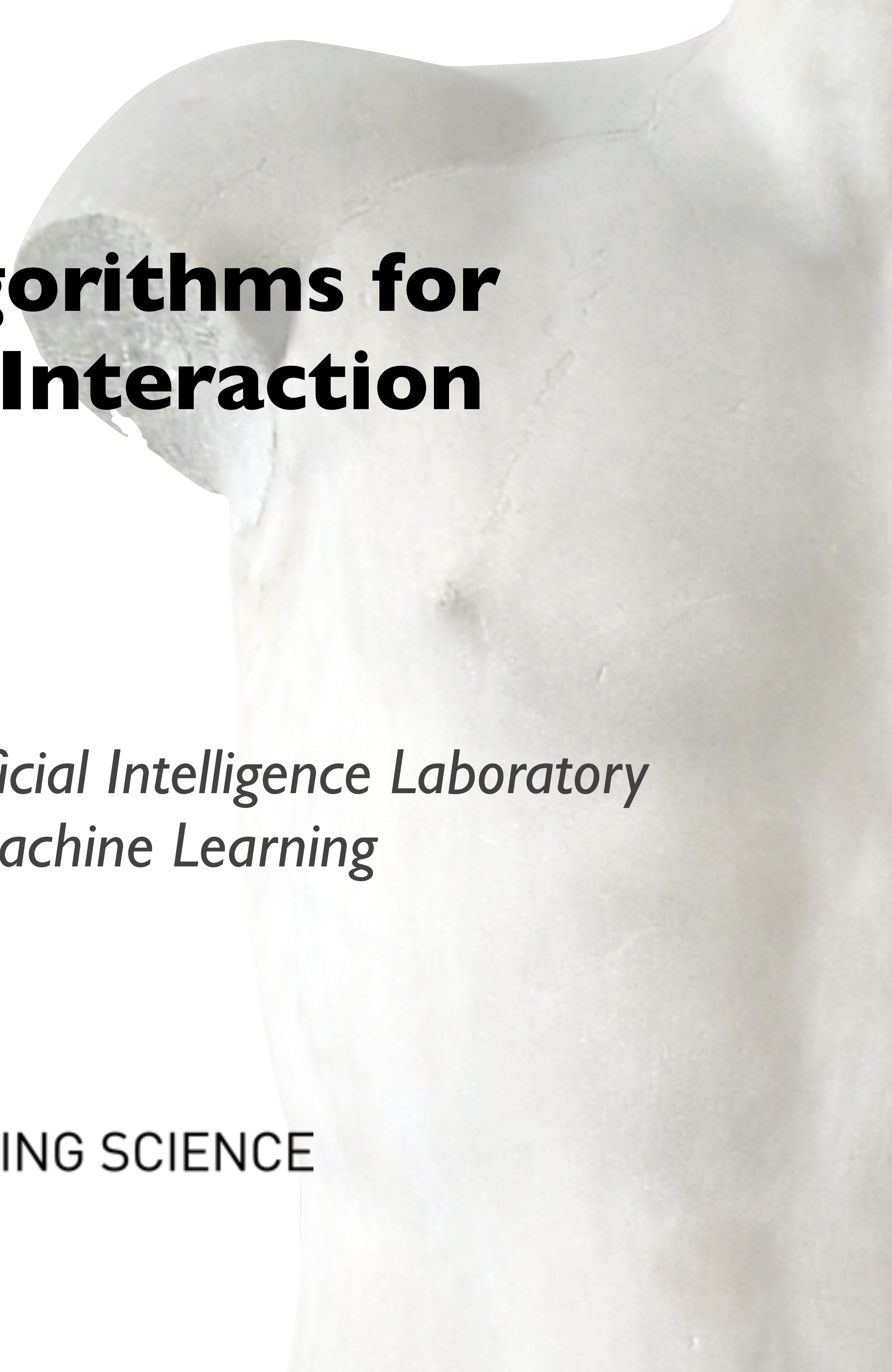
Collaborative Algorithms for Human-Machine Interaction

Patrick M. Pilarski

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

UNIVERSITY OF ALBERTA
DEPARTMENT OF COMPUTING SCIENCE



(Using Prediction to Streamline Prosthesis Control Policies)

Patrick M. Pilarski

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

UNIVERSITY OF ALBERTA
DEPARTMENT OF COMPUTING SCIENCE

Machine Learning for Assistive Devices

- Real-time RL methods applied to:
  - Rehabilitation robotics;
  - Assistive biomedical devices;
  - Human-machine (e.g. neural) interfaces.

- Direct human interaction with complex systems (without assumptions about H&M).

Team

- RLAI / AICML: new methods for improved control, feedback, and online interaction.

- Mec. Eng.: new mechanical limbs and platforms for amputee training (MTT).

- Glenrose / Medicine: new surgeries (TMR & TSR), patients, and clinical expertise.






Multifunction Myoelectric Prostheses

Three 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.

