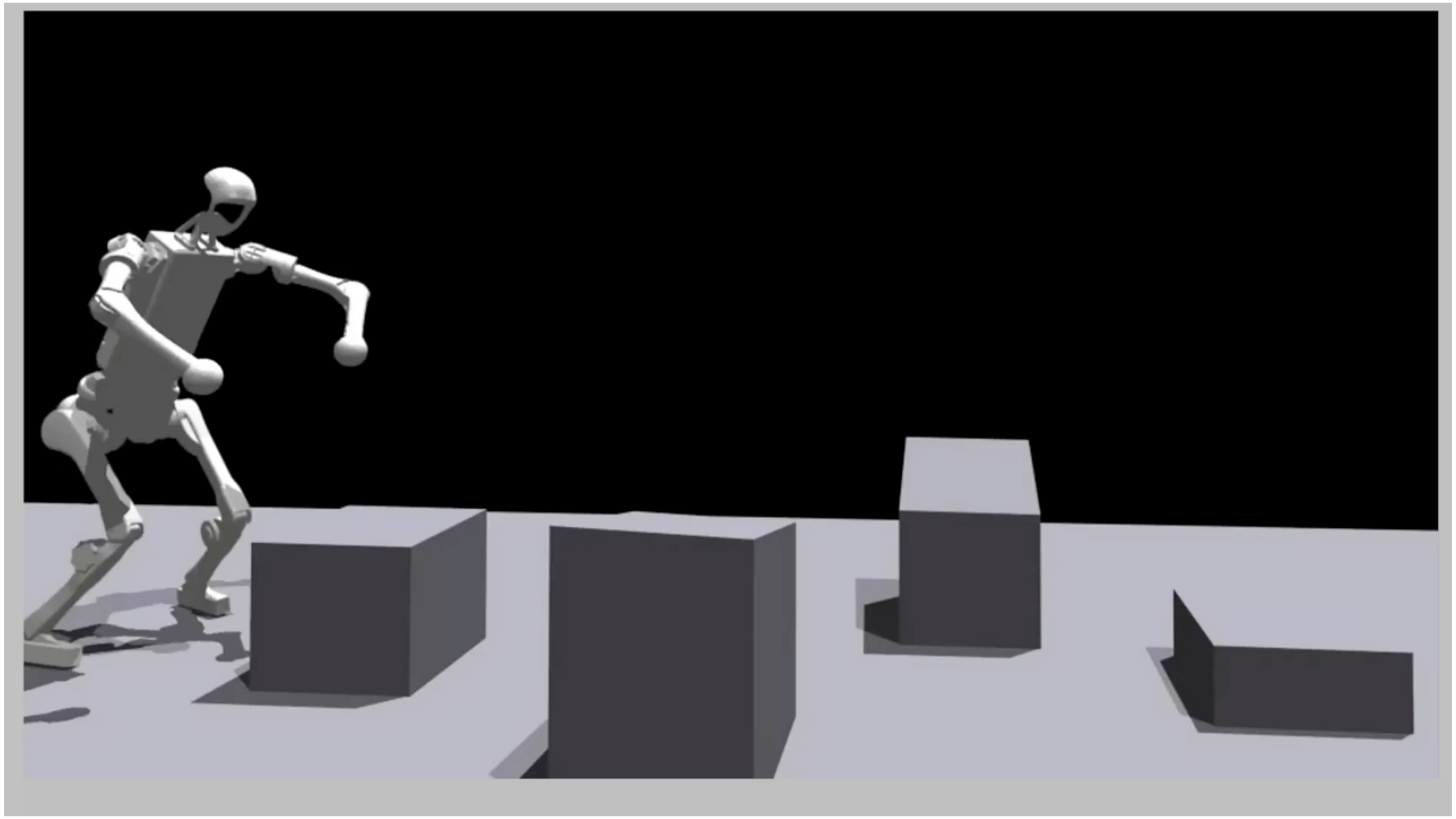
One teacher -> multiple students
HOVER from NVIDIA [He* and Xiao* et al., ICRA’25]



VideoMimic from Berkeley
[Allshire*, Choi*, Zhang*, McAllister*, et al.]

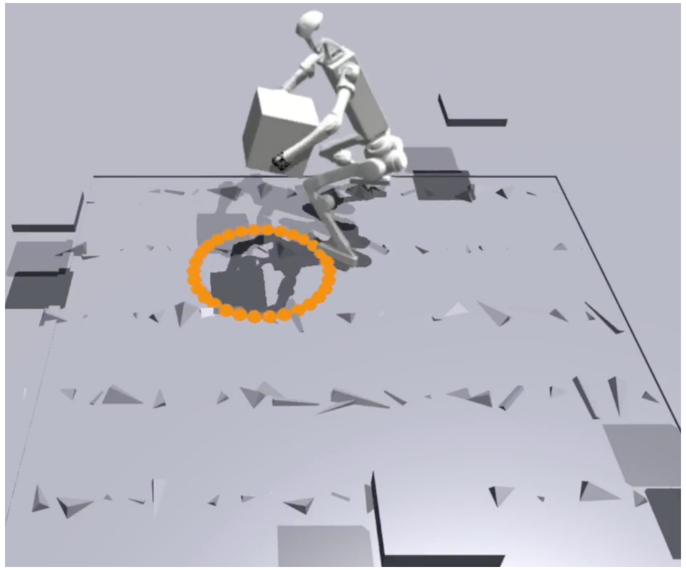
WoCoCo: Learning Whole-Body Control with Sequential Contacts

- **Goal**: learning long-horizon whole-body skills *without* any motion priors
- **Key idea**: decompose a long-horizon skill into a sequence of contact goals and task goals

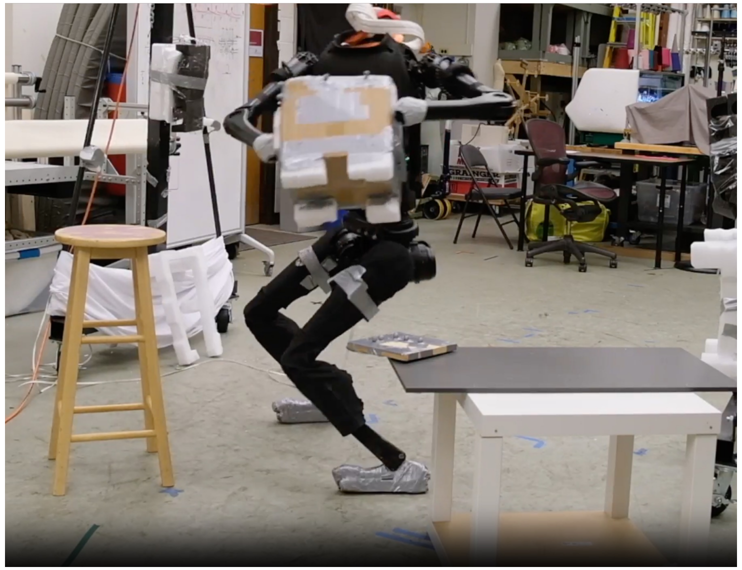


[WoCoCo: Learning Whole-Body Humanoid Control with Sequential Contacts, Zhang* & Xiao* et al., CoRL'24 (**Oral**)]

What is Wrong with Sim2Real 2.0?

