Robot Learning

Jan Peters
TU Darmstadt

TECHNISCHE UNIVERSITÄT DARMSTADT

Source: British TV (1969)

So where have robots been successful?

Whenever we adapt tasks to robots!

Analytical Robotics World View

Analytical robotics needs three components:

1. Accurate forward models (=Physics simulators)
   - *Great prior*: Physical principles yield simulators
   - "All models are wrong but some are useful!" (Cox)
   - Un-modelable nonlinearities (friction, actuator dynamics, contact, …)

2. Planning algorithms
   Exponential explosions, replanning is hard, optimization bias, …

3. Fast feedback control
   - The error killer!
   - Build "best bodies for control": stiff, power hungry, complex design…

Deep Learning World View

End-to-end deep learning needs:

1. A highly flexible representation with suitable algorithms (=Deep net)
   - We can learn anything ➜ often physically implausible solutions!
   - Small errors ➜ huge *optimization bias*
   - Black-box ➜ often little insight into the solution

2. Loads of data
   - Robots live in real-time ➜ Few episodes, fast state-action stream
   - Real-World ➜ Real damages
   - Physics simulators as data generators? Back to square one…

3. Loads of computation
   Online learning? Energy storage/communication problems?

How should Robot Learning differ?

1. Learn on the real system
2. Adapt online without replanning!
3. Avoid real-time bottle neck
4. Cope with little episodic data problem
5. At least partially explainable?
6. Be physically plausible!
7. Cope with simulation optimization bias
8. Build “best bodies” not “best bodies for feedback control”

I obviously don’t have all the solutions …
but I had to learn some good lessons!

Resulting Research Questions
Outline

1. Can we learn on a real system from little data?
2. How can we learn comprehensible, modular policies?
3. How can we learn physically plausible deep models?
4. How can we build the best bodies *and* learn on real systems?
5. Conclusion & Outlook

Imitation Learning

Model-Based Behavioral Cloning (Englert et al.)

Objective: Policy Similarity

$$\max_{\pi,\mu^\pi} J(\pi) = \sum_{\boldsymbol{s},\boldsymbol{a}} \mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}|\boldsymbol{s}) \log \frac{\mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}|\boldsymbol{s})}{q(\boldsymbol{s},\boldsymbol{a})}$$

Model-Free Behavioral Cloning (Michie & Chambers, Sammut et al.)

Constraints: Assumptions on the Policy

$$\mu^\pi(\boldsymbol{s}') = \sum_{\boldsymbol{s},\boldsymbol{a}} \mathcal{P}^{\boldsymbol{a}}_{\boldsymbol{s}\boldsymbol{s}'} \mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}|\boldsymbol{s})$$

$$1 = \sum_{\boldsymbol{s},\boldsymbol{a}} \mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}|\boldsymbol{s})$$

Dual Problem

Putermann (1998) implies: *IRL is harder than MBC!*

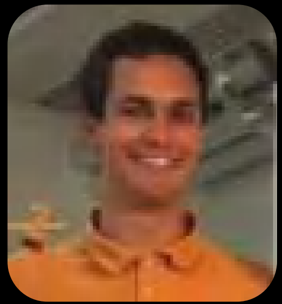
Dual Function for Minimal Physics

Inverse Reinforcement Learning

(Ziebart et al.; Boularias et al.)

Solve for the optimal parametric policy class: Motor primitives

(Schaal et al; Kober et al; Paraschos et al; Gomez-Gonzalez)


Jens Kober

Learning Perception-adapted Probabilistic Motor Primitives

Learning from human demonstrations

Sebastian Gomez-Gonzalez

Gomez-Gonzalez, S.; Neumann, G.; Schölkopf, B.; Peters, J. (2019). Adaptation and Robust Learning of Probabilistic Movement Primitives, IEEE Transactions on Robotics.

Reinforcement Learning

Dual: RL by Linear Programming

Objective: Expected Returns

$$\max_{\pi,\mu^\pi} J(\pi) = \sum_{s,a} \mu^\pi(s)\pi(a|s)\mathcal{R}_{sa}$$

Constraints: Assumptions on the Policy

$$\mu^\pi(s') = \sum_{s,a} \mathcal{P}^{a}_{ss'}\mu^\pi(s)\pi(a|s)$$

$$1 = \sum_{s,a} \mu^\pi(s)\pi(a|s)$$

Dual Problem

Bellman's Principle of Optimality

Putermann (1998): *Primal is harder than Dual!*

Primal: RL by Linear Programming

"Bellman Equation":

$$V^*(s) = \max_a E_{s'}\left\{r(s,a,s') + \gamma V(s')\right\}$$

No natural notion of data!

Relative Entropy Policy Search

Dual: RL by Linear Programming

Objective: Expected Returns

$$\max_{\pi,\mu^\pi} J(\pi) = \sum_{s,a} \mu^\pi(s)\pi(a|s)\mathcal{R}_{sa}$$

Constraints: Assumptions on the Policy

$$\mu^\pi(s') = \sum_{s,a} \mathcal{P}^a_{ss'}\mu^\pi(s)\pi(a|s)$$

$$1 = \sum_{s,a} \mu^\pi(s)\pi(a|s)$$

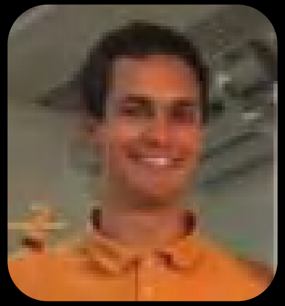
Peters (2007). Relative Entropy Policy Search, Tech. Rep.
Peters, Muelling, Altun (2010). Relative Entropy Policy Search, AAAI

Further Constraint: Policy Similarity

$$\epsilon \geq \sum_{s,a} \mu^\pi(s)\pi(a|s)\log\frac{\mu^\pi(s)\pi(a|s)}{q(s,a)}$$

Objective from Behavioral Cloning

Different *q* yield analytical solution, mellow/softmax, entropy regularization…
Natural policy gradients/TRPO are its approximations!

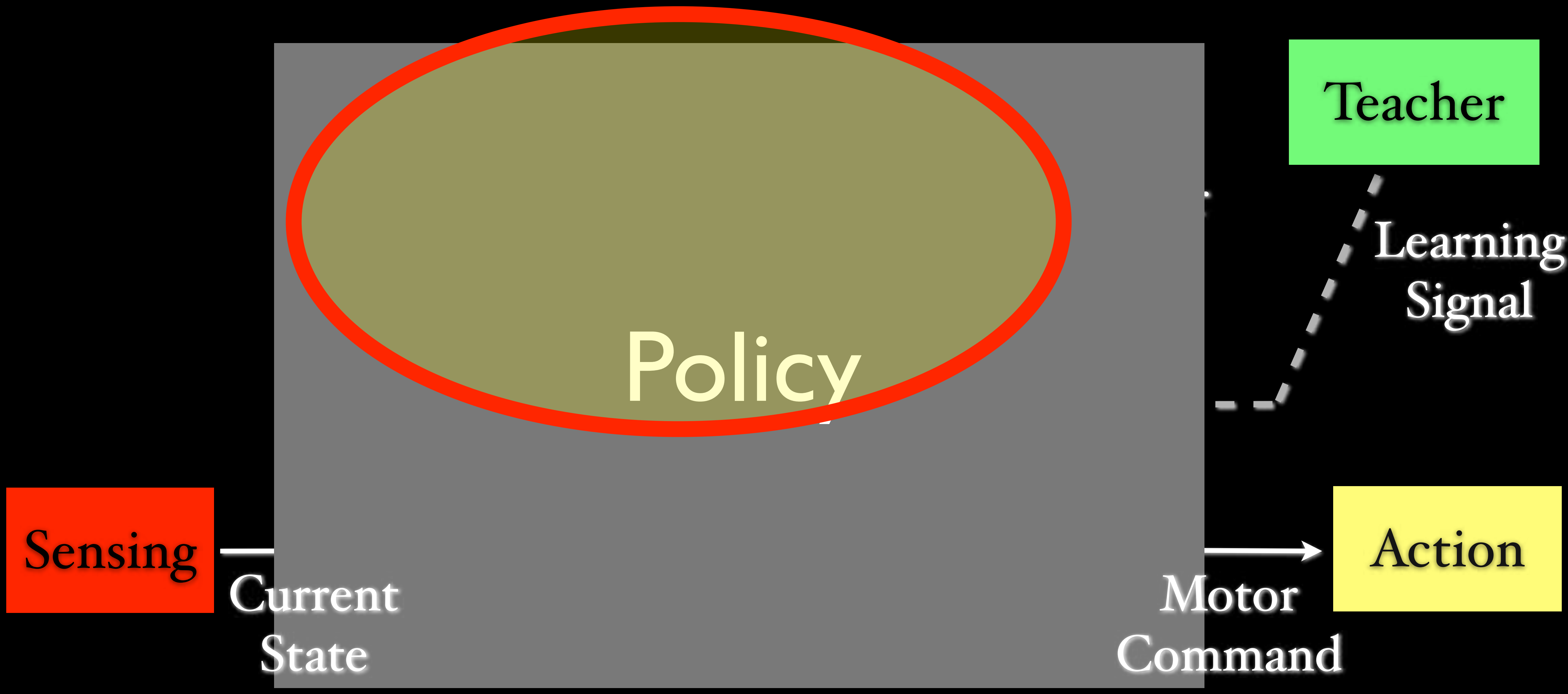


Jens Kober

Outline

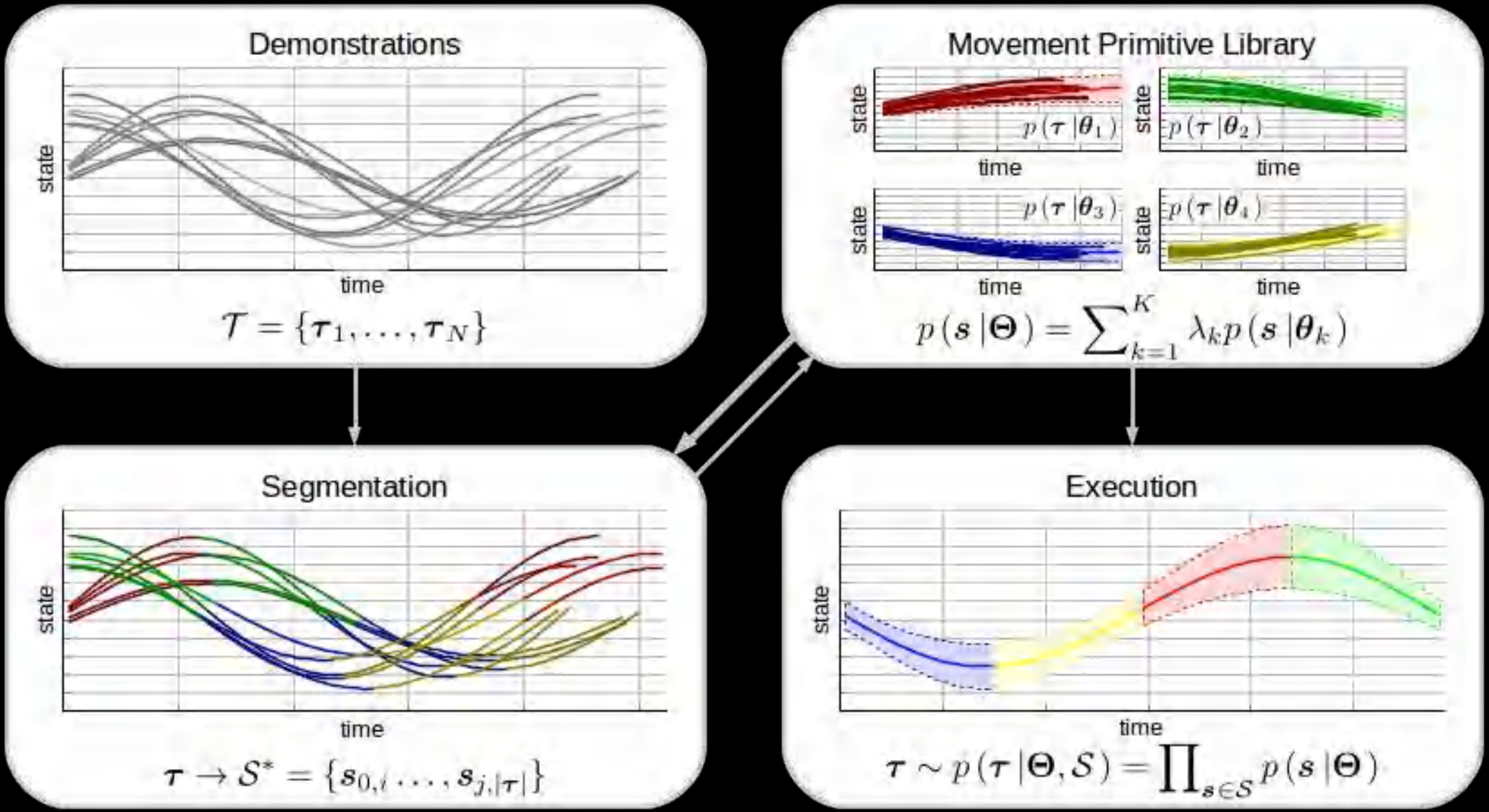
1. Can we learn on a real system from little data?
2. How can we learn comprehensible, modular policies?
3. How can we learn physically plausible deep models?
4. How can we build the best bodies *and* learn on real systems?
5. Conclusion & Outlook

