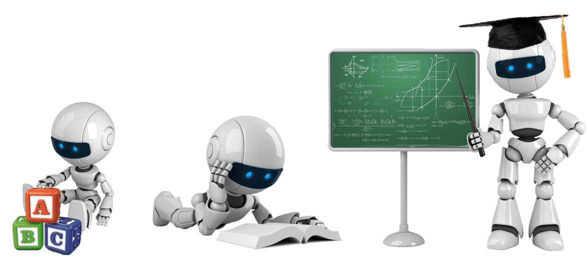
Using Task Features for Zero-Shot Knowledge Transfer in Lifelong Learning

David Isele*   Mohammad Rostami*   Eric Eaton

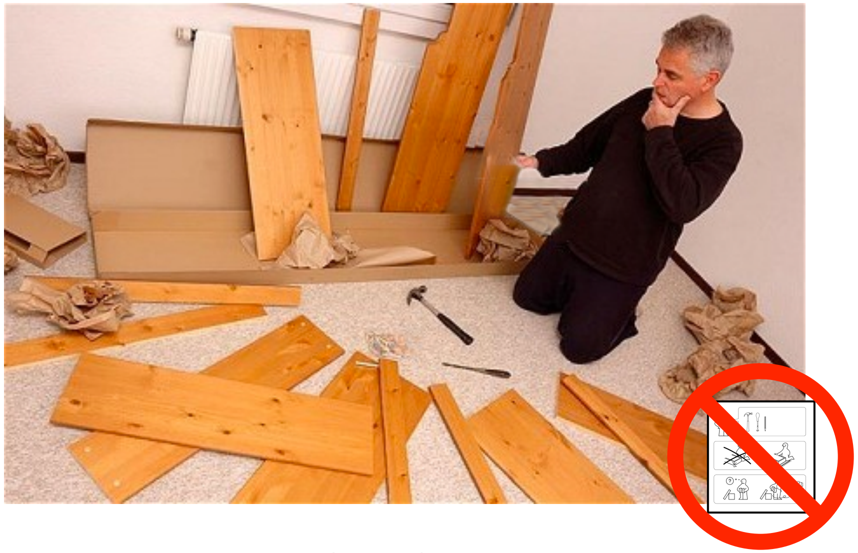
University of Pennsylvania



* Authors contributed equally

International Joint Conference on Artificial Intelligence 2016

Motivation



"Bookshelf with 5 shelves"

Need to transfer from prior experience

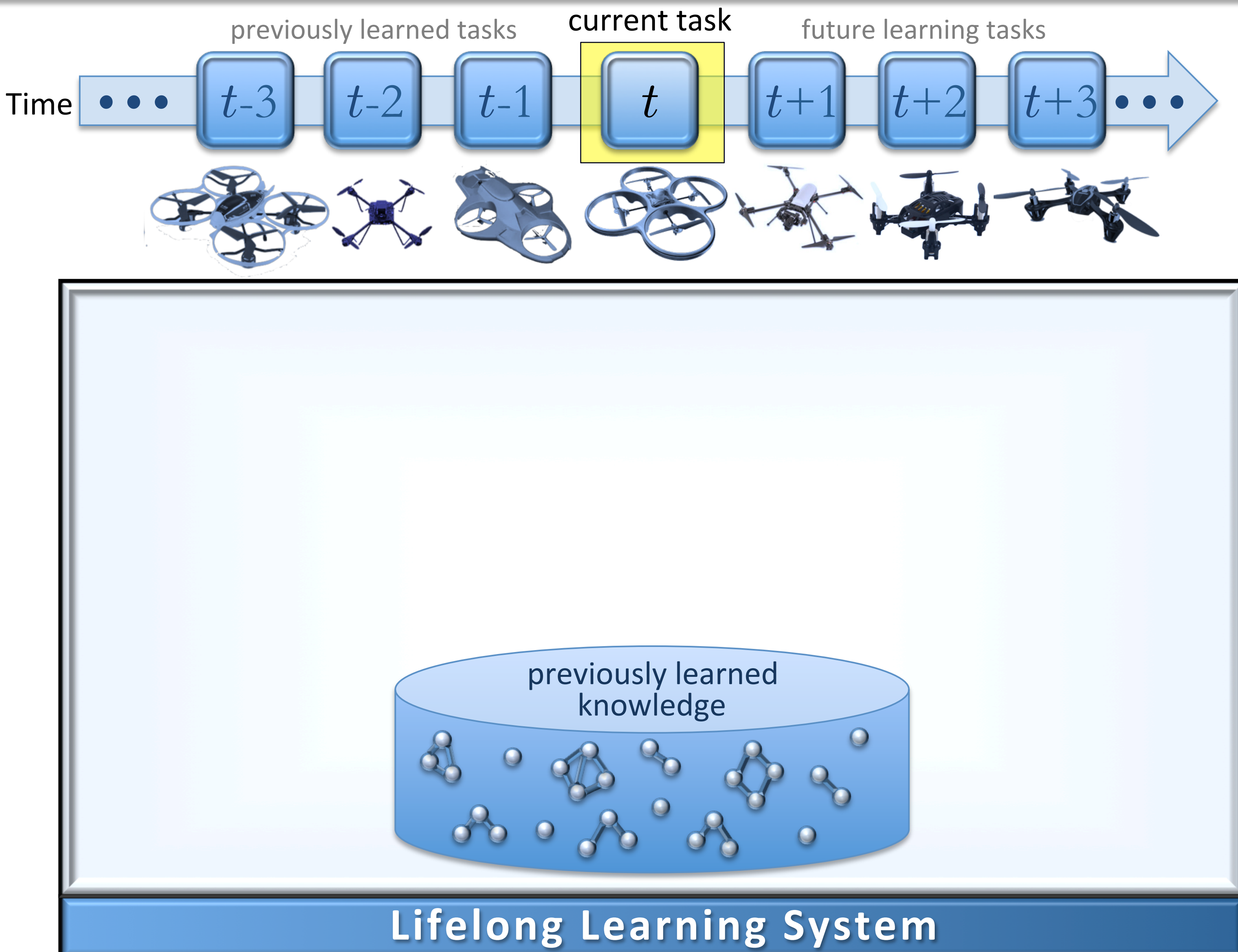
Key Idea: Use a high-level task description to identify relevant knowledge for transfer in lifelong learning

