Robot Learning from Multivalued Demonstration

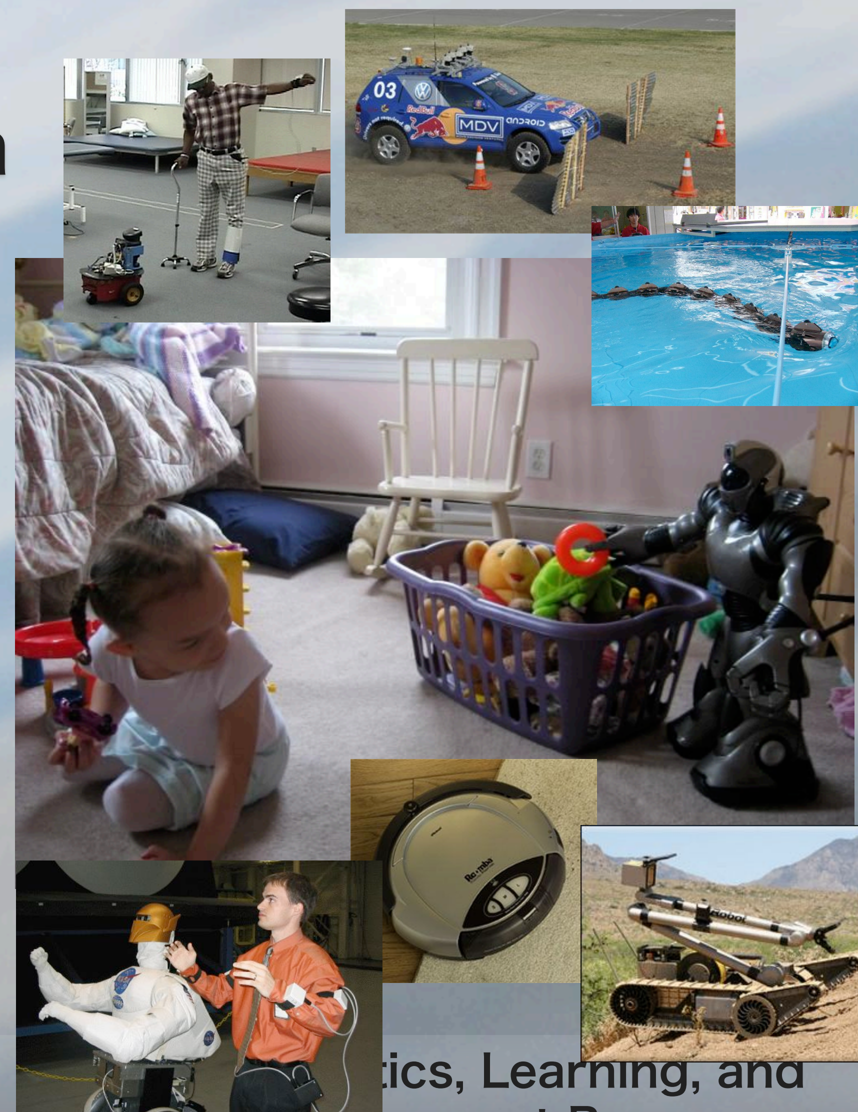
Chad Jenkins
Assistant Professor

Brown University
Computer Science

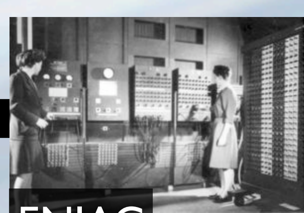
IROS 2008 Workshop on Robotics Challenges for Machine Learning
Sep 22, 2008
Nice, France

My Goal for Robotics

- Robot collaborators in human endeavors
    - Tools to enhance users' productivity
    - Become the path of least resistance
    - Critical path tasks?
    - Learning can broaden societal involvement



