


# Bayesian Machine Learning for Controlling Autonomous Systems

**Marc Deisenroth**
Department of Computing
Imperial College London

Department of Computer Science
TU Darmstadt

m.deisenroth@imperial.ac.uk

Large-scale Online Learning and Decision Making Workshop
Cumberland Lodge
September 24, 2013

# Motivation

- Three key challenges in autonomous systems:

  **Modeling. Predicting. Decision making.**


**Robotics**

# Motivation

- Three key challenges in autonomous systems:

  **Modeling. Predicting. Decision making.**

- Noisy signals and processes


**Robotics**

# Motivation

- Three key challenges in autonomous systems:

  **Modeling. Predicting. Decision making.**

- Noisy signals and processes


**Robotics**

Increase autonomy: deal with uncertainty

⏩ **Bayesian machine learning**

# Motivation

- Three key challenges in autonomous systems:

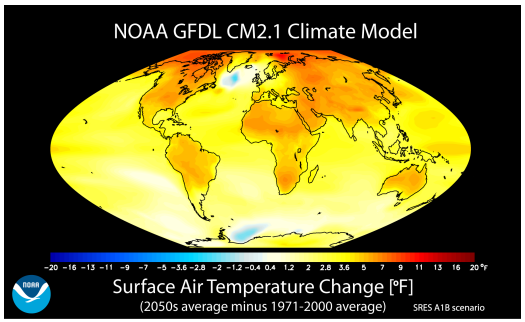
  **Modeling. Predicting. Decision making.**

- Noisy signals and processes


**Robotics**

Increase autonomy: deal with uncertainty

⏩ **Bayesian machine learning**




Climate Science

# Motivation

- Three key challenges in autonomous systems:

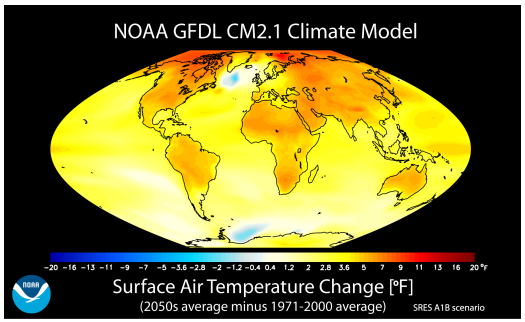
  **Modeling. Predicting. Decision making.**

- Noisy signals and processes

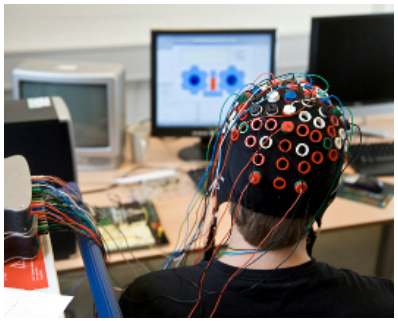

**Robotics**

Increase autonomy: deal with uncertainty

⏩ **Bayesian machine learning**




Climate Science



Brain-Computer Interface

# Motivation


**Robotics**

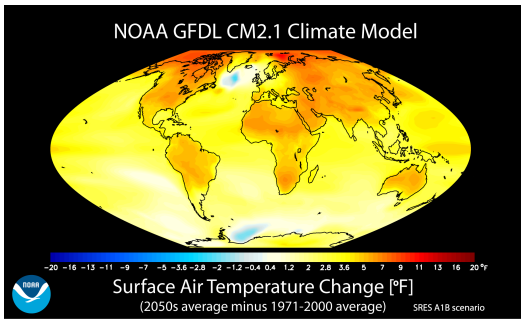
- Three key challenges in autonomous systems:

