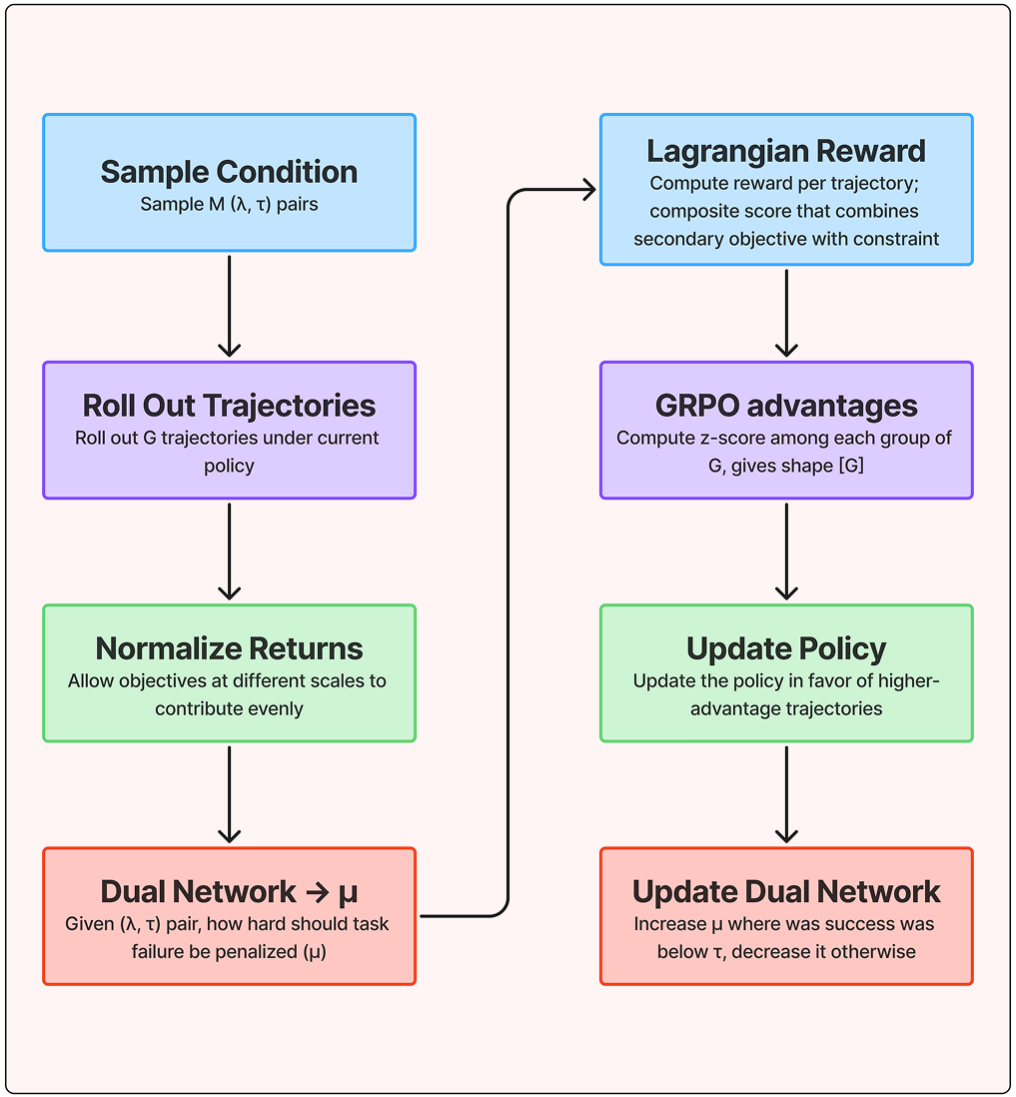
WEEK 6 UPDATE

CODY HILL • 6/29/26
DR. YUE WANG & ZAIN FARHAT



Problem Overview



Task

- GRPO with task-success floor
- Optimized secondary objectives
- Policy π with maximum returns

Problem Overview

	Bad	Feasible	Optimal
Success	0.64	0.85	0.85
Efficiency	0.84	0.41	0.77
Safety	0.73	0.51	0.68
Smoothness	0.75	0.43	0.75

$\tau = 0.8$

Progress: Environment

- LIBERO
- Franka Emika Panda arm
- Gripper: Franka Hand
- OSC_POSE
- Delta mode => velocity
  - Tiny action per timestep
- Reproduced baseline results