Personal Computing



ENIAC

Apple II

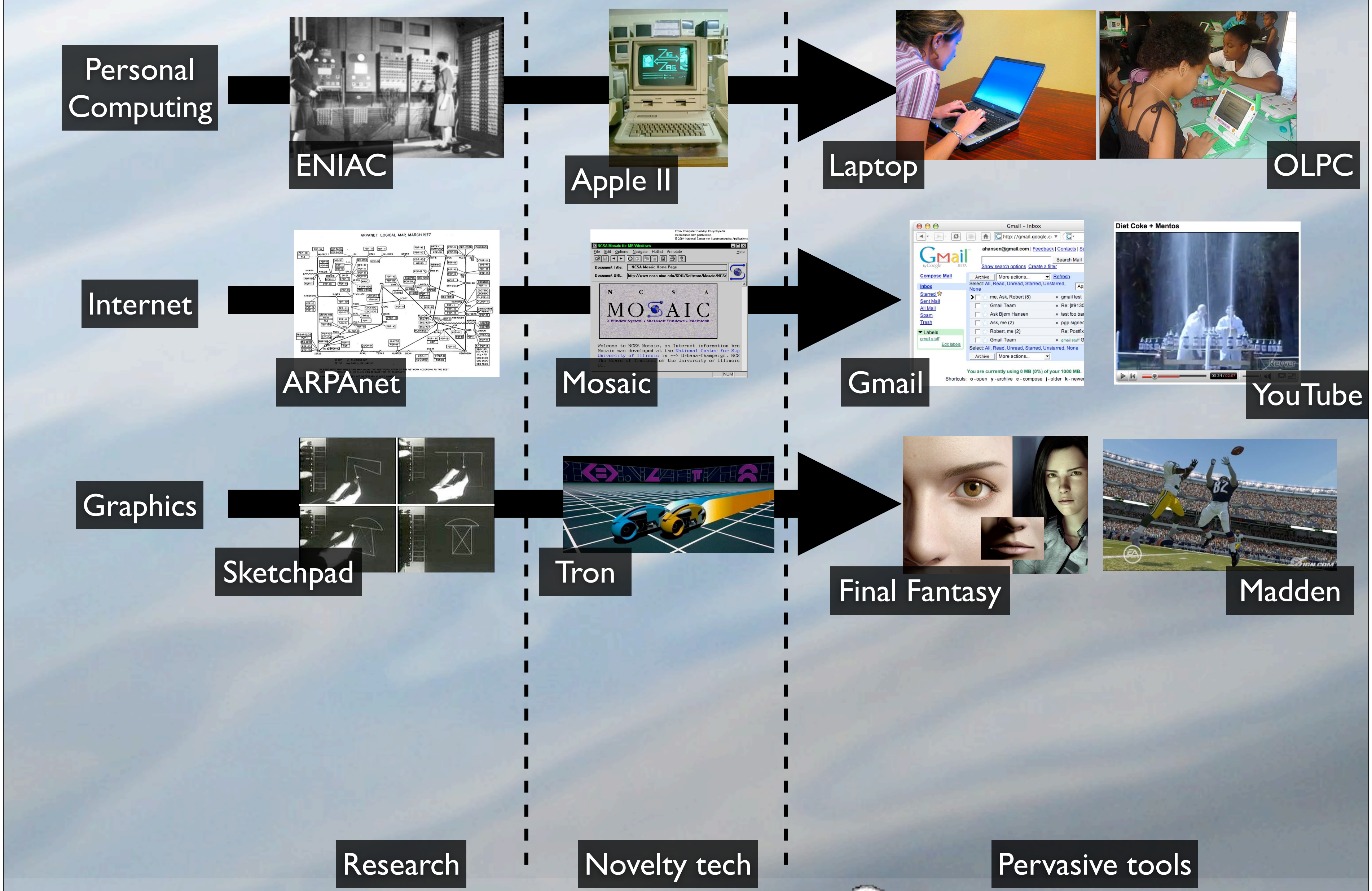
Laptop

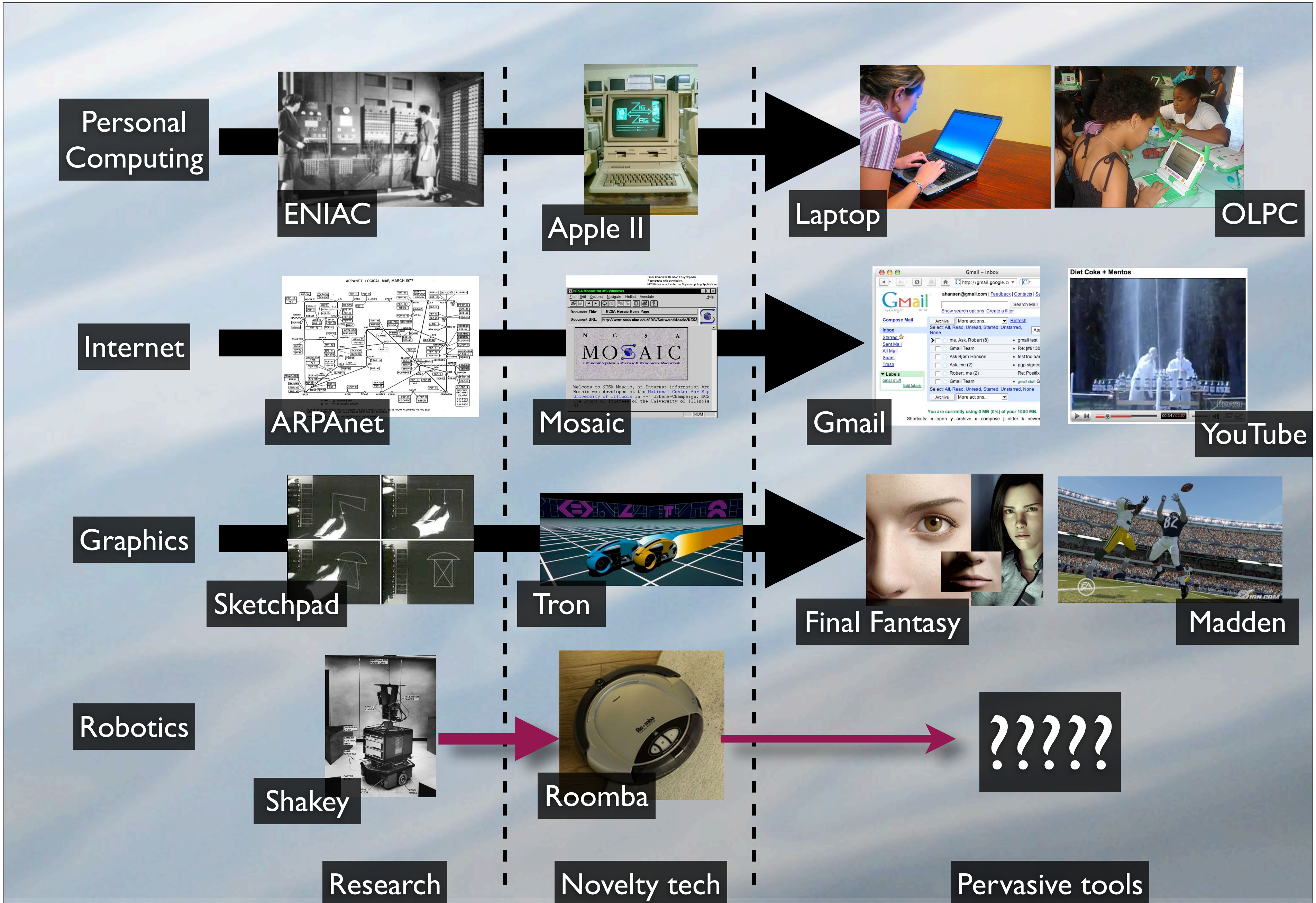
OLPC

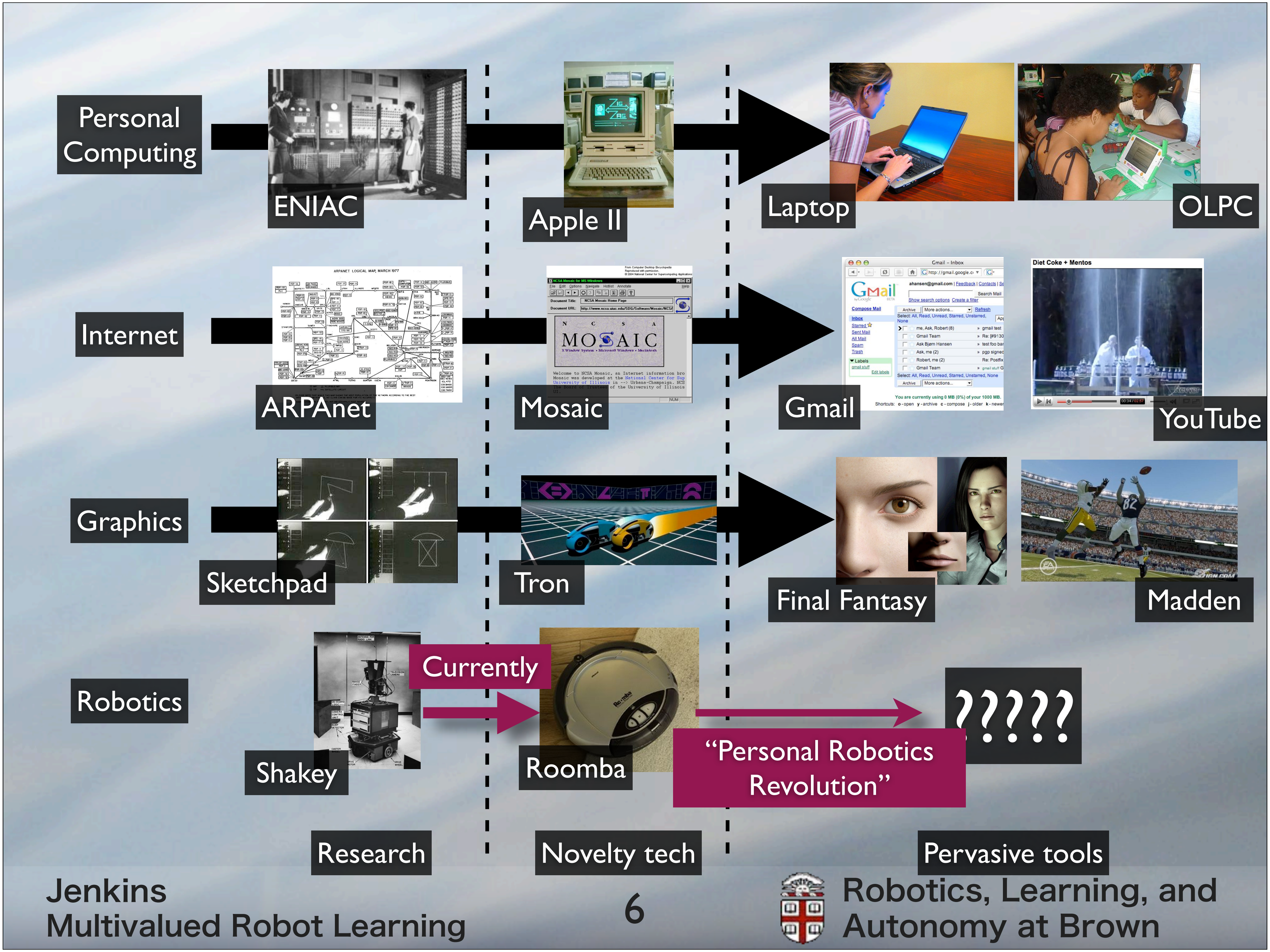
Research

Novelty tech

Pervasive tools


Personal Computing
ENIAC
Apple II
Laptop
OLPC
Internet
ARPAnet
Mosaic
Gmail
Diet Coke + Mentos
YouTube
Graphics
Sketchpad
Tron
Final Fantasy
Madden
Research
Novelty tech
Pervasive tools


Personal Computing
ENIAC
Apple II
Laptop
OLPC
Internet
ARPAnet
Mosaic
Gmail
YouTube
Graphics
Sketchpad
Tron
Final Fantasy
Madden
Robotics
Shakey
Roomba
?????
Research
Novelty tech
Pervasive tools


Personal Computing
ENIAC
Apple II
Laptop
OLPC
Internet
ARPAnet
Mosaic
Gmail
YouTube
Graphics
Sketchpad
Tron
Final Fantasy
Madden
Robotics
Shakey
Currently
Roomba
"Personal Robotics Revolution"
?????
Research
Novelty tech
Pervasive tools
Jenkins
Multivalued Robot Learning
6
Robotics, Learning, and Autonomy at Brown

Challenges for the Personal Robotics Revolution

- Programming Autonomous Robots?
  - developing policies that accord with user's intended behavior
- Manipulation?
  - achieving desired physical effects, uncertainty
- Communication?
  - coordination between many humans and robots
- Integration
  - reproducible evaluation, feedback loop w/usability