  **Modeling. Predicting. Decision making.**

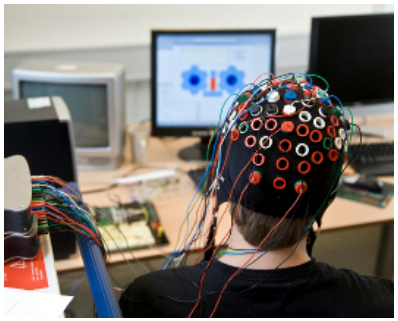
- Noisy signals and processes

Increase autonomy: deal with uncertainty

⏩ **Bayesian machine learning**



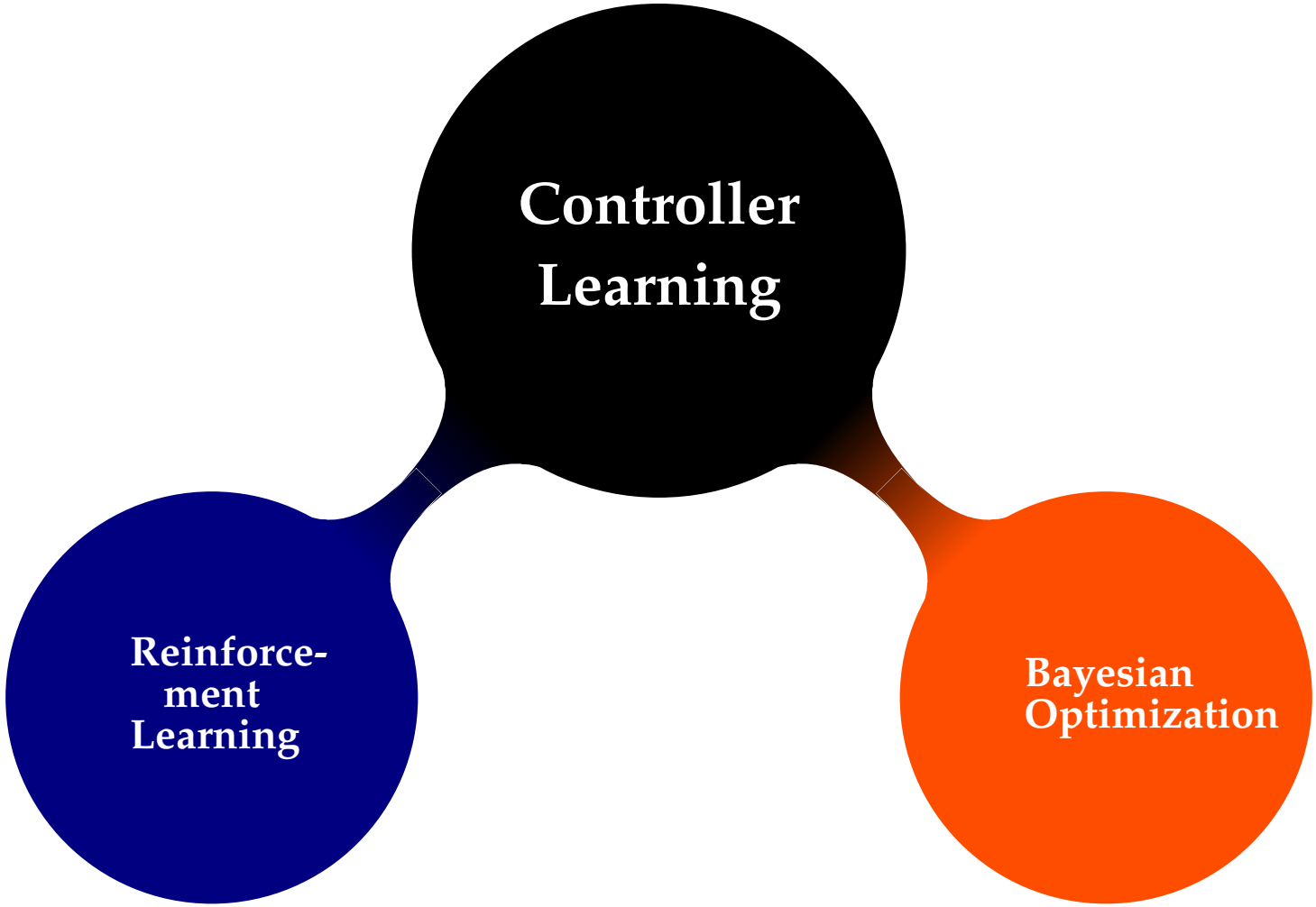

Climate Science



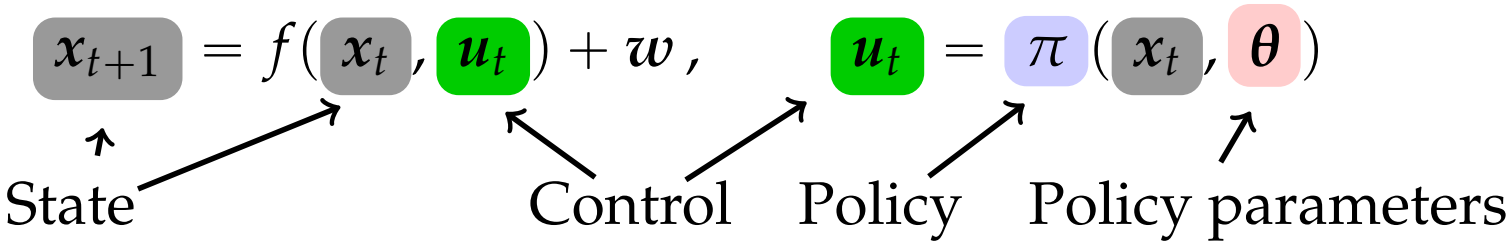
Brain-Computer Interface

Prosthetics

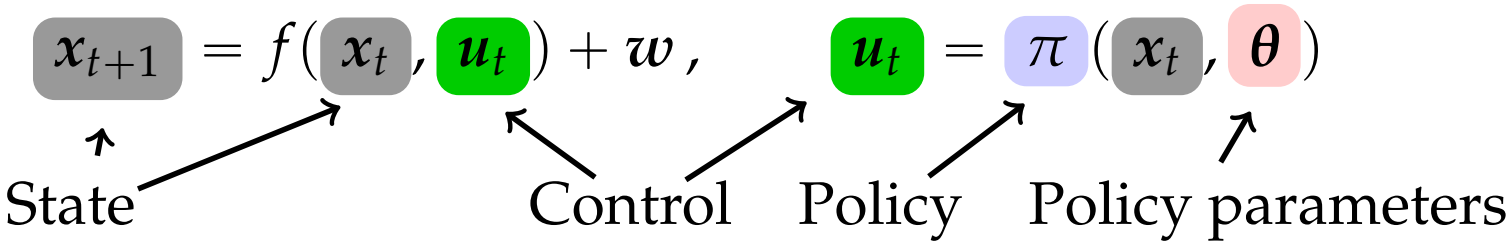
# Outline


Controller Learning
Reinforce-ment Learning
Bayesian Optimization

# Reinforcement Learning Set-up



# Reinforcement Learning Set-up




## Objective

Find policy parameters $\theta^*$ that minimize the expected long-term cost

$$J(\theta) = \sum_{t=1}^{T} \mathbb{E}[c(x_t)|\theta]\,, \qquad p(x_0) = \mathcal{N}(\mu_0, \Sigma_0)\,.$$

Instantaneous cost $c(x_t)$, e.g., $\|x_t - x_{\text{target}}\|^2$

- Typical objective in optimal control and reinforcement learning (Bertsekas, 2005; Sutton & Barto, 1998)

# Model-based Policy Search

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

**High-Level Steps:**

1. Probabilistic model for transition function $f$ to be robust to model errors

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Model-based Policy Search

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

**High-Level Steps:**

1. Probabilistic model for transition function $f$ to be robust to model errors
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Model-based Policy Search

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

**High-Level Steps:**

1. Probabilistic model for transition function $f$ to be robust to model errors
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Model-based Policy Search

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

**High-Level Steps:**

1. Probabilistic model for transition function $f$ to be robust to model errors
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement
4. Apply controller

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Model-based Policy Search

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

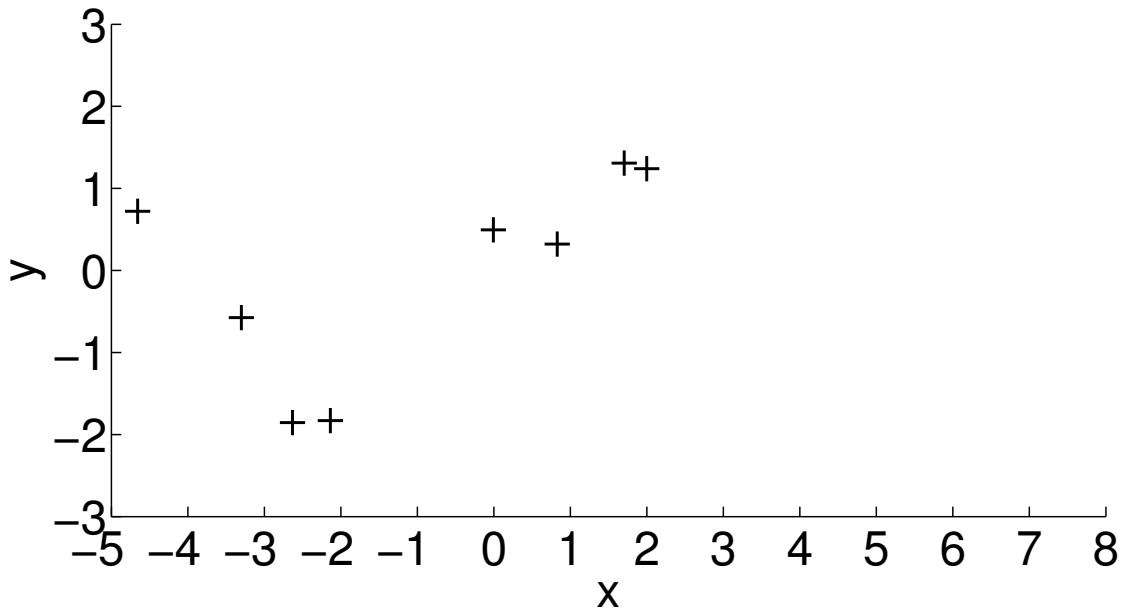
**High-Level Steps:**

1. **Probabilistic model for transition function $f$ to be robust to model errors**
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement
4. Apply controller

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Model Learning

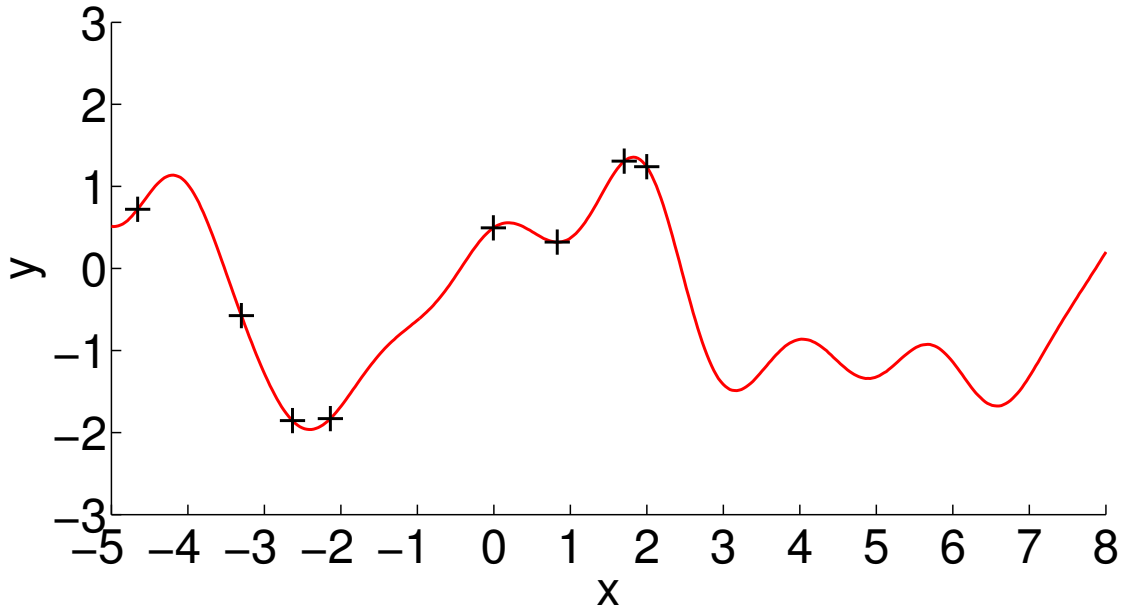
Model learning problem: Find a function $f : x \mapsto f(x) = y$




Observed function values

# Model Learning

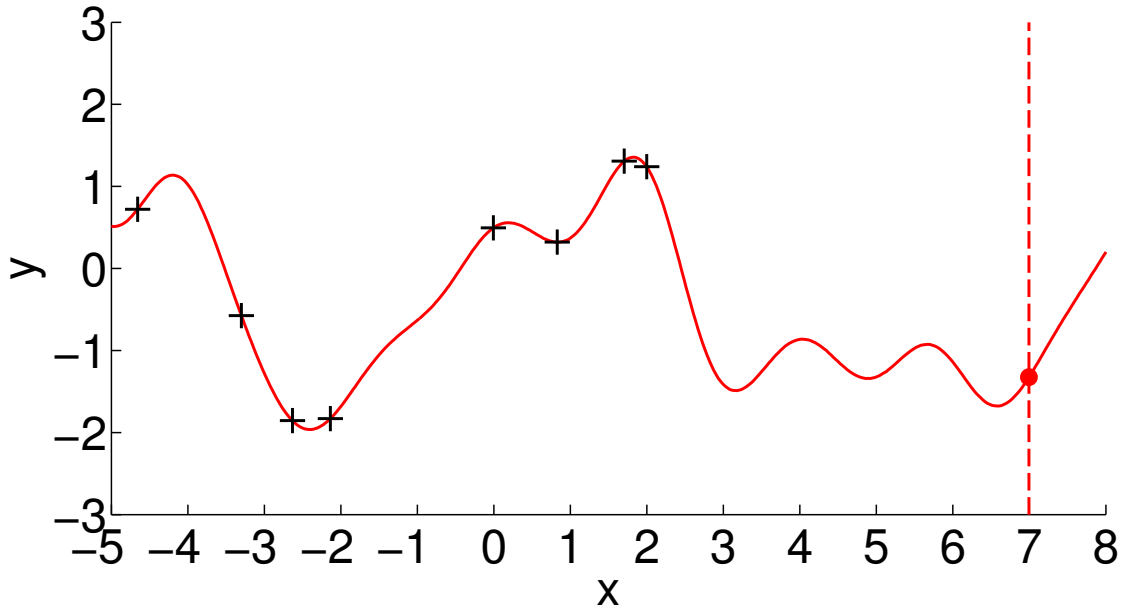
Model learning problem: Find a function $f : x \mapsto f(x) = y$