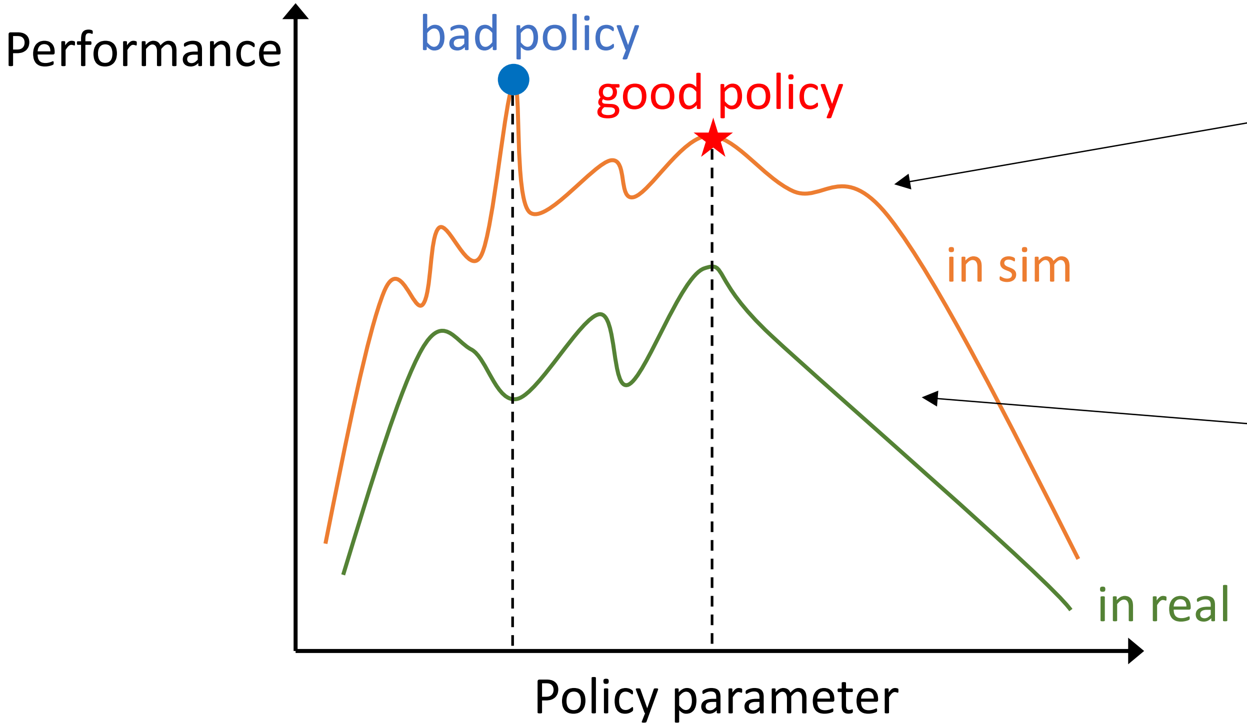


sim2real



- ❑ Sim2Real gap is large, unintuitive, and hard to quantify
- ❑ Tedious reward / curriculum / domain randomization tuning
- ❑ Hard to encode prior physics, poor sample complexity, unsafe
- ❑ No online reasoning: policies learned from sim are frozen in test time

From Sim2Real 2.0 to 3.0: Real2Sim and Structured RL



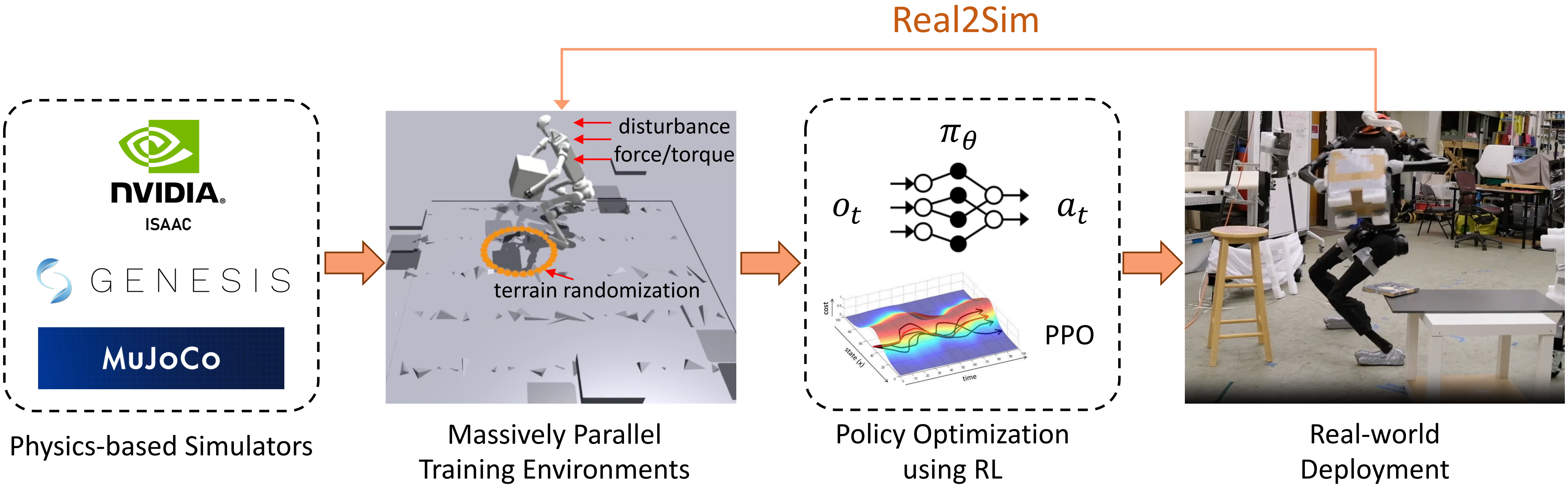
Challenge 2: super complex landscape and RL could easily “hack” it

Challenge 1: large sim2real gap (sim performance doesn’t transfer to real)

- ❑ **Real2sim**: reduce the sim2real gap
- ❑ **Structured RL**: add priors and inductive bias to have a smoother landscape

Sim2Real 3.0: Real2Sim

- ❑ learning "residuals" to bridge the gap between real and sim



Residual Dynamics Learning for Other Robotic Systems

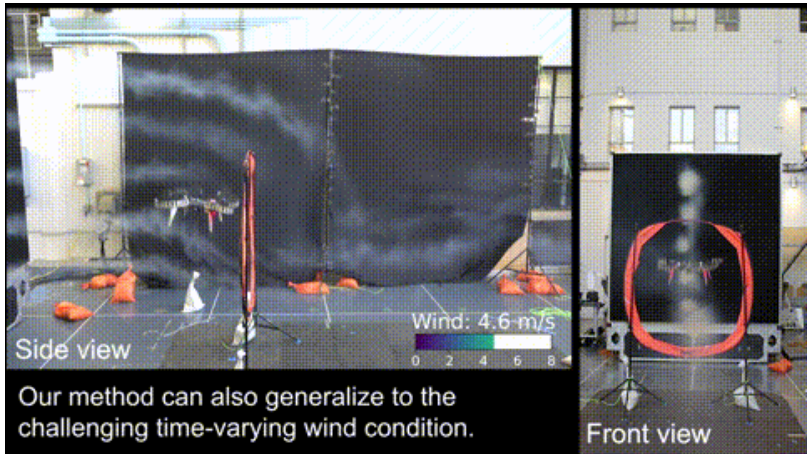
- Neural-Control Family
  - Key idea: Collect data *in real* and use a DNN $\hat{f}$ to approximate f

$$\boxed{M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) = u + f(q,\dot{q},a,t)}$$

unknown dynamics

  - Then design a nonlinear controller $u = \pi(q,\dot{q},\hat{f})$
  - Often need to regularize $\hat{f}$ for control-theoretic guarantees

Neural-Lander
[ICRA'19]