Where Learning Plays A Role

- Programming Autonomous Robots?
  - learn policies from demonstration (LfD) [Jenkins,Mataric 00-04] [Grollman,Jenkins 07-08]
- Manipulation?
  - manifold learning to form subspace priors [Jenkins,Peters 05-06] [Tsoli,Jenkins 07-08]
- Communication?
  - Multi-robot MRFs; [Butterfield, Jenkins, et al. 07-08]
  - Peer-to-peer HRI [Loper, Jenkins, et al. 07-08]
- Integration

Where Learning Plays A Role

Focus of this talk:
Multivalued functions in robot LfD

- Programming Autonomous Robots?
  - learn policies from demonstration (LfD)
    [Jenkins,Mataric 00-04] [Grollman,Jenkins 07-08]
- Manipulation?
  - manifold learning to form subspace priors
    [Jenkins,Peters 05-06] [Tsoli,Jenkins 07-08]
- Communication?
  - Multi-robot MRFs; [Butterfield, Jenkins, et al. 07-08]
  - Peer-to-peer HRI [Loper, Jenkins, et al. 07-08]
- Integration

Where Learning Plays A Role

Focus of this talk:
Multivalued functions in robot LfD

- Programming Autonomous Robots?
  - learn policies from demonstration (LfD)
    [Jenkins,Mataric 00-04] [Grollman,Jenkins 07-08]
- Manipulation?
  - manifold learning to form subspace priors
    [Jenkins,Peters 05-06] [Tsoli,Jenkins 07-08]
- Communication?

  Thursday at 13:30: "Neighborhood Denoising for Learning High-Dimensional Grasping Manifolds"

  - Multi-robot MRFs; [Butterfield, Jenkins, et al. 07-08]
  - Peer-to-peer HRI [Loper, Jenkins, et al. 07-08]
- Integration

Where Learning Plays A Role

Focus of this talk:
Multivalued functions in robot LfD

- Programming Autonomous Robots?
  - learn policies from demonstration (LfD) [Jenkins,Mataric 00-04] [Grollman,Jenkins 07-08]
- Manipulation?
  - manifold learning to form subspace priors [Jenkins,Peters 05-06] [Tsoli,Jenkins 07-08]

Thursday at 13:30: "Neighborhood Denoising for Learning High-Dimensional Grasping Manifolds"

- Communication?
  - Multi-robot MRFs; [Butterfield, Jenkins, et al. 07-08]
  - Peer-to-peer HRI [Loper, Jenkins, et al. 07-08]
- Integration

Wednesday during CPS session "MRFs for Cyber-Physical Systems"

Where Learning Plays A Role

Focus of this talk:
Multivalued functions in robot LfD

- **Programming Autonomous Robots?**
  - learn policies from demonstration (LfD)
    [Jenkins,Mataric 00-04] [Grollman,Jenkins 07-08]
- **Manipulation?**
  - manifold learning to form subspace priors
    [Jenkins,Peters 05-06] [Tsoli,Jenkins 07-08]

Thursday at 13:30: "Neighborhood Denoising for Learning High-Dimensional Grasping Manifolds"

- **Communication?**
  - Multi-robot MRFs; [Butterfield, Jenkins, et al. 07-08]
  - Peer-to-peer HRI [Loper, Jenkins, et al. 07-08]
- **Integration**

Wednesday during CPS session
"MRFs for Cyber-Physical Systems"

Learning from Demonstration

- Assume: true policy is a function $\pi(S) \rightarrow A$
  - maps states S to actions A, plus noise
- Training: dataset of state-action pairs (s_i, a_i)
  - instances of user policy $\pi(s_i) \rightarrow a_i$
  - latent task objective G (known to user)
- Estimate: policy $\hat{\pi}$ to perform task, meet G:
  - bootstrapped reinforcement learning
  - function approximation (regression)

[Mackay 98; Csato,Opper 02]

GP Regression Basics

$$C_{ij} = q(\mathbf{x}_i, \mathbf{x}_j) + \delta_{ij}\sigma_\nu^2$$

Compute similarity matrix *C*

$$q(\mathbf{x}_i, \mathbf{x}_j) = \exp \frac{||\mathbf{x}_i = \mathbf{x}_j||^2}{2\sigma_k^2}$$