- ➢ Improve task performance
- ➢ Zero-shot transfer

- Task descriptors used for pairwise transfer by Sinapov et al. (2015)

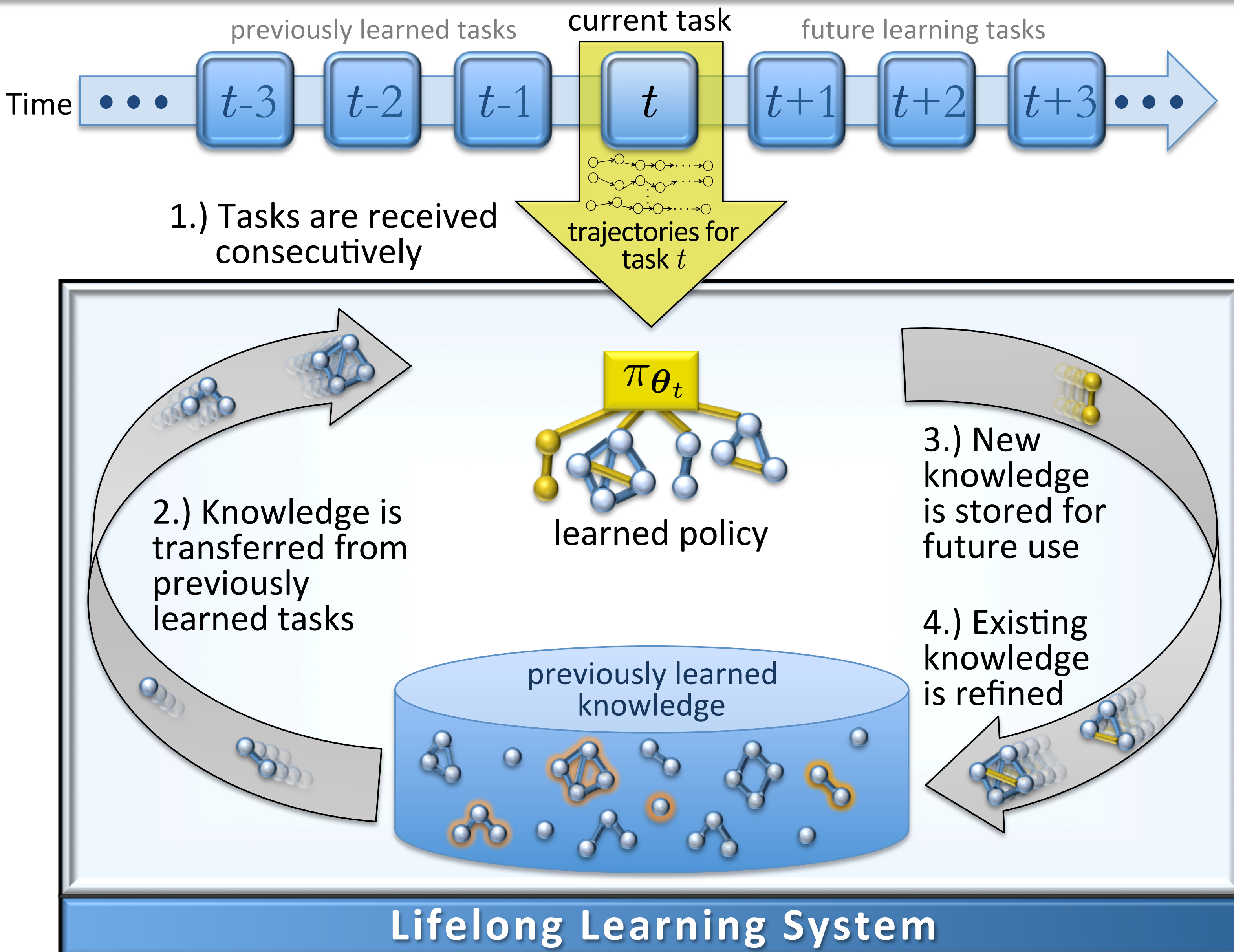
Lifelong Machine Learning

[Bou Ammar, Eaton, et al. ICML14]