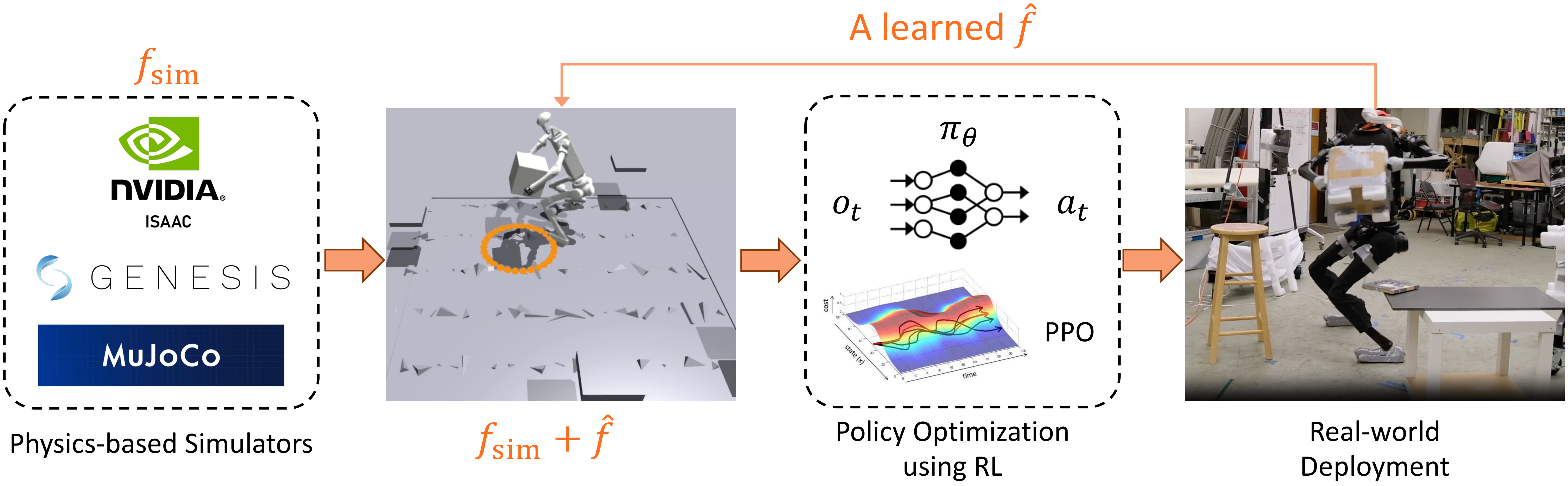


Neural-Fly: f is time-variant
[NeurIPS'21][Science Robotics'22]

Aerial Manipulations
[Guo* and He* et al., RAL'24]
[He* and Guo* et al., RSS'25]

Residual Dynamics Learning for Humanoids?

- Directly learning dynamics may not be a good idea for humanoids:
  - $\hat{f}$ needs to generalize well (requiring a lot of real-world data)
  - Need to regularize $\hat{f}$ heavily to ensure $f_{\text{sim}} + \hat{f}$ still "makes sense"
  - $\hat{f}$ will be exploited by RL



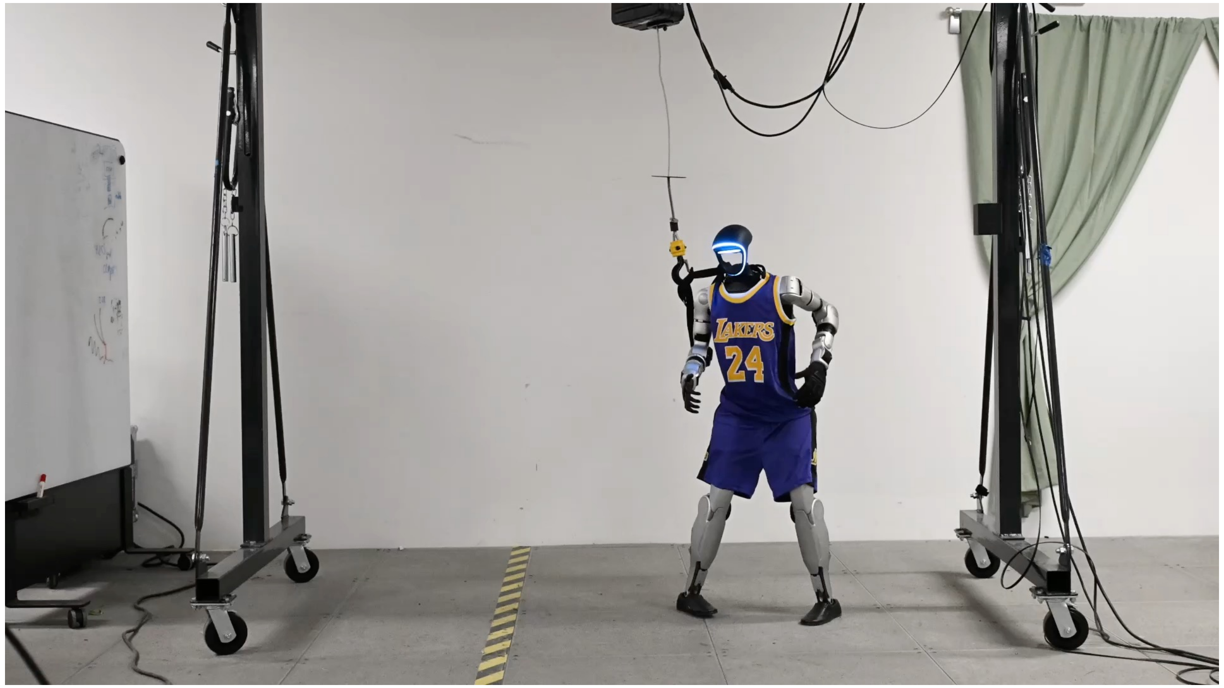
An Alternative Solution: Learning a Delta Action Model

- The ASAP framework: learn a *delta action model* to match sim and real
  - Pretrain a policy π in sim, rollout in real: $\{x_1^r, a_1^r, \cdots, x_T^r\}$
  - Replay $\{a_1^r, \cdots\}$ in sim: $x_{1:T}^s$. Due to the sim2real gap, $x_{1:T}^s \neq x_{1:T}^r$
  - Train a delta action model $\Delta a(x, a, \cdots)$ in sim such that $a_t^r + \Delta a_t$ yields $x_{1:T}^s \approx x_{1:T}^r$
  - Rollout $\pi + \Delta a$ in sim to fine-tune π. Finally deploy π in real.

[ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills, RSS'25]

Performance in Agile Whole-Body Control Tasks

- Similar to the human-to-humanoid pipeline but each policy focuses on one motion



[ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills, RSS'25]