Input similarity kernel

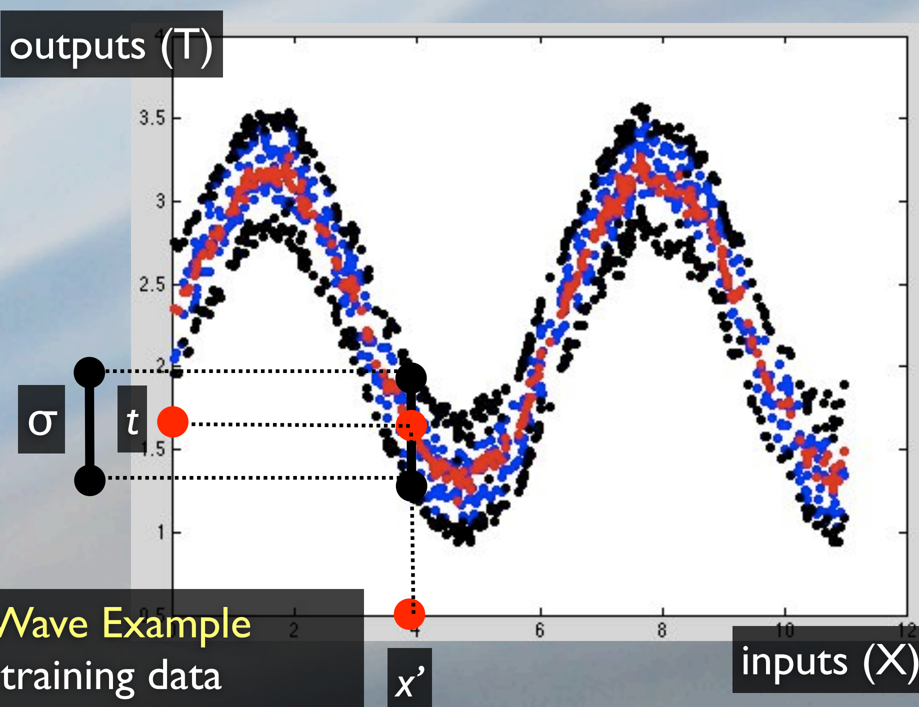
$$k_i = q(\mathbf{x}', \mathbf{x}_i)$$

$$\hat{\mathbf{t}} = \mathbf{k}^\top \mathbf{C}^{-1} \mathbf{T}$$

Predict output *t* from similarity matrix

$$\sigma_{\hat{\mathbf{t}}}^2 = (1 + \sigma_\nu^2) - \mathbf{k}^\top \mathbf{C}^{-1} \mathbf{k}$$

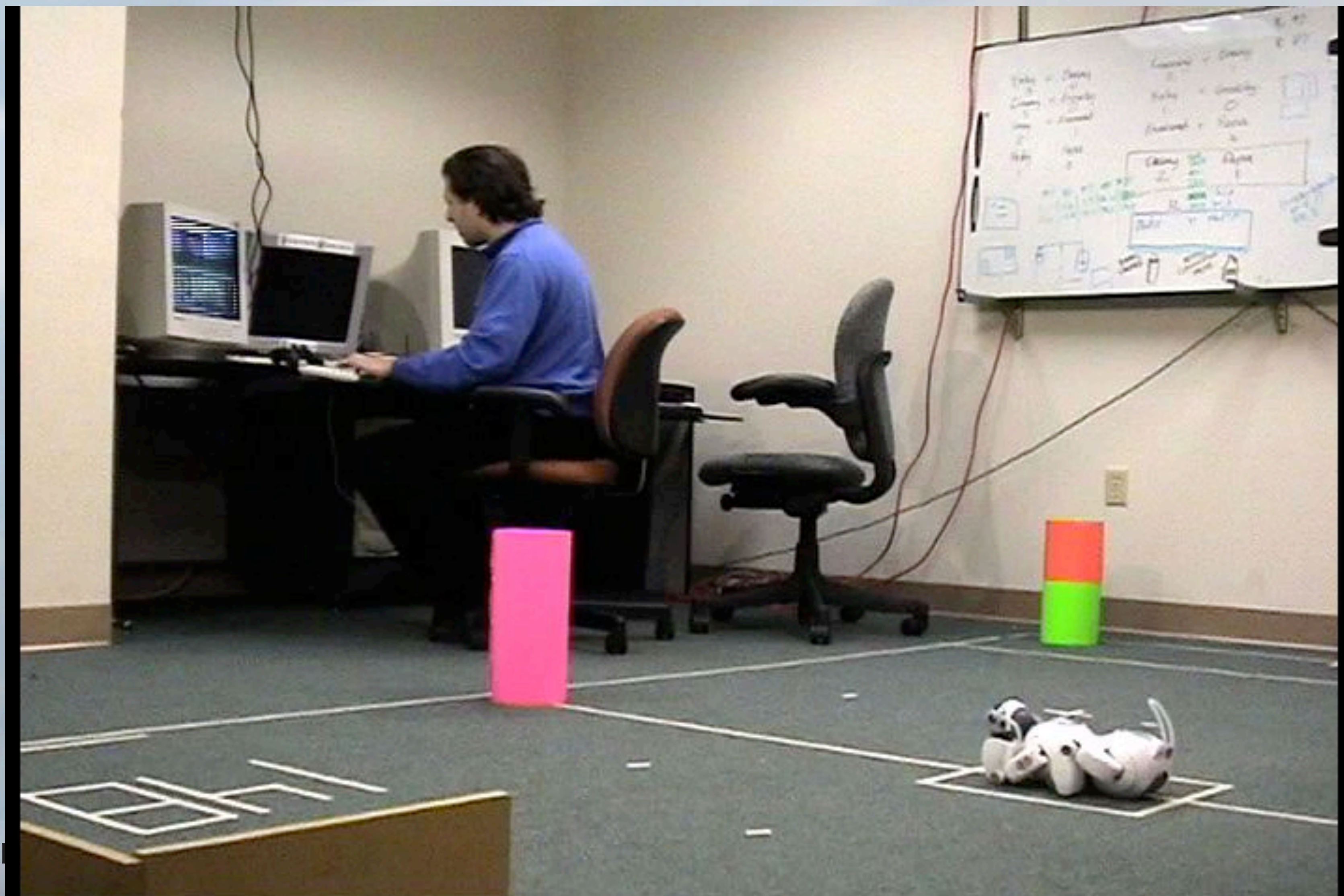
Compute confidence variance



Sine Wave Example
blue: training data
red: predicted outputs
black: confidence bounds

Object Tracking Example Using Gaussian Process Regression

[Grollman, Jenkins 07]