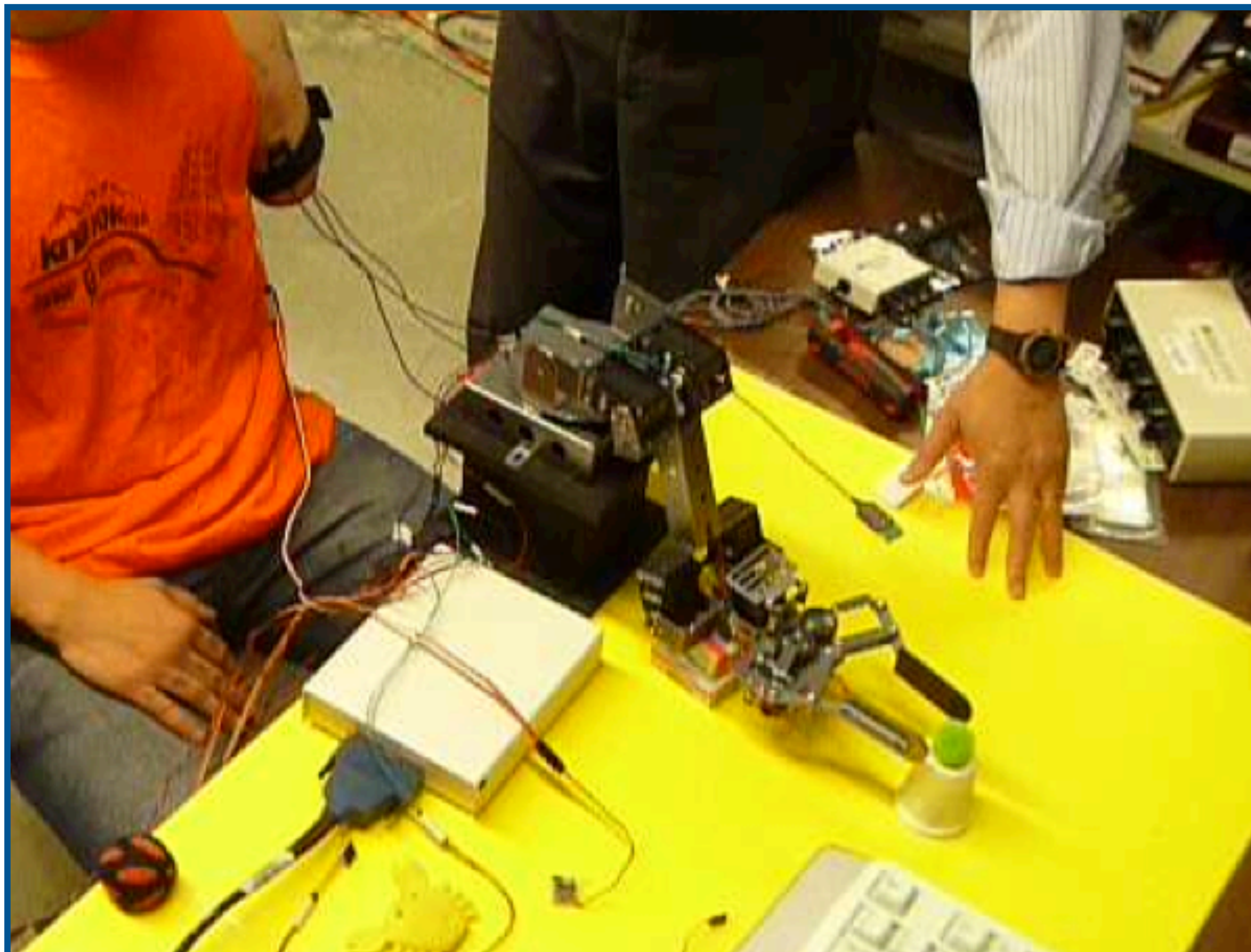
Conventional Control

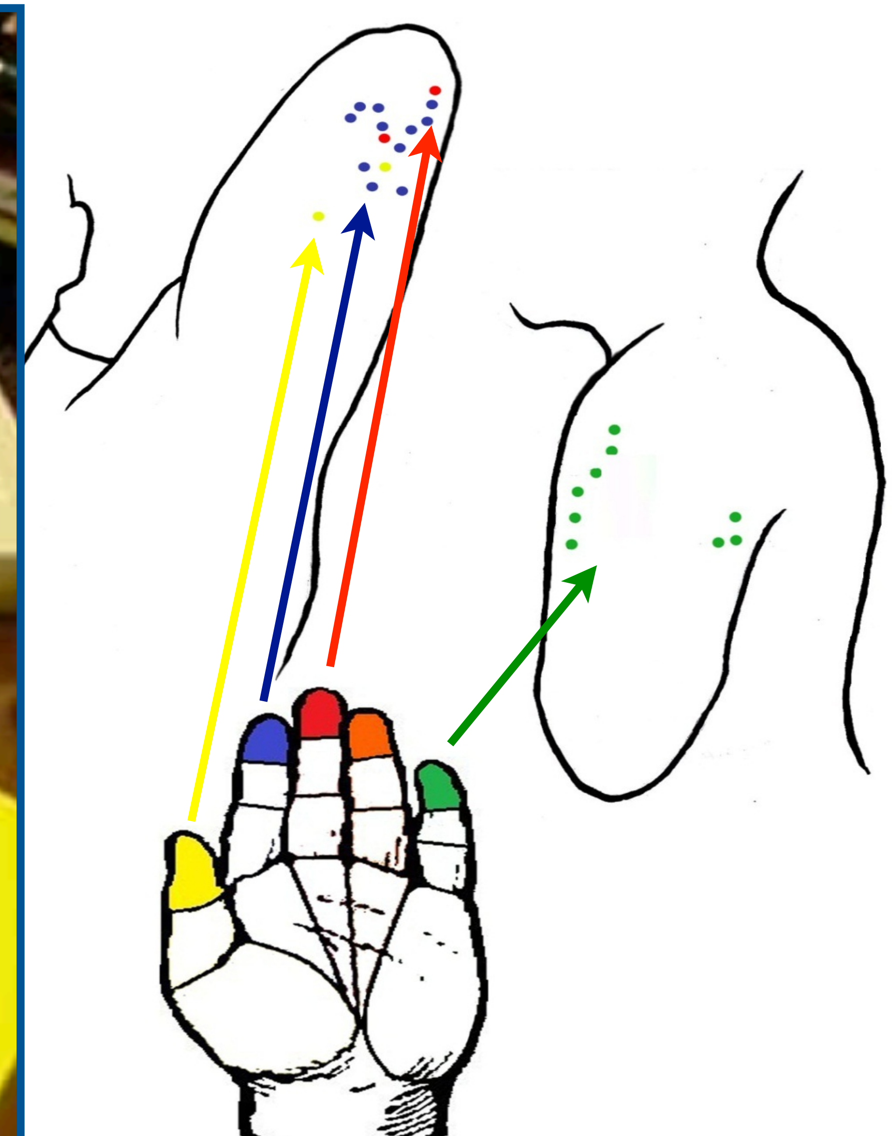
- Conventional myoelectric controllers typically control a single degree of freedom with a single residual muscle pair.
- Unfortunately, as the amputation level increases, the number of muscle sites available for use as input signals to control schemes decreases.
- Growing disparity between the sensing/actuation capability and control system ability.

Learning Approaches

- Developing literature of machine learning work on classifying EMG patterns for use in limb control (e.g. Oskoei and Hu 2008, Parker et al. 2006, Scheme 2011, Sensinger et al. 2009).
- Most contemporary learning approaches rely on external knowledge of their domain to guide learning, and function primarily in offline or batch learning scenarios.
- Robust online adaptation is an open problem (Sensinger et al. 2009, Scheme and Englehart 2011)

Targeted Reinnervation





Our Ongoing Projects

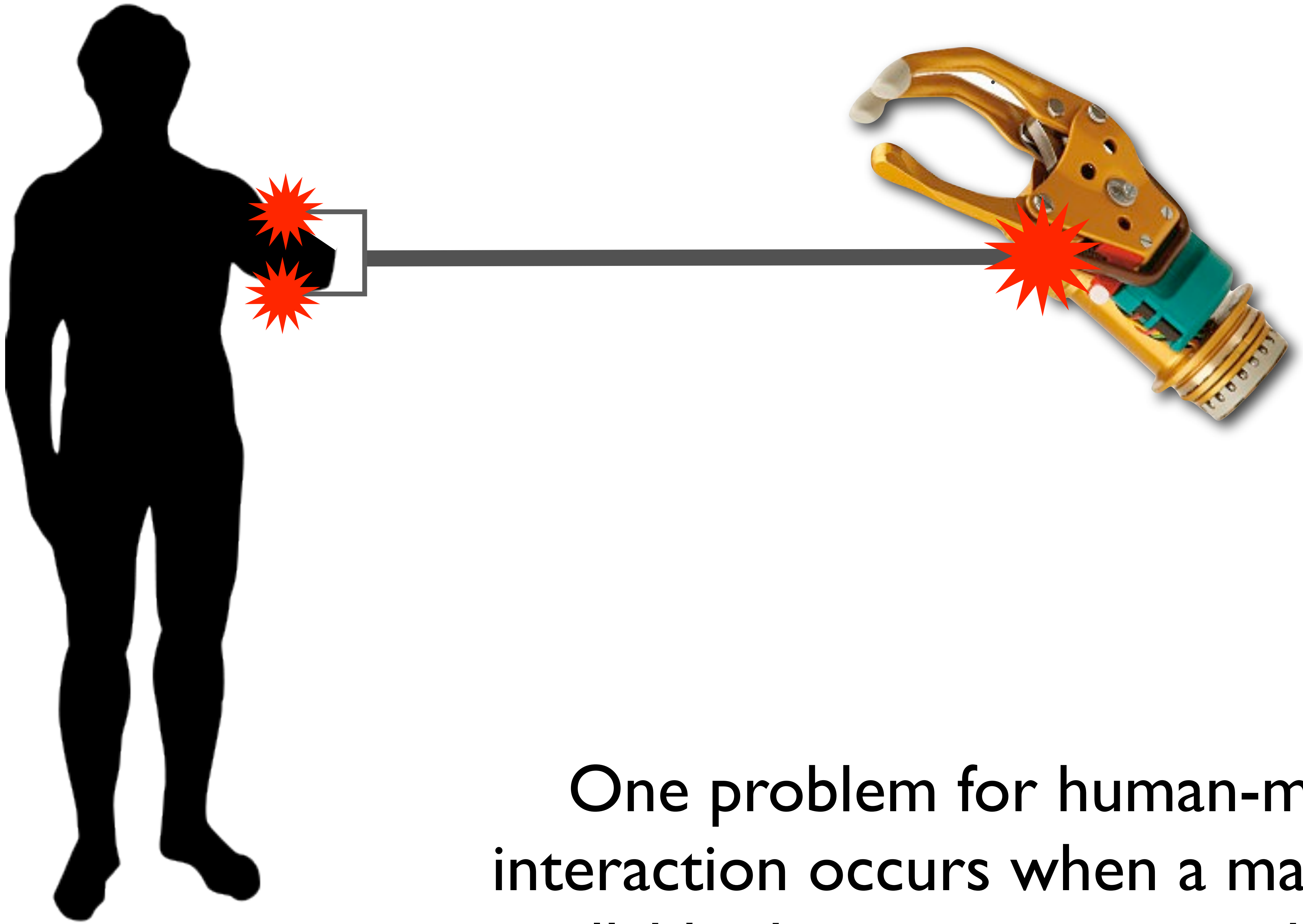
- Real-time control learning without *a priori* information about a user or device.
- Prediction and anticipation of signals during patient-device interaction.
- Collaborative algorithms for the online human improvement of limb controllers.