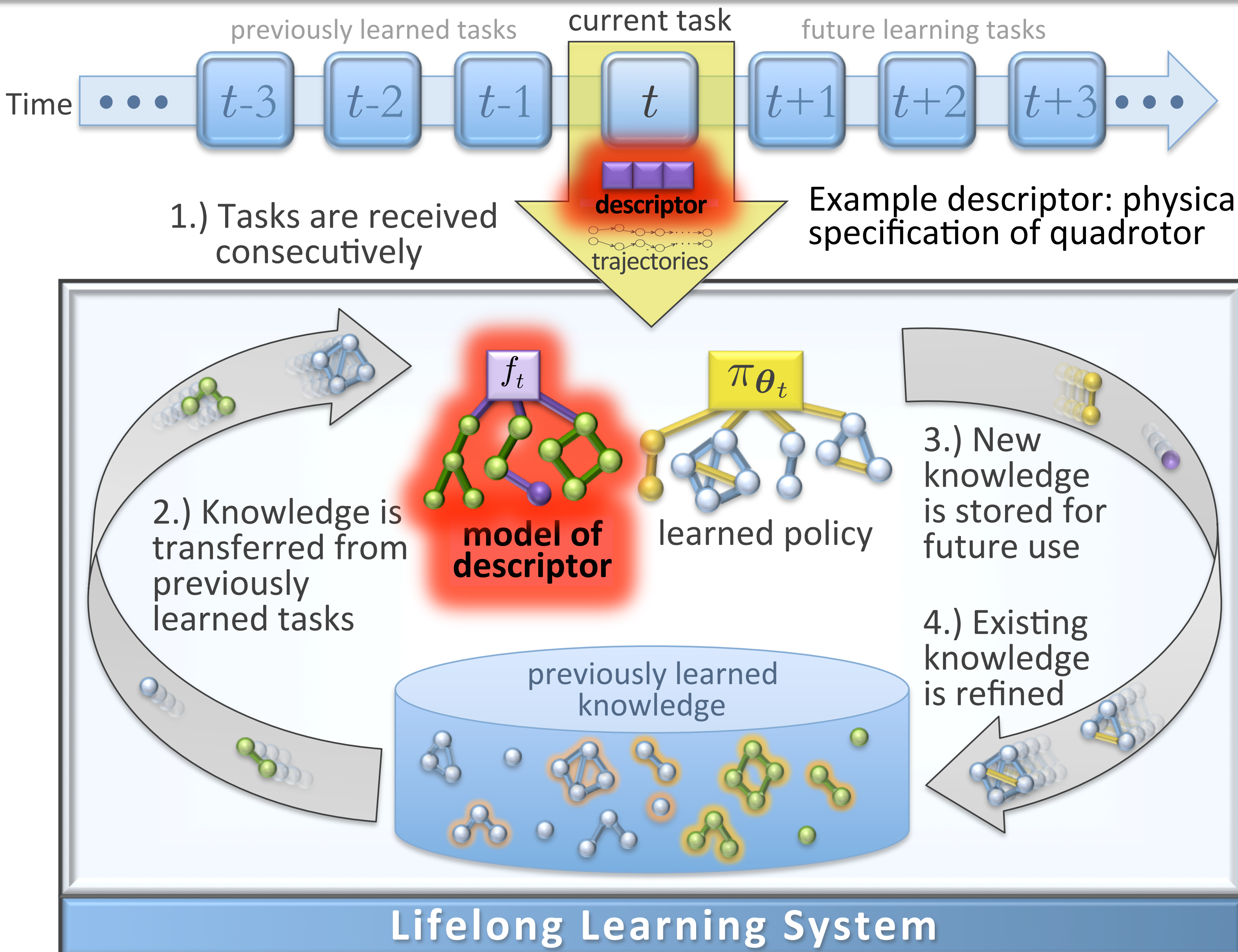


Lifelong Machine Learning

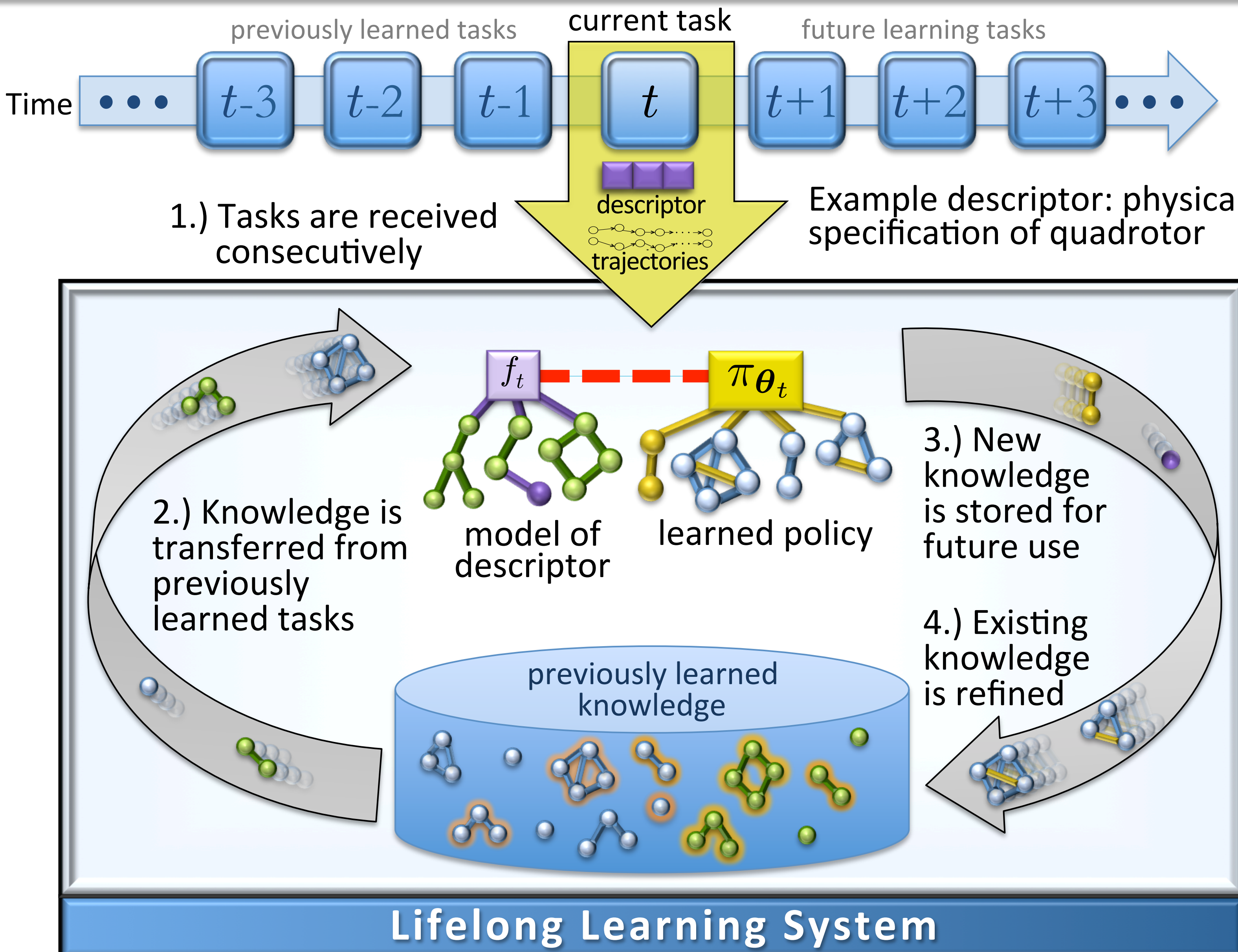
[Bou Ammar, Eaton, et al. ICML14]