Conjecture

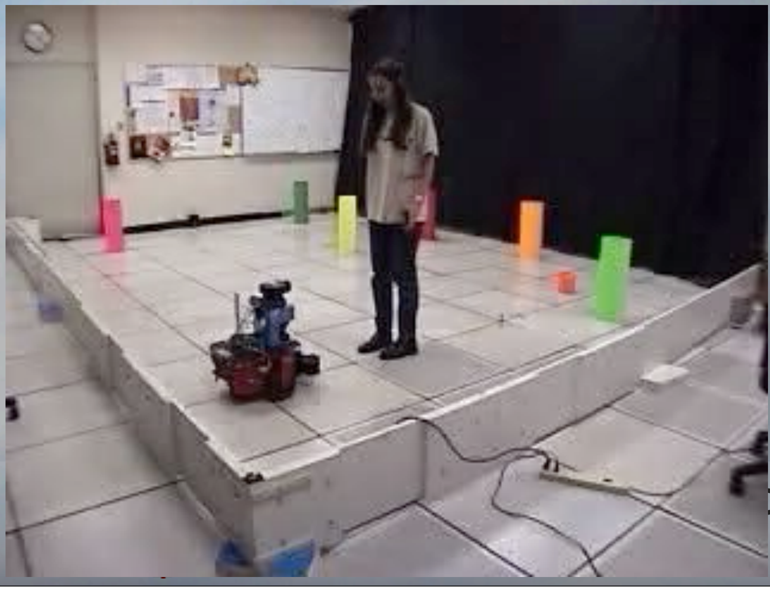
- Desired controllers $\not\Rightarrow$ math. functions
  - demonstrated control can be <u>multivalued</u>
    - single perceived state $\rightarrow$ multiple actions
- When would this occur?
  - training has multiple underlying policies
    - learn a FSM? or robot soccer?
  - ambiguous state spaces
    - "state space race"

Open Area? Multivalued LfD

- "Multivaluedness" partitions LfD work:
  - skill-level: learn policy for single objective
    - [Schaal,Atkeson 98; Abbeel,Ng 04; Smart,Kaelbling 02]

Open Area? Multivalued LfD

- "Multivaluedness" partitions LfD work:
  - skill-level: learn policy for single objective
    - [Schaal,Atkeson 98; Abbeel,Ng 04; Smart,Kaelbling 02]
  - task-level: given skills, learn transitions
    - [Nicolescu,Mataric 03; Bentivegna et al 01; Lockerd,Breazeal 06]



Multivalued Demonstration

- **Partition true world state** $S = \{S_O, S_U\}$
  - S_O **: robot observable state variables**
  - S_U **: unobservable by robot, maybe human**
    - **objectives, task context, other features**

- **Demonstration data:** (s_{Oi}, a_i) ; $\pi(s_{Oi}, s_{Ui}) \rightarrow A$

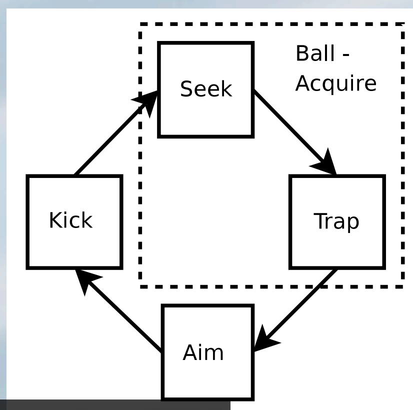
Multivalued Demonstration

- **Partition true world state** $S = \{S_O, S_U\}$
  - S_O **: robot observable state variables**
  - S_U **: unobservable by robot, maybe human**
    - **objectives, task context, other features**

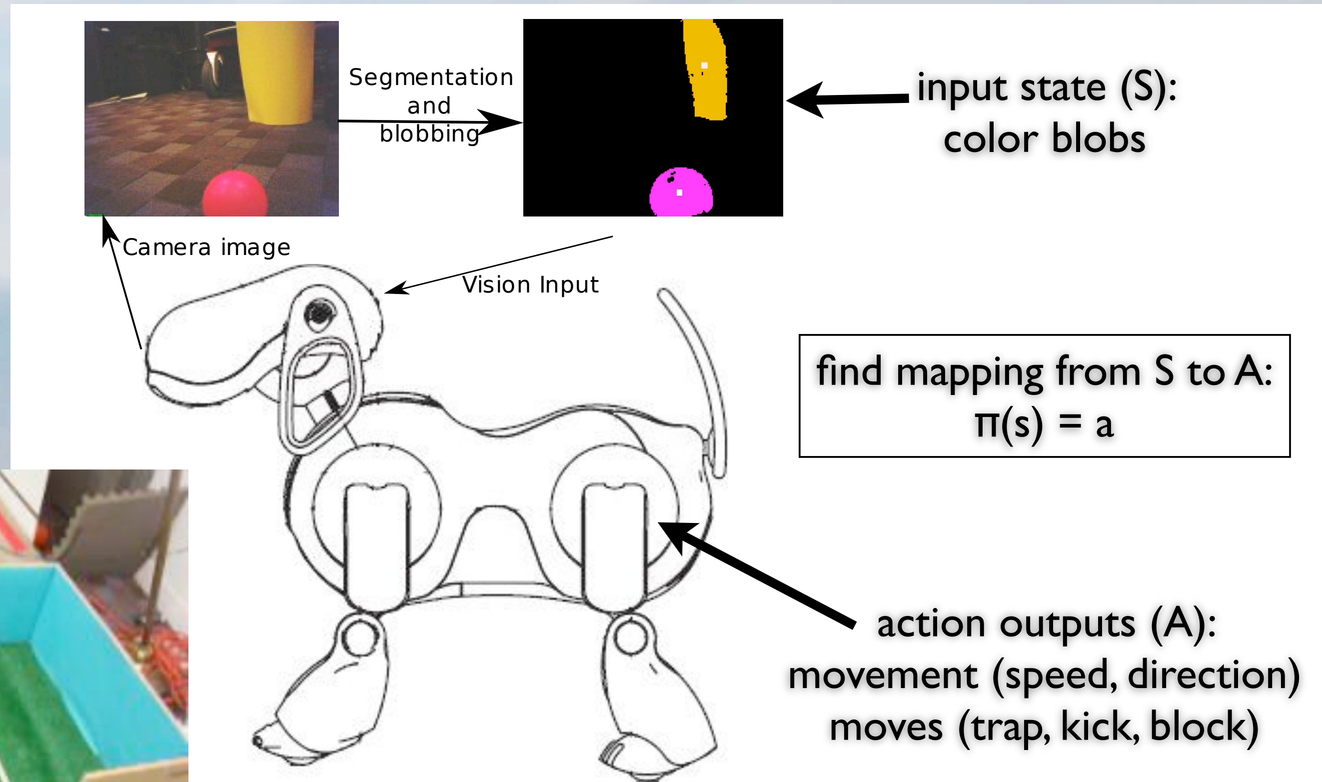
- **Demonstration data:** (s_{Oi}, a_i) ; $\pi(s_{Oi}, s_{Ui}) \rightarrow A$