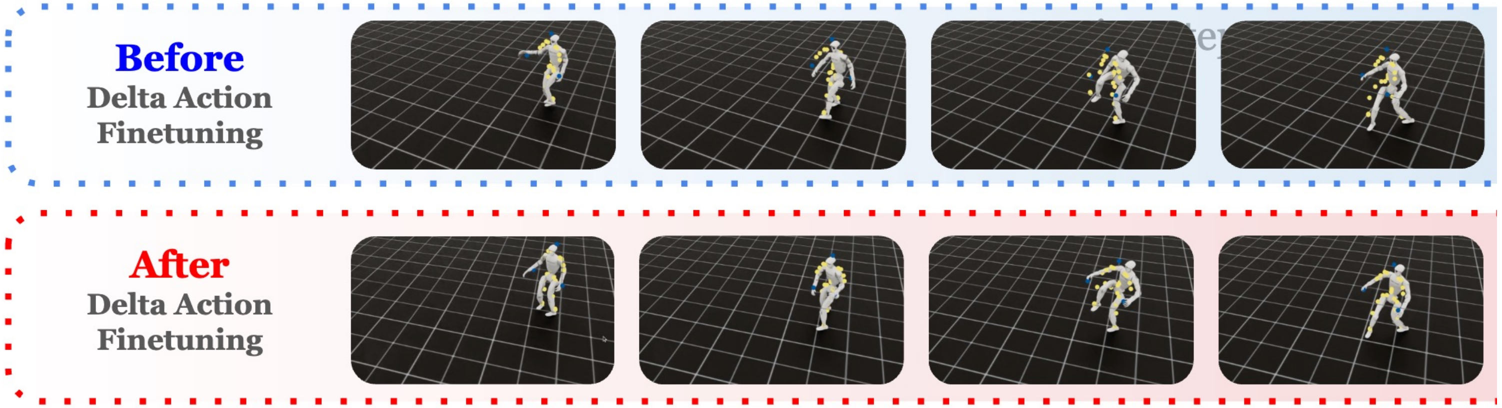
More Detailed Analysis of ASAP

- How to train Δa?
  - Another RL problem – the objective is to get $x_{1:T}^{s} \approx x_{1:T}^{r}$
- Why it makes sense?
  - $\pi + \Delta a$ in sim $\approx \pi$ in real. Δa effectively aligns sim and real dynamics
- Why is it different from Iterative Learning Control (ICL)?
  - The idea is very similar. ASAP is “deeper” and learns a closed-loop Δa
- Why don’t we call Δa a *residual policy*?
  - $\Delta a(x, a, \cdots)$ is closed-loop, but shared by all tasks: $\pi_1, \cdots \pi_N$ share the same Δa
- Example: in real, the motor is 80% as strong as sim: $a^r = 0.8\pi(x^r)$ but $a^s = \pi(x^s)$
  - In this case, our algorithm will learn $\Delta a(x, a) = -0.2\mathrm{a}$

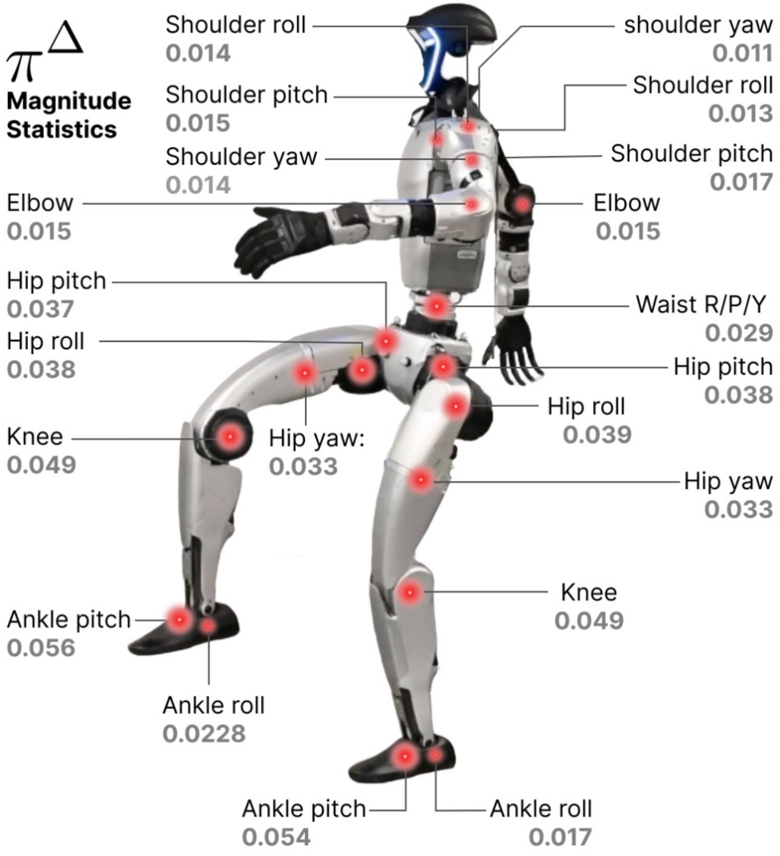
[ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills, RSS’25]

More Detailed Analysis of ASAP

- Quantitative results in sim2sim setting: Isaac Gym -> Isaac Sim



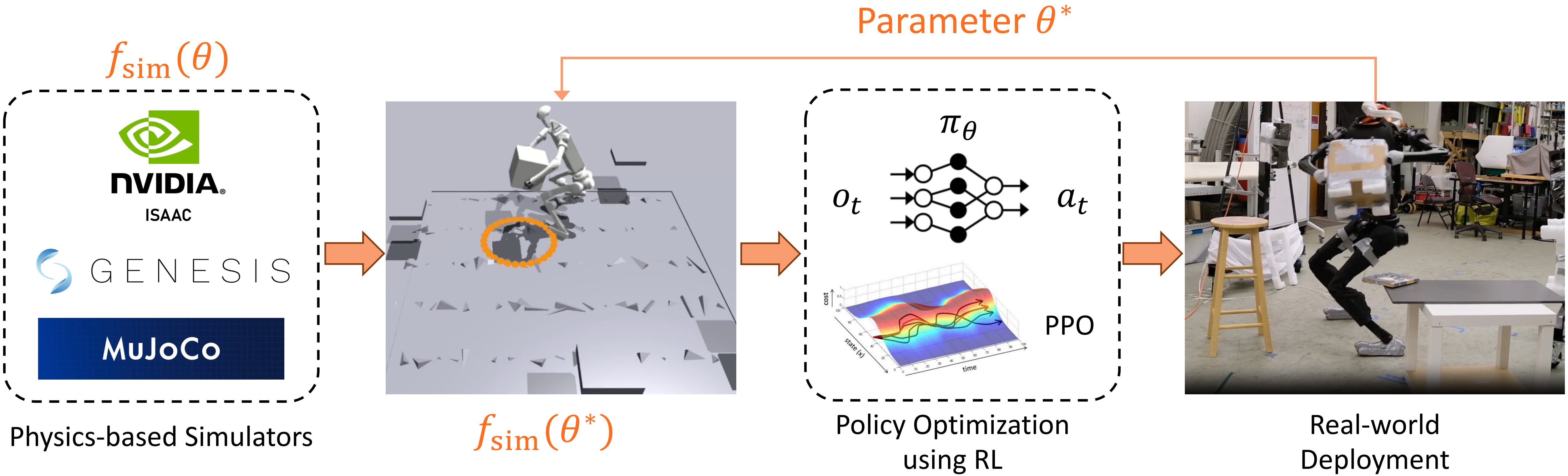
- Visualization of $||\Delta a||$ for each DoF
    - Lower body has bigger gaps
    - Ankle pitch has the biggest gap
    - In real world, we only learned a 4-DoF Δa for ankle



[ASAP: Aligning Simulation and Real-World Physics for Learning Agile Humanoid Whole-Body Skills, RSS'25]