Policy Composition by Selection, Superposition & Sequencing



Initialize both supervisor and primitives by imitation

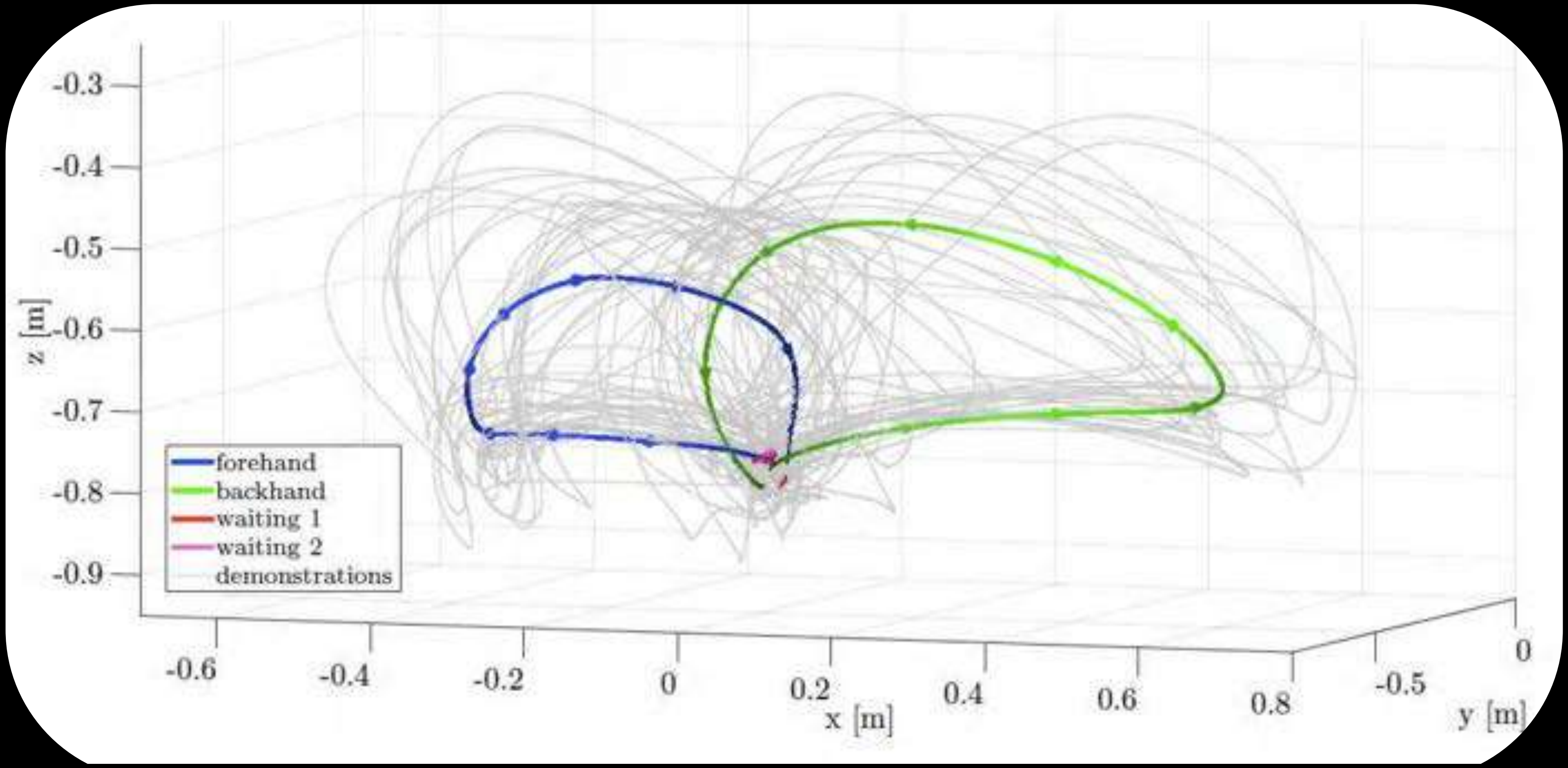
Rudolf Lioutikov



Lioutikov, R.; Neumann, G.; Maeda, G.; Peters, J. (2017). Learning Movement Primitive Libraries through Probabilistic Segmentation, International Journal of Robotics Research (IJRR).

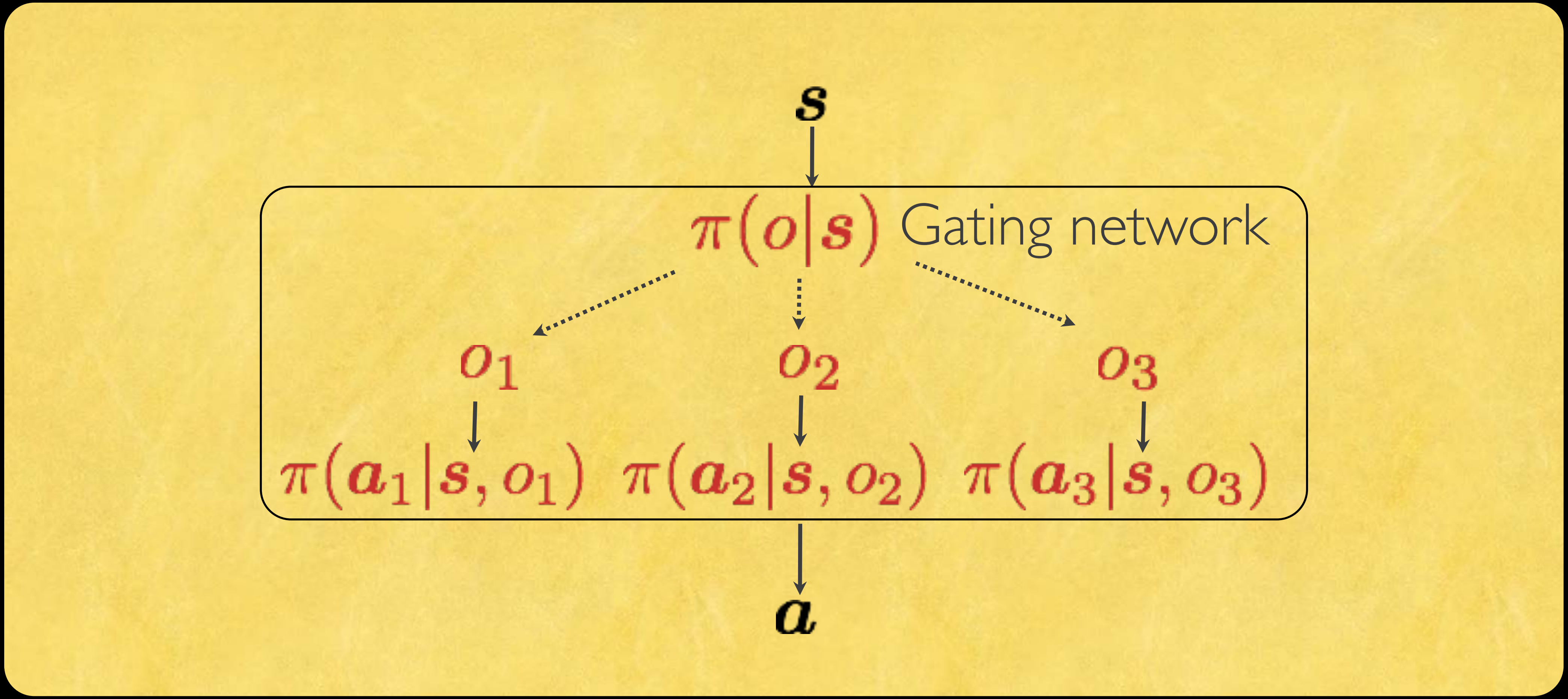
Learn both supervisor and sub-policies by imitation

Rudolf Lioutikov



Lioutikov, R.; Neumann, G.; Maeda, G.; Peters, J. (2017). Learning Movement Primitive Libraries through Probabilistic Segmentation, International Journal of Robotics Research

Modular Control Policies



Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Daniel, Neumann & Peters (2016). Hierarchical Relative Entropy Policy Search, JMLR

"Naïve" Extension of REPS

Relative Entropy Policy Search (REPS)

$\max_{\pi,\mu^\pi} J(\pi) = \sum_{\boldsymbol{s},\boldsymbol{a}} \mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}\vert\boldsymbol{s})\mathcal{R}_{\boldsymbol{sa}}$	Maximize reward
$1 = \sum_{\boldsymbol{s},\boldsymbol{a}} \mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}\vert\boldsymbol{s})$	Probability distribution
$\mu^\pi(\boldsymbol{s}') = \sum_{\boldsymbol{s},\boldsymbol{a}} \mathcal{P}^{\boldsymbol{a}}_{\boldsymbol{ss}'}\mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}\vert\boldsymbol{s})$	Follow system dynamics
$\epsilon \geq \sum_{\boldsymbol{s},\boldsymbol{a}} \mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}\vert\boldsymbol{s}) \log \frac{\mu^\pi(\boldsymbol{s})\pi(\boldsymbol{a}\vert\boldsymbol{s})}{q(\boldsymbol{s},\boldsymbol{a})}$	Close to training data (no wild exploration)

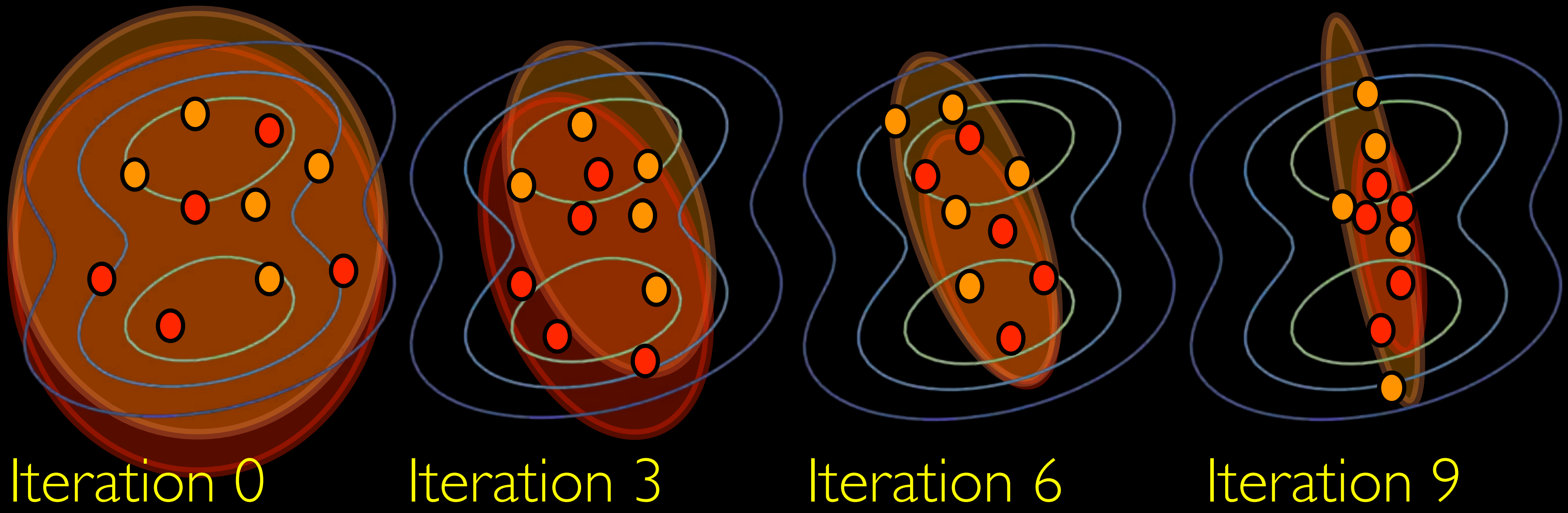
Mülling, K.; Kober, J.; Kroemer, O.; Peters, J. (2013). Learning to Select and Generalize Striking Movements in Robot Table Tennis, International Journal on Robotics Research.

