Real-time Control with Temporally Extended Predictions

(A Sensorimotor Approach to Planning?)

Patrick M. Pilarski

Reinforcement Learning & Artificial Intelligence Laboratory
Alberta Innovates Centre for Machine Learning

Joint work with Travis B. Dick and Richard S. Sutton

UNIVERSITY OF ALBERTA
DEPARTMENT OF COMPUTING SCIENCE

Adaptive Prosthetics Project

Algorithm 1 Learning General Value Functions with TD(λ)

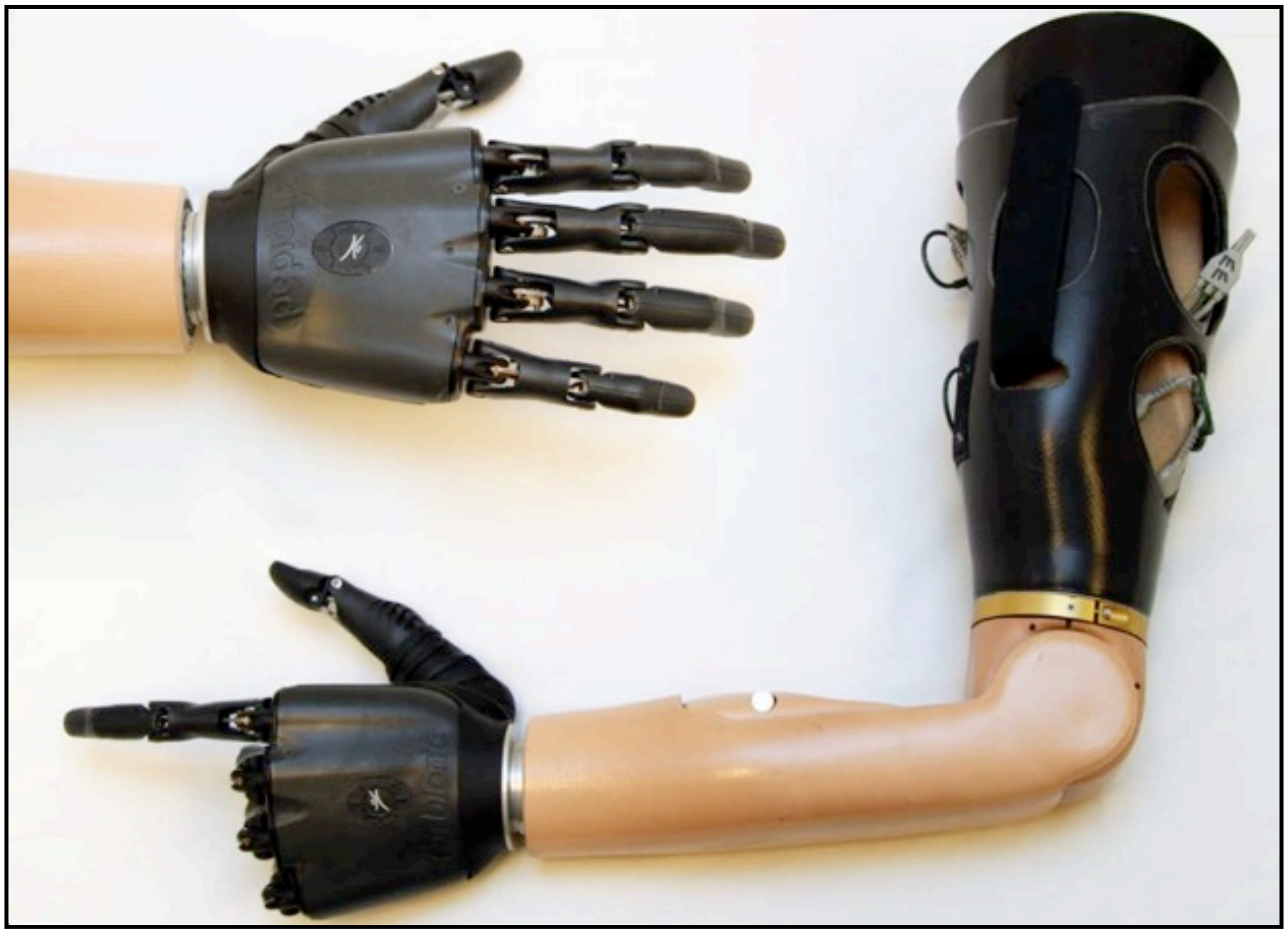
initialize: $\mathbf{w}, \mathbf{e}, \mathbf{s}, \mathbf{x}$
repeat:
  observe $\mathbf{s}$
  $\mathbf{x}' \leftarrow \mathsf{approx}(\mathbf{s})$
  for all joints j **do**
    observe joint activity signal r_j
    $\delta \leftarrow r_j + \gamma \mathbf{w}_j^T \mathbf{x}' - \mathbf{w}_j^T \mathbf{x}$
    $\mathbf{e}_j \leftarrow \mathsf{min}(\lambda \mathbf{e}_j + \mathbf{x}, 1)$
    $\mathbf{w}_j \leftarrow \mathbf{w}_j + \alpha \delta \mathbf{e}_j$
    $\mathbf{x} \leftarrow \mathbf{x}'$

The prediction of future joint activity p_j at any given time is sampled using the linear combination: $p_j \leftarrow \mathbf{w}_j^T \mathbf{x}$

- Develop new machine learning methods to improve human-machine interaction.
- Translate these techniques to preliminary use by amputee and non-amputee subjects.
- Demonstrate clinical impact in studies with amputee participants.

Multifunction Myoelectric Prostheses

Commercial State-of-the-Art



Known Barriers

"Three main problems were mentioned as reasons that amputees stop using their ME prostheses: *nonintuitive control*, *lack of sufficient feedback*, and *insufficient functionality*."

— Peerdeman et al., JRRD, 2011.

Also: cost!

Adaptation & Scalability

"Supervised adaptation should be considered for incorporation into any clinically viable pattern recognition controller, and unsupervised adaptation should receive renewed interest in order to provide transparent adaptation."

— Sensinger et al., 2009.