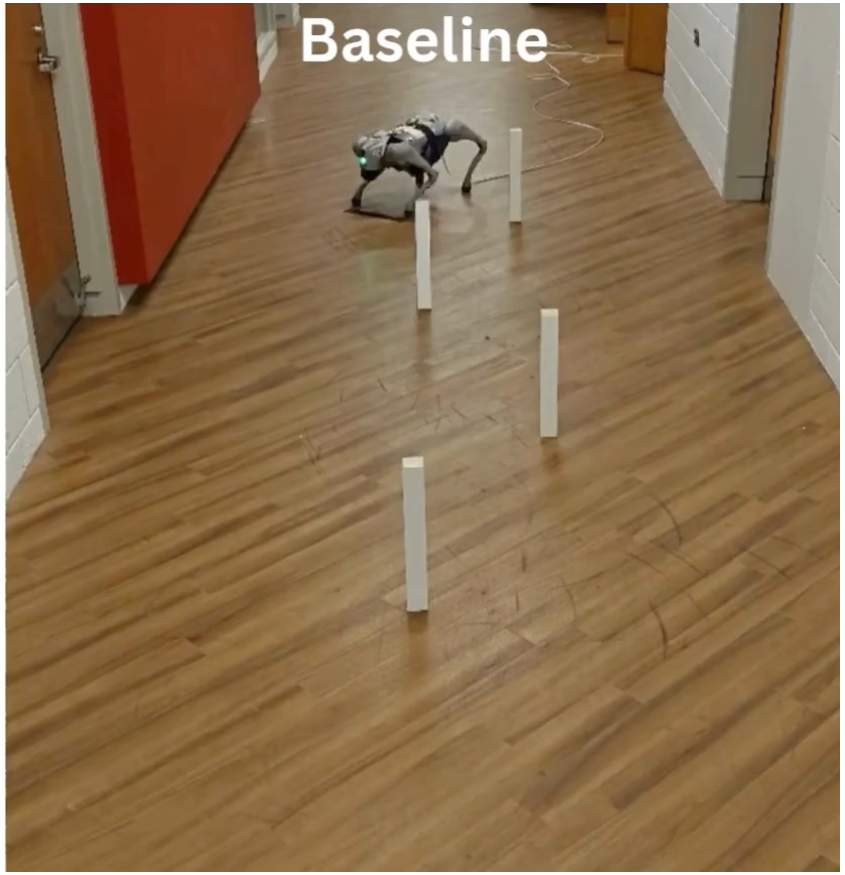
Another Solution for Real2Sim: System ID

- System identification (ID) is the oldest real2sim!
- Challenging for humanoids: $f_{\text{sim}}(\theta)$ is not differentiable or smooth



Another Solution for Real2Sim: System ID

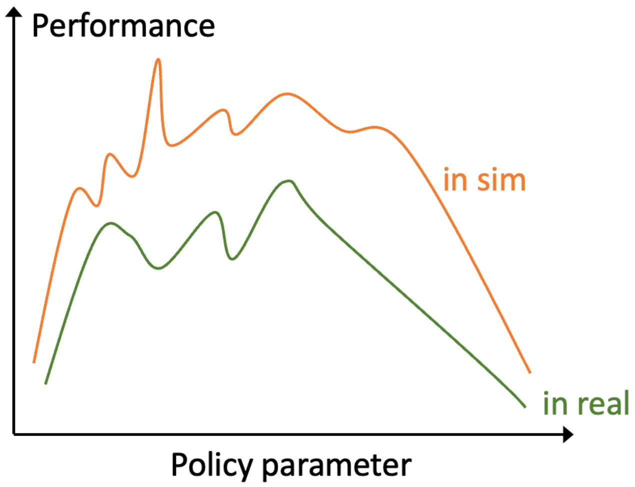
- SPI-Active: sampling-based system ID + active exploration
  - Use the policy that maximizes the Fisher Information to collect data
- https://lecar-lab.github.io/spi-active_/



[Sampling-Based System Identification with Active Exploration for Legged Robot Sim2Real Learning, Sobanbabu* and He* et al., 2025]

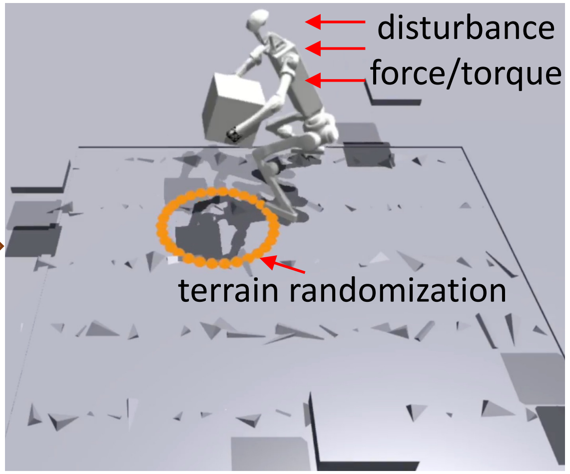
Sim2Real 3.0: Structured RL

- ❑ Leverage humanoid structure to design better policy architecture
- ❑ Goal: have a smoother RL optimization landscape