the switching problem

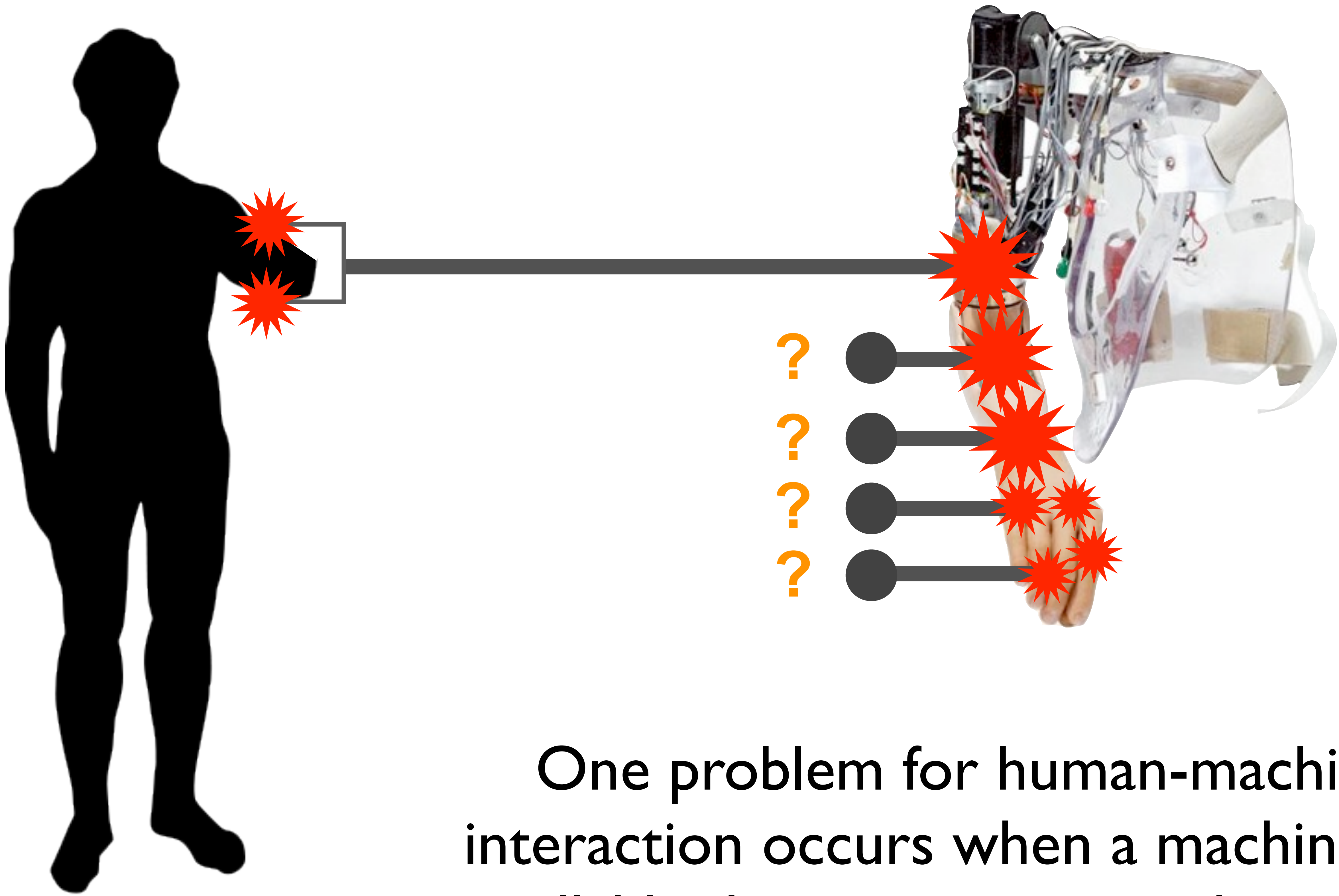
for assistive biomedical devices

Switching in Practice

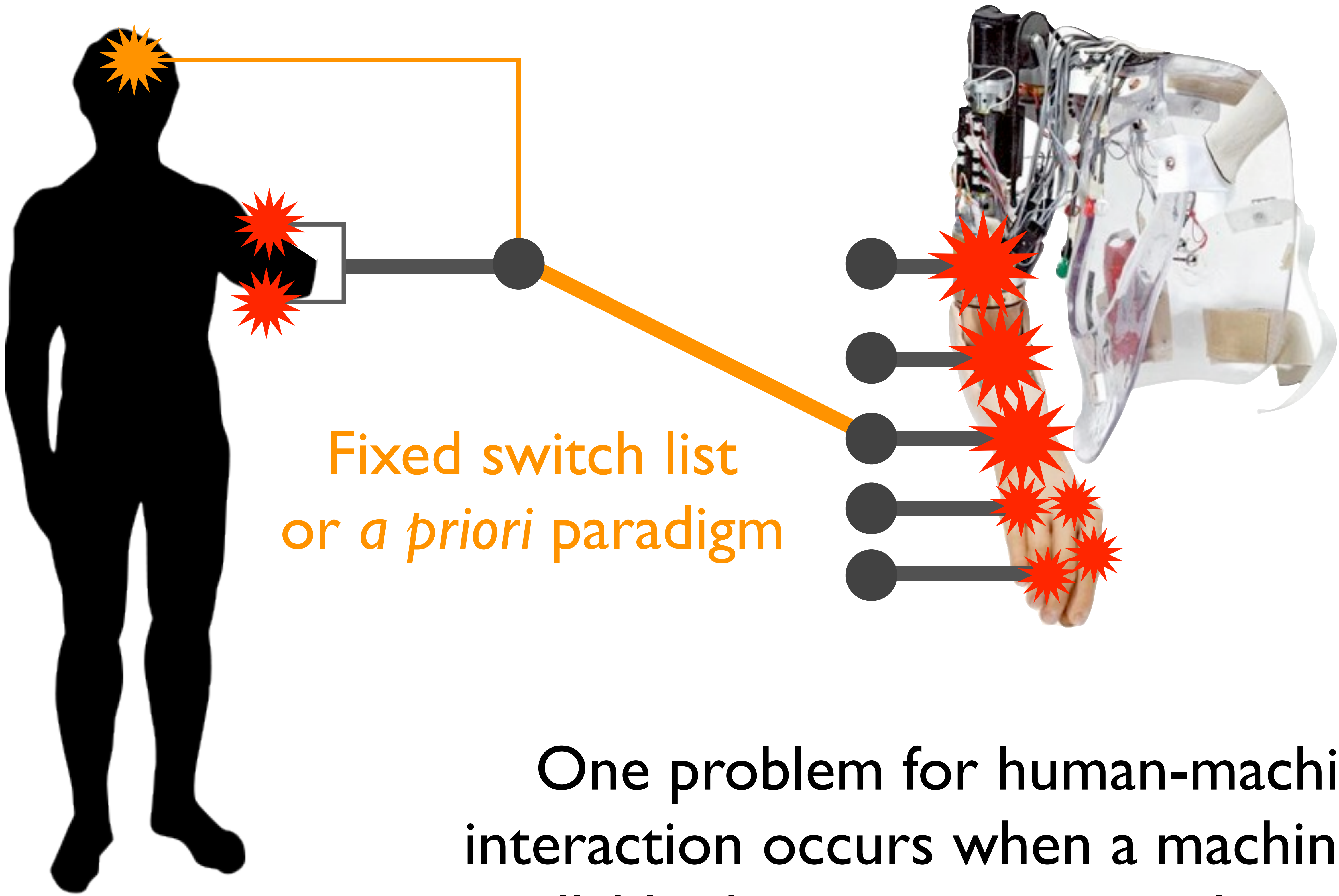
- Most commerical multifunction prostheses use some form of function switching (1 site to 1 DoF).
- In order to increase the number of cotrollable DoFs, conventional controllers are often extended using a voluntary switch.
- It is challenging to form a link between the human and the robot that enables high levels of robot functionality while simultaneously providing an intuitive, learnable control scheme for the user.



One problem for human-machine interaction occurs when a machine's controllable dimensions outnumber the control channels available to its human user



One problem for human-machine interaction occurs when a machine's controllable dimensions outnumber the control channels available to its human user



One problem for human-machine interaction occurs when a machine's controllable dimensions outnumber the control channels available to its human user

predictions

dynamic (adaptive) switching order
for improved control

predictions

dynamic (adaptive) switching order
for improved control

P.M. Pilarski, M.R. Dawson, T. Degris, J.P. Carey, and R.S. Sutton,
4th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob),
June 24-28, Roma, Italy, 7 pages, 2012.