Progress: Objective Measurement

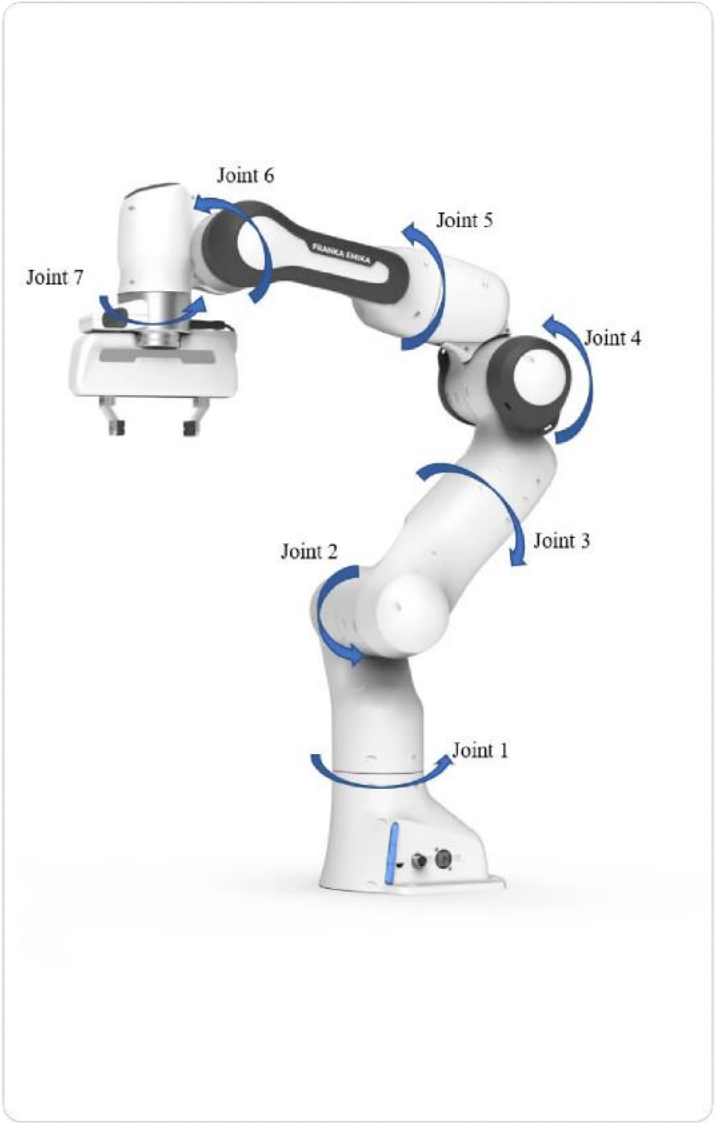
- Task success*
- Energy efficiency
- Motion smoothness
- Safety/contact

Objective	Reward Signal
Task Success*	Task Completion
Energy Efficiency	$-\sum_j \lvert\tau_j \dot{q}_j\rvert$
Motion Smoothness	$-\lVert\ddot{a}_t\rVert^2$
Safety/Contact	$-\max(0, f_{\text{contact}} - f_{\max})$