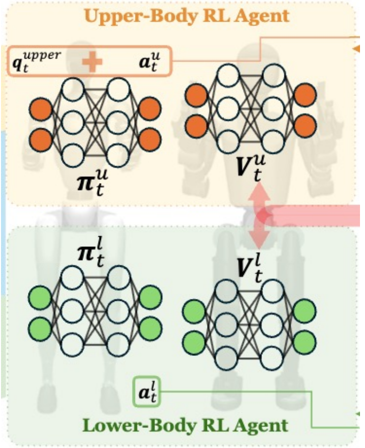


NVIDIA ISAAC
GENESIS
MuJoCo

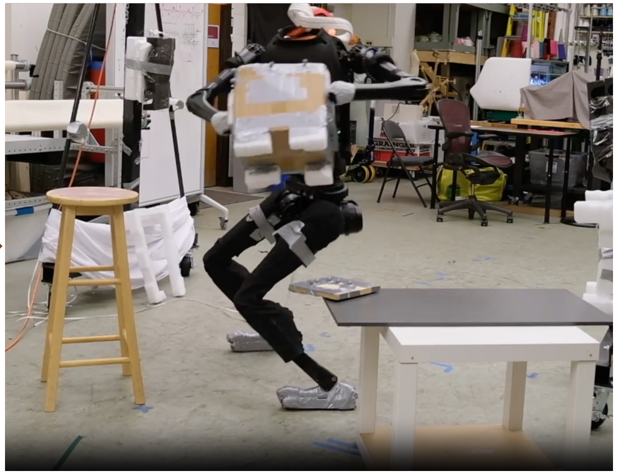
Physics-based Simulators



Massively Parallel Training Environments



Structured RL

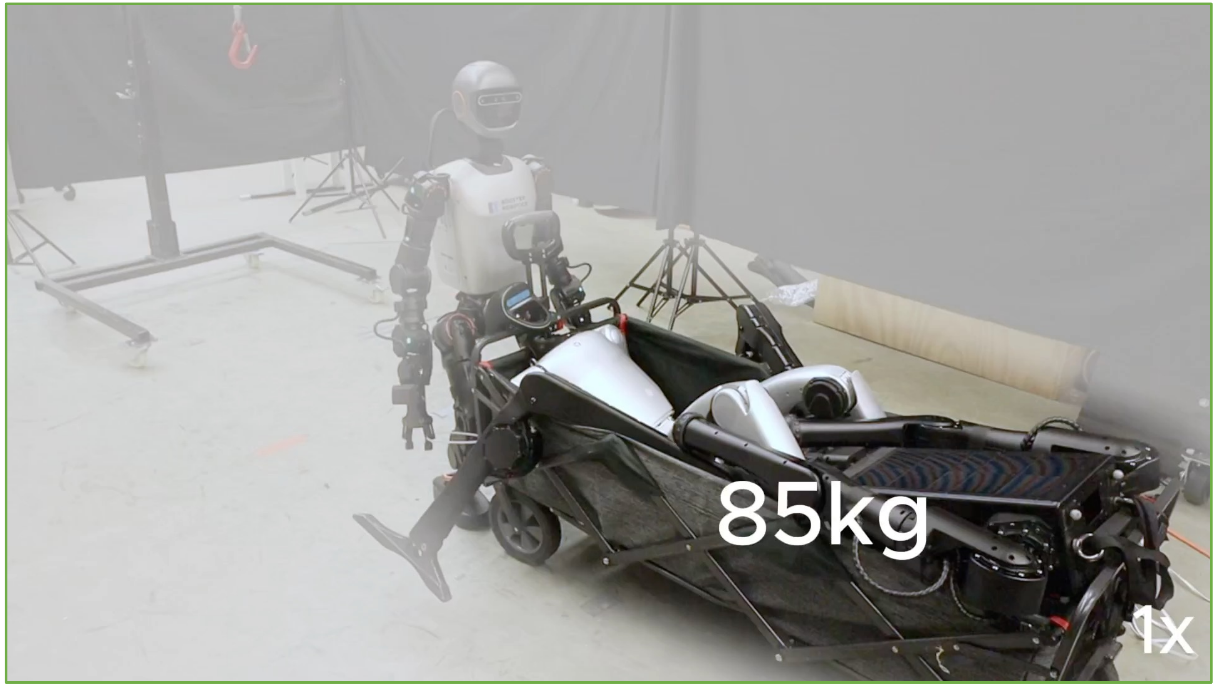


Real-world Deployment

FALCON: Dual-Agent RL for Force-Adaptive Loco-Manipulation

- ❑ Tasks: heavy-duty loco-manipulation

Baseline

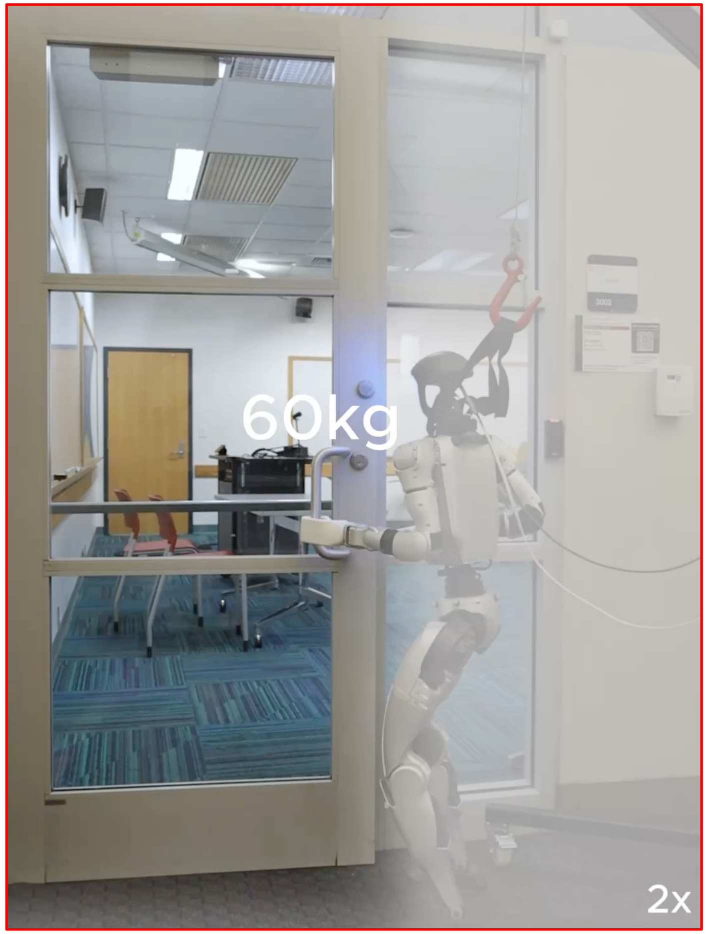


FALCON

[FALCON: Learning Force-Adaptive Humanoid Loco-Manipulation, Zhang et al., under review]

FALCON: Dual-Agent RL for Force-Adaptive Loco-Manipulation

- ❑ Tasks: heavy-duty loco-manipulation



Baseline

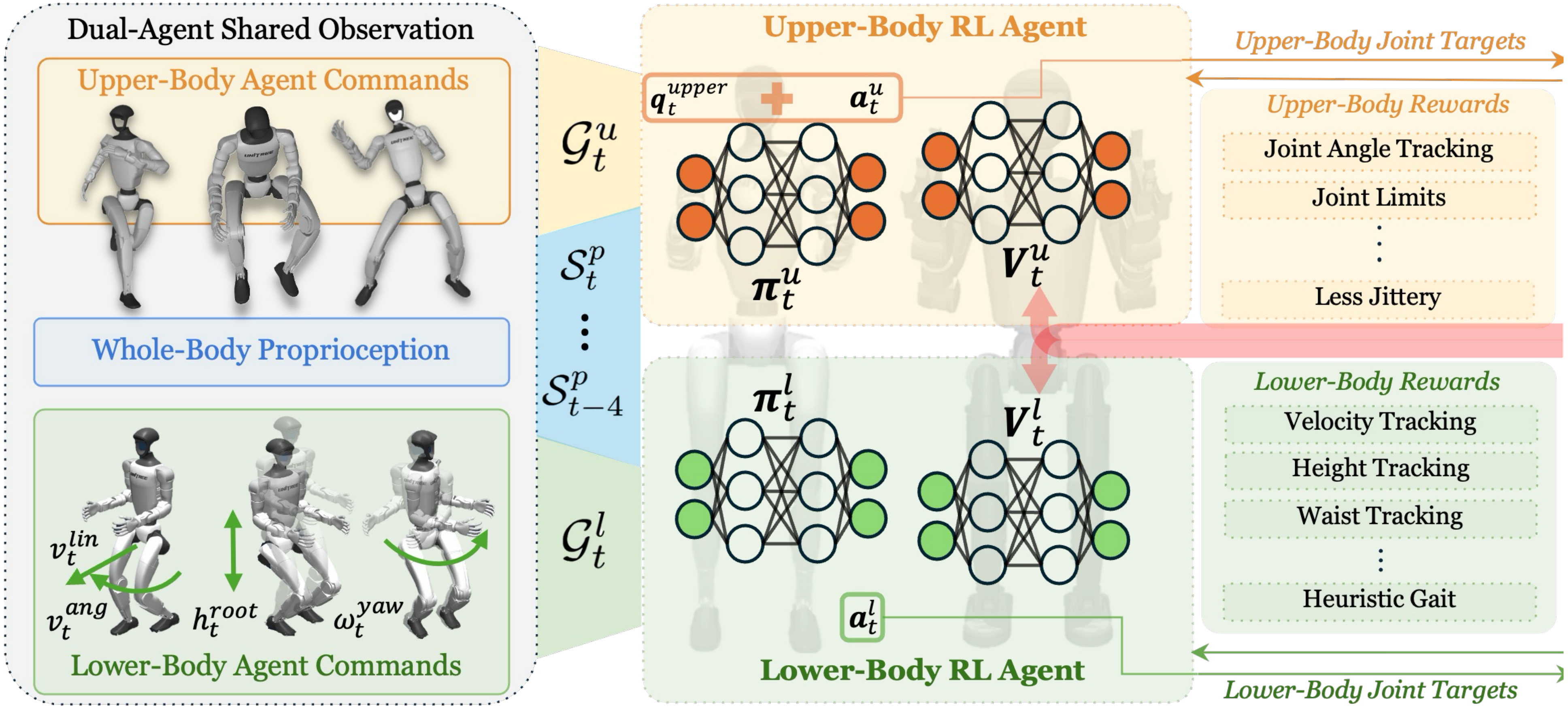
FALCON

[FALCON: Learning Force-Adaptive Humanoid Loco-Manipulation, Zhang et al., under review]

FALCON: Dual-Agent RL for Force-Adaptive Loco-Manipulation

❑ **Key structure 1**: dual-agent RL

- Two policies, two value functions (critics), two sets of rewards
- Jointly trained and both have whole-body proprioception input



[FALCON: Learning Force-Adaptive Humanoid Loco-Manipulation, Zhang et al., under review]

FALCON: Dual-Agent RL for Force-Adaptive Loco-Manipulation

- **Key structure 2**: adaptive and feasible 3D force curriculum on the end-effector
  - Apply random external forces f^{ee} on two end-effectors
  - Make sure f^{ee} is feasible with the motor torque limit