Approach

- Learning system streamlines user switching.
- Intuition: switching order should refelct context, and adapt to changes in the task, changes in the user.
- Learn (and adapt) predictions about user control interactions in real-time.
- Dynamically reorder DoFs in the switching list (in an online, ongoing fashion).

Experimental Domain

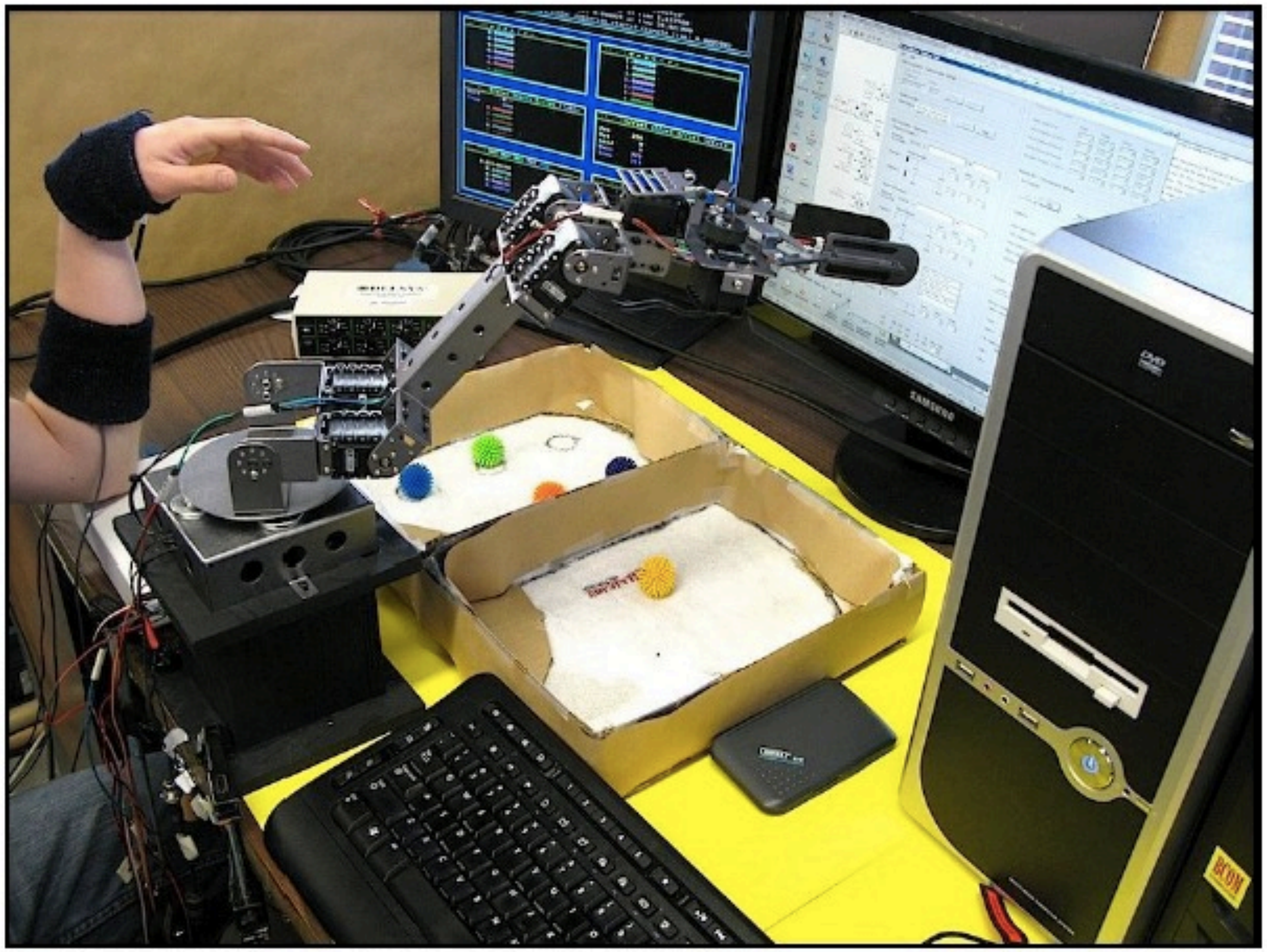
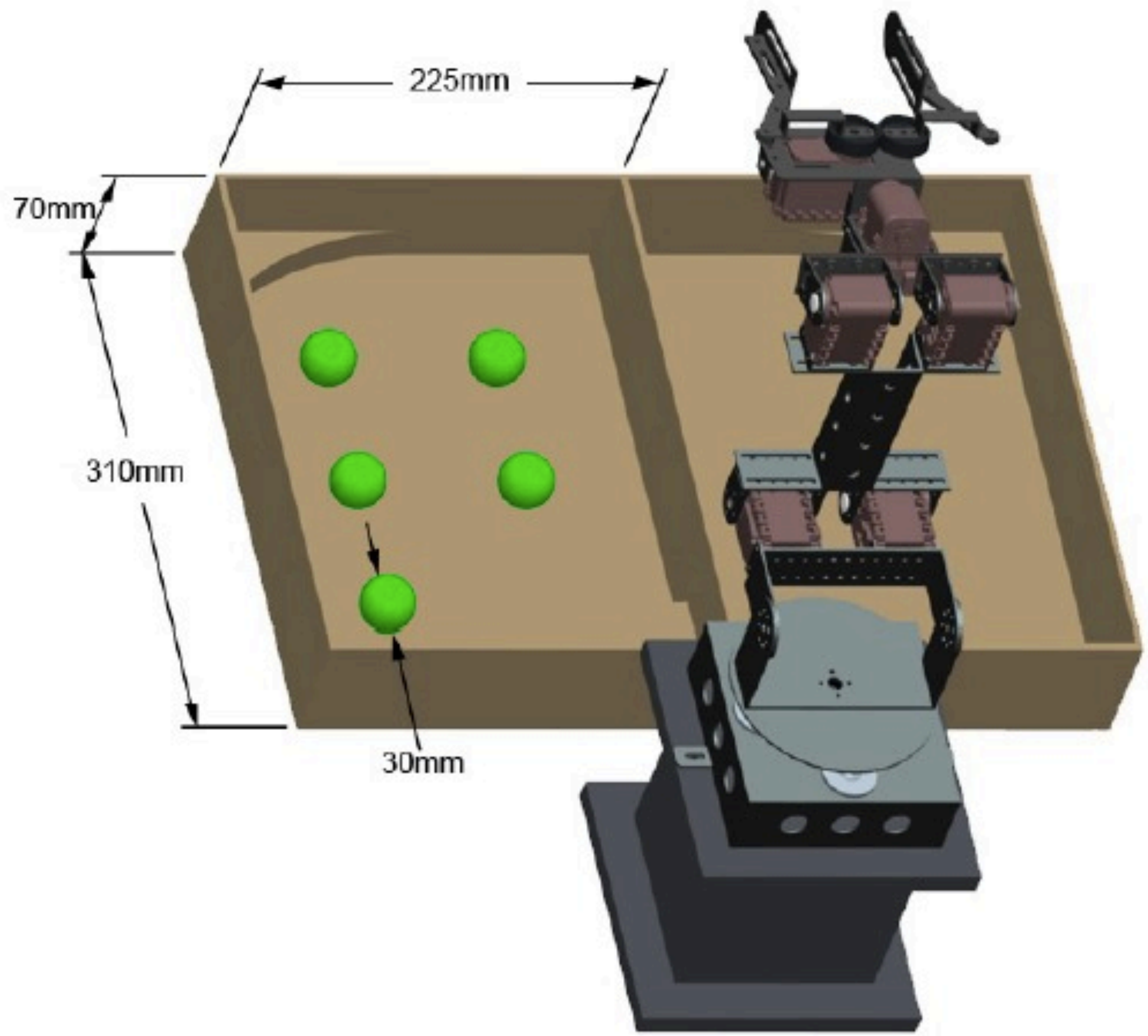
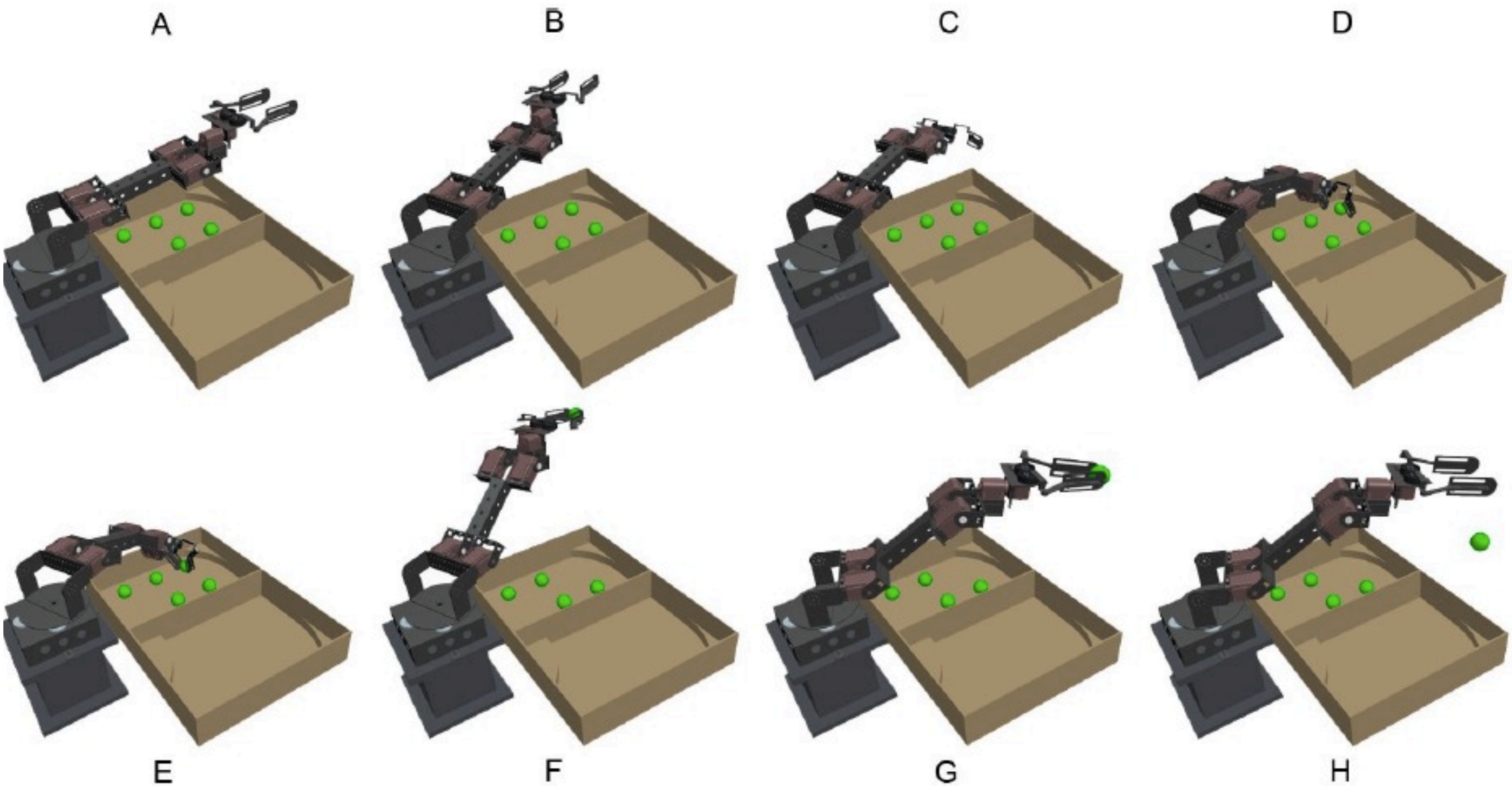