Daniel, Neumann & Peters (2016). Hierarchical Relative Entropy Policy Search, JMLR


WTSP
K. Muelling, Z. Wang, J. Peters of TU Darmstadt and MPI Intelligen

Problems with Naïvety

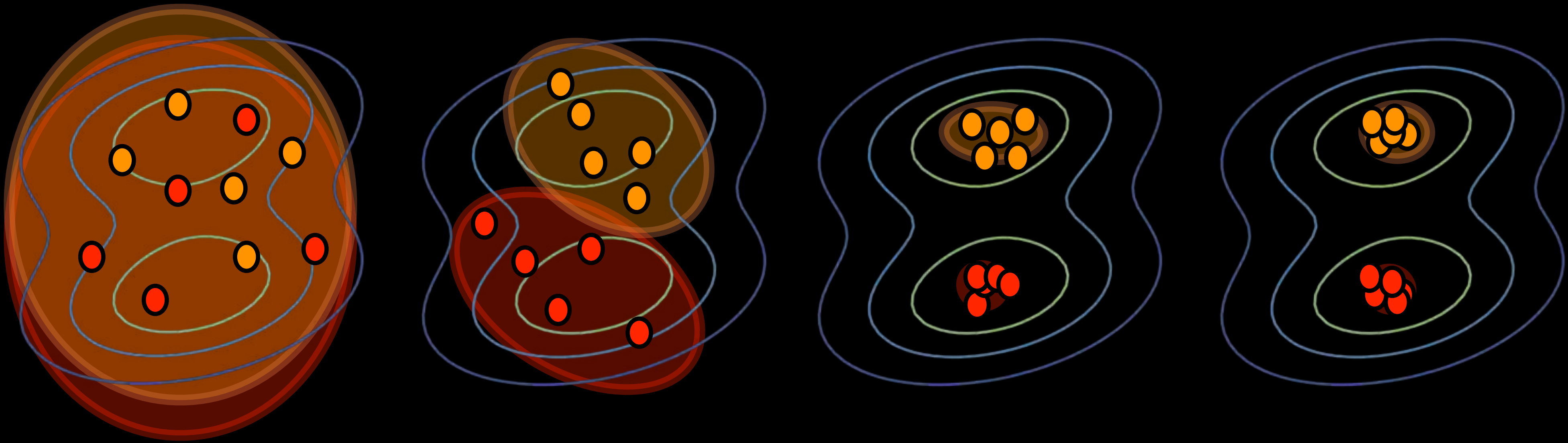
Christian Daniel



Daniel, Neumann & Peters (2016). Hierarchical Relative Entropy Policy Search, JMLR

Localized behavior can be learned efficiently!

Christian Daniel



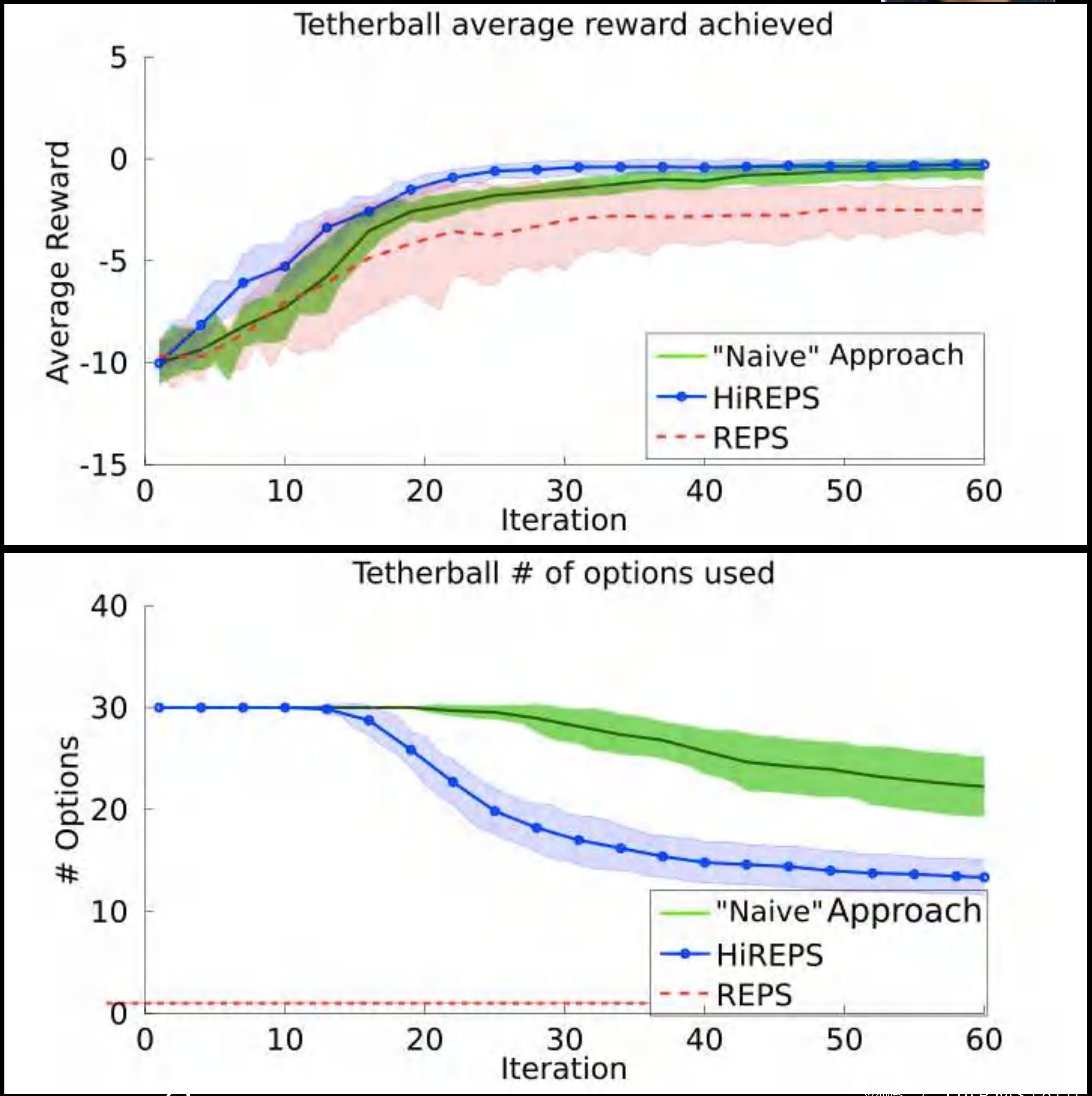
$$\kappa \geq \mathbb{E}_{s,a}\left[\sum_{o} -p(o|s,a)\log p(o|s,a)\right]$$ Force the primitives to limited responsibility

Daniel, Neumann & Peters (2016). Hierarchical Relative Entropy Policy Search, JMLR

Localized behavior can be learned efficiently!

Good performance

Fast reduction in the number of primitives

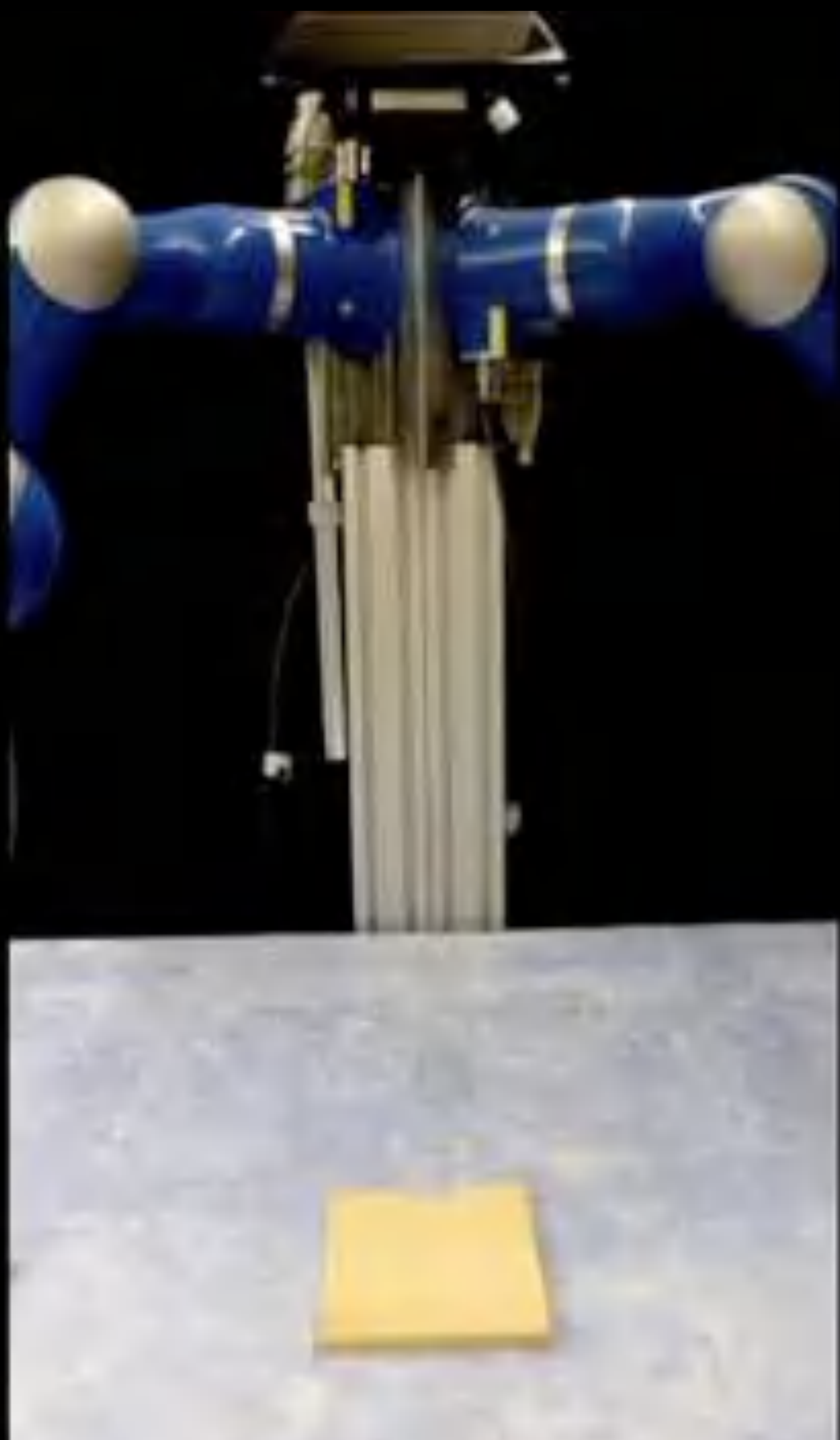


Daniel, Neumann & Peters (2016). Hierarchical Relative Entropy Policy Search, JMLR