Plausible function approximators

# Model Learning

Model learning problem: Find a function $f : x \mapsto f(x) = y$

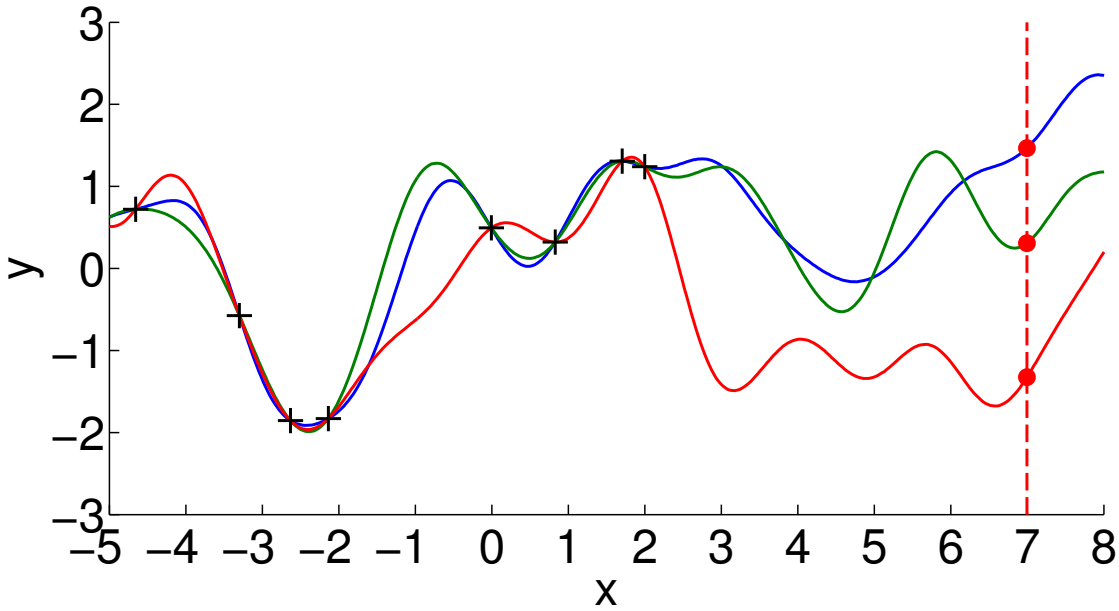



Plausible function approximators

**Predictions? Decision Making?**

# Model Learning

Model learning problem: Find a function $f : x \mapsto f(x) = y$

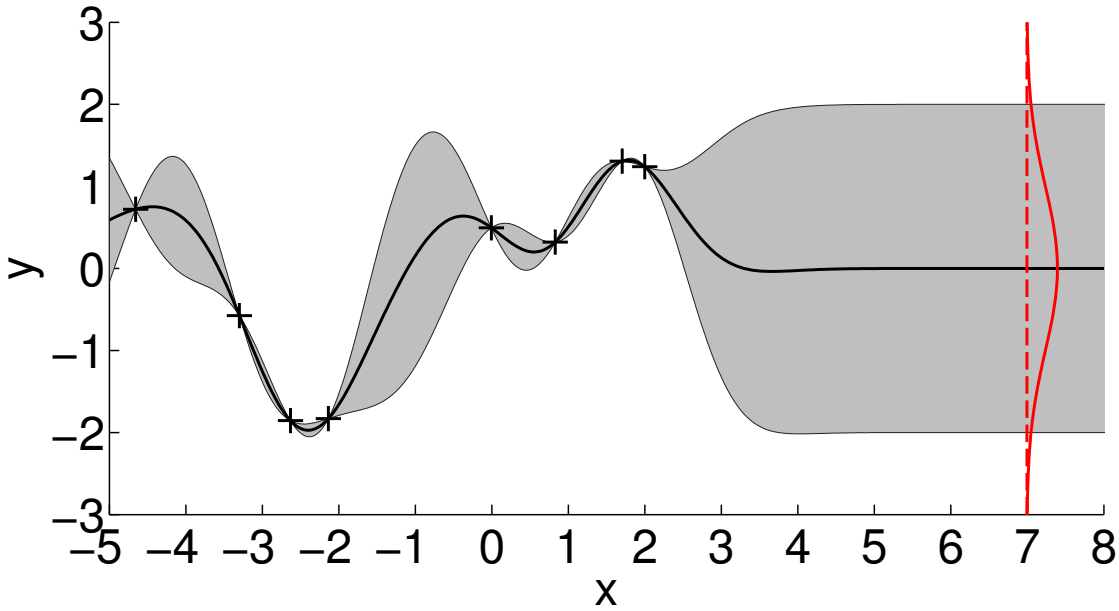



Plausible function approximators

**Predictions? Decision Making? Model Errors!**

# Model Learning

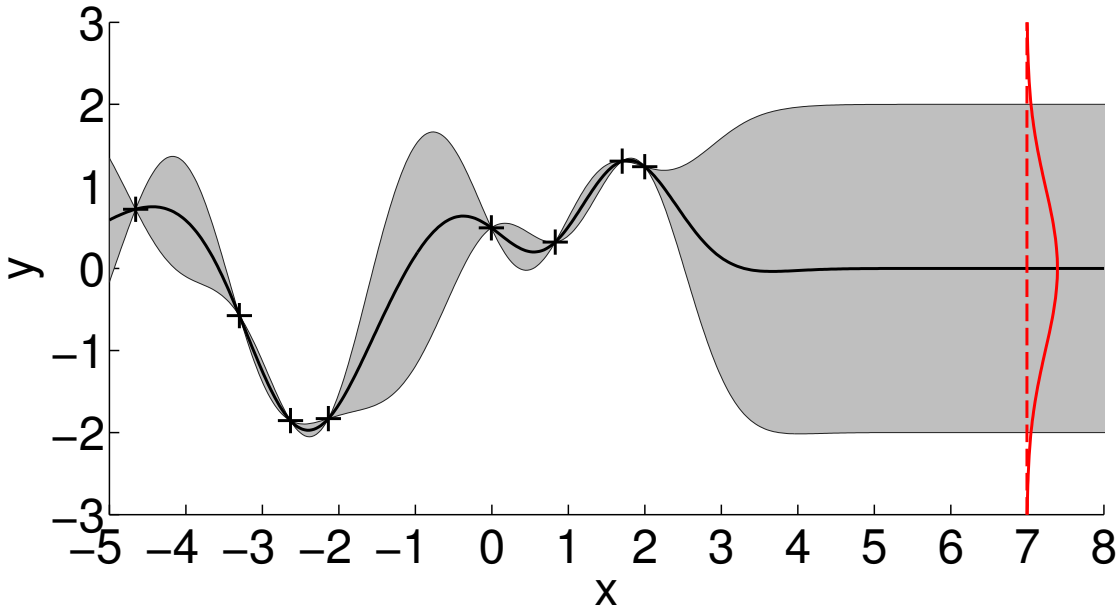
Model learning problem: Find a function $f : x \mapsto f(x) = y$




Distribution over plausible functions

# Model Learning

Model learning problem: Find a function $f : x \mapsto f(x) = y$




Distribution over plausible functions

- Express uncertainty about the underlying function
- Gaussian process for model learning (Rasmussen & Williams, 2006)