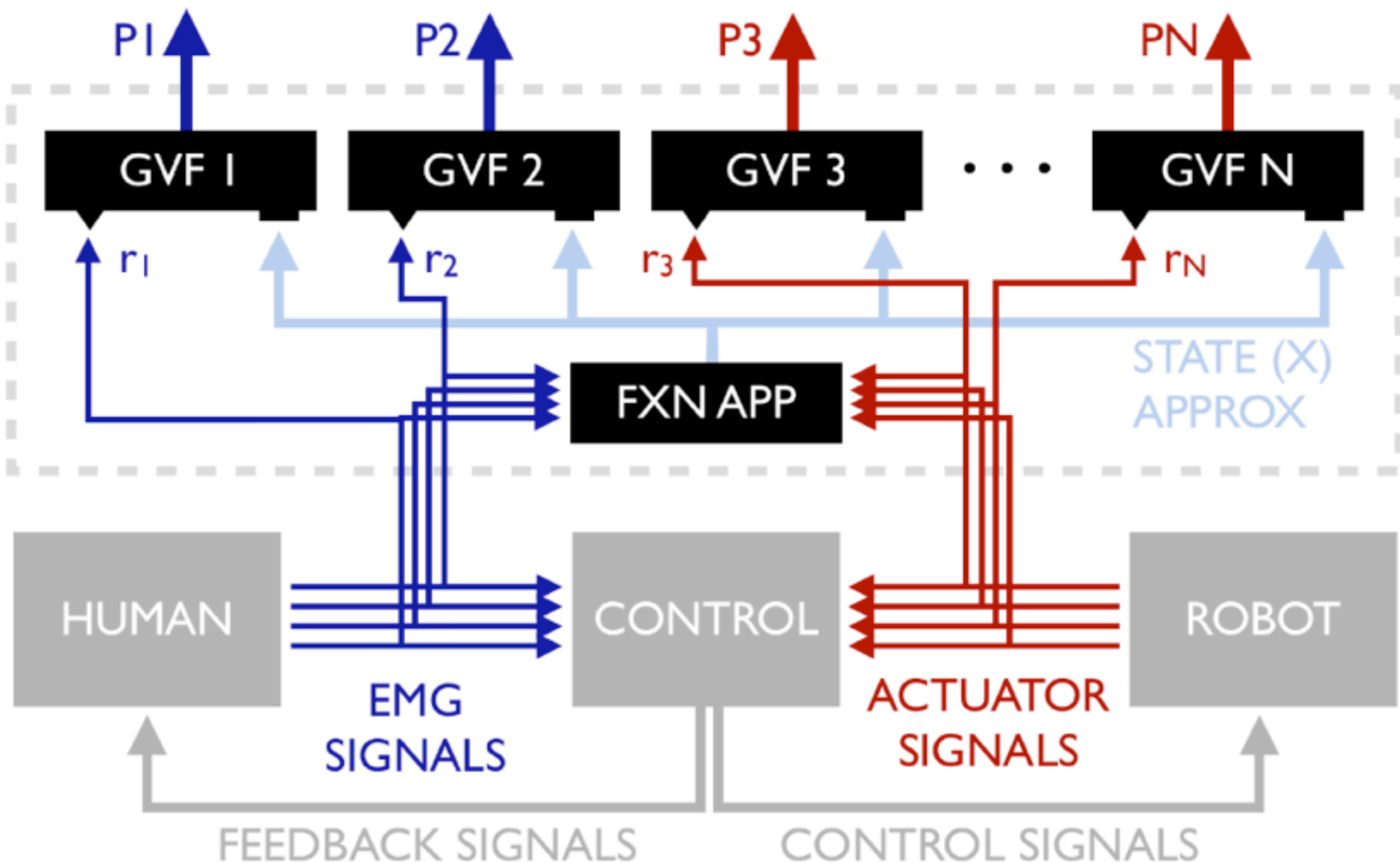
"Completely stable, unsupervised [adaptation] has yet to be realized but is of great clinical interest."

— Scheme and Englehart, 2011.

Our Ongoing Approaches

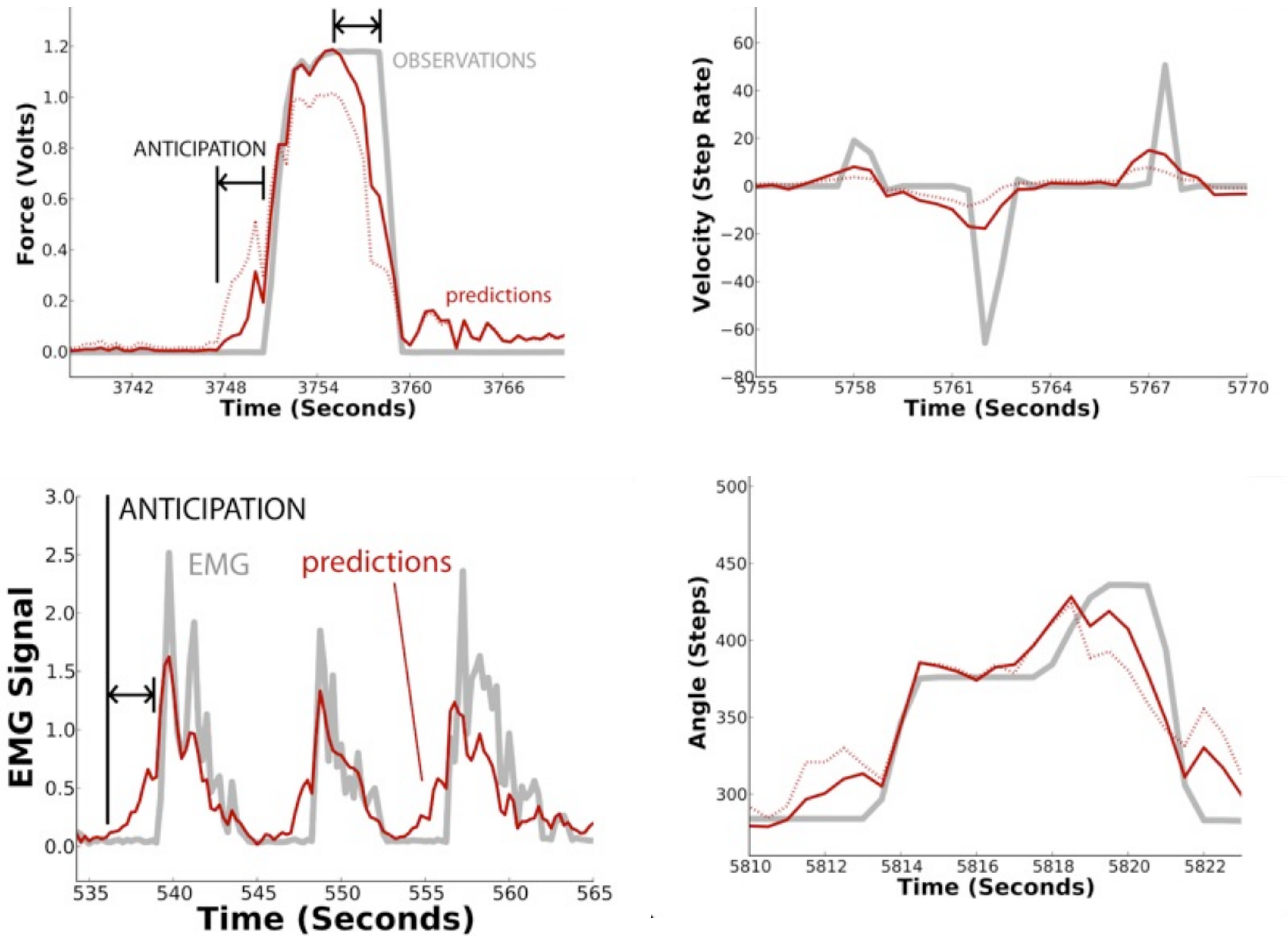
- Real-time control learning without *a priori* information about a user or device.
- Prediction and anticipation of signals during amputee-device interaction.
- Collaborative algorithms for the online human improvement of limb controllers.