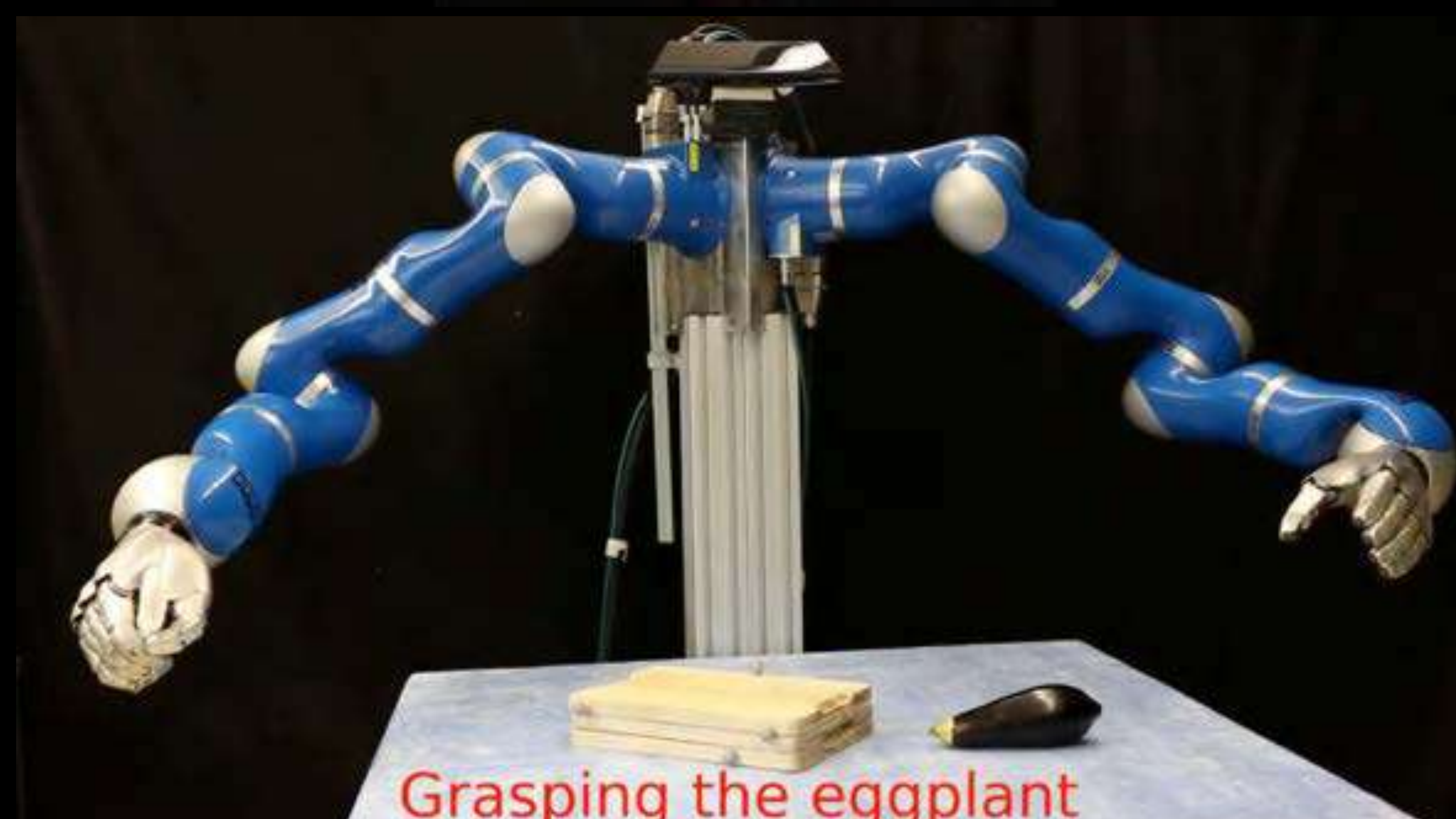
Oliver Kroemer

Sequencing in Manipulation

Rudolf Lioutikov

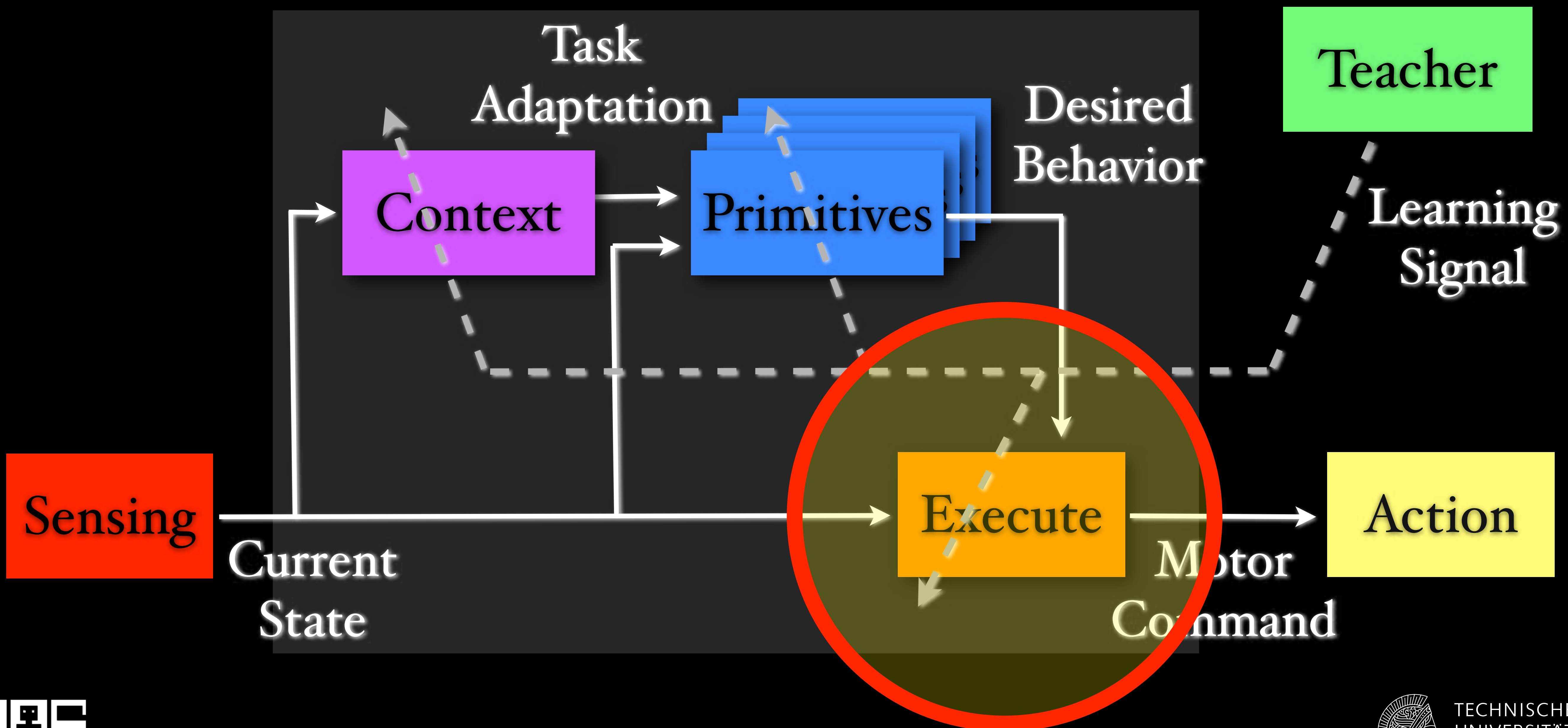


Phase: 1



CHE
ITÄT
ADT

Policy Composition by
Selection, Superposition & Sequencing



TECHNISCHE UNIVERSITÄT DARMSTADT

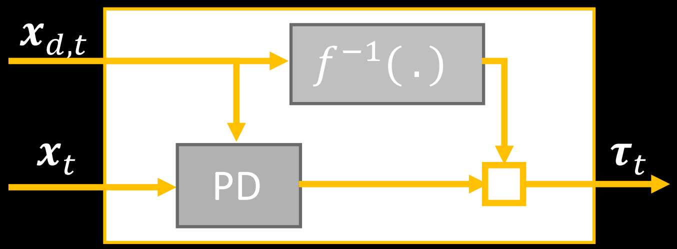
Outline

1. Can we learn on a real system from little data?
2. How can we learn comprehensible, modular policies?
3. How can we learn physically plausible deep models?
4. How can we build the best bodies *and* learn on real systems?
5. Conclusion & Outlook

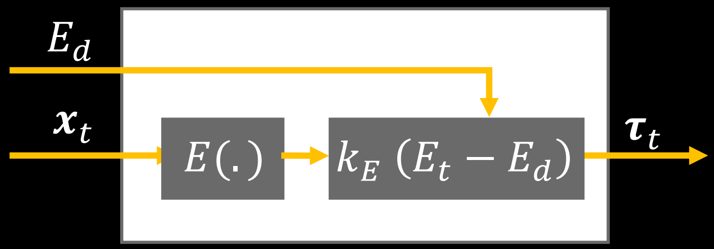
Models are important for Execution

Michael Lutter

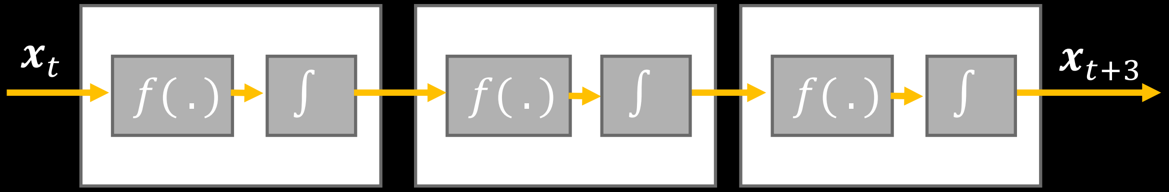
Inverse Model



Energy Model



Forward Model

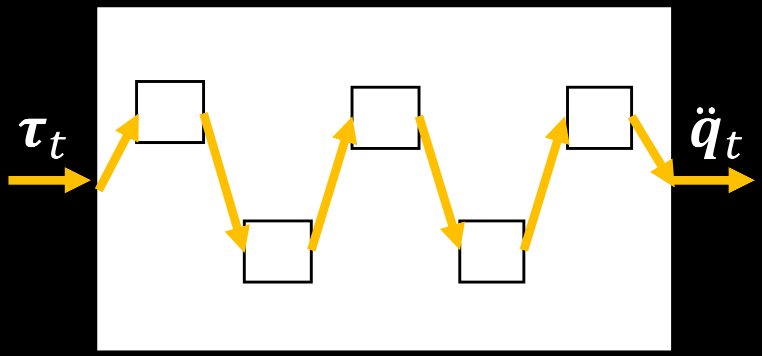


TECHNISCHE UNIVERSITÄT DARMSTADT

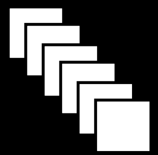
Engineers prefer engineering to model learning due to plausibility

Michael Lutter

Model Engineering[1]



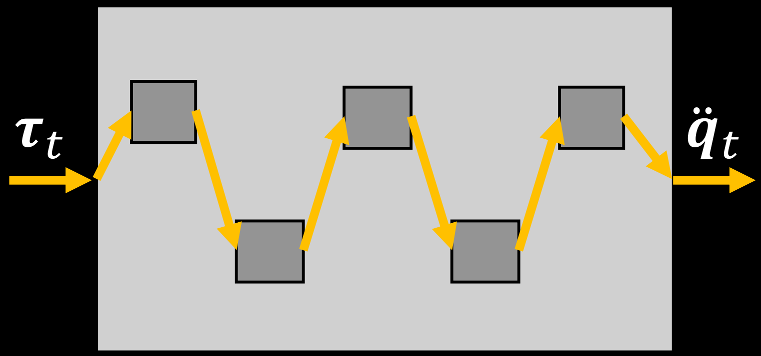
Identify the parameters by taking apart and measuring,



Center of Gravity,
Mass, Inertia, etc.

Model can be used as to compute forward, inverse and energy model

System Identification[1]



Learn parameters by minimizing the MSE with handcrafted features,

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \| \boldsymbol{\tau}_i - \boldsymbol{A}(\boldsymbol{q}_i, \dot{\boldsymbol{q}}_i, \ddot{\boldsymbol{q}}_i)\, \theta \|_2^2$$

Model can be used as to compute forward, inverse and energy model

Black-box Model Learning

Learn parameters by minimizing the naïve MSE,

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \| \boldsymbol{\tau}_i - f^{-1}(\ddot{\boldsymbol{q}}_i; \theta) \|_2^2$$

Model can only be used to either compute forward OR inverse model

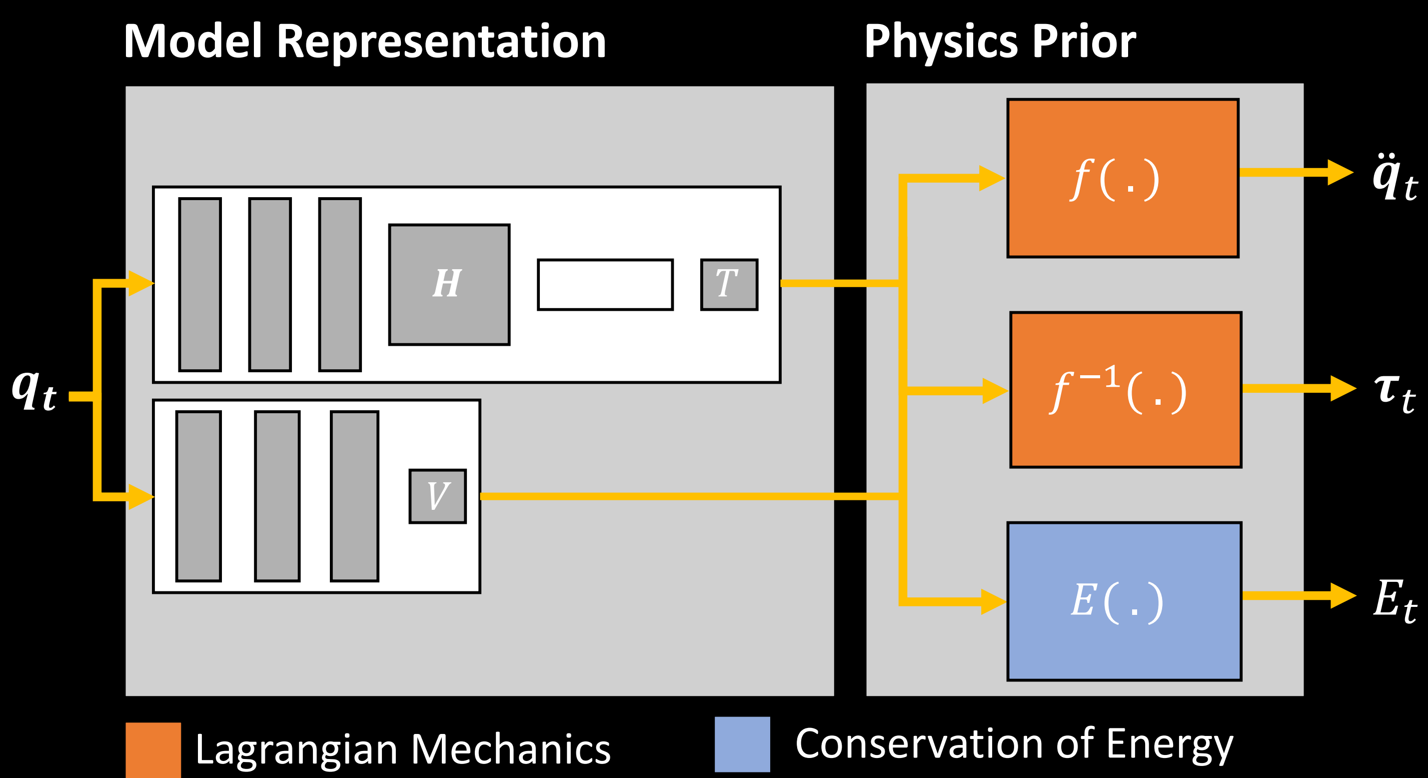
[1]Mass-matrix, Coriolis-, centrifugal- & gravitational force can be computed using the Featherstone algorithm.