Our Contribution



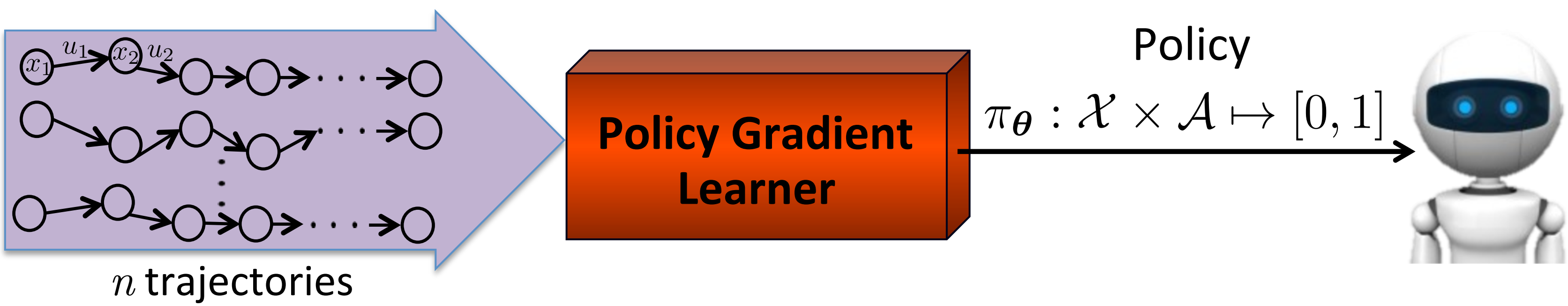
Our Contribution



Background: Policy Gradient Methods for Control

Agent interacts with environment, taking consecutive actions

- Continuous state and action spaces
- Demonstrated in robotic control [Kober & Peters '11; Peters & Schaal '08; Sutton '00]



Goal: find policy π_θ that maximizes $\mathcal{J}(\boldsymbol{\theta}) = \int_{\mathbb{T}} p_\theta(\boldsymbol{\tau})\mathcal{R}(\boldsymbol{\tau})d\boldsymbol{\tau}$

$$p_\theta(\boldsymbol{\tau}) = p_0(\mathbf{x}_0) \prod_{h=1}^{H} p(\mathbf{x}_{h+1}|\mathbf{x}_h, \mathbf{a}_h)\pi_{\boldsymbol{\theta}}(\mathbf{a}_h|\mathbf{x}_h)$$

probability of trajectory

$$\mathcal{R}(\boldsymbol{\tau}) = \frac{1}{H}\sum_{h=0}^{H} r_{h+1}$$

reward function

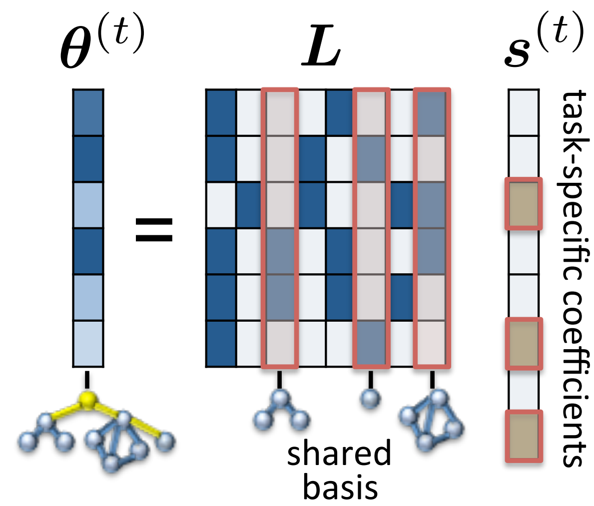
Sharing Knowledge Between Tasks

Policy for task t:

$$\pi_{\boldsymbol{\theta}^{(t)}} : \mathcal{X} \times \mathcal{A} \mapsto [0, 1]$$

Factor the policy as

$$\boldsymbol{\theta}^{(t)} = \mathbf{L}\mathbf{s}^{(t)}$$



Objective Function

$$e_T(\boldsymbol{L}) = \frac{1}{T}\sum_{t=1}^{T} \min_{\boldsymbol{s}^{(t)}} \Big[\underbrace{-\mathcal{J}\left(\boldsymbol{\theta}^{(t)}\right)}_{\text{fit to each task}} + \mu \underbrace{\left|\left|\mathbf{s}^{(t)}\right|\right|_1}_{\text{sparsity of coefficients}} \Big] + \underbrace{\lambda ||\mathbf{L}||_{\mathsf{F}}^2}_{\text{regularize basis complexity}}$$

Batch Optimization
(all tasks are given)