* = primary objective

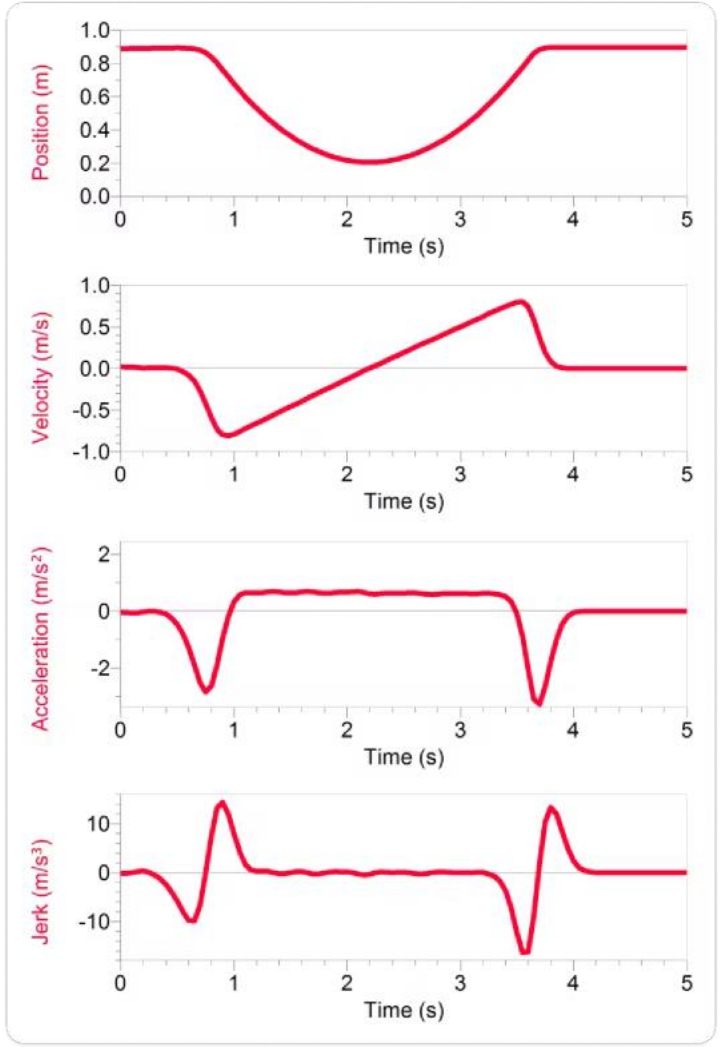
Progress: Energy Efficiency

- Mechanical work by motors
- Efficiency => low energy
- $-\sum_j |\tau_j \dot{q}_j|$
- Power at single joint
- Sum across arm



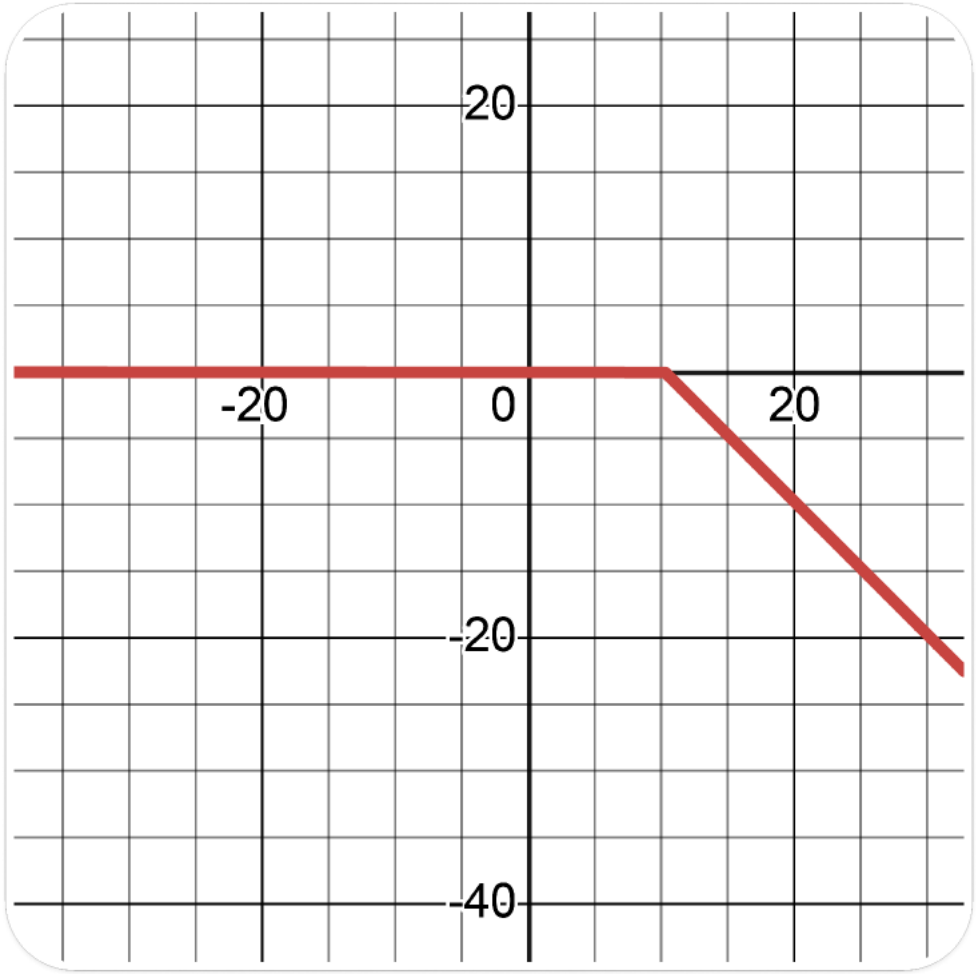
Progress: Smoothness

- Action jerkiness
- Low jerk => less wear
- Acceleration rate of change
  - Second action difference
- $-\|\ddot{a}_t\|^2$



Progress: Safety/Contact

- Is the arm pressing too hard
- Gentle manipulation
- Determined fmax; 95th %
- Contact unavoidable
  - Linear hinge
- $-\max(0, f_{\text{contact}} - f_{\text{max}})$