Lutter, M. et al. (2019). HJB Optimal Feedback Control with Deep Differential Value Functions and Action Constraints, CoRL
Lutter, M. et al. (2019). Deep Lagrangian Networks for end-to-end learning of energy-based control for under-actuated systems, IROS
Lutter, M. et al. (2019). Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning, ICLR
Lutter, M. et al. (2019). Deep Optimal Control: Using the Euler-Lagrange Equation to learn an Optimal Feedback Control Law, Multi-disciplinary Conference on Reinforcement Learning and Decision Making (RLDM).

TECHNISCHE UNIVERSITÄT DARMSTADT

Deep Lagrangian Networks (DeLaN)

Michael Lutter



$$f(.) = \boldsymbol{H}^{-1}\left(\boldsymbol{\tau} - \dot{\boldsymbol{H}}\dot{\boldsymbol{q}} + \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}^T\boldsymbol{H}\dot{\boldsymbol{q}}\right) - \frac{\partial V}{\partial \boldsymbol{q}}\right)$$ with $\boldsymbol{H}$ being p.d.

$$f^{-1}(.) = \boldsymbol{H}\ddot{\boldsymbol{q}} + \dot{\boldsymbol{H}}\dot{\boldsymbol{q}} - \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}^T\boldsymbol{H}\dot{\boldsymbol{q}}\right) + \frac{\partial V}{\partial \boldsymbol{q}}$$

$$E(.) = T + V$$

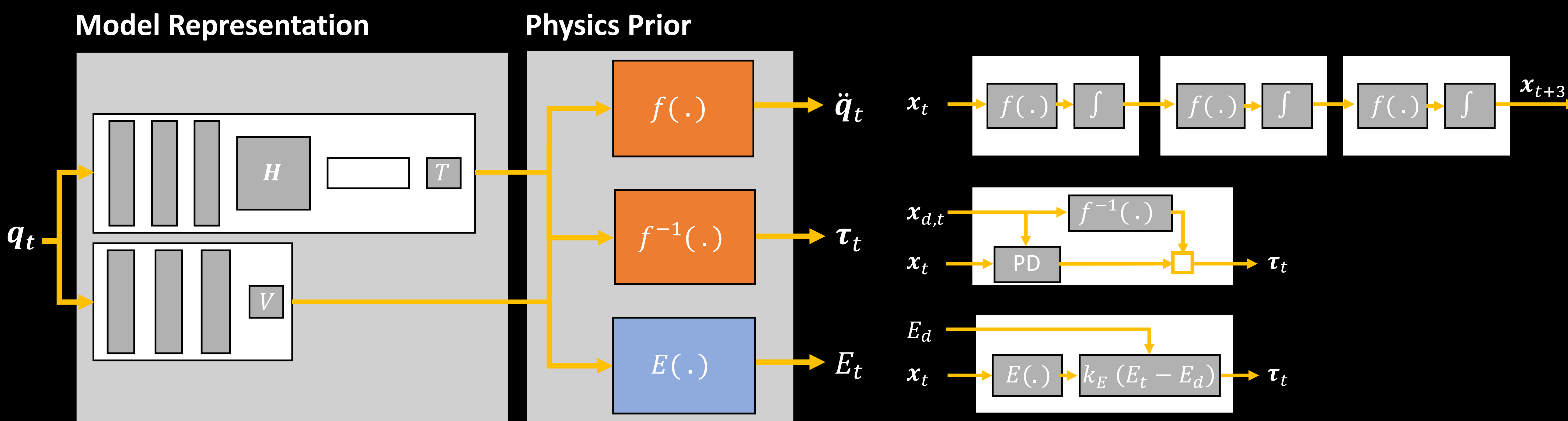
Deep Lagrangian Networks (DeLaN)
Guarantee physically-plausible models by constraining the model with priors.

Physical plausibility means that every possible parameter configuration is a mechanical system.

The structured models enables the usage as forward, inverse & energy model

DeLaN enables the simultaneous learning of forward, inverse & energy

Michael Lutter



Lagrangian Mechanics

Conservation of Energy

$$f(.) = \boldsymbol{H}^{-1}\left(\boldsymbol{\tau} - \dot{\boldsymbol{H}}\dot{\boldsymbol{q}} + \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}^T\boldsymbol{H}\dot{\boldsymbol{q}}\right) - \frac{\partial V}{\partial \boldsymbol{q}}\right)$$ with $\boldsymbol{H}$ being p.d.

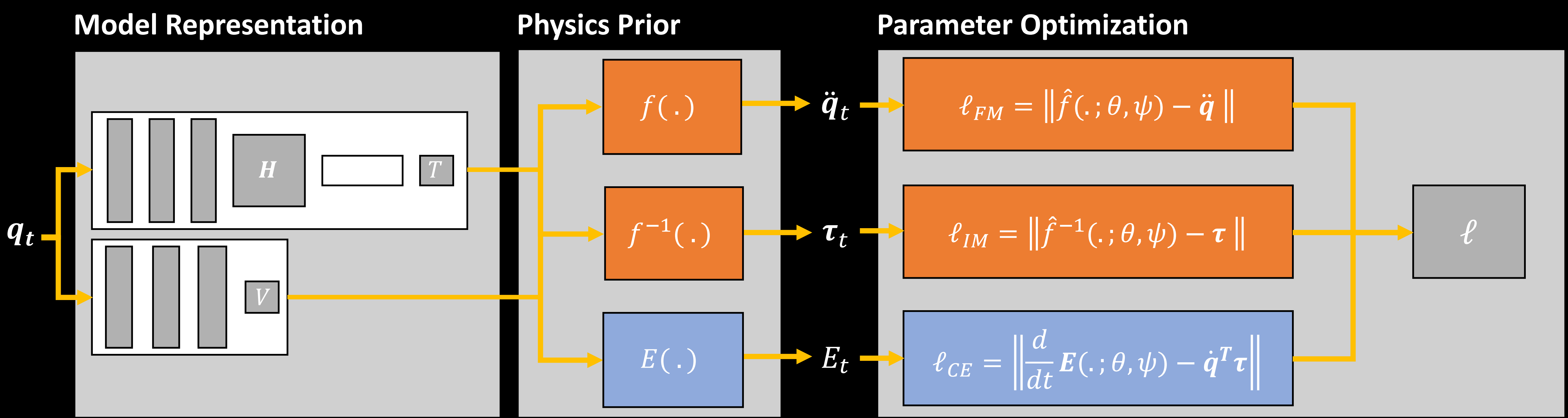
$$f^{-1}(.) = \boldsymbol{H}\ddot{\boldsymbol{q}} + \dot{\boldsymbol{H}}\dot{\boldsymbol{q}} - \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}^T\boldsymbol{H}\dot{\boldsymbol{q}}\right) + \frac{\partial V}{\partial \boldsymbol{q}}$$

$$E(.) = T + V$$

TECHNISCHE UNIVERSITÄT DARMSTADT

Energies are learned by minimising the residual of the differential equations

Michael Lutter



Lagrangian Mechanics

Conservation of Energy

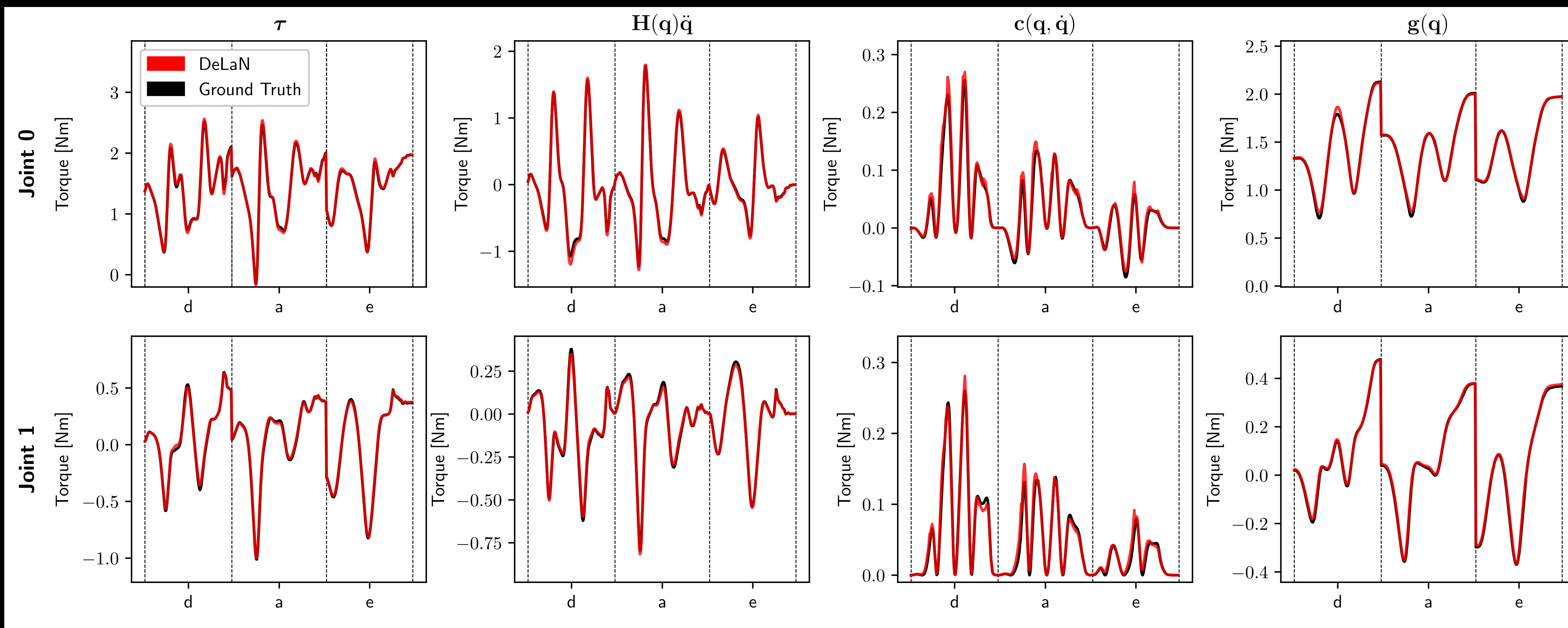
$$f(.) = \boldsymbol{H}^{-1}\left(\boldsymbol{\tau} - \dot{\boldsymbol{H}}\dot{\boldsymbol{q}} + \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}^T\boldsymbol{H}\dot{\boldsymbol{q}}\right) - \frac{\partial V}{\partial \boldsymbol{q}}\right)$$ with $\boldsymbol{H}$ being p.d.

$$f^{-1}(.) = \boldsymbol{H}\ddot{\boldsymbol{q}} + \dot{\boldsymbol{H}}\dot{\boldsymbol{q}} - \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}}^T\boldsymbol{H}\dot{\boldsymbol{q}}\right) + \frac{\partial V}{\partial \boldsymbol{q}}$$