Fig. 1. **Able-bodied subject interacting interacting with the Myoelectric Training Tool (MTT);** experimental setup also includes a Bagnoli 8-channel EMG system, real-time control computer, and task workspace.

Box and Blocks Task



Example Sequence



There are many ways to achieve this task.

Rich Data

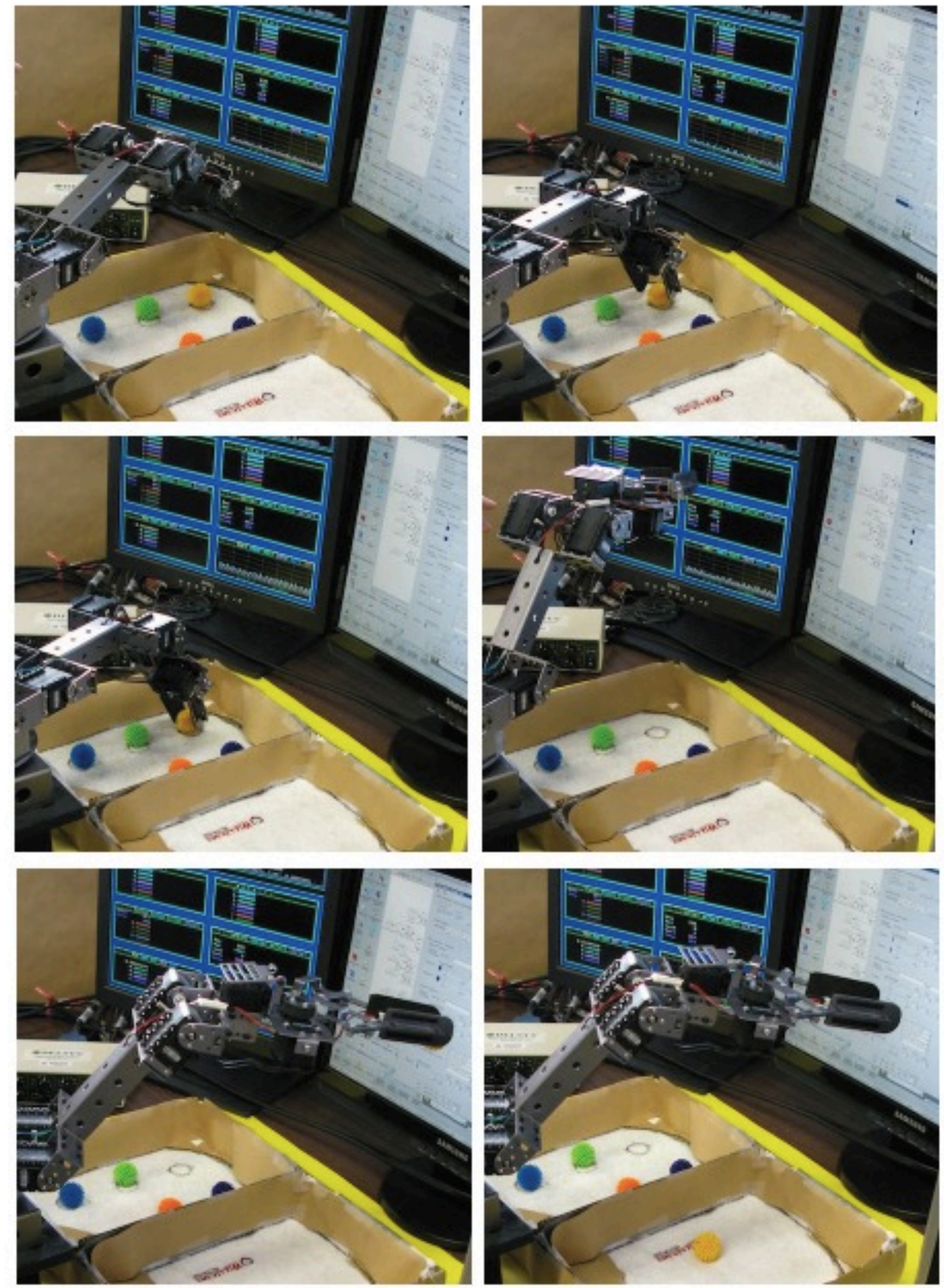
(A) Array Dimensions 1 through 4

Shoulder Servo Position	Elbow Servo Position
Wrist Servo Position	Hand Servo Position

(B) Array Dimension 5

One of:

ShoulderServoVelocity	ShoulderServoLoad
ShoulderServoVoltage	ShoulderServoTemperature
ElbowServoVelocity	ElbowServoLoad
ElbowServoVoltage	ElbowServoTemperature
WristServoVelocity	WristServoLoad
WristServoVoltage	WristServoTemperature
HandServoVelocity	HandServoLoad
HandServoVoltage	HandServoTemperature
HandForceSensor	EmgSwitchMav
Emg1Mav	Emg2Mav
HandControlState	WristControlState
ElbowControlState	ShoulderControlState
HandActivityTrace	WristActivityTrace
ElbowActivityTrace	ShoulderActivityTrace



Interesting Questions

- **Predictions regarding user control:**
  - Which function will the user select when they perform their next switching action?
  - How much activity will be observed on a DoF over the next few seconds?
  - Will the voluntary switch be activated in the next few timesteps?