Multi-Task Learning

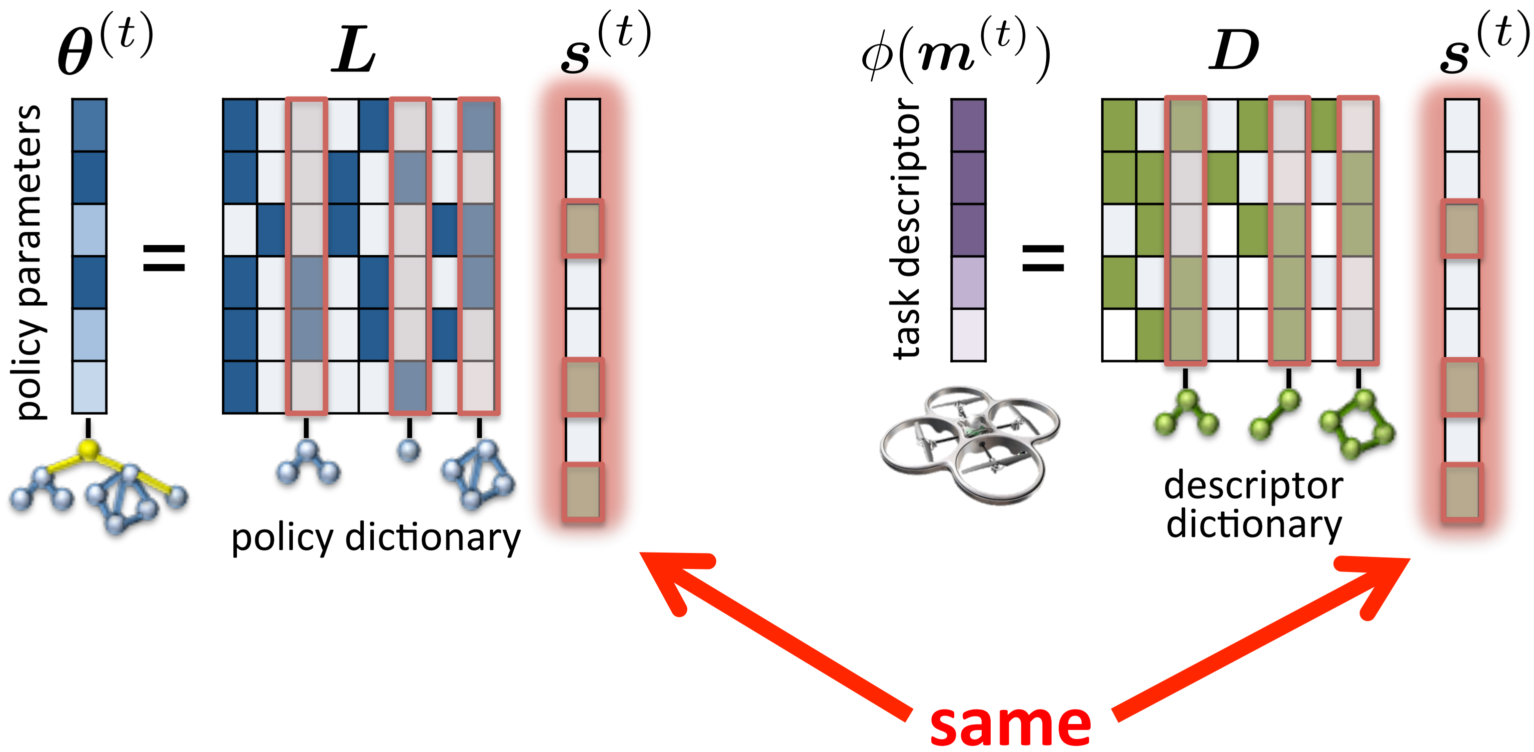
Online Optimization
(tasks arrive consecutively)

Lifelong Learning (PG-ELLA)

[Bou Ammar, Eaton, et al., ICML '14]

