# Deep Reinforcement Learning in a Handful of Trials using Probabilistic Dynamics Models

Kurtland Chua, Roberto Calandra, Rowan McAllister, Sergey Levine

***UC Berkeley***

## Motivation

- Model-free reinforcement learning (RL) often requires thousands or millions of trials to reach good policies.
- To apply RL in real-world, and especially in robotics, we are often limited by the number of trials that can be performed due to cost and time constraints.
- How can we increase the data-efficiency of current RL algorithms?
- **Contribution:**

  1) We propose a model-based RL approach based on learning deep probabilistic dynamics models. Our approach significantly outperforms the SOTA of both model-based and model-free RL methods.

  2) We perform a thorough ablation study of the importance of uncertainty for model-based RL approaches based on neural networks.

## Neural Network Dynamics Models

- To model the true forward dynamics $f$, we assume that the distribution of the next state is given by

$$p(s_{t+1}|s_t, a_t) = \mathcal{N}(\mu_\theta(s_t, a_t), \Sigma_\theta(s_t, a_t)). \qquad (1)$$

- **Probabilistic models** assume a diagonal covariance matrix (i.e. uncorrelated output dimensions).

$$l(\theta) = \sum_{i=1}^{n} \Delta_i^T \Sigma_\theta^{-1}(s_i, a_i)\Delta_i + \log\det\, \Sigma_\theta(s_i, a_i). \qquad (2)$$

  Compared to models without variance output, probabilistic models are better equipped to capture aleatoric uncertainty since they can model heteroscedastic noise.

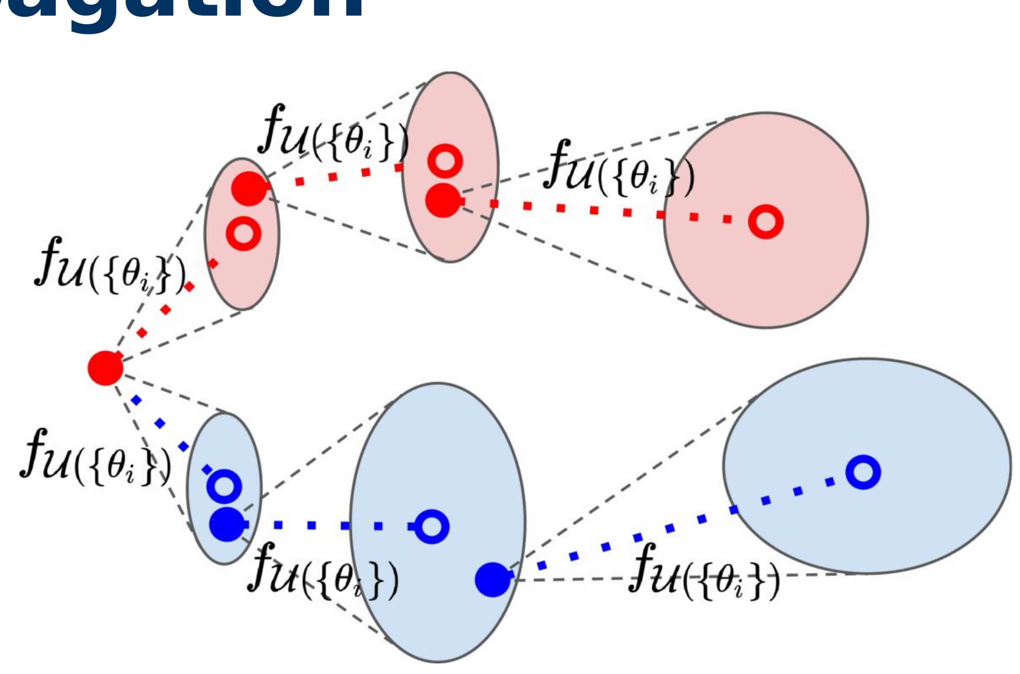
- **Probabilistic ensembles** consist of probabilistic models trained with $(2)$. For ensembles with $N$ networks, the output distribution mean $\mu$ is the output mean, while the covariance is given by

$$\frac{1}{N}\sum_{i=1}^{N} \Sigma_{\theta_i}(s, a) + \mathrm{diag}\,[\mu_{\theta_i}(s, a) - \mu]^2.$$