# Model-based Policy Search

## Objective

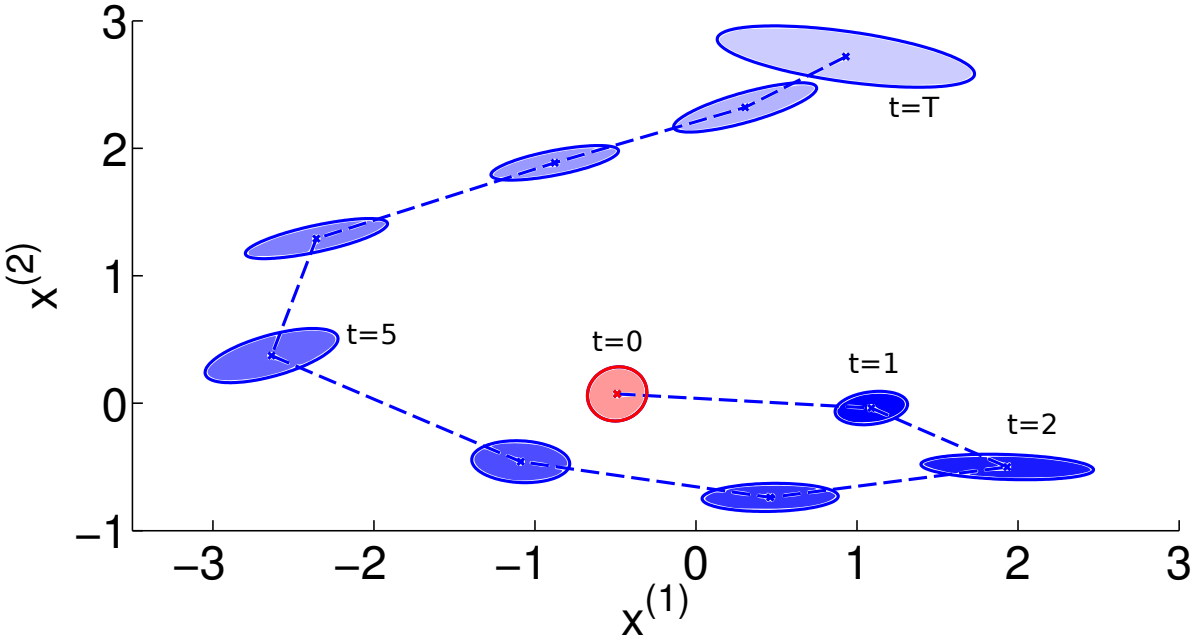
Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

**High-Level Steps:**

1. Probabilistic model for transition function $f$ to be robust to model errors
2. **Compute long-term predictions** $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. Policy improvement
4. Apply controller

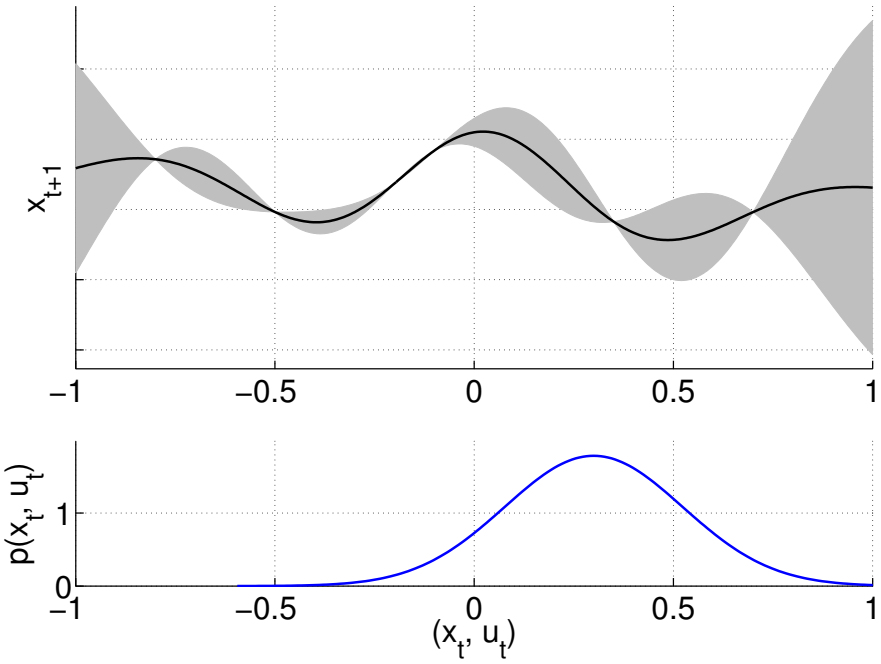
Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Long-Term Predictions




- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

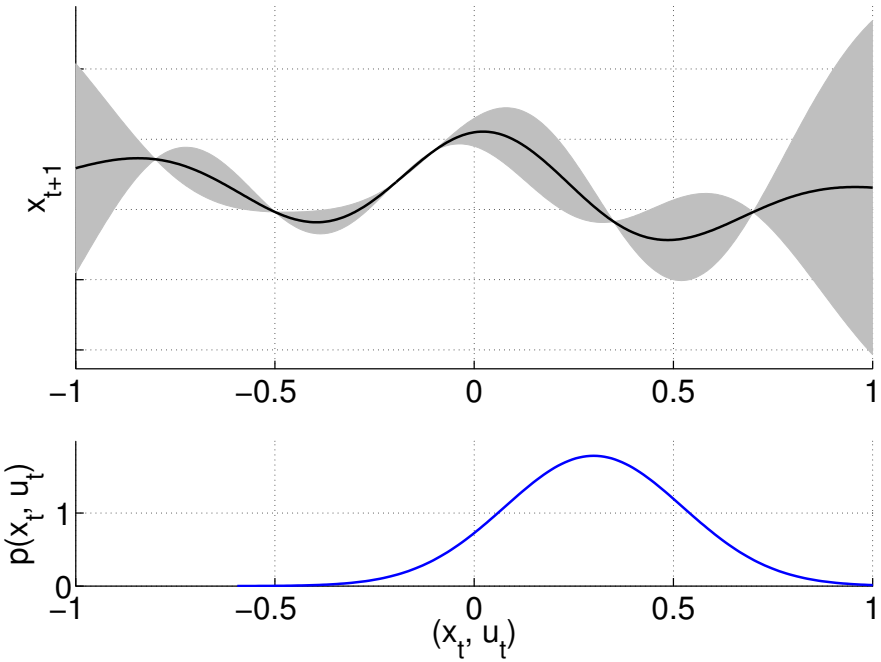
# Long-Term Predictions




- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

$$\underbrace{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)}_{\text{GP prediction}} \quad \underbrace{p(\boldsymbol{x}_t, \boldsymbol{u}_t|\boldsymbol{\theta})}_{\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})}$$

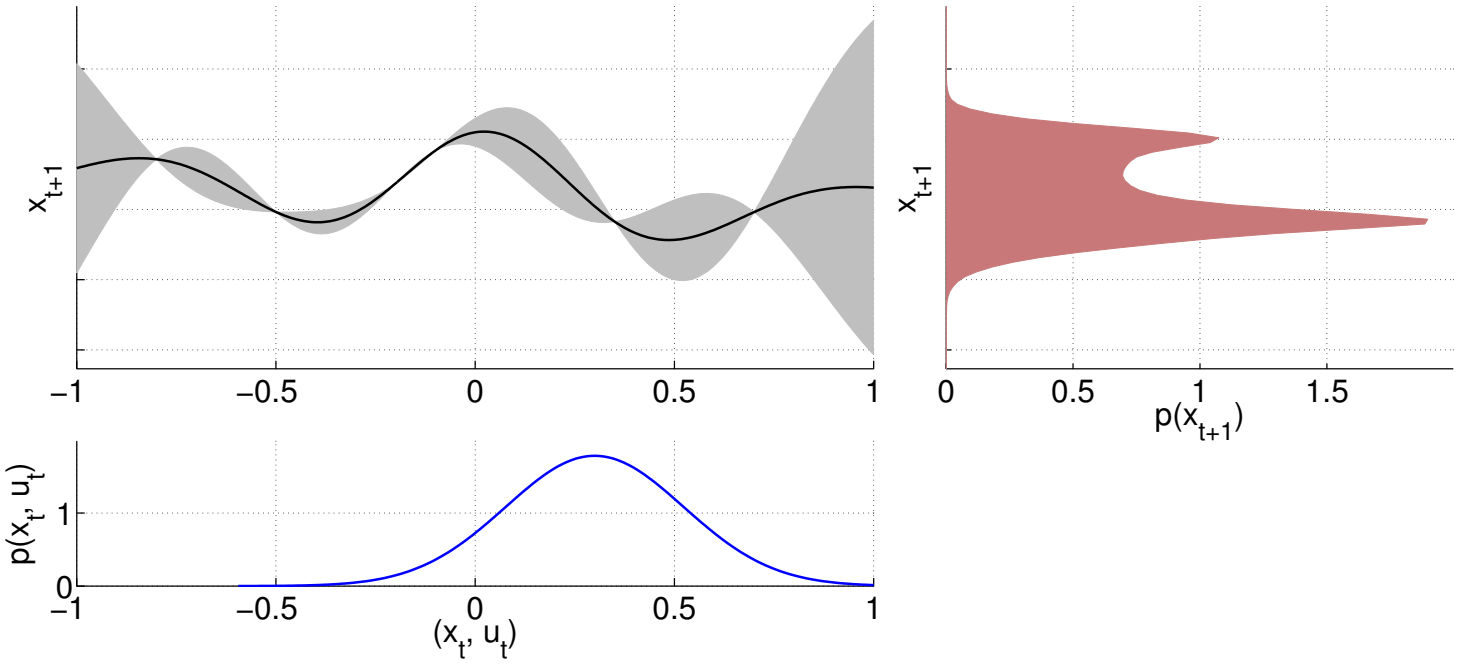
# Long-Term Predictions




- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}) = \iiint \underbrace{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)}_{\text{GP prediction}} \underbrace{p(\boldsymbol{x}_t, \boldsymbol{u}_t|\boldsymbol{\theta})}_{\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathrm{d}f \, \mathrm{d}\boldsymbol{x}_t \, \mathrm{d}\boldsymbol{u}_t$$