Online Nexting

- General Value Functions. (Sutton et al., 2011, AAMAS)
- GVFs form questions; "what will happen next?" (Nexting; Modayil et al. 2012)
- In brief: instead of reward, learn anticipations (expectations of real-valued signals).
- Can learn many temporally extended predictions in parallel.

Why GVFs?

- Thousands of accurate predictions can be made and learned in real time (i.e., 10hz)
- A single state representation be used to accurately predict many different sensors at many different time scales.
- A model-free algorithm that can learn fast enough to be useful.

Multi-timescale Nexting in a Reinforcement Learning Robot, Modayil, White, and Sutton. ArXiv preprint 1112.1133, 2012.

Sutton et al., AAMAS, 2011.

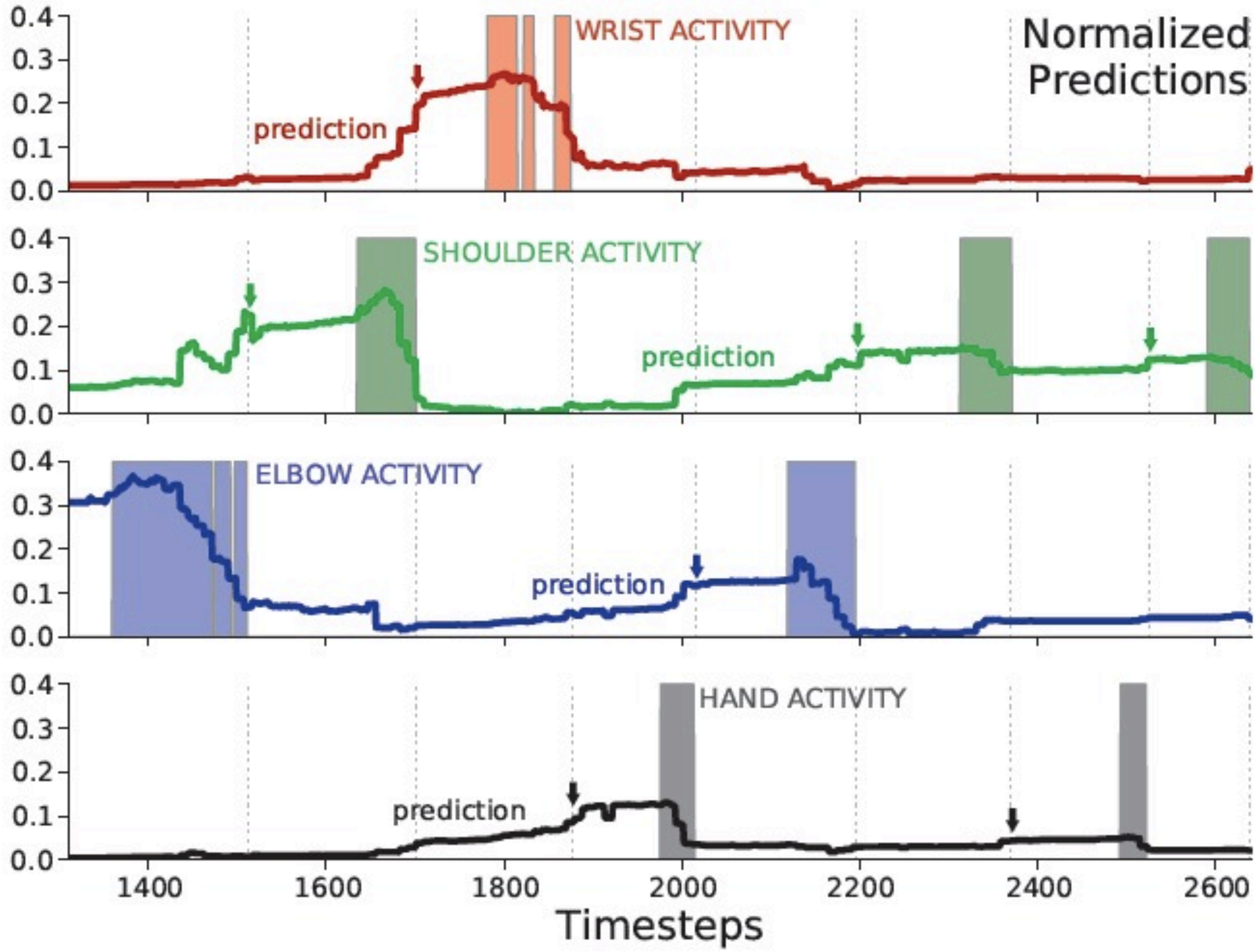
Learning Algorithm

Algorithm 1 Learning General Value Functions with TD(λ)

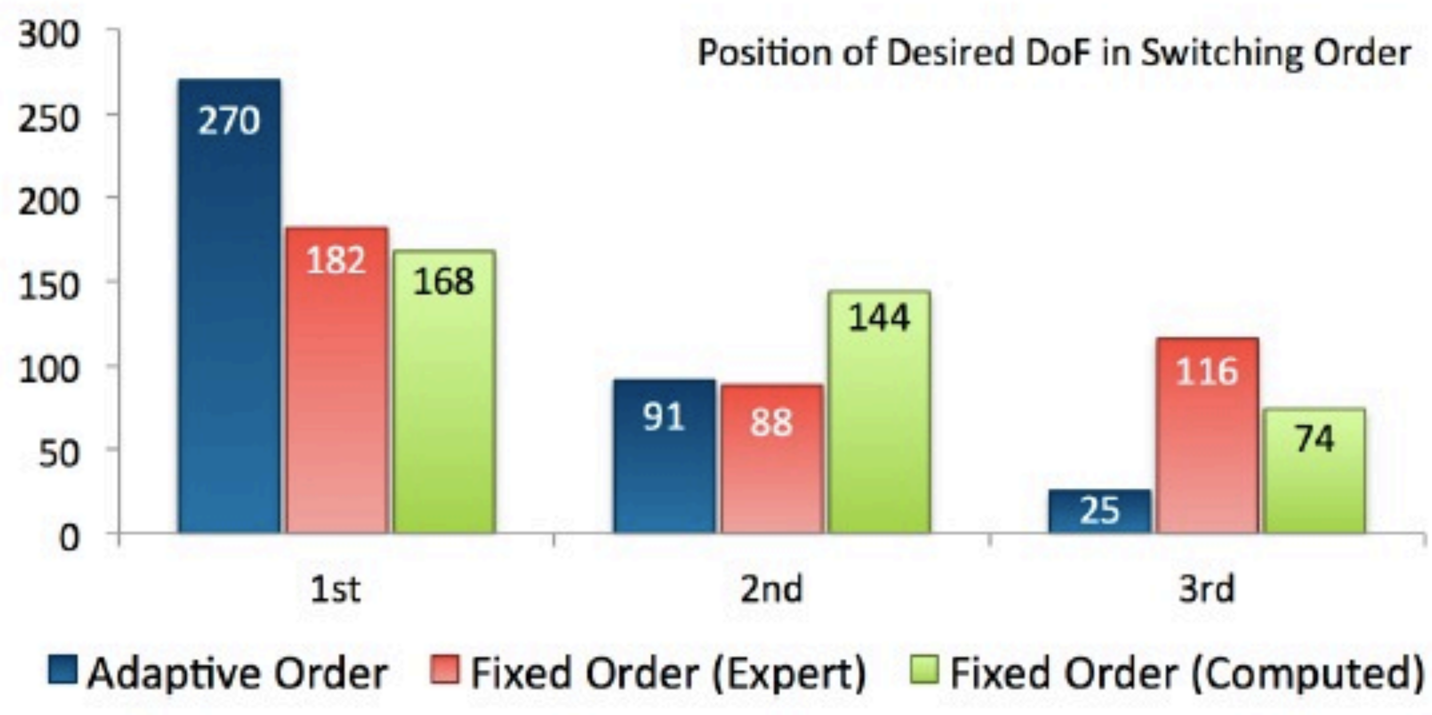
initialize: $\mathbf{w}, \mathbf{e}, \mathbf{s}, \mathbf{x}$
repeat:
  observe $\mathbf{s}$
  $\mathbf{x}' \leftarrow \mathsf{approx}(\mathbf{s})$
  for all joints j **do**
    observe joint activity signal r_j
    $\delta \leftarrow r_j + \gamma \mathbf{w}_j^T \mathbf{x}' - \mathbf{w}_j^T \mathbf{x}$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

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

Accurate Anticipations



Switching Improvement



Increase in the number of ideal switching suggestions (+23%)

Switching Improvement

Transition with 1 switching actions, mean time:	1.09 sec
Transition with 2 switching actions, mean time:	1.75 sec
Transition with 3 switching actions, mean time:	2.21 sec
Net experiment time:	20.66 min
Net observed transition time:	10.40 min
Net transition time(projected for best fixed order):	9.98 min
Net transition time(projected for adaptive order):	8.49 min

Potential time savings with adaptive control:	1.49 min
Potential time savings on transitions:	14.3%
Potential time savings on full experiment:	7.2%

Fig. 8. Additional performance numbers for the switching task, including projected time savings from nexting predictions.