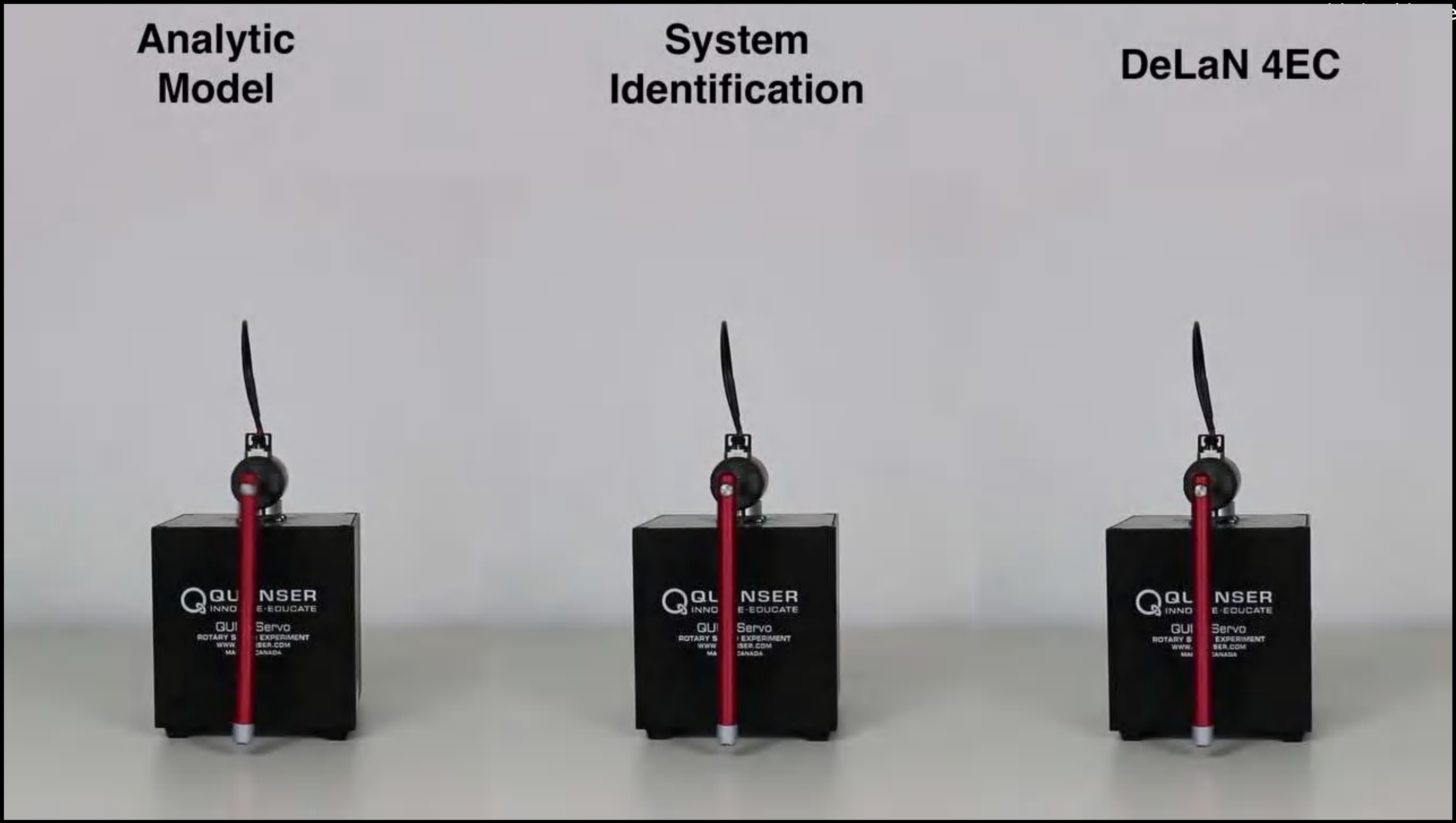
$$E(.) = T + V$$

DeLaN can learn the force decomposition unsupervised

Michael Lutter

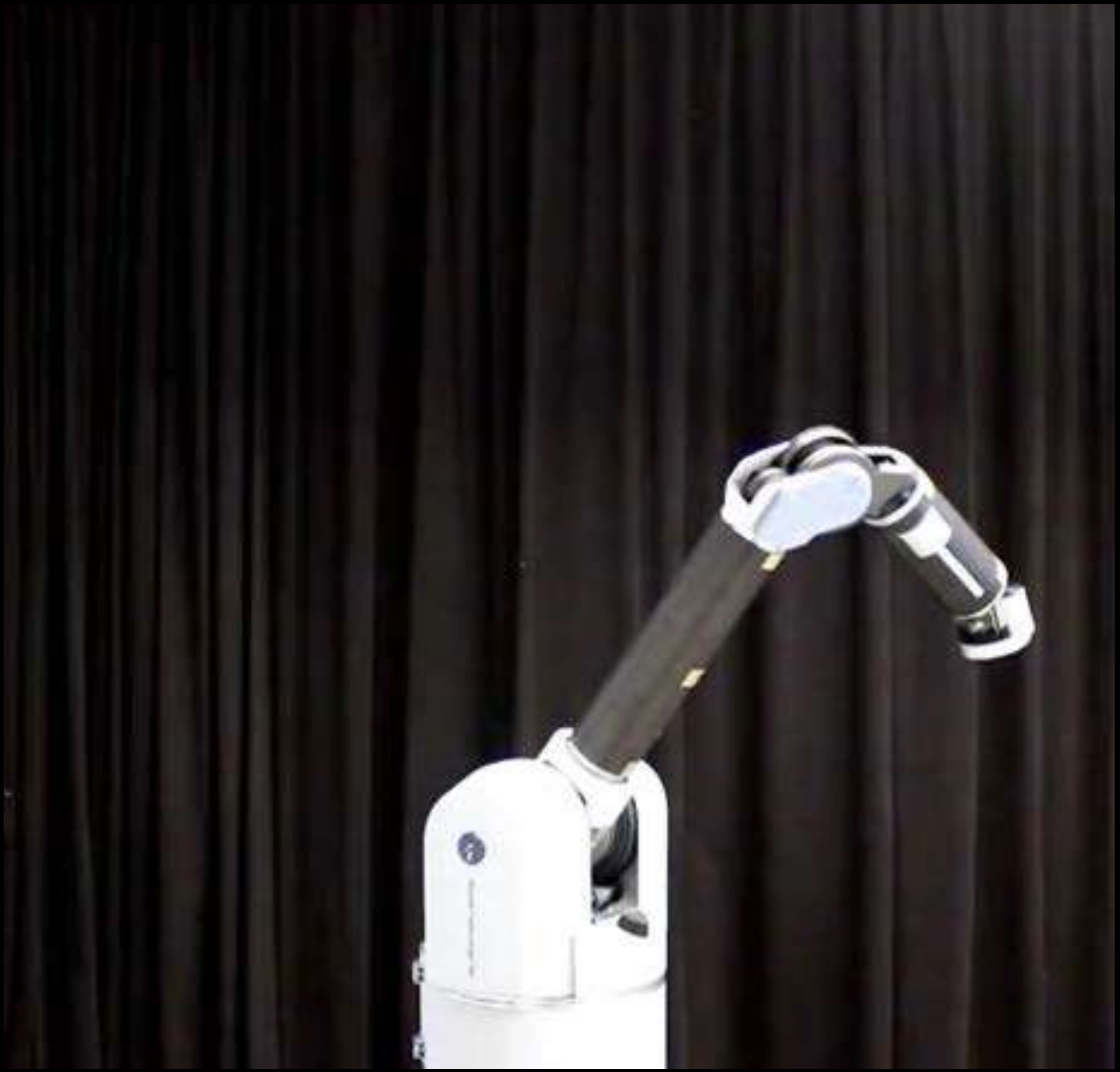


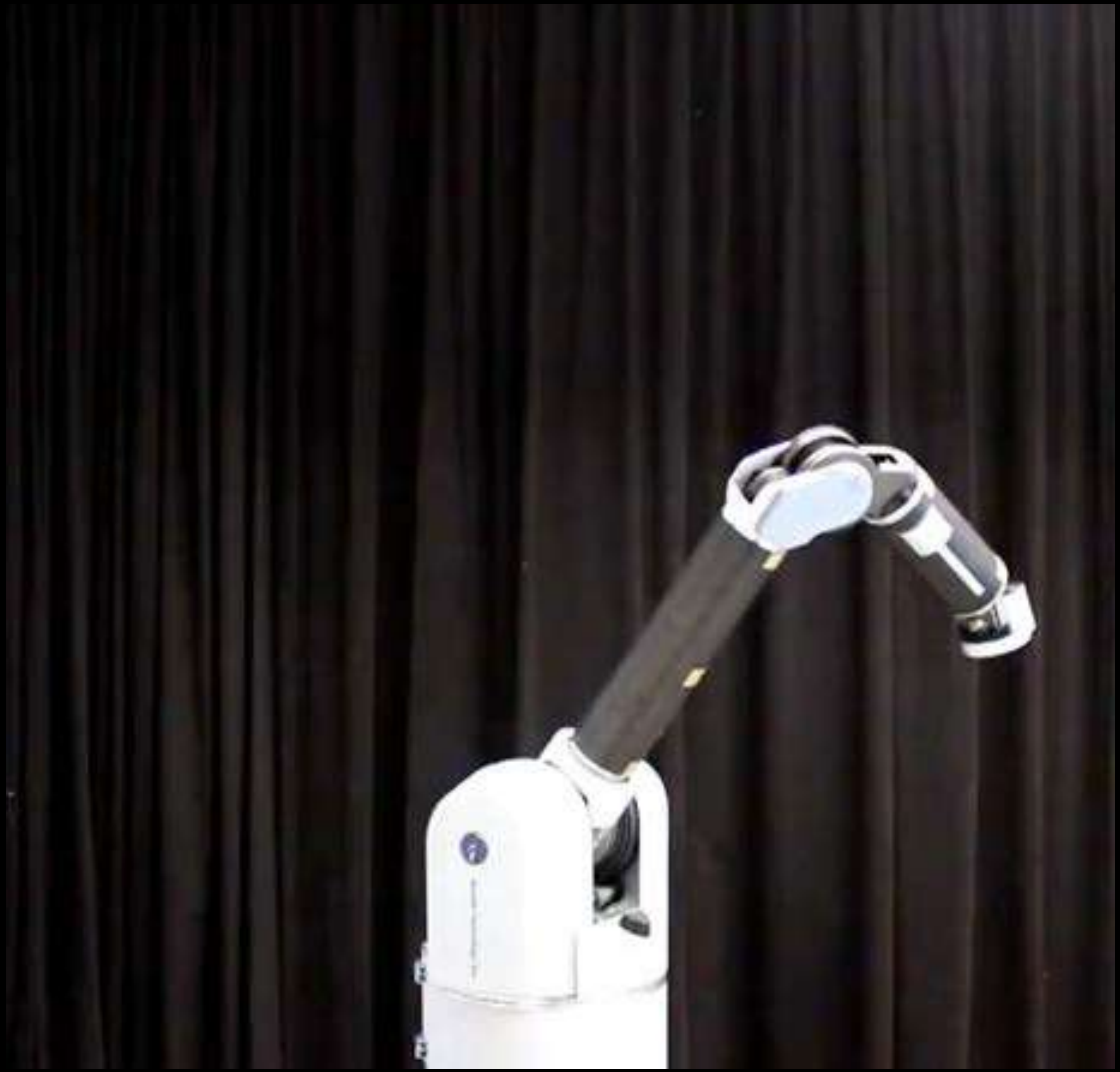
Energy Control of the Furuta Pendulum



Non-linear feed-forward control of the Barrett WAM

Michael Lutter





TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

1. Can we learn on a real system from little data?
2. How can we learn comprehensible, modular policies?
3. How can we learn physically plausible deep models?
4. How can we build the best bodies *and* learn on real systems?
5. Conclusion & Outlook

Robot Bodies for Learning?

Dieter Büchler

Classical robotics builds the best body that can be controlled with classical approaches!

Human bodies would defy such an approach but generate high accelerations in order to

- reach high velocities
- perform skillful motions

Humans learn (typically) without breaking!