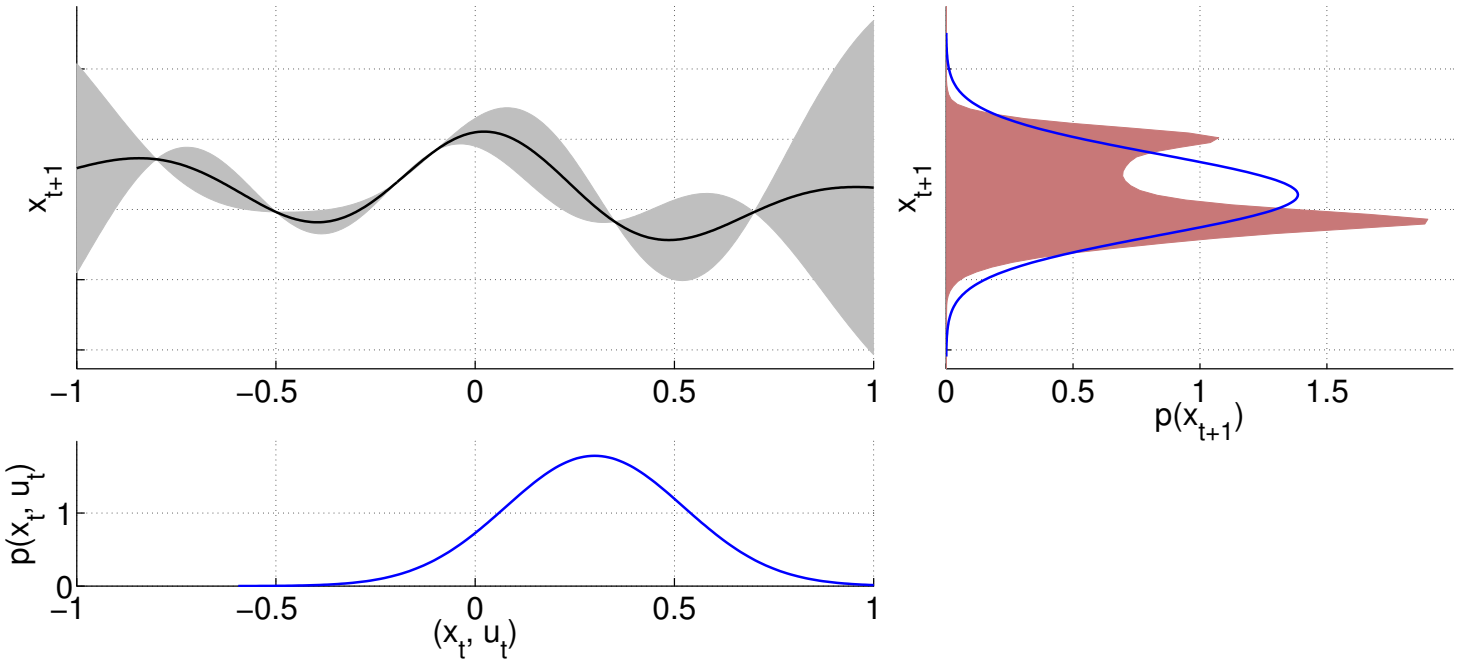
# Long-Term Predictions




- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}) = \iiint \underbrace{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)}_{\text{GP prediction}} \underbrace{p(\boldsymbol{x}_t, \boldsymbol{u}_t|\boldsymbol{\theta})}_{\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathrm{d}f \, \mathrm{d}\boldsymbol{x}_t \, \mathrm{d}\boldsymbol{u}_t$$

# Long-Term Predictions




- Iteratively compute $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

$$p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}) = \iiint \underbrace{p(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, \boldsymbol{u}_t)}_{\text{GP prediction}} \underbrace{p(\boldsymbol{x}_t, \boldsymbol{u}_t|\boldsymbol{\theta})}_{\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})} \mathrm{d}f \, \mathrm{d}\boldsymbol{x}_t \, \mathrm{d}\boldsymbol{u}_t$$

  - ⏩ Approximate inference
- Moment matching (Quiñonero-Candela et al., 2003)

# Model-based Policy Search

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

**High-Level Steps:**

1. Probabilistic model for transition function $f$ to be robust to model errors
2. Compute long-term predictions $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
3. **Policy improvement**
   - Compute expected long-term cost $J(\boldsymbol{\theta})$
   - Find parameters $\boldsymbol{\theta}$ that minimize $J(\boldsymbol{\theta})$
4. Apply controller

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Policy Improvement

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

- Know how to predict $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Policy Improvement

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

- Know how to predict $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
- Compute

$$\mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}] = \int c(\boldsymbol{x}_t)\mathcal{N}\left(\boldsymbol{x}_t \,|\, \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)\mathrm{d}\boldsymbol{x}_t\,, \quad t = 1, \ldots, T\,,$$

and sum them up to obtain $J(\boldsymbol{\theta})$

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Policy Improvement

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

- Know how to predict $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
- Compute

$$\mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}] = \int c(\boldsymbol{x}_t)\mathcal{N}\left(\boldsymbol{x}_t \,|\, \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)\mathrm{d}\boldsymbol{x}_t\,, \quad t = 1, \ldots, T\,,$$

  and sum them up to obtain $J(\boldsymbol{\theta})$
- Analytically compute gradient $\mathrm{d}J(\boldsymbol{\theta})/\mathrm{d}\boldsymbol{\theta}$
- Standard gradient-based optimizer (e.g., BFGS) to find $\boldsymbol{\theta}^*$

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Policy Improvement

## Objective

Minimize expected long-term cost $J(\boldsymbol{\theta}) = \sum_t \mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}]$

- Know how to predict $p(\boldsymbol{x}_1|\boldsymbol{\theta}), \ldots, p(\boldsymbol{x}_T|\boldsymbol{\theta})$
- Compute

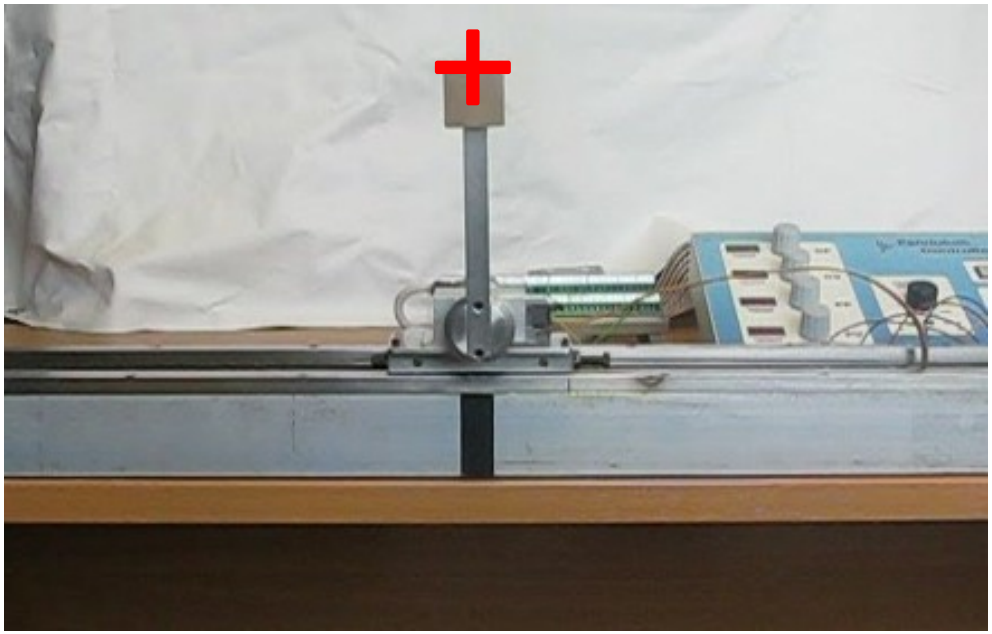
$$\mathbb{E}[c(\boldsymbol{x}_t)|\boldsymbol{\theta}] = \int c(\boldsymbol{x}_t)\mathcal{N}\left(\boldsymbol{x}_t \,|\, \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t\right)\mathrm{d}\boldsymbol{x}_t\,, \quad t = 1, \ldots, T\,,$$

  and sum them up to obtain $J(\boldsymbol{\theta})$
- Analytically compute gradient $\mathrm{d}J(\boldsymbol{\theta})/\mathrm{d}\boldsymbol{\theta}$
- Standard gradient-based optimizer (e.g., BFGS) to find $\boldsymbol{\theta}^*$