Prediction Learning



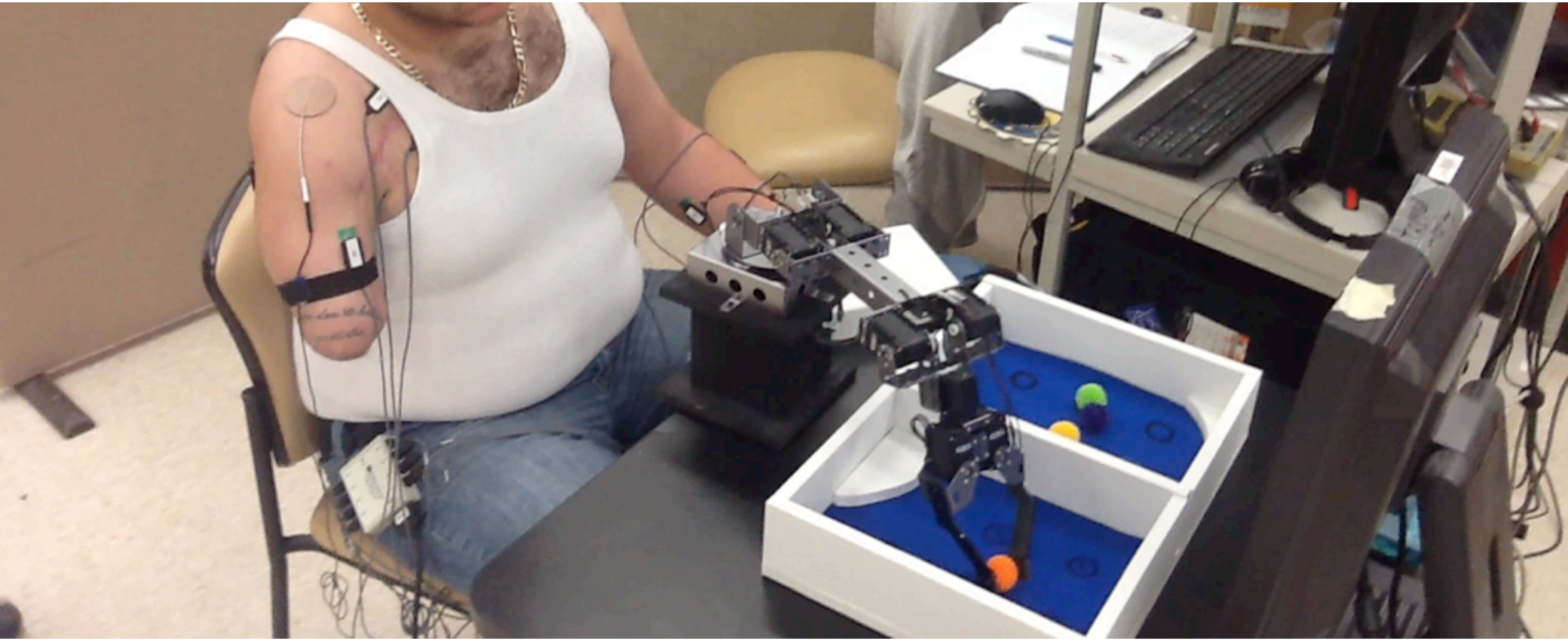
Pilarski et al., IEEE RAM, 2013.