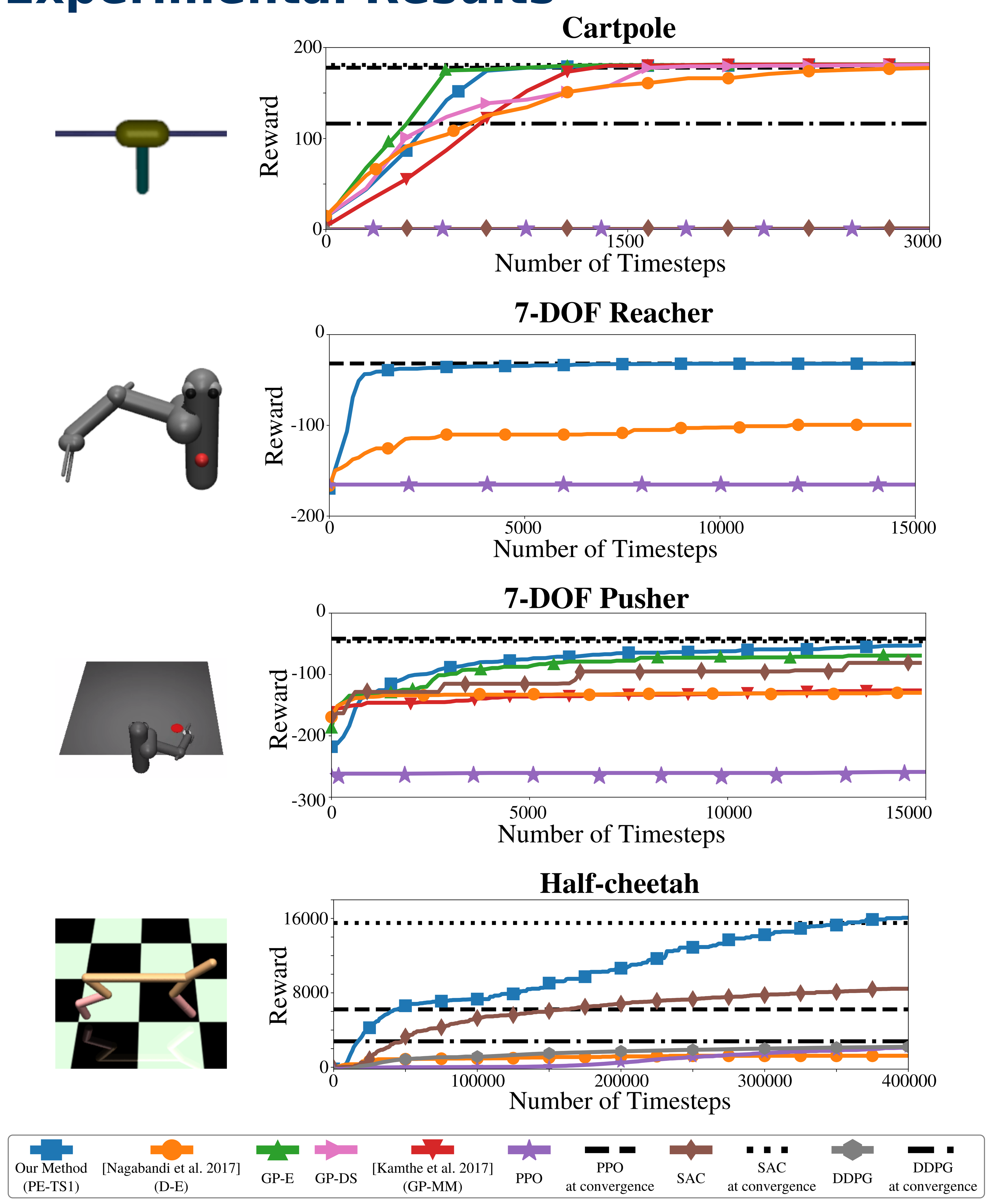
  Unlike their non-ensembled counterparts, these models can represent epistemic uncertainty (model uncertainty).