Contact Examples



Aggressive Contact
Pushes too hard



Optimal Contact
Completes task with minimal force

Progress: CMO-GRPO MiniVLA

- CMO-GRPO LIBERO-90
- Baseline maintained
- Foundation for SFT use
- Swept λ
- Objectives respond to λ

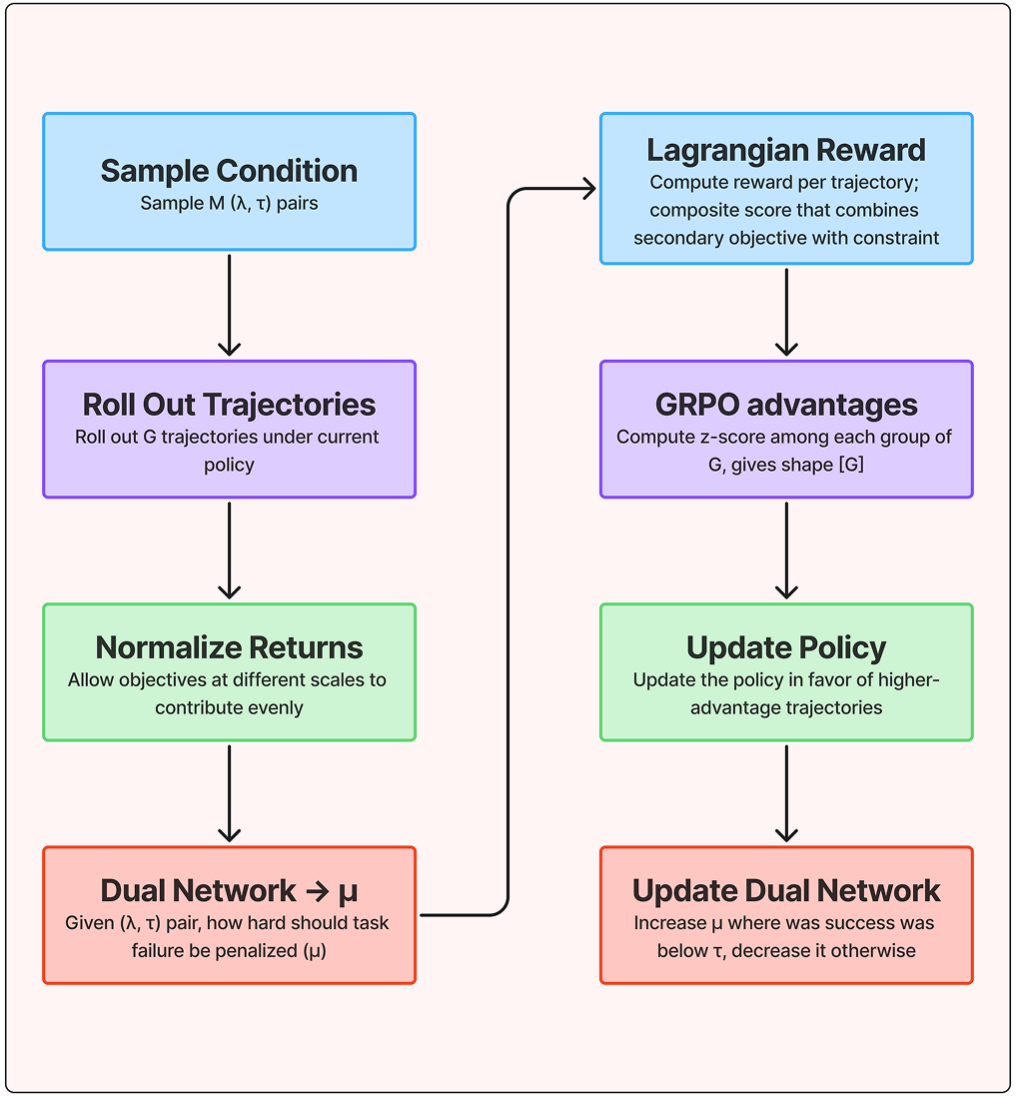
My Results

Preference Vector (λ)	Success	Reward Delta
(0.33, 0.33, 0.33)	0.53	All improved
(1.0, 0.0, 0.0)	0.53	Energy (r2) improved
(0.0, 1.0, 0.0)	0.73	No changes

Next Steps

Next Steps

- SFT
- True implementation



WEEK 6 UPDATE

CODY HILL • 6/29/26
DR. YUE WANG & ZAIN FARHAT