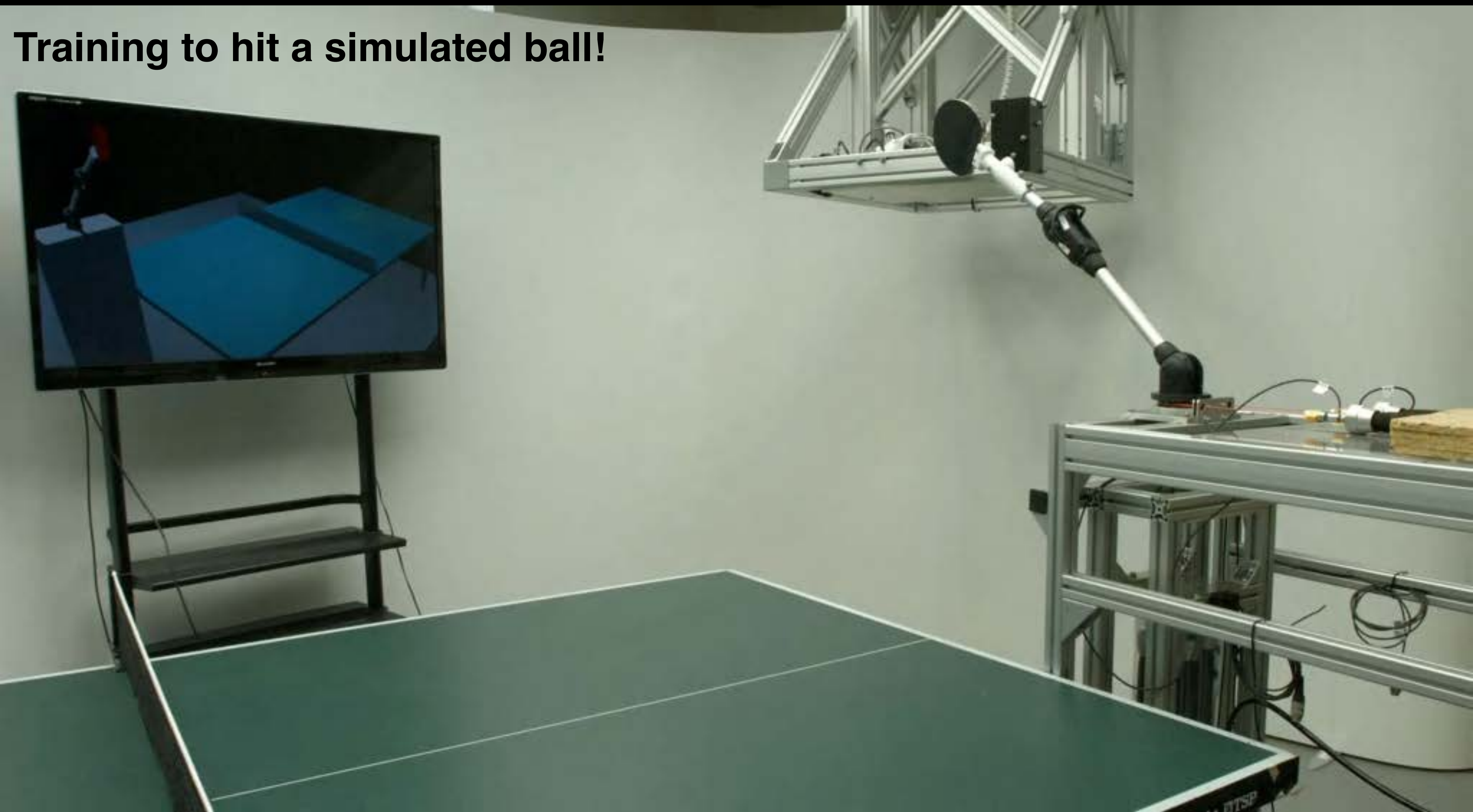
➡ Human performance robot learning needs better bodies!

Learning Robot Table Tennis *from Scratch*

Dieter Büchler

Simon Guist

Training to hit a simulated ball!



Learning Robot Table Tennis *from Scratch*

Dieter Büchler

Simon Guist

Training to hit a real ball!

Outline

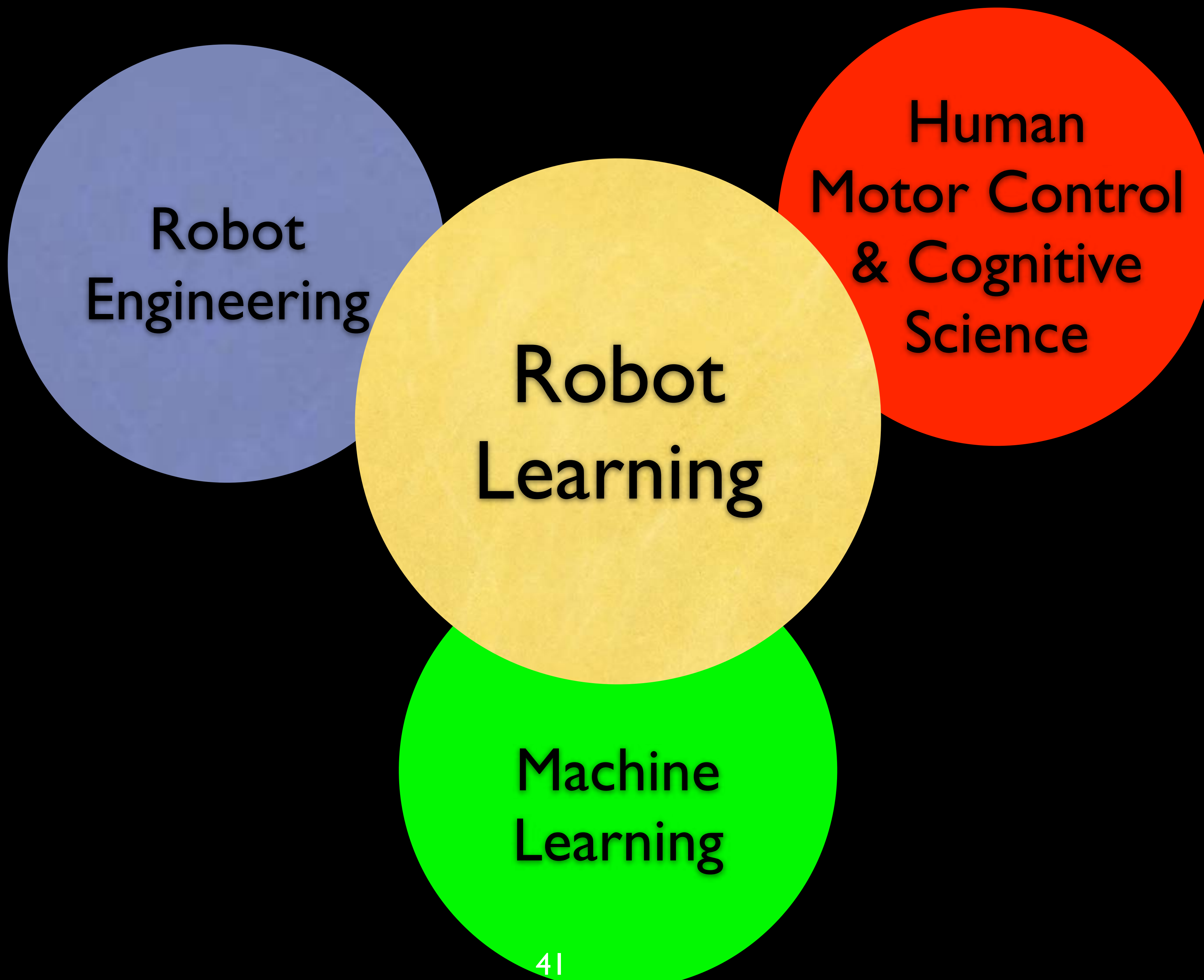
1. Can we learn on a real system from little data?
2. How can we learn comprehensible, modular policies?
3. How can we learn physically plausible deep models?
4. How can we build the best bodies *and* learn on real systems?
5. Conclusion & Outlook

TECHNISCHE UNIVERSITÄT DARMSTADT

Lessons for Robot Learning

1. Learn on the real system ➜
   (i) Start with imitation, then RL
   (ii) Find safe model-learning methods for model-based RL
   (iii) Build bodies for learning from scratch
2. Adapt online without replanning! ➜ Use perception modulated movement primitives (DMPs, ProMPs, …)
3. Avoid real-time bottle neck ➜ modularity & parametrized policies
4. Cope with little episodic data problem ➜ modularity, smart data re-use
5. At least partially explainable? ➜ read (Lioutikov et al., IJRR 2019)
6. Be physically plausible! ➜ Use DeLaN
7. Cope with simulation optimization bias ➜ Use SPOTA / Entropic Gradients
8. Build "best bodies" not "best bodies for feedback control" ➜ small moving masses, antagonistic variable stiffness actuation, robot learning

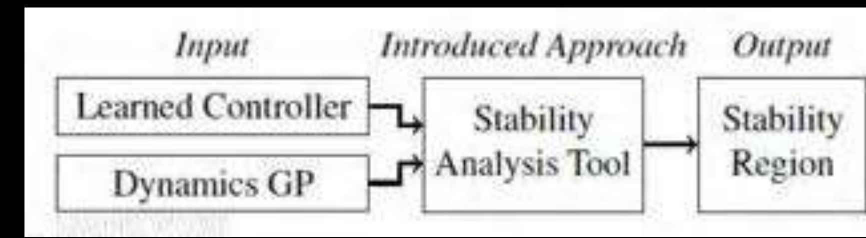
Outlook


Robot Engineering
Human Motor Control & Cognitive Science
Robot Learning
Machine Learning

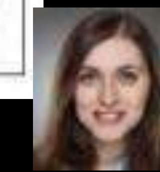
Learning State Representations for Robotics

(Simone Parisi @ MLJ 2019)

Automated Stability Proofs



(Julia Vinogradska @ JMLR 2017)



Robot Engineering

Self-Paced Robot Reinforcement Learning

SELF-PACED CONTEXTUAL REINFORCEMENT LEARNING (SPRL)

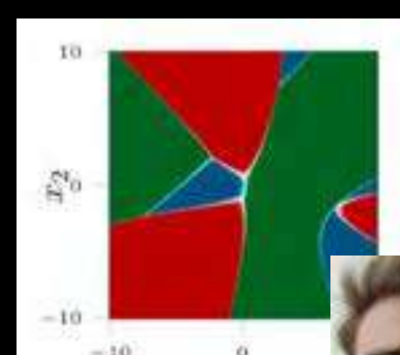
Pascal Klink, Hany Abdulsamad, Boris Belousov, Jan Peters

Intelligent Autonomous Systems, TU Darmstadt

SPARSE BALL IN A CUP TASK

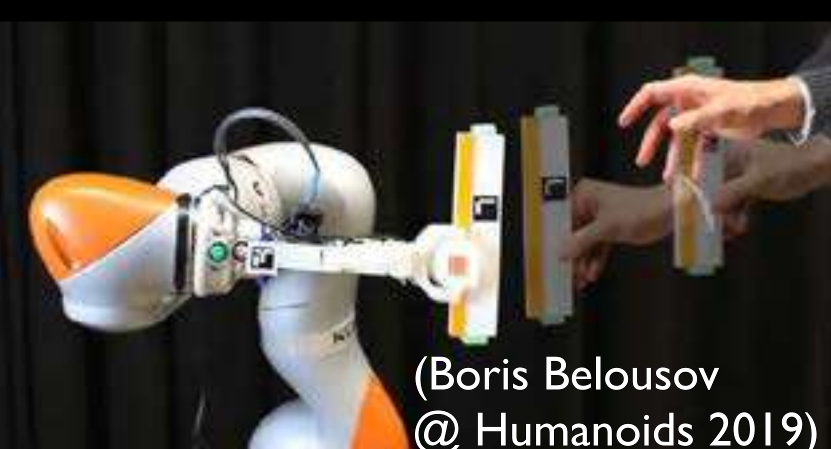
(Pascal Klink @ CoRL 2019)

Inferring Hybrid Control From Data



(Hany Abdulsamad @ Under Review)

Tactile Skill Libraries

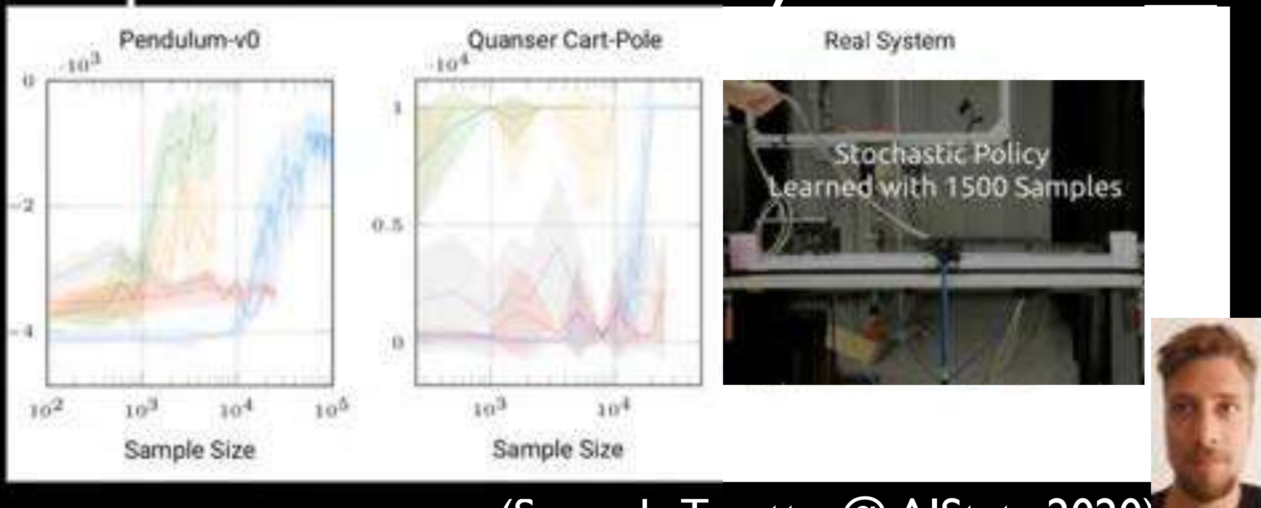


(Boris Belousov @ Humanoids 2019)