Anticipating Human and Robot Dynamics



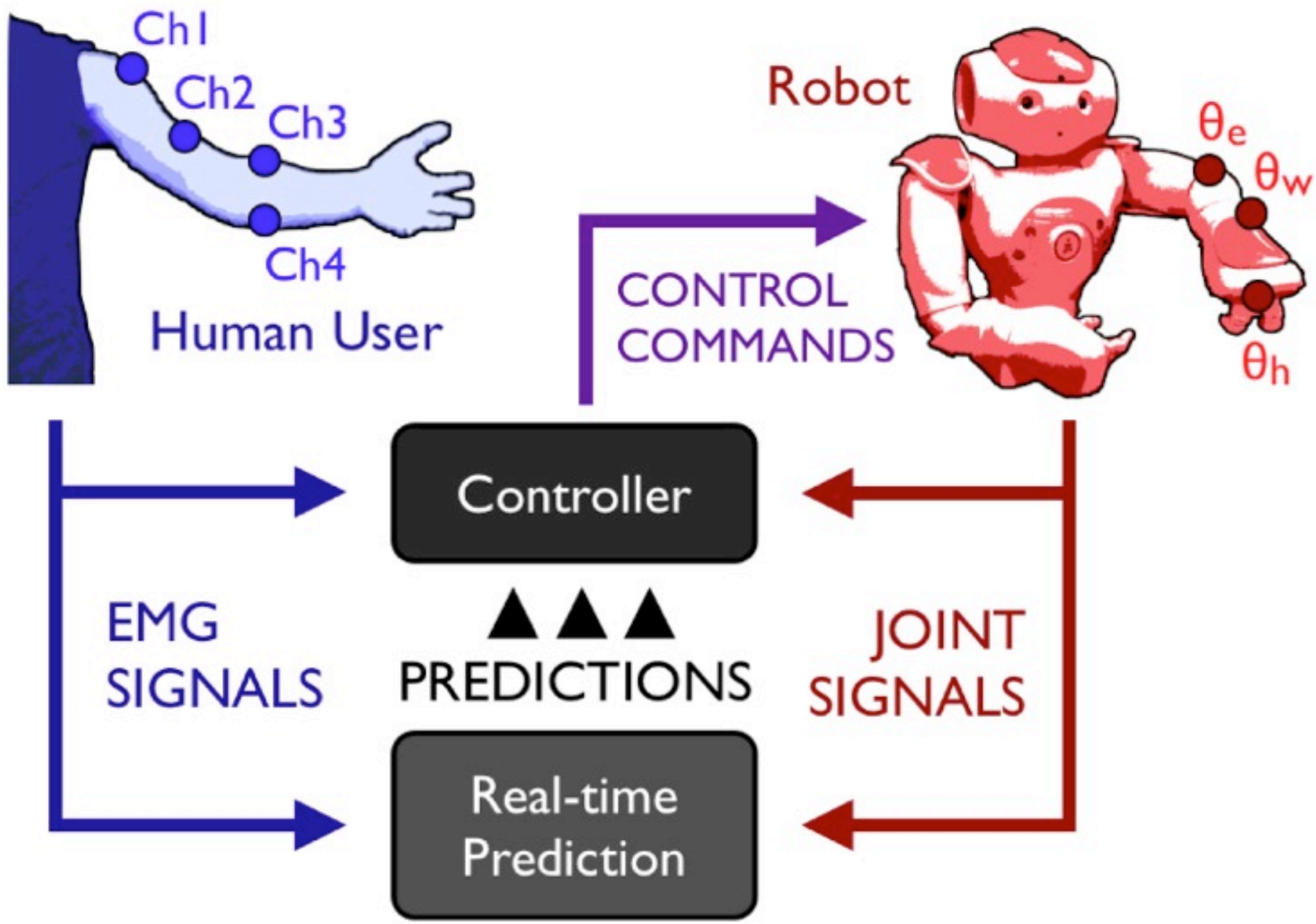
Pilarski et al., IEEE RAM, 2013.

Prediction-based Improvement of a Control Interface



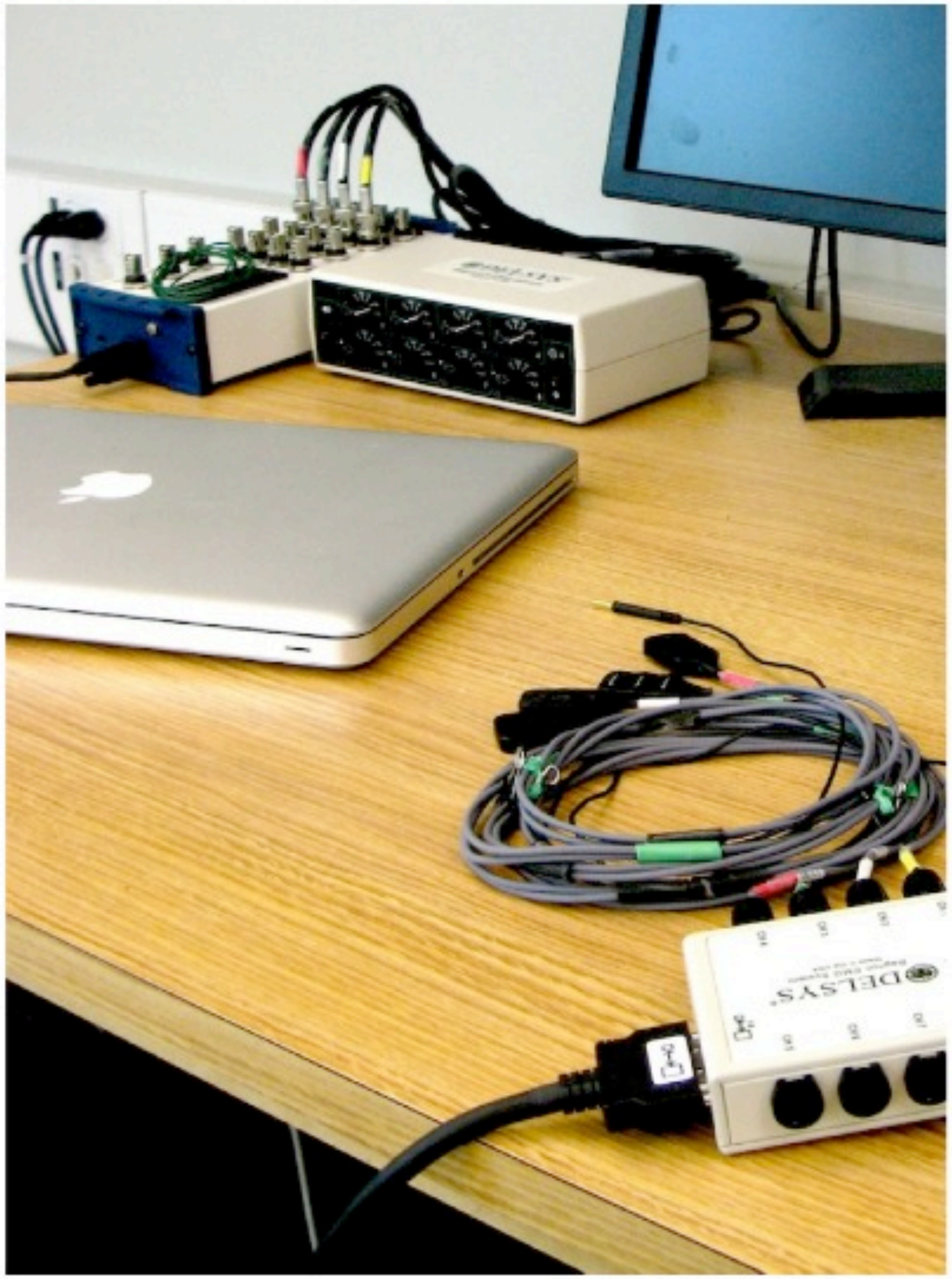
Pilarski et al., BioRob, 2012.
Pilarski and Sutton, AAAI-FS, 2012.