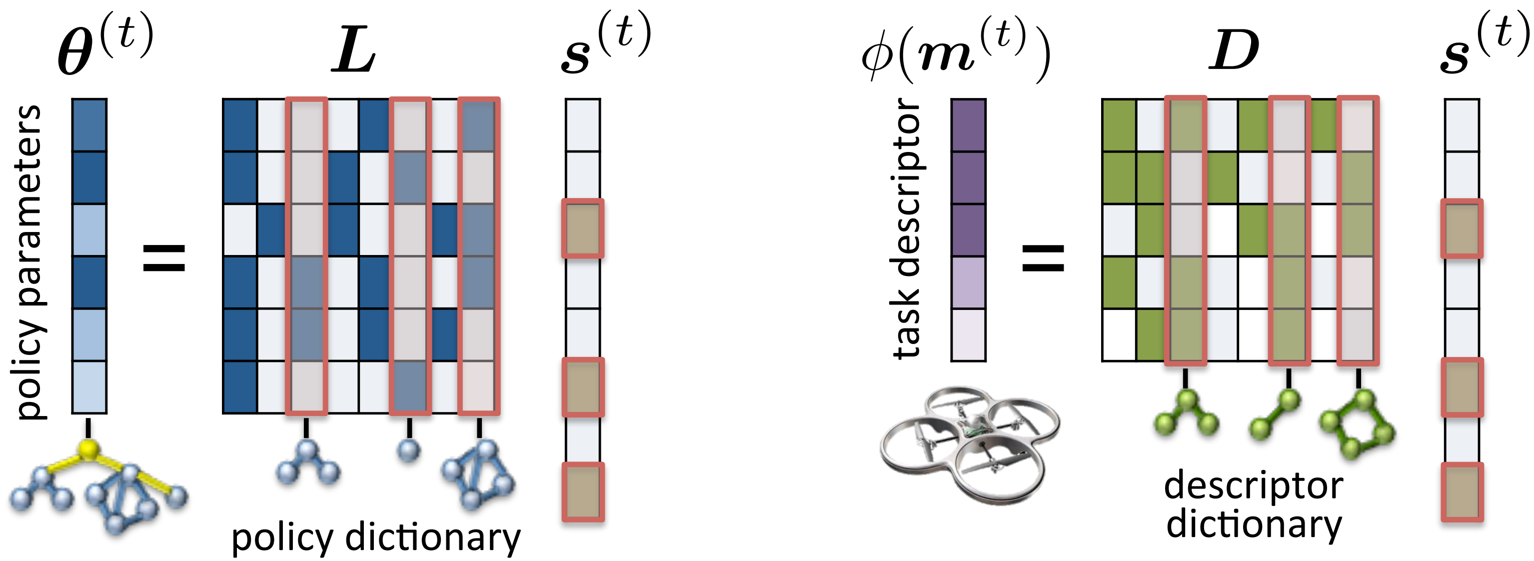
Incorporating Task Descriptors

Coupled dictionaries relate policy parameters and task descriptors



Incorporating Task Descriptors

Coupled dictionaries relate policy parameters and task descriptors

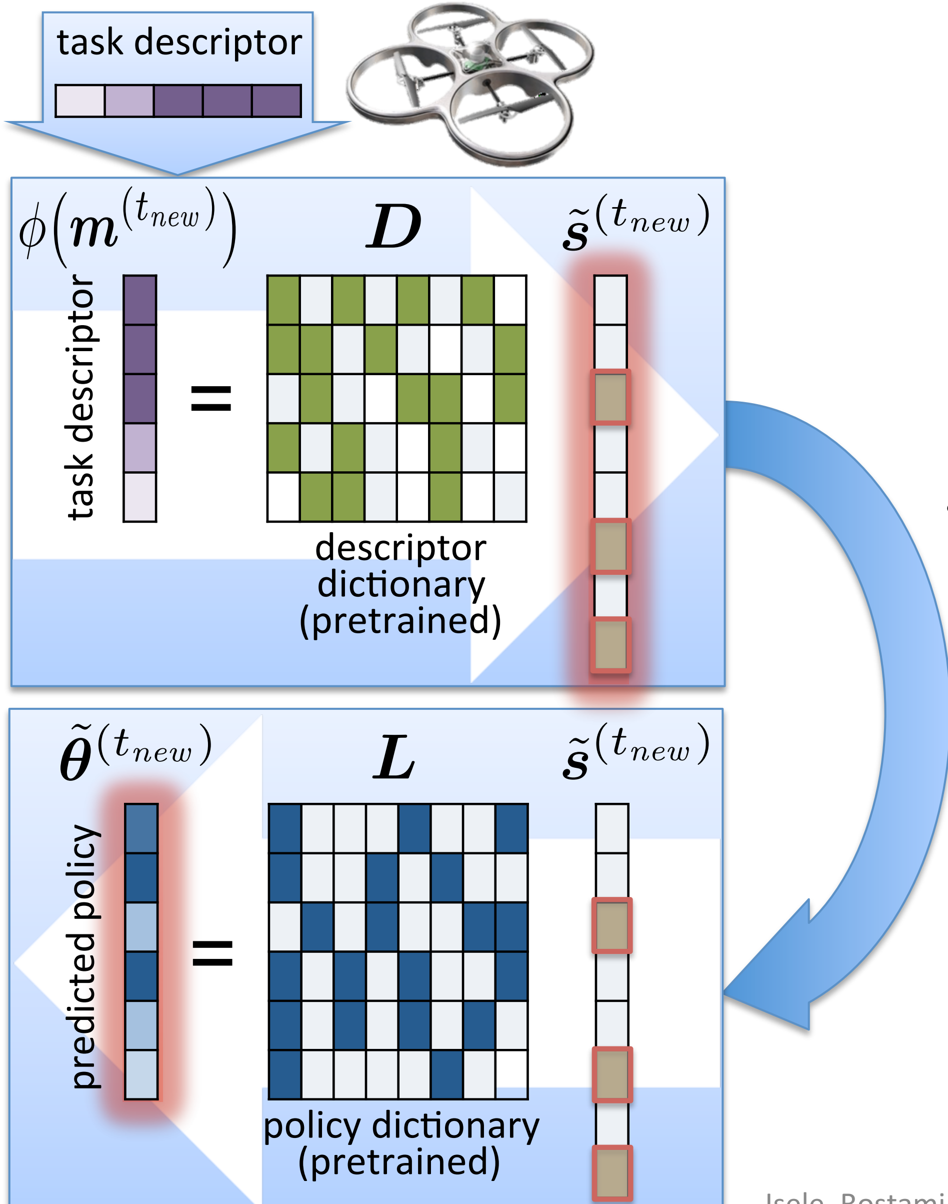


Solve:

$$\min_{\boldsymbol{L},\boldsymbol{D},\boldsymbol{S}} \frac{1}{T}\sum_t \left[\overbrace{\left\|\boldsymbol{\alpha}^{(t)} - \boldsymbol{L}\boldsymbol{s}^{(t)}\right\|^2_{\boldsymbol{\Gamma}^{(t)}}}^{\text{policy fit}} + \overbrace{\rho \left\|\phi(\boldsymbol{m}^{(t)}) - \boldsymbol{D}\boldsymbol{s}^{(t)}\right\|_2^2}^{\text{descriptor fit}} + \overbrace{\mu \left\|\boldsymbol{s}^{(t)}\right\|_1}^{\text{sparsity}} \right] + \overbrace{\lambda(\|\boldsymbol{L}\|_{\mathsf{F}}^2 + \|\boldsymbol{D}\|_{\mathsf{F}}^2)}^{\text{complexity}}$$