⏩ PILCO framework for controller learning

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Standard Benchmark Problem: Cart-Pole Swing-up

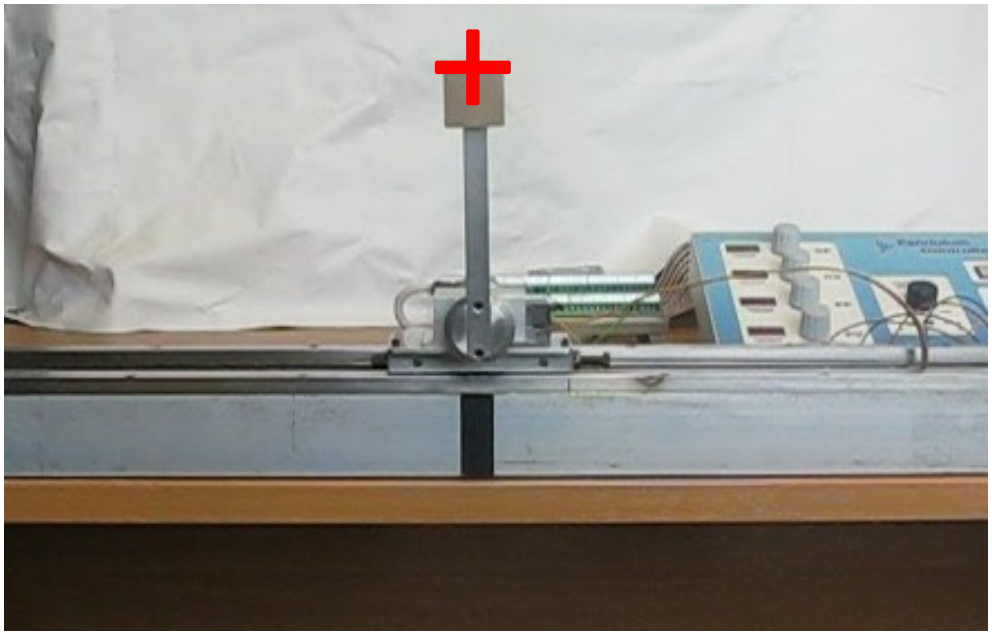


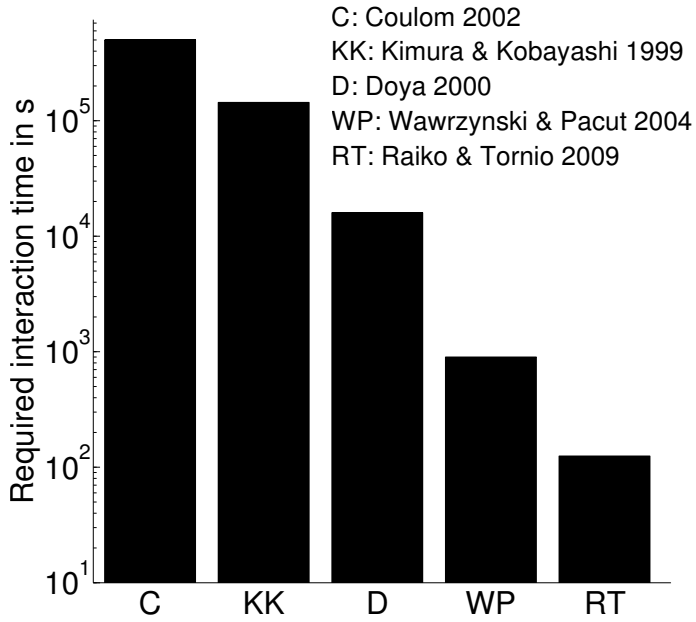
- Swing up and balance a freely swinging pendulum on a cart
- Cost function $c(\boldsymbol{x}) = -\exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$

- Code available at `http://mlg.eng.cam.ac.uk/pilco`

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Standard Benchmark Problem: Cart-Pole Swing-up

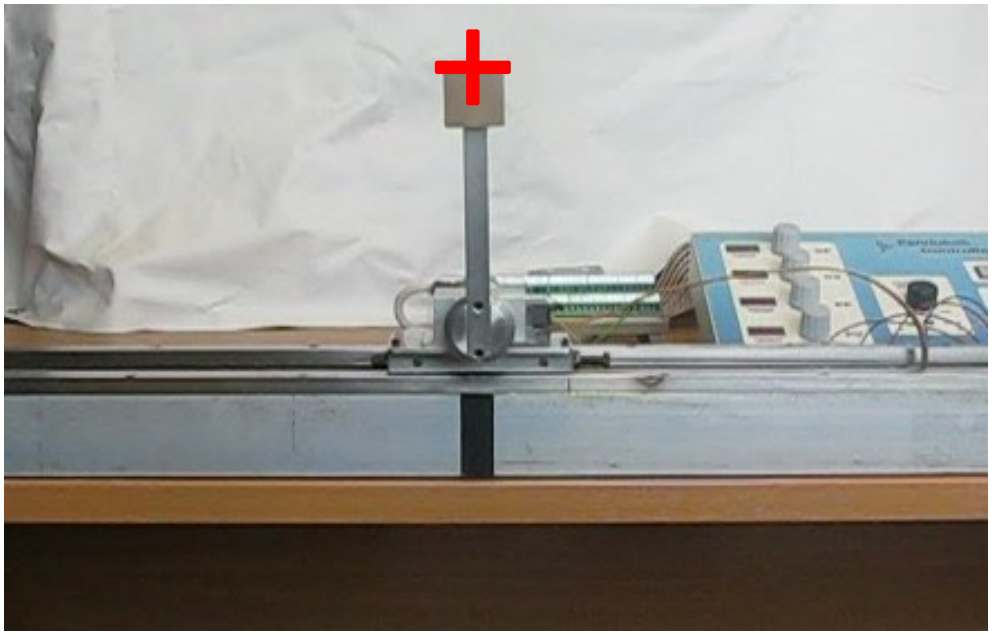


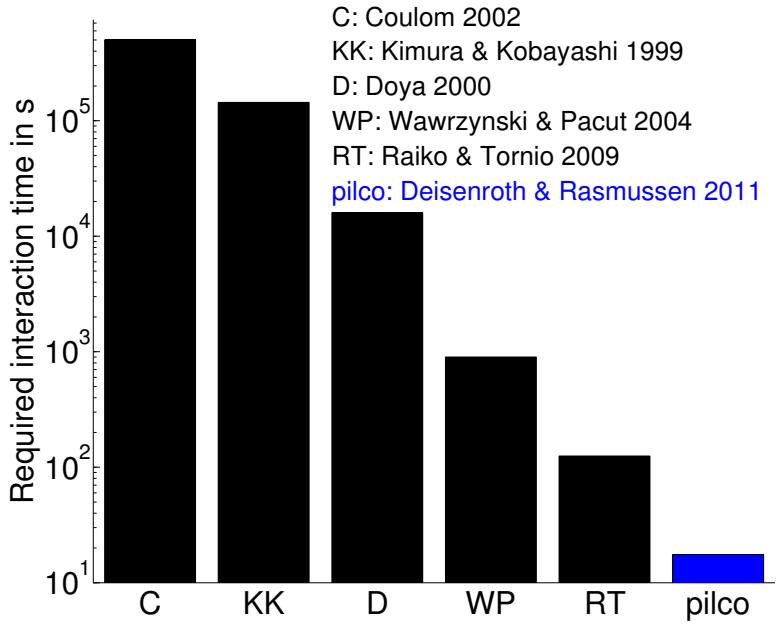



- Swing up and balance a freely swinging pendulum on a cart
- Cost function $c(\boldsymbol{x}) = -\exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- Code available at `http://mlg.eng.cam.ac.uk/pilco`