Simultaneous Control of Multiple Joints by using Predictions as Observations



30Hz

Pilarski, Dick, and Sutton, ICORR, 2013.



Free actuation of elbow and hand using conventional control.
Dependent wrist actuator, with desired targets (poses).
~2min online prelearning, ~21min online learning.

Pilarski, Dick, and Sutton, ICORR, 2013.

Wrist Joint Controllers

- Direct W-**Reactive** Control: θW set to θW^*
- Direct W-**Predictive** Control: θW set to PW^*
- **ACRL** Reactive Control: S = {θE,θH,vE,vH,dEMGx2,W})
- **ACRL** EH-Predictive Control: S = {PE, PH, W})
- **ACRL** W-Predictive Control: S = {PW*, W})
- **Prediction Learner**: S = {θE,θH,vE,vH,dEMG x 2})

Both ACRL and TD(lambda) use eligibility traces, and function approximation via tile coding.

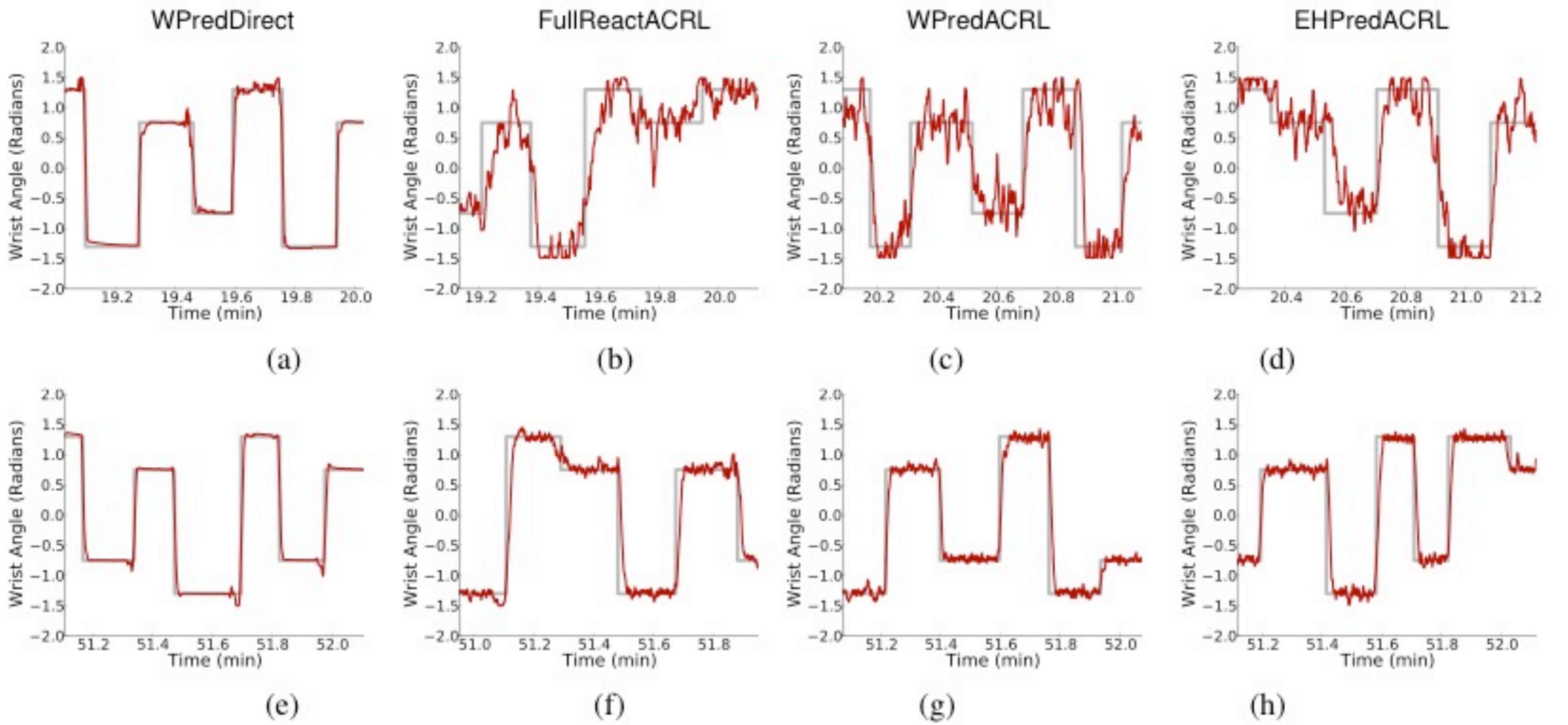
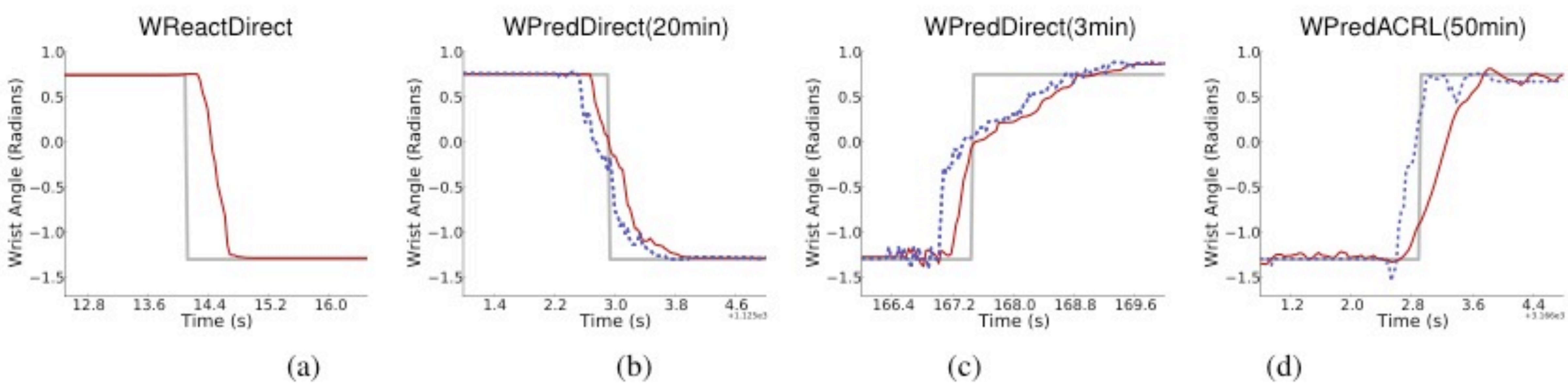
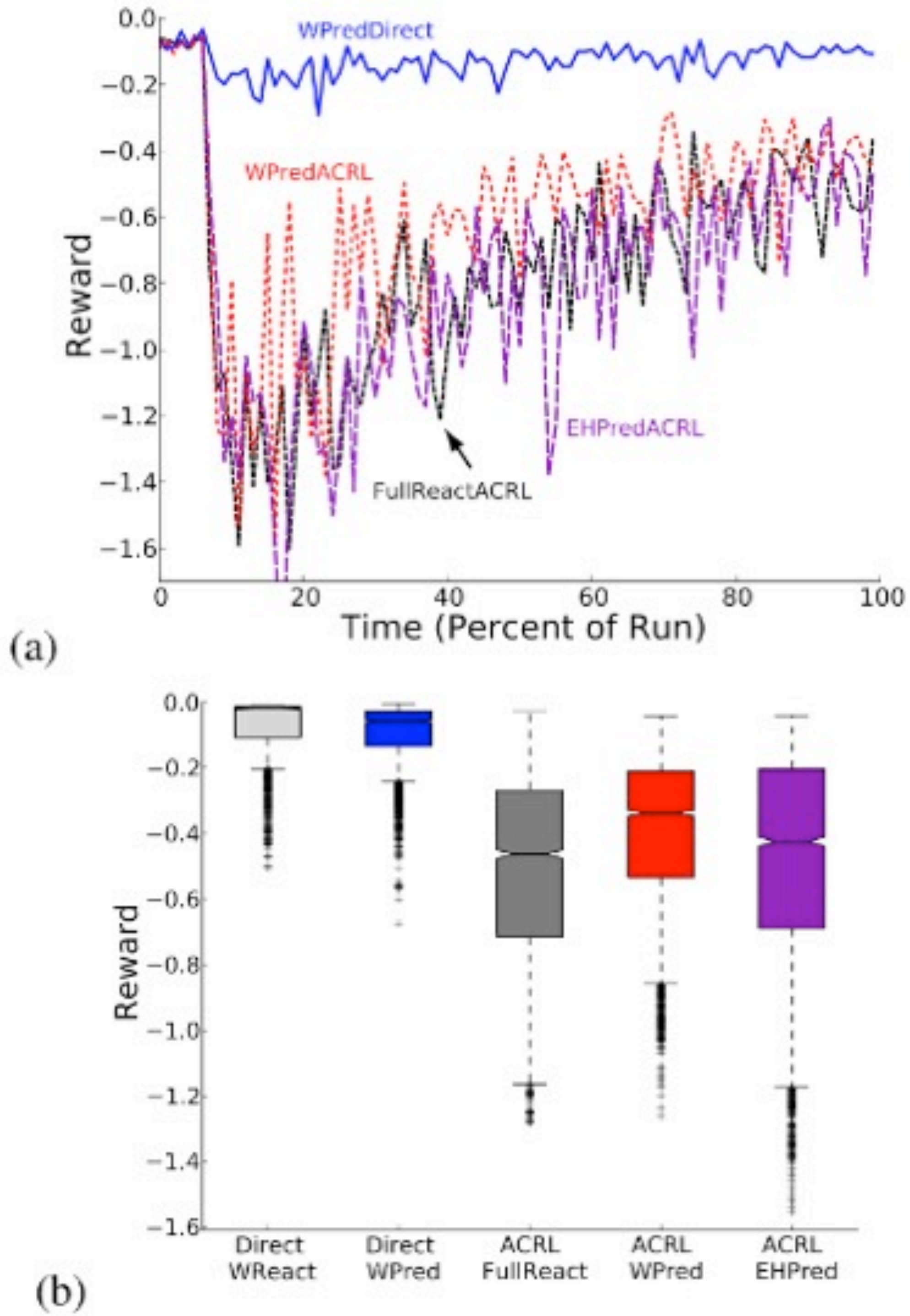


Fig. 5. Comparison of target (grey line) and achieved (red line) wrist trajectories after (a–d) ~20min of online learning and (e–h) ~50min of offline learning. Shown for (a/e) Direct W-Predictive control, (b/f) Full-Reactive ACRL, (c/g) W-Predictive ACRL, and (d/h) EH-Predictive ACRL.



Pilarski, Dick, and Sutton, ICORR, 2013.