Multi-Task Learning: TaDeMTL **Lifelong Learning:** TaDeLL

Zero-Shot Transfer



Given: descriptor for new task

1. Use descriptor and descriptor dictionary to recover sparse coefficients via LASSO:

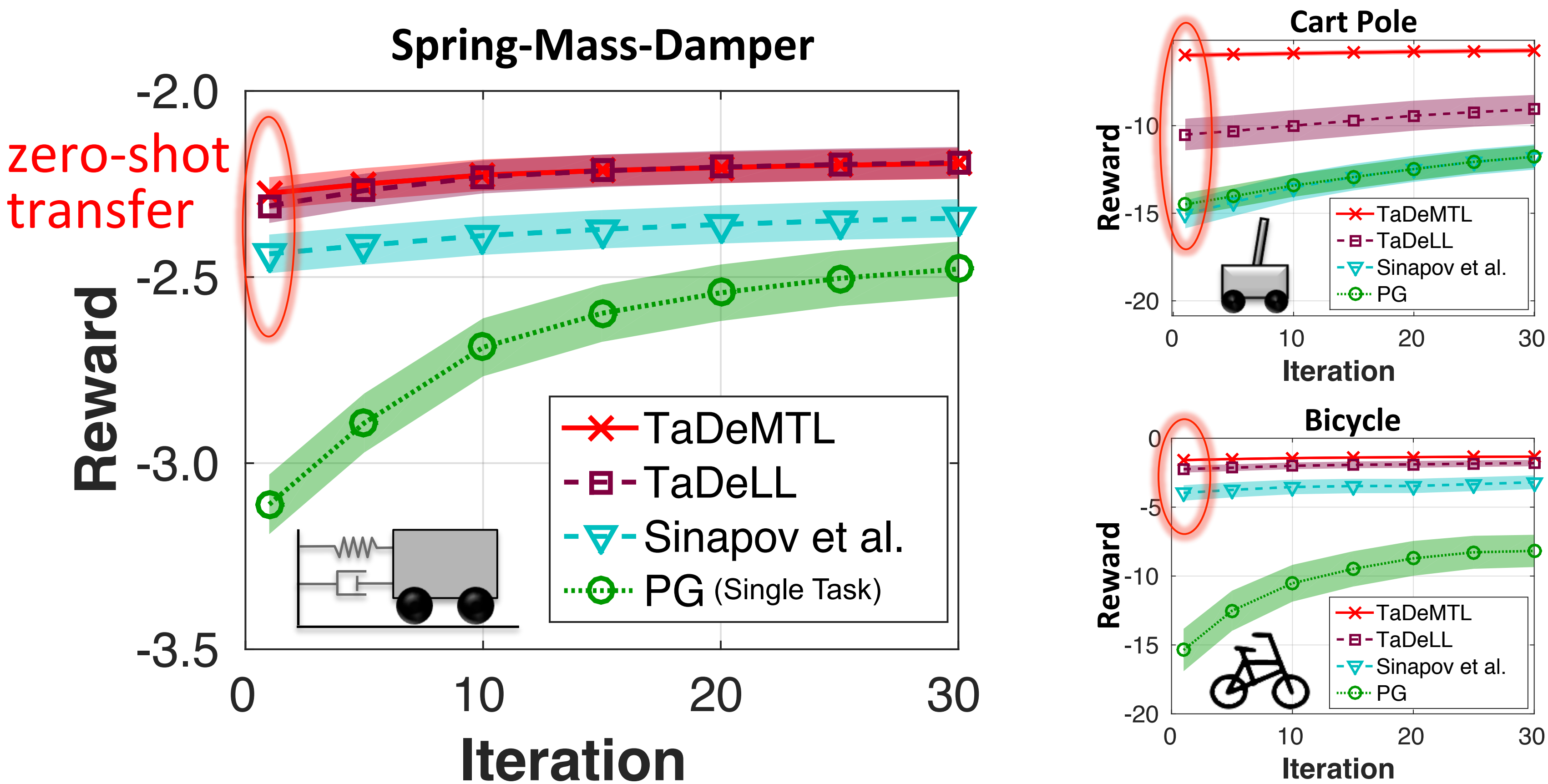
$$\tilde{\boldsymbol{s}}^{(t_{new})} \leftarrow \arg\min_{\boldsymbol{s}} \left\| \phi\left(\boldsymbol{m}^{(t)}\right) - \boldsymbol{D}\boldsymbol{s} \right\|_2^2 + \mu \left\| \boldsymbol{s} \right\|_1$$

2. Use recovered coefficients and policy dictionary to predict policy parameters

$$\tilde{\boldsymbol{\theta}}^{(t_{new})} = \boldsymbol{L}\tilde{\boldsymbol{s}}^{(t_{new})}$$