Improvement and Error

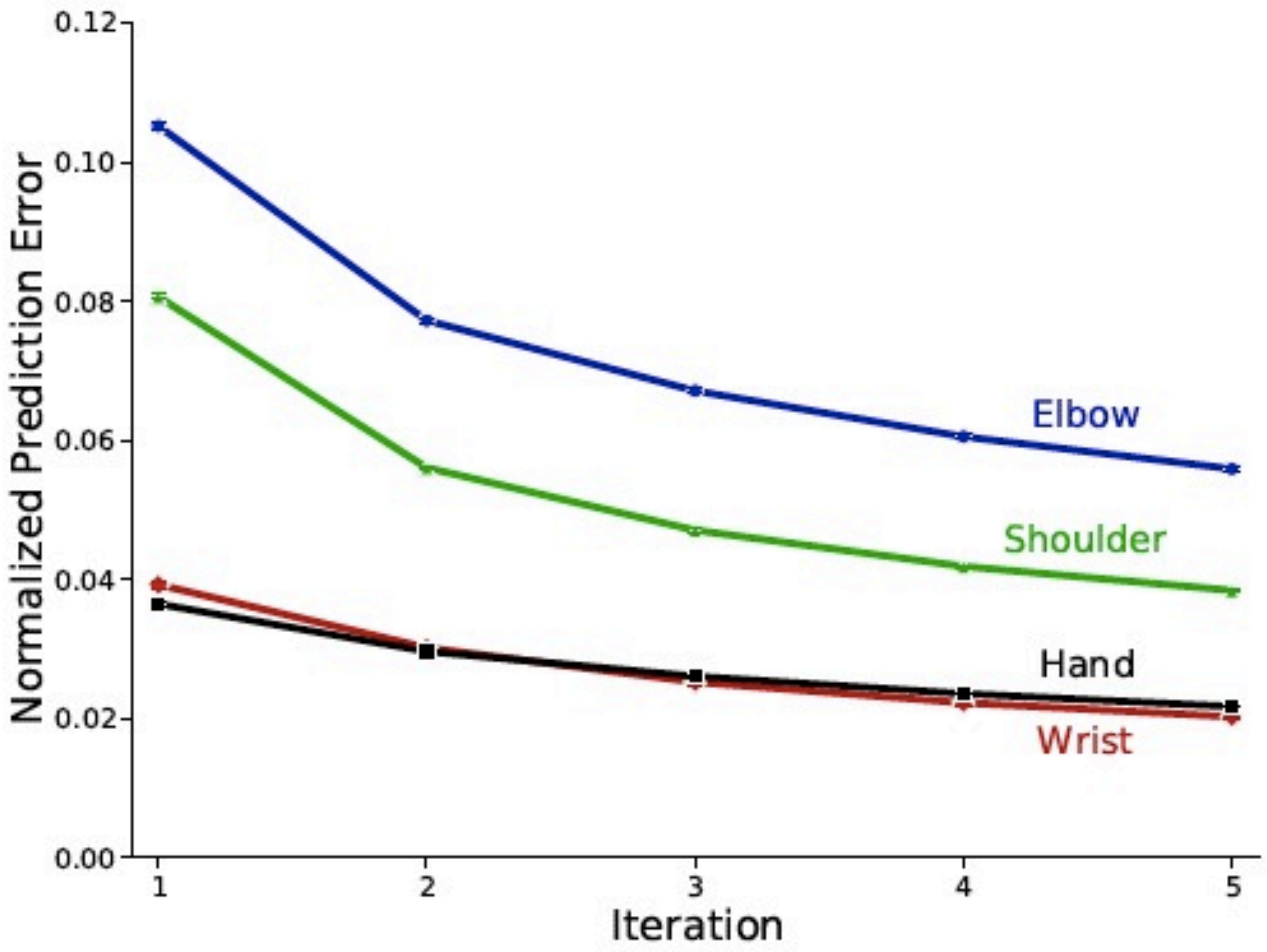


Fig. 7. **Improvement in prediction error after multiple learning iterations over the training data**, in terms of the mean absolute normalized error between learned predictions and the true, post-hoc predictions.

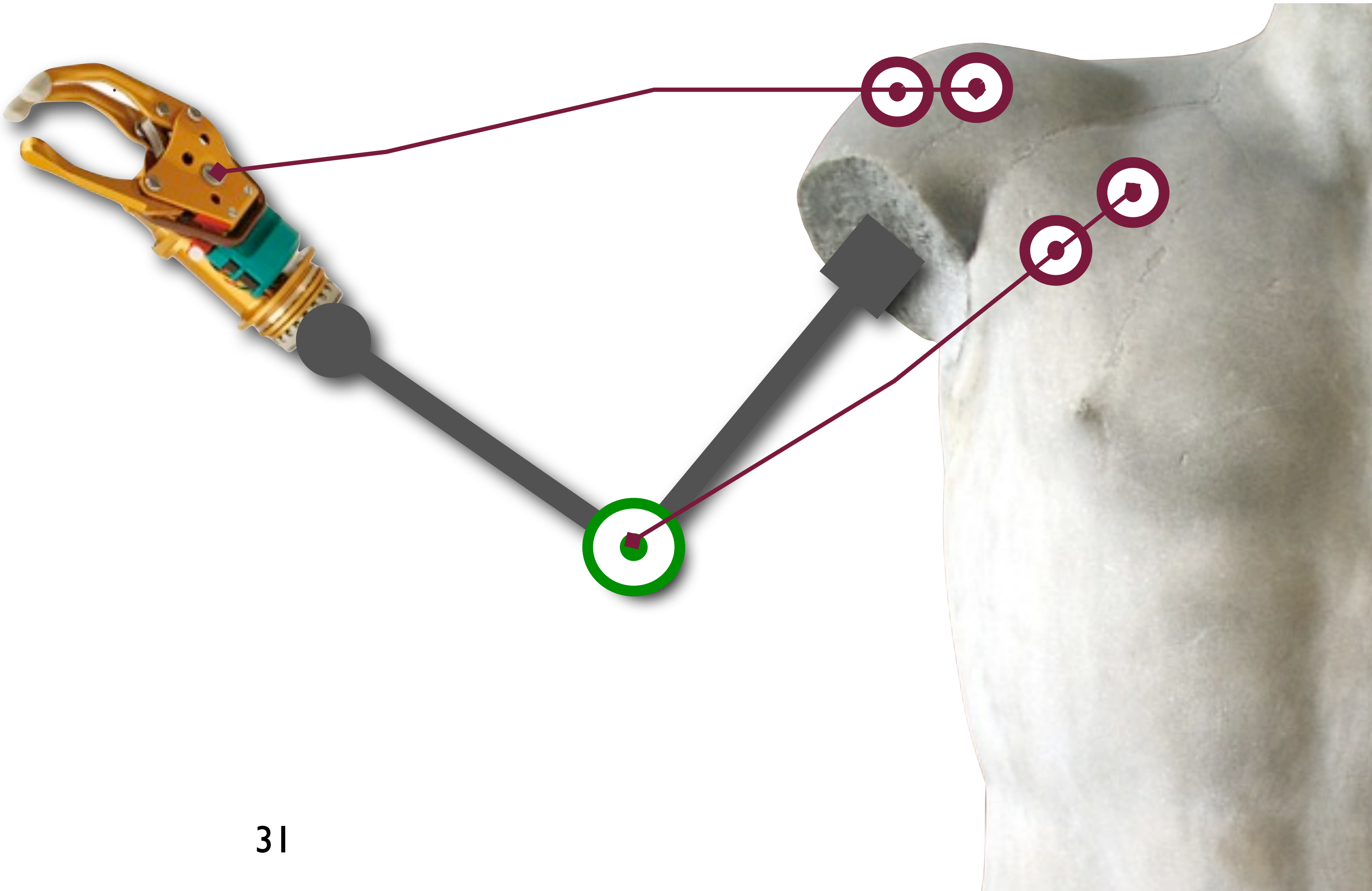
Other Interesting Predictions

- Assuming we continue as usual (on-policy):
  - What will the force sensor report over the next few seconds? *(Slippage/gripping.)*
  - Where will the limb be in the next 30s? *(Safety; fluid multi-joint motion.)*
  - How strong will each user EMG signal be in 250ms? *(User intent; preemptive motion.)*