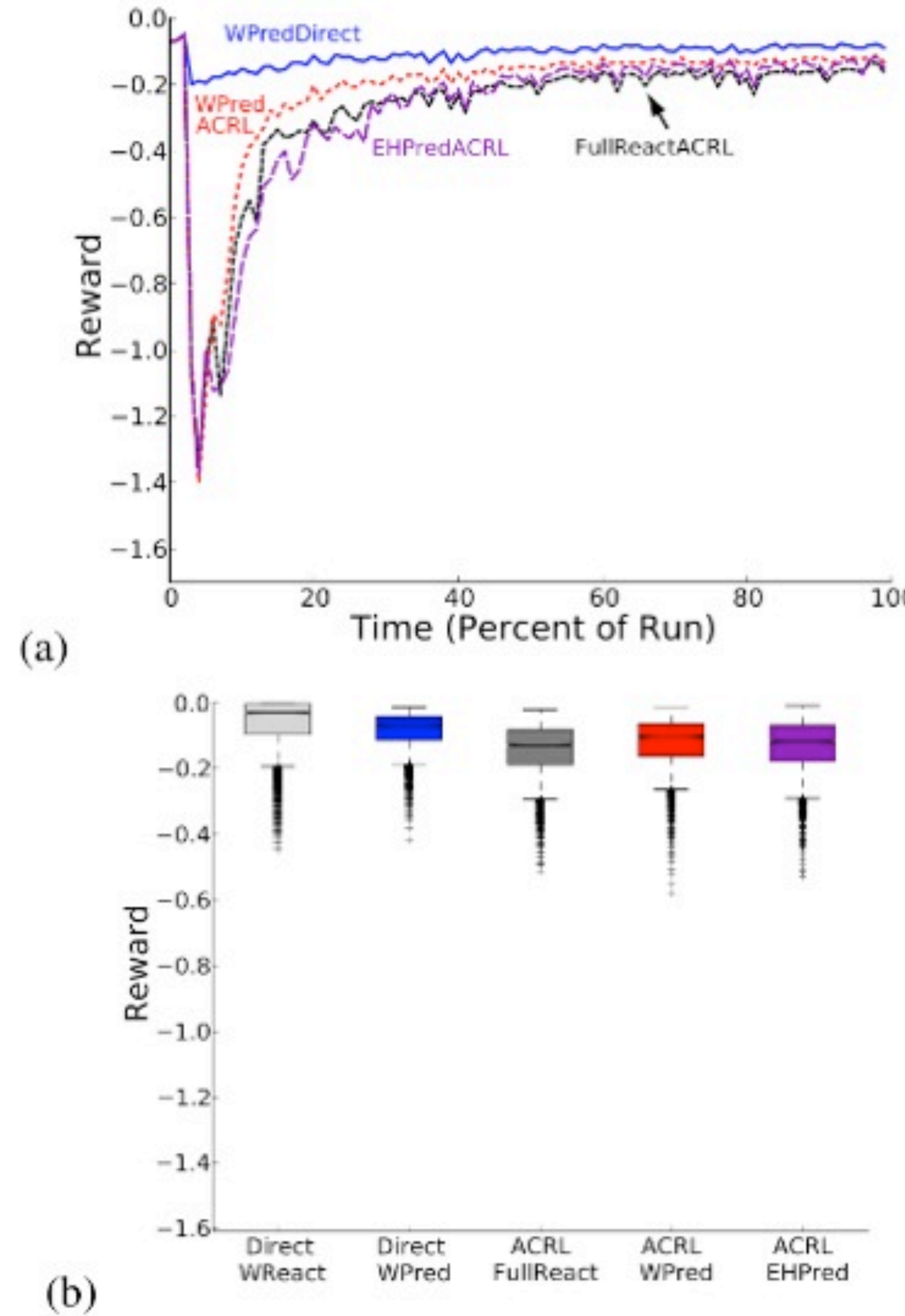
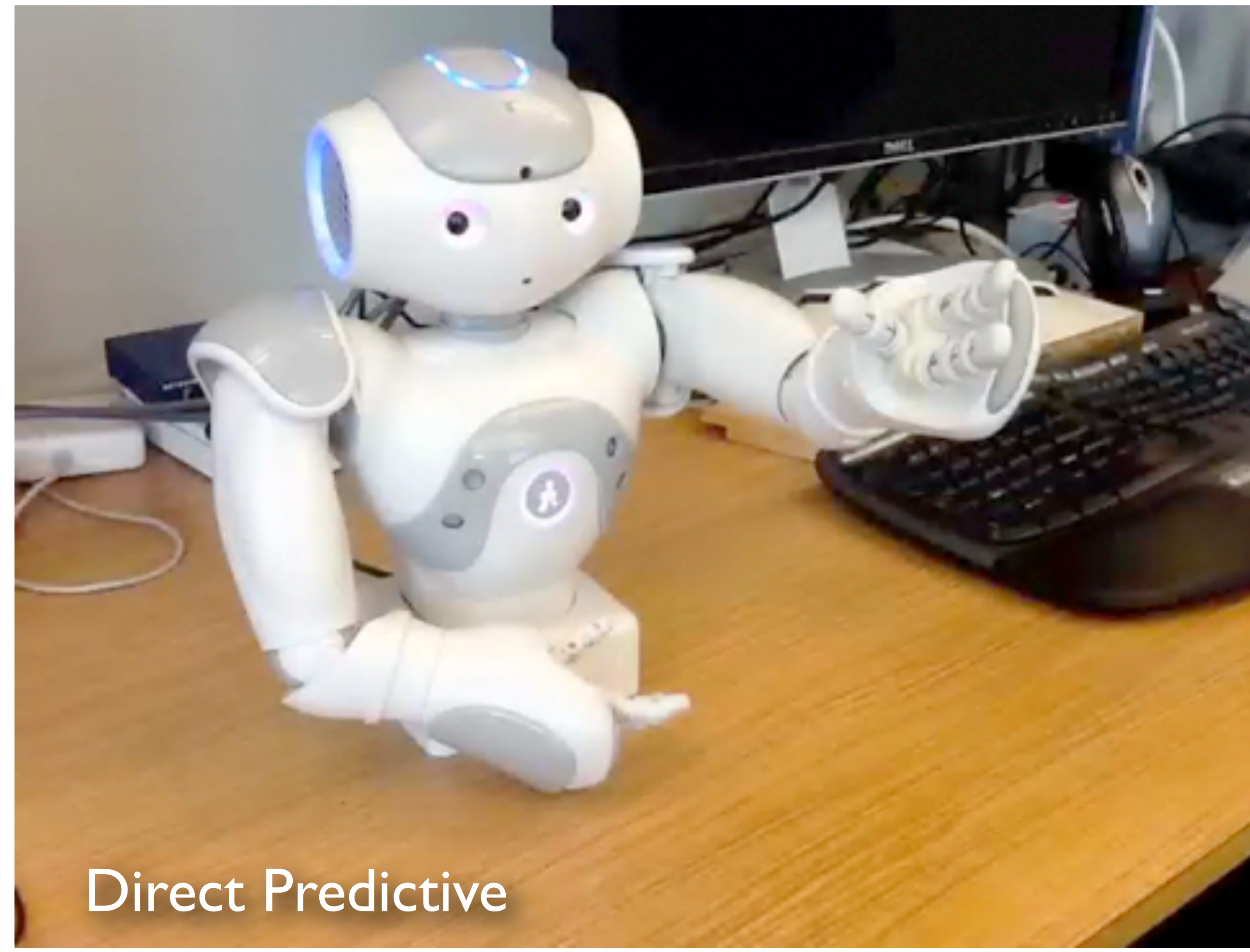