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

# Standard Benchmark Problem: Cart-Pole Swing-up

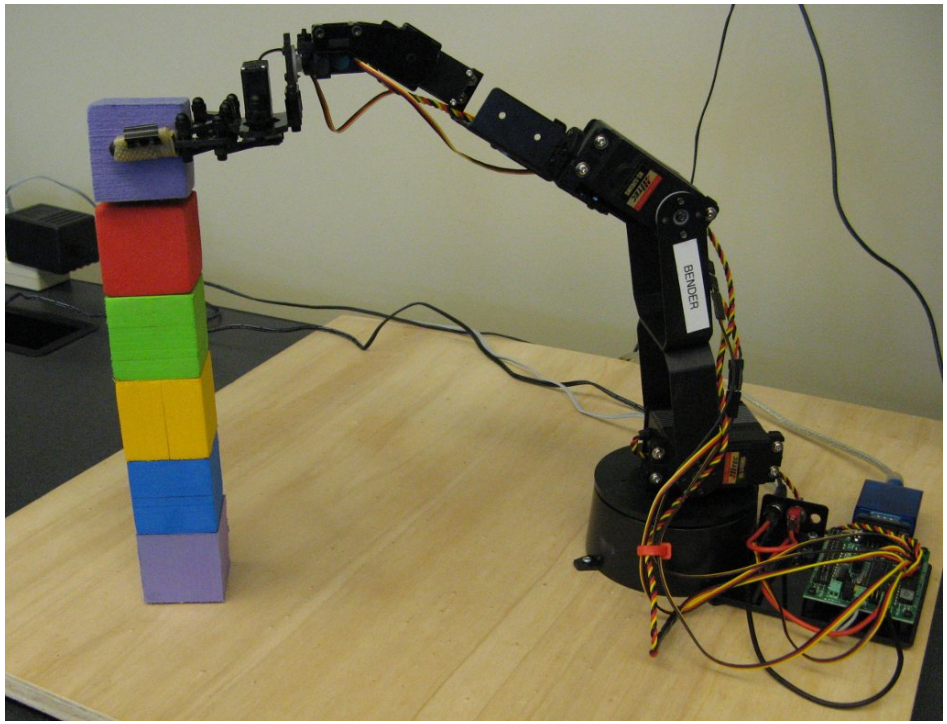





- Swing up and balance a freely swinging pendulum on a cart
- Cost function $c(\boldsymbol{x}) = -\exp(-\|\boldsymbol{x} - \boldsymbol{x}_{\text{target}}\|^2)$
- **Unprecedented learning speed** compared to state-of-the-art
- Code available at `http://mlg.eng.cam.ac.uk/pilco`

Deisenroth & Rasmussen (ICML, 2011): *PILCO: A Model-based and Data-efficient Approach to Policy Search*

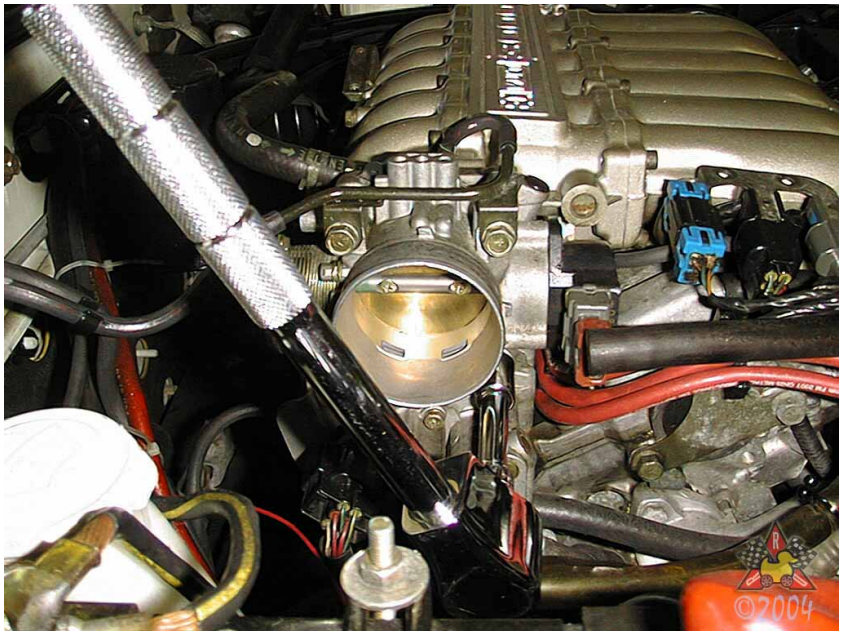
# Learning to Control an Off-the-Shelf Robot




- Autonomously learn block-stacking with a low-cost robot
- Robot very noisy
- Learn forward model and controller from scratch

Deisenroth et al. (RSS, 2011): *Learning to Control a Low-Cost Manipulator using Data-efficient Reinforcement Learning*

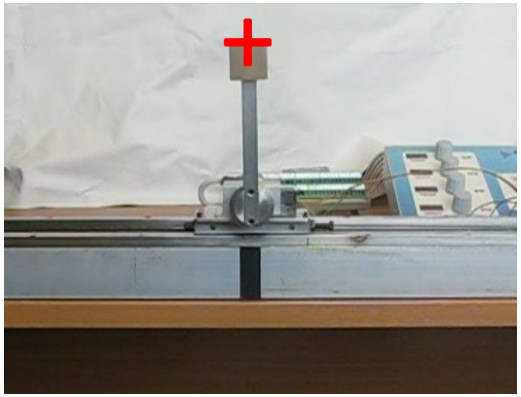
# Controlling Throttle Valves in Combustion Engines

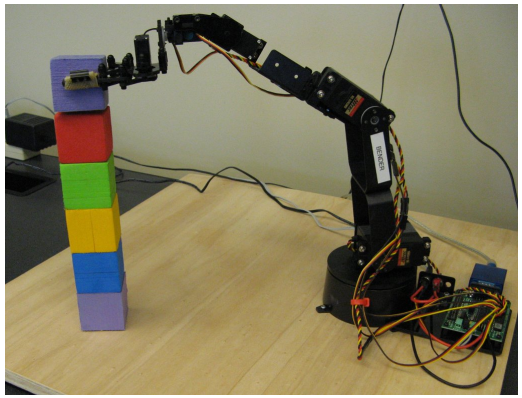



Bischoff et al., ECML 2013

- More videos at `http://www.youtube.com/user/PilcoLearner`

# Summary (1)

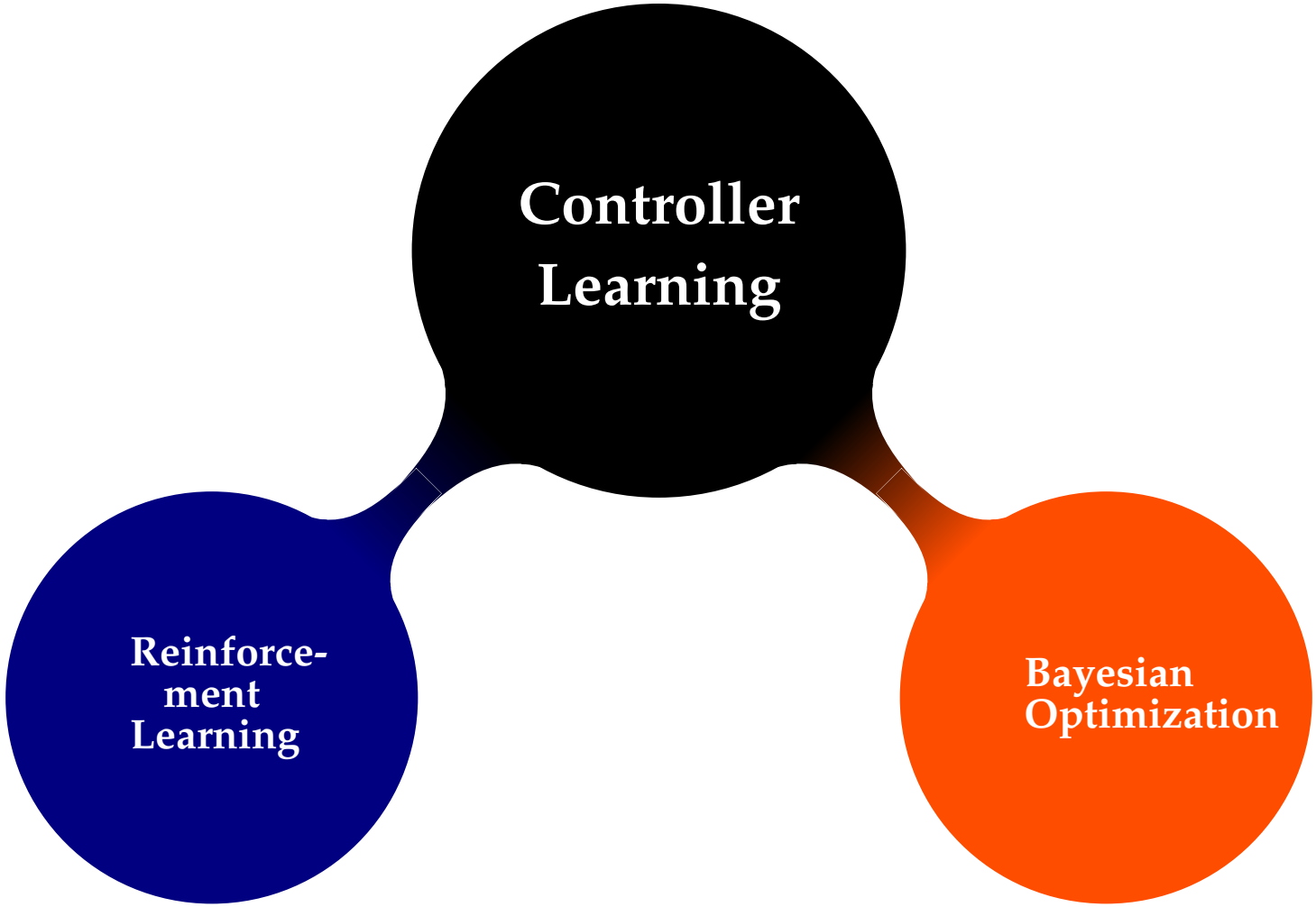




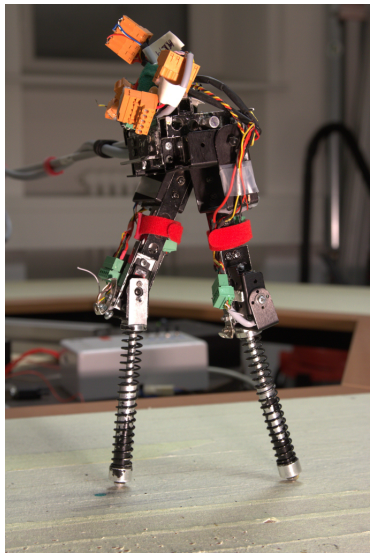
## Practical Framework for Autonomous Learning