** Address key issues, as per Scheme and Englehart, JRRD, 2011; Peerdeman et al., JRRD, 2011.*

Toward Complex Interfaces

Prostheses that approach and someday exceed the abilites of a biological limb.



MyoHand VariPlus Speed
Otto Bock

Toward Complex Interfaces

Prostheses that approach and someday exceed the abilites of a biological limb.

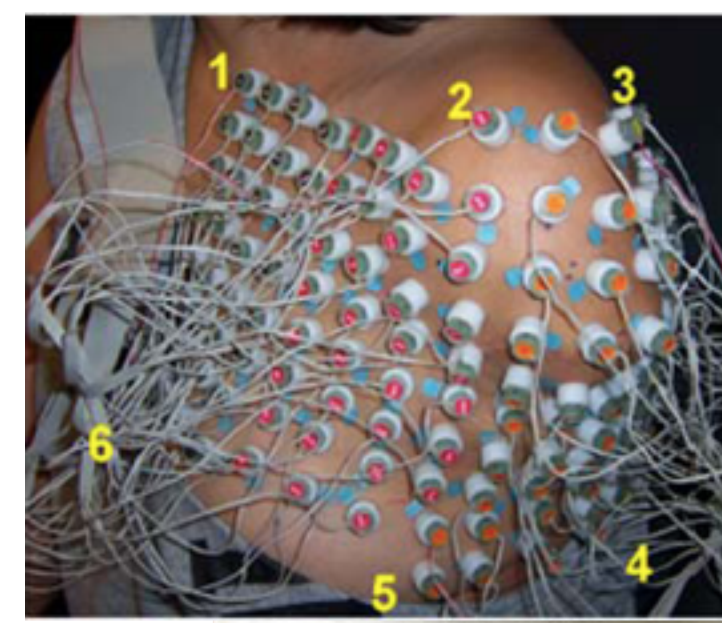
?

Modular Prosthetic Limb
Johns Hopkins University

Toward Complex Interfaces

Prostheses that approach and someday exceed the abilites of a biological limb.

?



Huang et al., *Ann. Biomed. Eng.*, 37:9 (2009).

Modular Prosthetic Limb
Johns Hopkins University

Three Important Elements

- **Real-time machine learning.**
  (Online, adaptive control algorithms; noted by Sensinger et al. '09, Scheme & Englehart '11)

- **Generalized interfaces.**
  (Blank-slate human-machine interaction & collaboration; e.g. Pilarski et al. 2011)

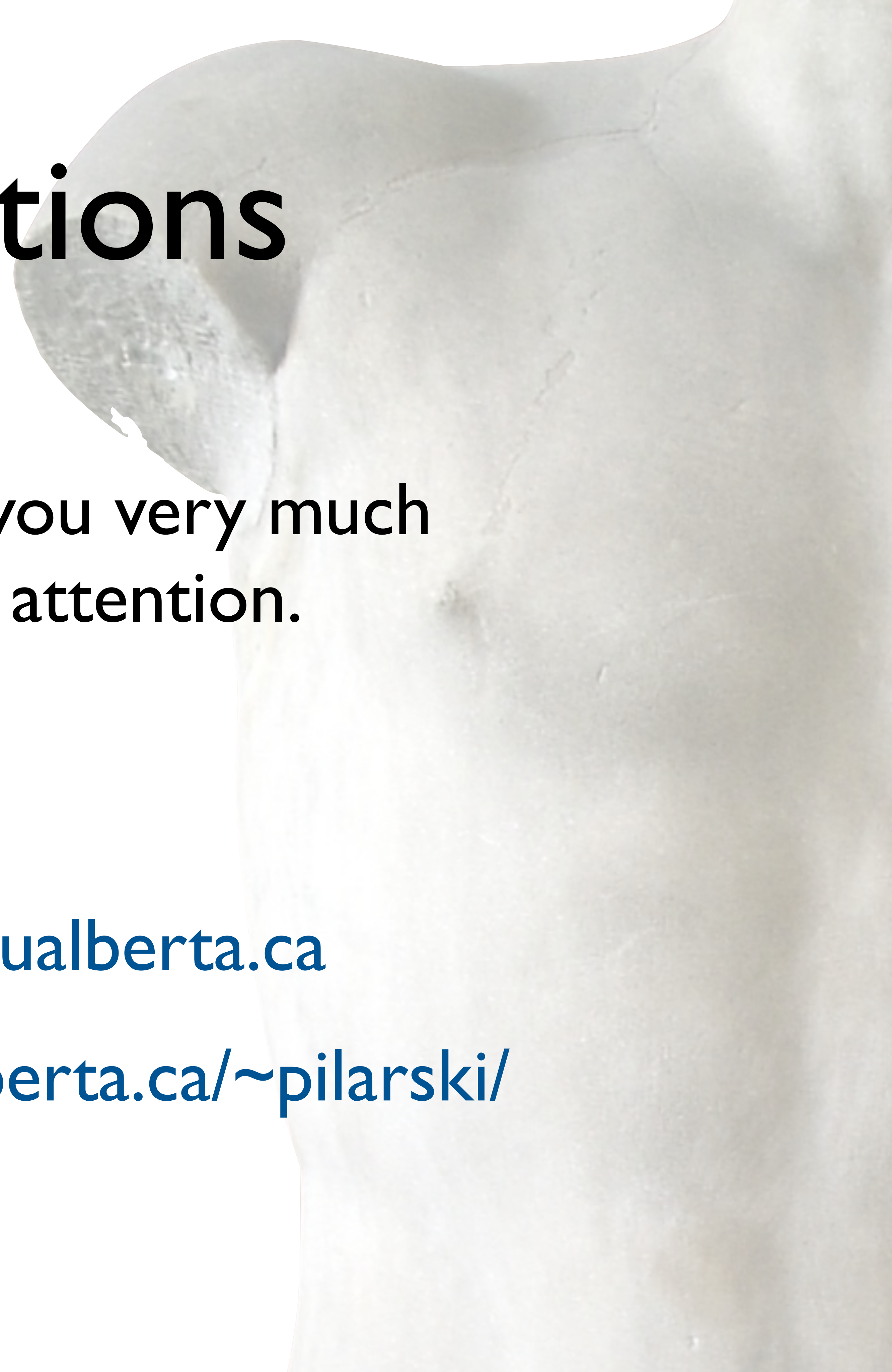
- **Data-respecting biomedical pattern analysis.**
  (Complexity is good: interpreting myriad signals without reducing the sensorimotor space)

Summary

- Real-time machine learning can help remove barriers to using assistive devices.
- Prediction and anticipation can be used to improve control of switchable artificial limbs.
- Results: on-policy nexting enables context-sensitive, adaptive switching (time savings).
- Big picture: artifical limbs that learn/improve through ongoing collaboration with a user.

- Dr. Richard S. Sutton, Dr. Thomas Degris
  RLAI, Dept. Computing Science, University of Alberta
- Michael R. Dawson, Dr. Jacqueline S. Hebert, Dr. K. Ming Chan
  Glenrose Rehabilitation Hospital & University of Alberta
- Dr. 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.



Questions

... and thank you very much
for your attention.

pilarski@ualberta.ca

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

Discussion Session

- Human training of RL agents.
- Foucs: maintaing policies when rewards are sporadic / reducing the need for constant reinforcement by the human.