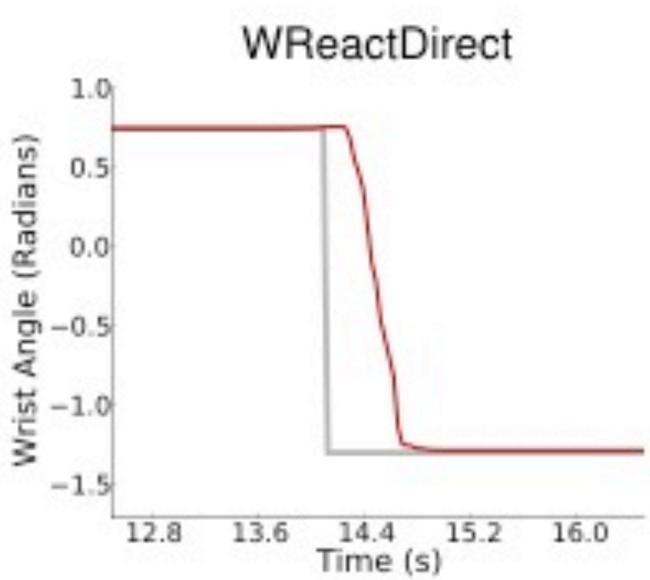
Fig. 3. Comparison of predictive and reactive control learning approaches (n=4) over the course of ~20min of online learning, following a 1.7min pre-learning phase: (a) binned per-time-step reward over time, and (b) quartile analysis of median values shown over the last 1.7min of learning, as compared to 1.7min of the direct reactive policy during pre-learning.

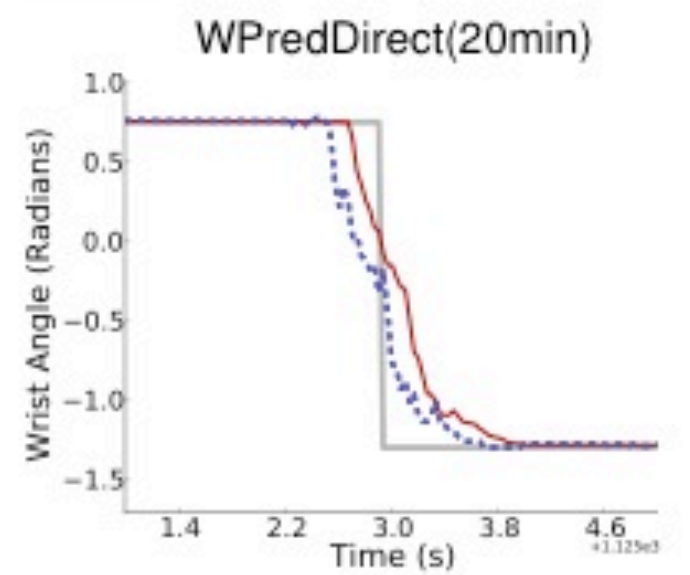
Fig. 4. Comparison of predictive and reactive control learning approaches (n=16) over the course of ~50min of offline learning (2.5 passes through 21min of logged online learning data, following 1.7min of pre-learning): (a) binned per-time-step reward over time, and (b) quartile analysis of median values shown over the last 1.7min of learning.

Pilarski, Dick, and Sutton, ICORR, 2013.

Example of Direct Predictive Actuator Control (0.25x Speed)







Advanced Artificial Limbs

(NON-INVASIVE)

Rehab. Institute of Chicago: Kuiken et al.

Summary

- When is it pragmatic to use learned, temporally extended predictions in picking robot actions in real-time (in effect, a model made of learned VFs)?
- Can we combine prediction learning with continuous action ACRL in a useful way? (compress, abstract)
- Can this approach be grounded in an incremental, sensorimotor approach to planning? (RS diagram.)
- **Results:** Simultaneous actuation of extra joints and demonstrated preemptive actuation.

Also: general value functions with TD-learning are a practical way to build up diverse predictive model during the real-time operation of a system.

QUESTIONS

pilarski@ualberta.ca

http://www.ualberta.ca/~pilarski/

ALBERTA INNOVATES CENTRE FOR MACHINE LEARNING

- Richard S. Sutton, Travis Dick, RLAI, Dept. Computing Science, Univeristy of Alberta
- Thomas Degris, INRIA, Bordeaux, France
- Michael R. Dawson, Jacqueline S. Hebert, K. Ming Chan
Glenrose Rehabilitation Hospital & University of Alberta
- Jason P. Carey
Dept. of Mechanical Engineering, University of Alberta
- Funders: Alberta Ingenuity Centre for Machine Learning (AICML), the Natural Sciences and Engineering Research Council (NSERC), Alberta Innovates – Technology Futures (AITF), and the Glenrose Rehabilitation Hospital Foundation.