- Key: Explicit incorporation of model uncertainty into long-term predictions and decision making
- Applied to real systems

# Outline


Controller Learning
Reinforce-ment Learning
Bayesian Optimization

# Bayesian Optimization for Learning Controllers

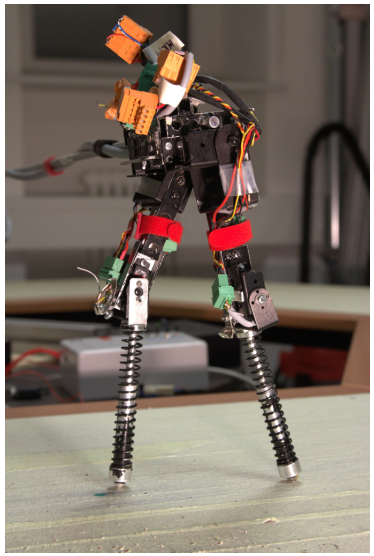


- Learning forward models is not always easy
- Legged locomotion: ground contacts

**Objective**

Find parameters $\theta$ of controller $\pi(\theta)$

# Bayesian Optimization for Learning Controllers



- Learning forward models is not always easy
- Legged locomotion: ground contacts

**Objective**

Find parameters $\theta$ of controller $\pi(\theta)$

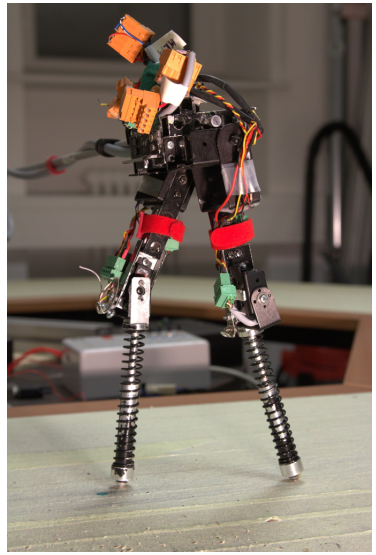
**Challenges:**

- No forward model
- No analytic cost function, no demonstrations
- Still need to be data efficient (fragile robot)
- Manual parameter search is tedious

⏩ Bayesian optimization (e.g., Jones 1998; Brochu et al., 2009)

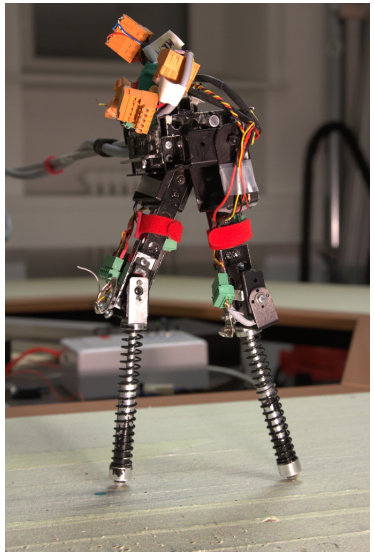
# Bayesian Gait Optimization for Legged Locomotion

- Maximize robustness and walking speed
- 4 motors:
  2 actuated hips + 2 actuated knees
- Controller implemented as a finite-state-machine (8 parameters)
- Good parameters found after 100 experiments



Calandra et al. (submitted): *Bayesian Gait Optimization for Bipedal Locomotion*

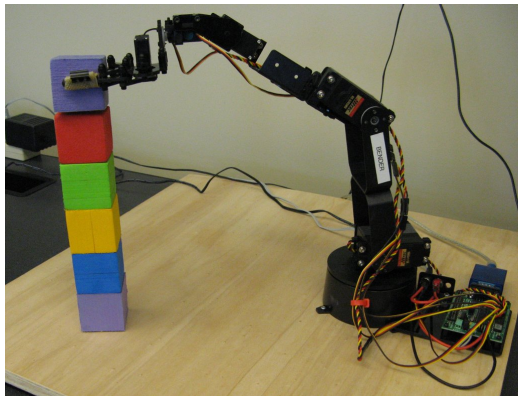
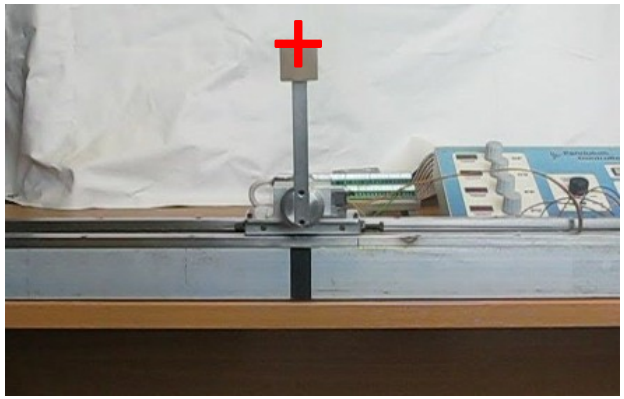
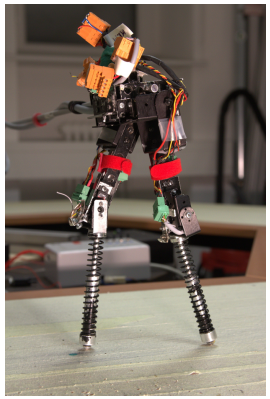
# Summary (2)



## Bayesian Gait Optimization

- Bayesian optimization for learning controllers in a few experiments
- General framework
  (no assumptions on dynamics, no explicit cost required)
- Limited to few parameters ($\approx$ 10–20)

# Wrap-up

- Controller learning for autonomous systems
  - Reinforcement learning
  - Bayesian optimization
- Key to success: Probabilistic modeling and Bayesian inference

m.deisenroth@imperial.ac.uk

**Thank you for your attention**

# References


[1] D. P. Bertsekas. Dynamic Programming and Optimal Control, volume 1 of Optimization and Computation Series. Athena Scientific, Belmont, MA, USA, 3rd edition, 2005.

[2] B. Bischoff, D. Nguyen-Tuong, T. Koller, H. Markert, and A. Knoll. Learning Throttle Valve Control Using Policy Search. In Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, 2013.

[3] E. Brochu, V. M. Cora, and N. de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. Technical Report TR-2009-023, Department of Computer Science, University of British Columbia, 2009.

[4] M. P. Deisenroth. Efficient Reinforcement Learning using Gaussian Processes, volume 9 of Karlsruhe Series on Intelligent Sensor-Actuator-Systems. KIT Scientific Publishing, November 2010. ISBN 978-3-86644-569-7.

[5] M. P. Deisenroth and C. E. Rasmussen. PILCO: A Model-Based and Data-Efficient Approach to Policy Search. In Proceedings of the International Conference on Machine Learning, pages 465–472, New York, NY, USA, June 2011. ACM.

[6] M. P. Deisenroth, C. E. Rasmussen, and D. Fox. Learning to Control a Low-Cost Manipulator using Data-Efficient Reinforcement Learning. In Proceedings of the International Conference on Robotics: Science and Systems, Los Angeles, CA, USA, June 2011.

[7] P. Hennig and C. J. Schuler. Entropy Search for Information-Efficient Global Optimization. Journal of Machine Learning Research, 13:1809–1837, 2012.

[8] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization, 13(4):455–492, December 1998.

[9] J. Quiñonero-Candela, A. Girard, J. Larsen, and C. E. Rasmussen. Propagation of Uncertainty in Bayesian Kernel Models—Application to Multiple-Step Ahead Forecasting. In IEEE International Conference on Acoustics, Speech and Signal Processing, volume 2, pages 701–704, April 2003.

[10] C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA, 2006.

[11] R. S. Sutton and A. G. Barto. Reinforcement Learning: An Introduction. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA, 1998.