- **If** $S_O \approx S$**, then** $p(a|s_O')$ **is unimodal distribution**
- **Otherwise,** $p(a|s_O')$ **potentially multimodal**

Can Robot Soccer Be Learned from Demonstration?



Attacker FSM



Sony AIBO

Basic robot soccer attack move

Internal Context

- Ball Acquire needs a context bit
  - during trap, ball goes out of view
  - similar states produce different actions
    - based on unobserved context
    - chest sensor?

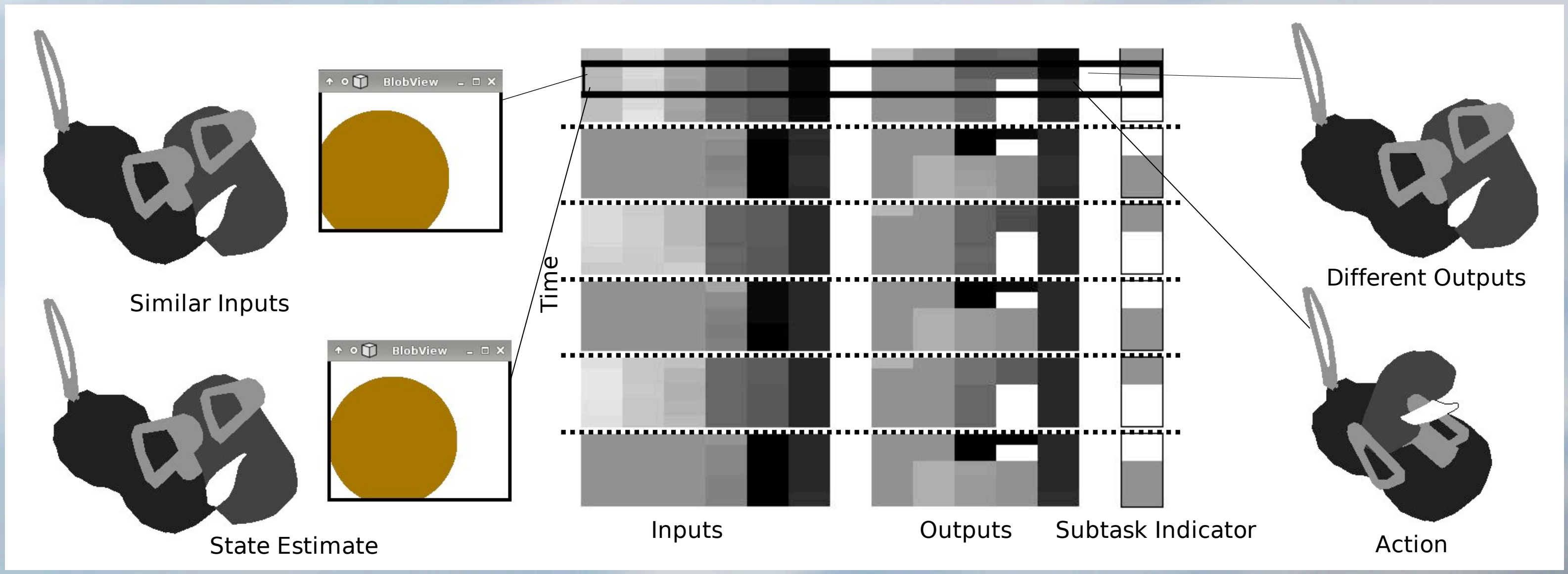
ball out of view:
seek or trap?

(a) Walk (b) Approach (c) Transition (d) Trap

action in seek context

action in trap context

Internal Context

- 

ball out of view: seek or trap?

(a) Walk (b) Approach (c) Transition (d) Trap

action in seek context

action in trap context

Multivalued LfD

- Our approach: extract multiple policies
  - Cause of multivalued data is unimportant
  - Only distinguishing modes of actions|state