## Uncertainty Propagation

- **Trajectory Sampling-1 (TS1)** initializes particles and propagates each of them using a model sampled from an ensemble every time step.



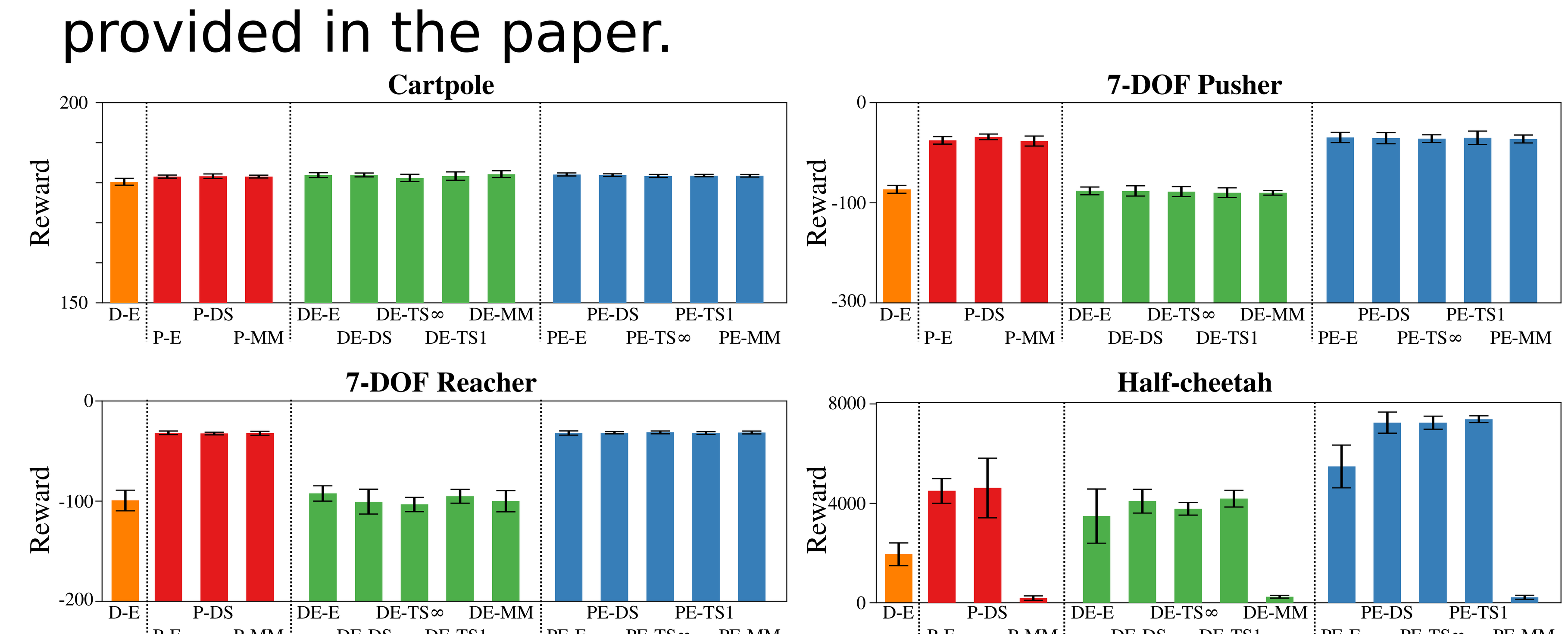

## Experimental Results




- As illustrated by the HalfCheetah results, our method can achieve better performance than most model-free methods, while using significantly less data.

## Ablation Study

- We also studied how the choice of model and propagation method affects performance; details are provided in the paper.




## Conclusion & Future Work

- We presented a model-based RL method that makes use of deep probabilistic dynamics models.
- Our approach is significantly more data-efficient than SOTA model-free approaches (25x faster), and can scale to high-dimensional tasks.
- In future work, we plan to evaluate our approach on real-robots.

Presented by Kurtland Chua - kchua@*berkeley.edu*