Feasible 3D Force Curriculum

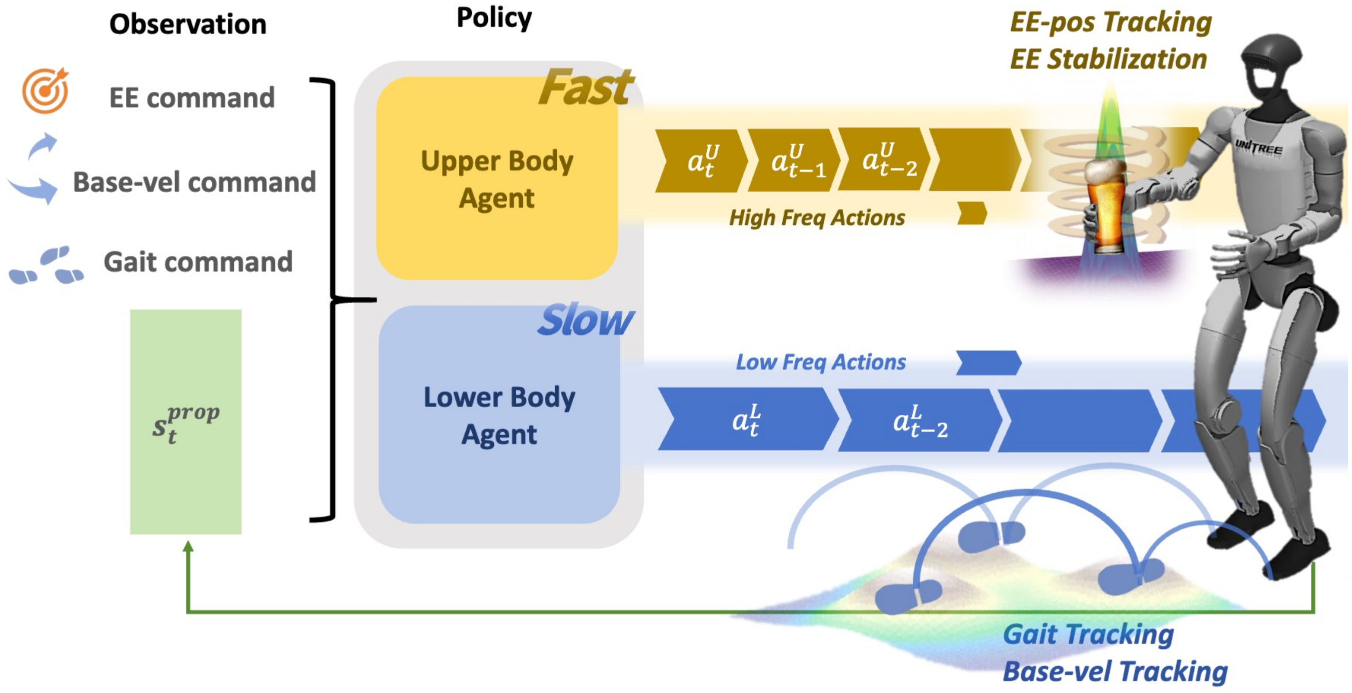
$$-\boldsymbol{\tau}^{\mathrm{lim}} \leq \boldsymbol{\tau}^{g} + \boldsymbol{J}_{EE}^{T} \boldsymbol{f}^{ee} \leq \boldsymbol{\tau}^{\mathrm{lim}}$$

$$\boldsymbol{\tau}^{\mathrm{lim}} \geq \mathbf{0}, \ \boldsymbol{\tau}^{\mathrm{lim}} - \boldsymbol{\tau}^{g} \geq \mathbf{0}$$

[FALCON: Learning Force-Adaptive Humanoid Loco-Manipulation, Zhang et al., under review]

Slow-Fast Two-Agent for “Hold My Beer”

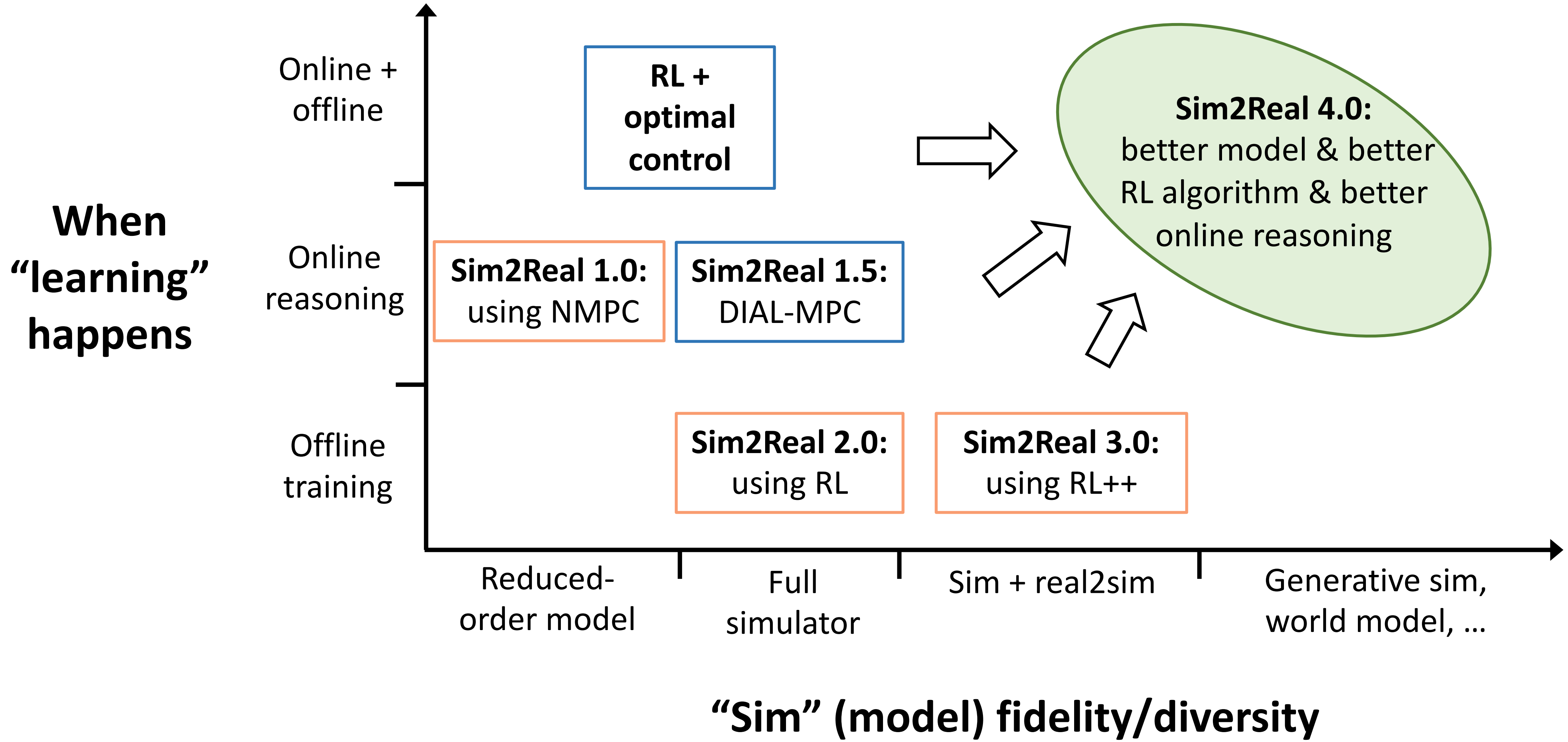
❑ Slow-fast two-agent framework for humanoid end-effector stabilization
- Upper body: ”fast” dynamics, high-precision corrections
- Lower body: “slow” dynamics, robust locomotion



[Hold My Beer: Learning Gentle Humanoid Locomotion and End-Effector Stabilization Control, Li et al., under review]

Zooming Out: Towards Sim2Real 4.0

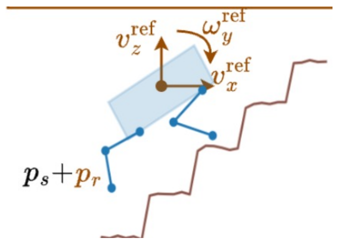
- ❑ Offline + online could be powerful!