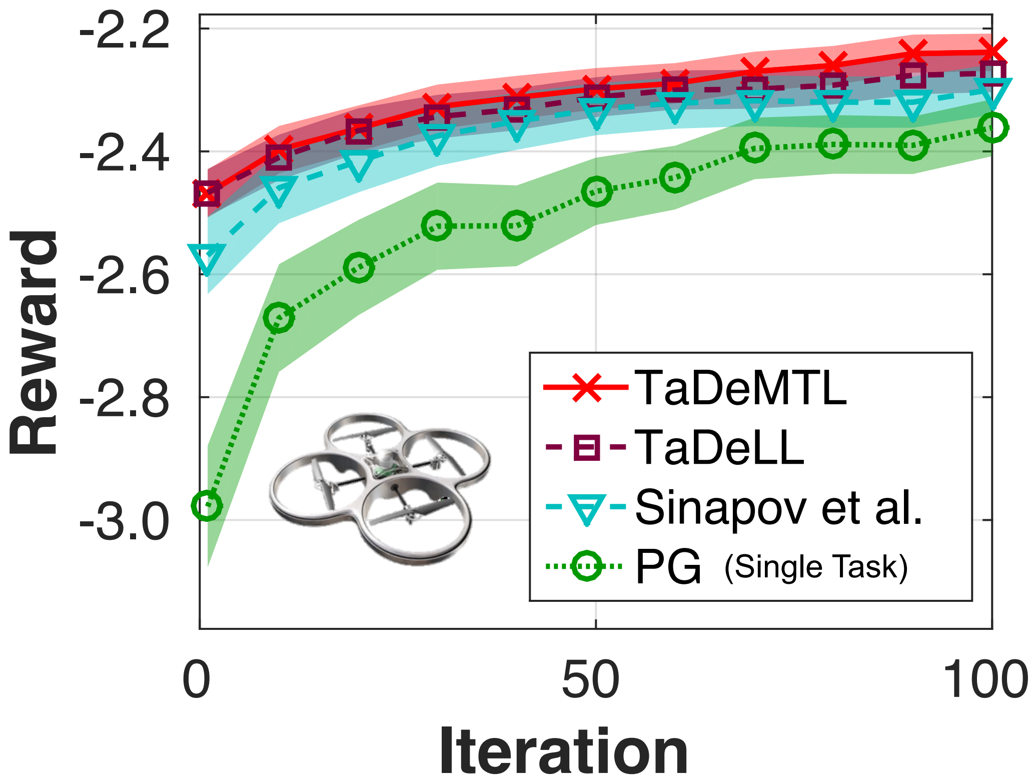
Lifelong Learning on Dynamical Systems

- Train on 40 tasks, predict the policy on a new task
- **Warm Start:** Zero-shot predicted policy used as an initialization



TaDeLL predicts effective policies for unseen tasks

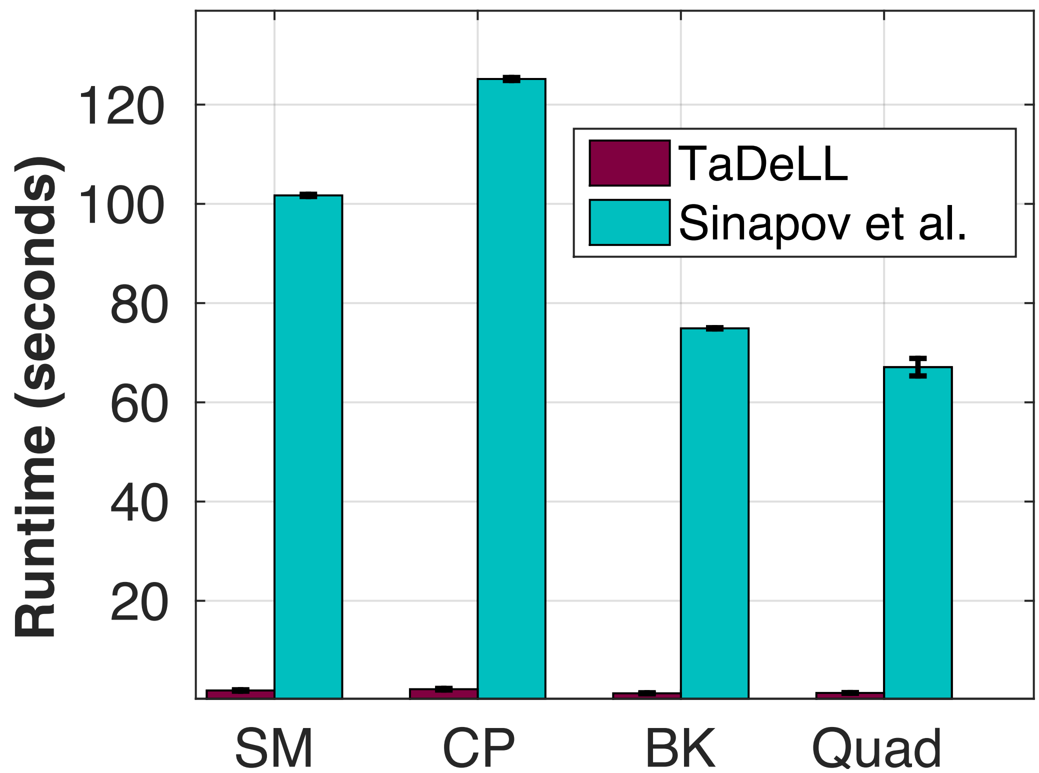
Application to Quadroter Control



Effective zero-shot transfer to controlling new quadrotor systems

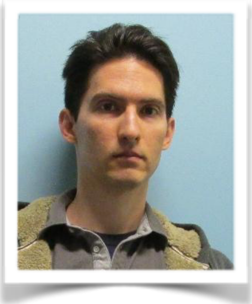
Runtime Comparison

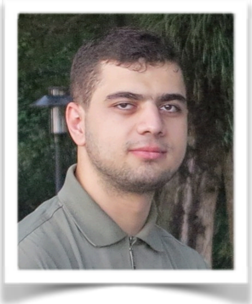
- TaDeLL scales effectively to numerous tasks
- Sinapov et al. has quadratic complexity in the number of tasks



Thank you!
Questions?

Using Task Features for Zero-Shot Knowledge Transfer in Lifelong Learning





David Isele*

Mohammad Rostami*

Eric Eaton

* Authors contributed equally

This research was supported by ONR grant #N00014-11-1-0139
and AFRL grant #FA8750-14-1-0069