(Learning) Control for Table Tennis

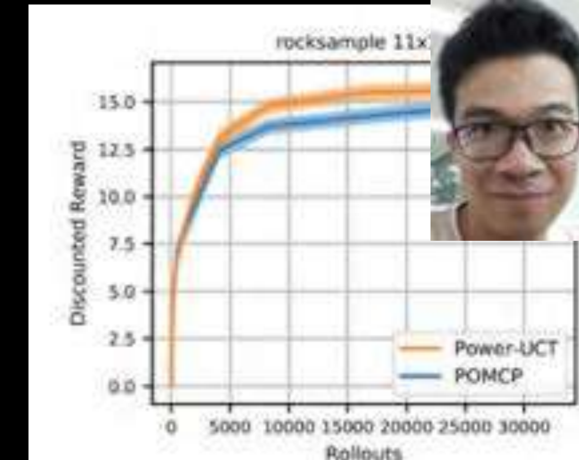
(Okan Koç @ R-AL/ ICRA 2019)

Sample Efficient Off-Policy Gradients



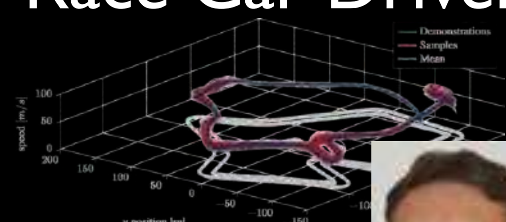
(Samuele Tosatto @ AIStats 2020)

Generalized Mean Estimation with MCTS



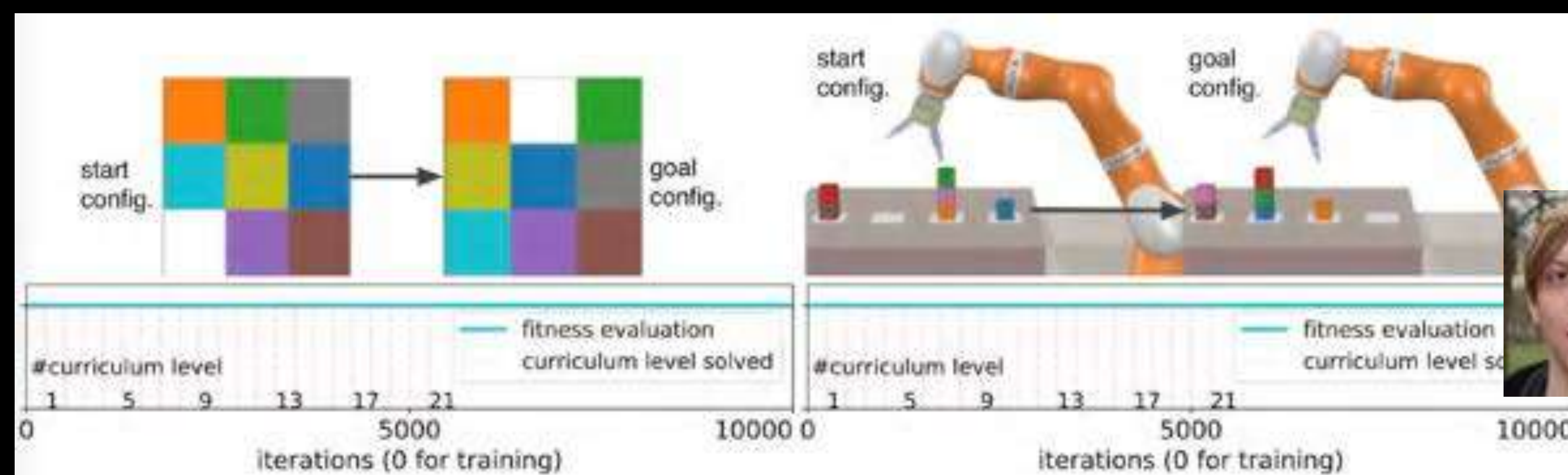
(Tuan Dam @ IJCAI 2020)

Imitation of Race Car Drivers



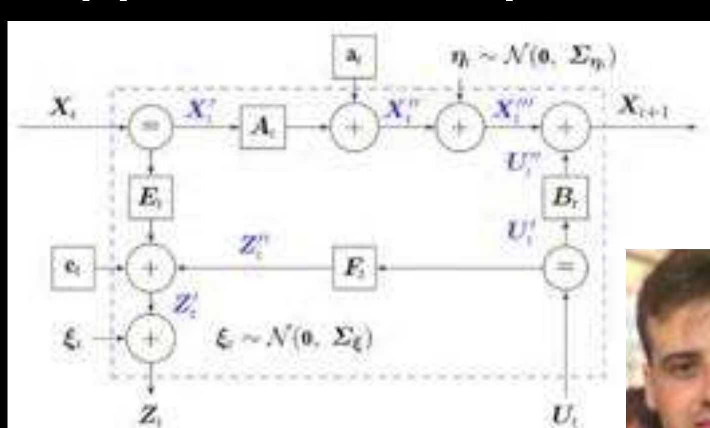
(Stefan Löckel @ ICRA 2020)

Learning Abstract Strategies independent of the Task Domain



(Daniel Tanneberg @ Under Review)

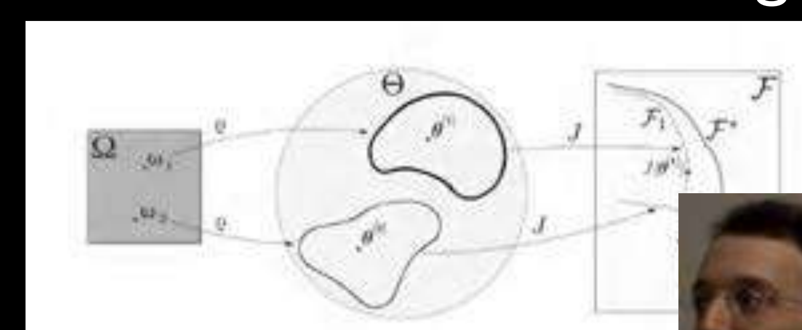
Stochastic Optimal Control by Approximate Input Inference



(Joe Watson @ CoRL 2019)

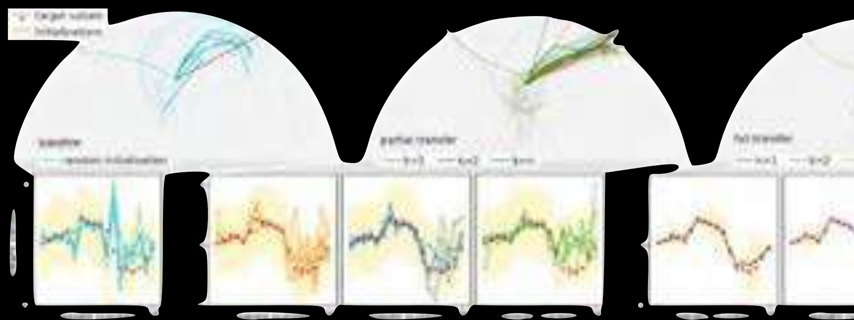
Machine Learning

Multi-Objective Reinforcement Learning



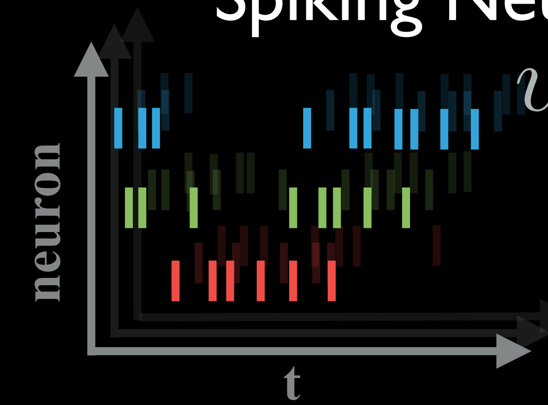
(Simone Parisi @ NECO 2017)

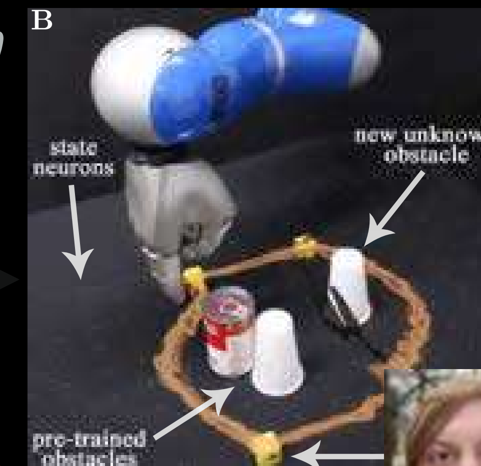
Human-like Experience Reuse

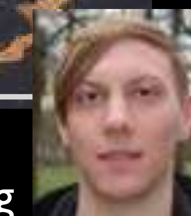


(Svenja Stark @ IROS 2019)

Spiking Neural Models

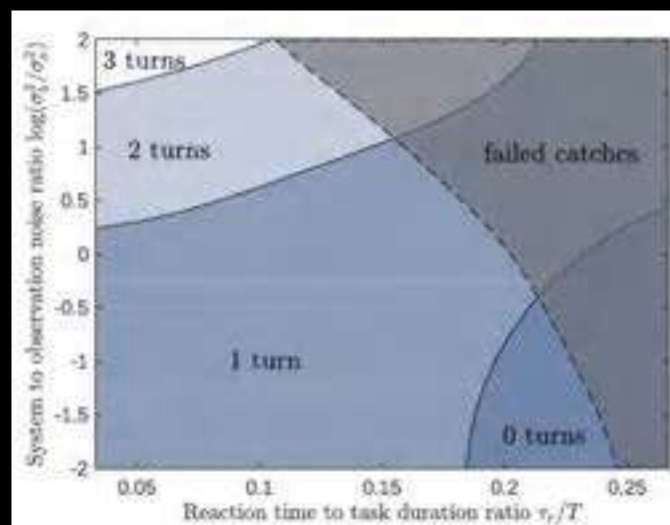






(Daniel Tanneberg @ Neural Networks, 2019)

Human Ball Catching



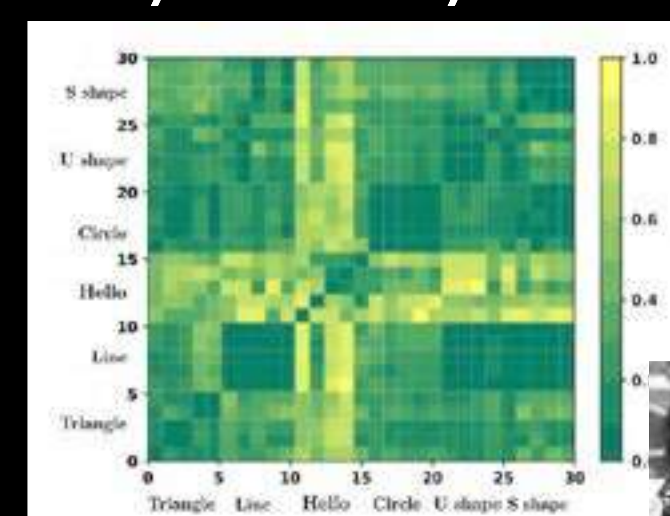
(Boris Belousov @ NeurIPS 2016)

Human Motor Control & Cognitive Science

Human Intent Prediction

(Dorothea Koert @ R-AL/IROS 2019)

Trajectory Similarity Measures



(Julen Urain @ IROS 2019)

Veiga, F. F.; Edin B.B; Peters, J. (submitted). Grip Stabilization through Independent Tactile Feedback Control, Submitted to Advanced Robotics.

Filipe
Veiga


Robot Juggling
Thanks for your Attention!

Thanks for your Attention!

Robot Beer Pong

Thanks for your Attention!

Robot Coffee Making

Thanks for your Attention!

Robot Dart Games

Changing a Light-Bulb

Thanks for your Attention!

Cooperative Throwing and Catching

Thanks for your Attention!

Thanks for your Attention!

Demonstration of Pouring

Robot
Pouring


Robot Juggling
Thanks for your Attention!

Thanks for your Attention!

Robot Beer Pong

Thanks for your Attention!

Robot Coffee Making

Thanks for your Attention!

Robot Dart Games


Changing a Light-Bulb
Thanks for your Attention!

Cooperative Throwing and Catching

Thanks for your Attention!

Thanks for your Attention!

Demonstration of Pouring

Robot
Pouring