RL (full-order) + Online Optimal Control (Reduced-order)

- π_{RL} outputs center of mass refs $\ddot{q}^{\text{ref}}$; π_{QP} optimizes ground reaction force (GRF)
- Fully onboard & autonomous (depth camera for sensing)

$\pi_{RL}(x)$



$\pi_{\text{QP}}(x, \pi_{RL})$

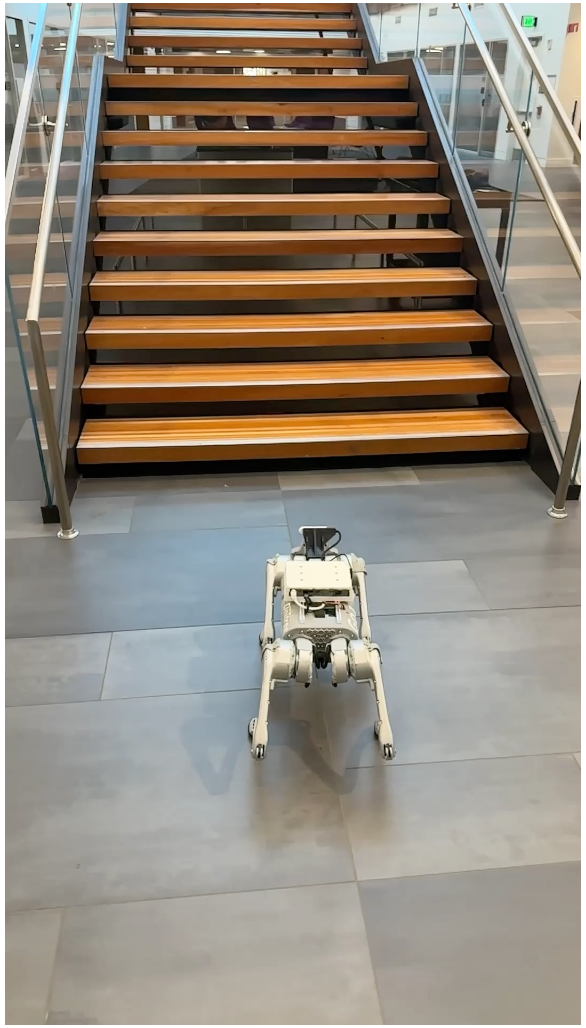
$$\min_{\boldsymbol{f}} \|\ddot{\boldsymbol{q}} - \ddot{\boldsymbol{q}}^{\text{ref}}\|_{\mathbf{U}} + \|\boldsymbol{f}\|_{\mathbf{V}}$$

$$\text{subject to: } \ddot{\boldsymbol{q}} = \mathbf{A}\boldsymbol{f} + \boldsymbol{g}$$

$$f_{i,z} = 0$$

$$f_{\min} \leq f_{i,z} \leq f_{\max}$$

$$-\mu f_{i,z} \leq f_{i,x} \leq \mu f_{i,z}, \quad -\mu f_{i,z} \leq f_{i,y} \leq \mu f_{i,z}$$



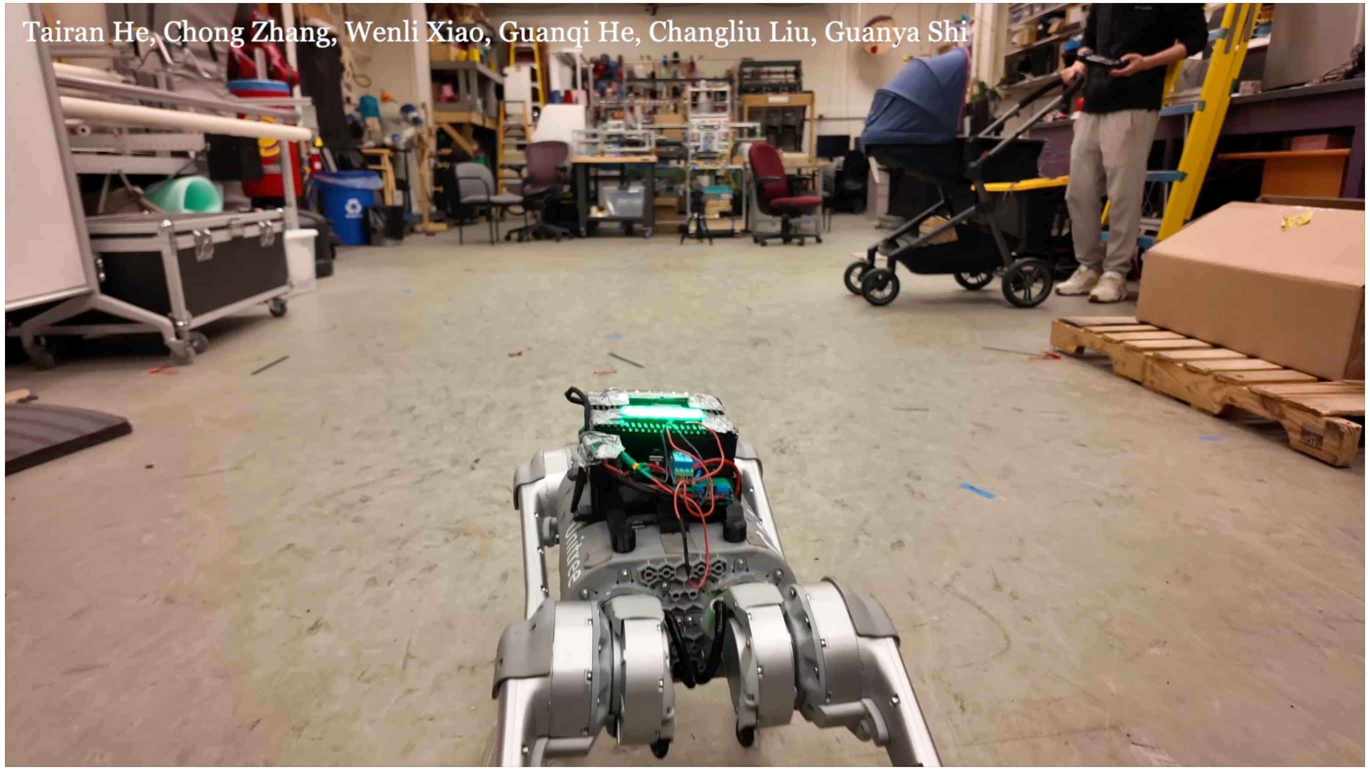
[Agile Continuous Jumping in Discontinuous Terrains, Yang et al., ICRA'25]
[CAJun: Continuous Adaptive Jumping using a Learned Centroidal Controller, Yang et al., CoRL'23]

RL (full-order) + Online Optimal Control (Reduced-order)

[RAMBO: RL-augmented Model-based Optimal Control for Whole-body Loco-manipulation, Cheng et al., under review]

Hierarchical RL and Safe Control Layer

Reinforcement Learning $\xrightarrow{\pi_{RL}(x)}$ Safety Filter $\xrightarrow{\pi_s(x, \pi_{RL})}$ 



- Fully onboard & autonomous
- Fast (up to 3.1m/s)
- Safe & robust

[Agile But Safe: Learning Collision-Free High-Speed Legged Locomotion, He* and Zhang* et al., RSS’24 (**Outstanding Student Paper Finalist**)]

Thank You!

All projects I presented are open-sourced:
https://lecar-lab.github.io/publications.html

Learning and Control for Agile Robotics Lab

Carnegie Mellon University
Robotics Institute