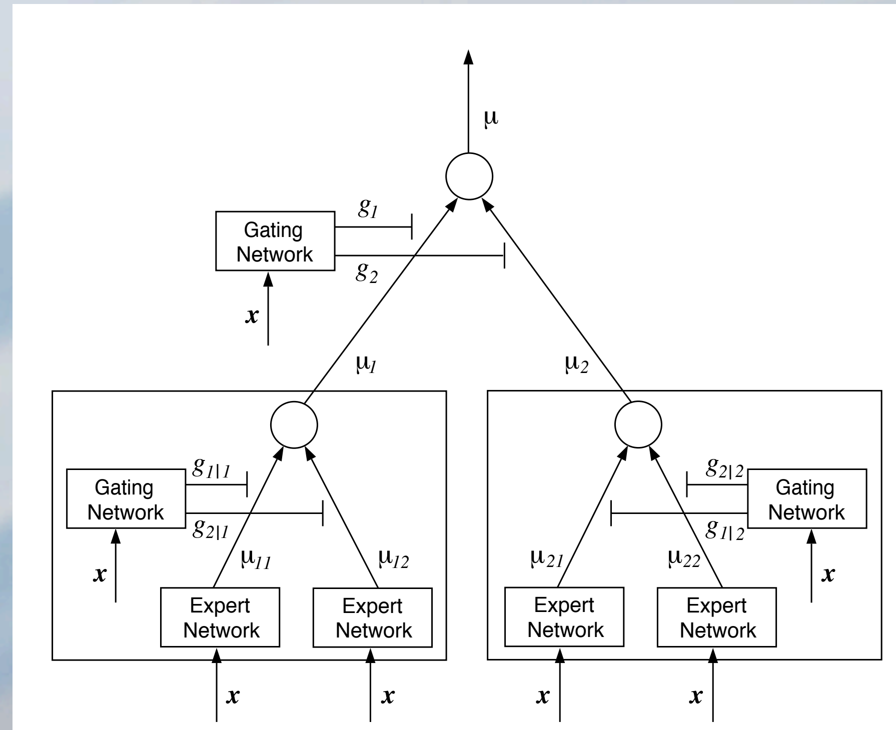
- Research roadmap
  - Uncover multiple policies as “experts”
    - Infinite “multimap” regression [Grollman poster]
  - Estimate FSMs over experts [Nicolescu,Mataric 03]
  - Optimize individual experts [Abbeel, Ng 04]

[Wood, Grollman et al. 08]

Multimap Regression

- Learn data gating function partition
  - based on input-output likelihood
- Expert learns policy for each gate
  - Sparse Online Gaussian Processes

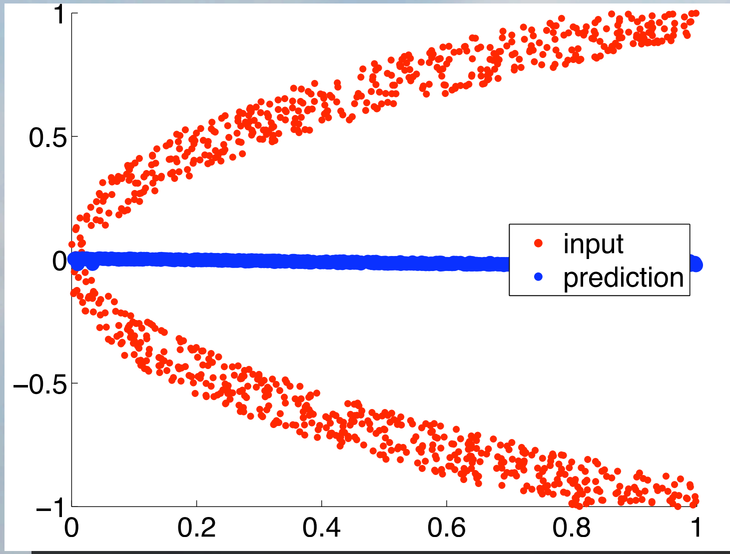
[Csato,Opper 02]



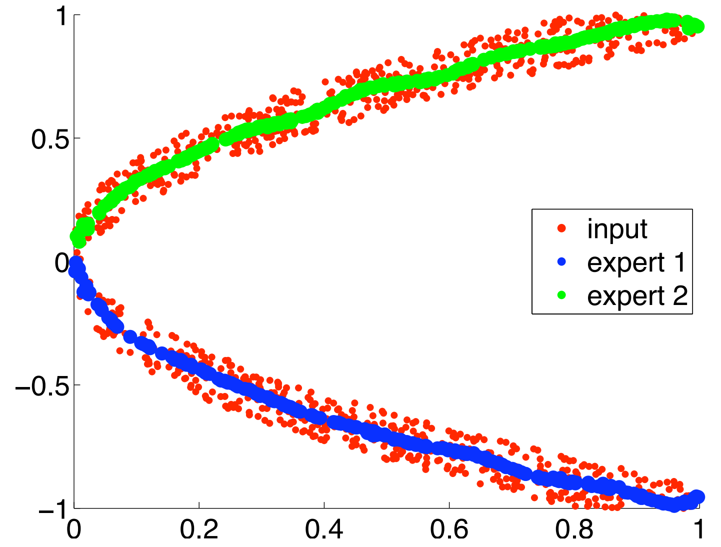
Hierarchical Mixtures of Experts
[Jordan, Jacobs 94]:
Experts learn pieces of function
Gates blend these pieces together

Square Root Example

- Consider y = sqrt(x)
  - averaging outputs will be incorrect
  - 2 regressors needed for pos. and neg.



Locally Weighted Projection Regression



Multimap Regression

Preliminary Results

- Multimap regression on Ball Acquire data
  - (+) Data partitions into "seek" & "trap"
  - (-) Both experts do both seek and trap
  - (+) Experts perform correctly from initialized partition

Gate Partitioning of Ball Acquire Training Data

True Partition

Multimap Partition

Summary

- Learning from demonstration as a means to program autonomous robots
- Multivalued demonstration data
  - perceived state maps to multiple actions
- Infinite mixtures of experts regression
  - promising results for learning robot soccer

Acknowledgments

Funding:

ONR YIP

ONR PECASE

NSF IIS

Brown Salomon

Dan Grollman

Brendan Dickinson

Micah Lapping-Carr

Mark Moseley

Dan Byers

Jessica Chermayeff

Frank Wood

Katherine Heller