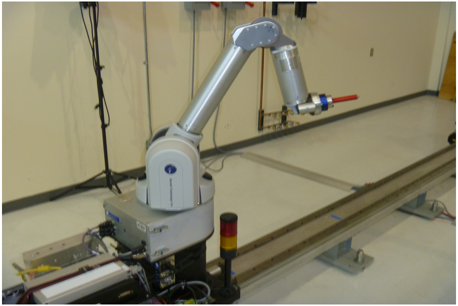
Implementation of an Embodied General Reinforcement Learner on a Serial Link Manipulator

Nicholas Malone[1] Brandon Rohrer[2] Lydia Tapia[1] Ron Lumia[1] John Wood[1]

[1]University of New Mexico

[2]Sandia National Laboratories



UNM

Sandia National Laboratories

Problem

- Every robot is different
  - Path planner must be tailord to the robot
  - Structures must de designed specificially
- Our goal
  - Automate the process
  - Apply reinforcement learning
  - Address RL scalability

Proposed Solution

- BECCA (a Brain Emulating Cognition and Control Architecture)
  - Optimistically based on biological brains
  - Feature creator with reinforcement learner
- PRMs for handeling scaling issues

Robotic Platform

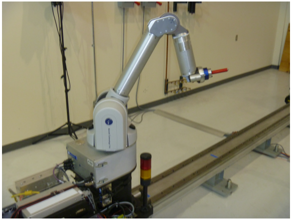


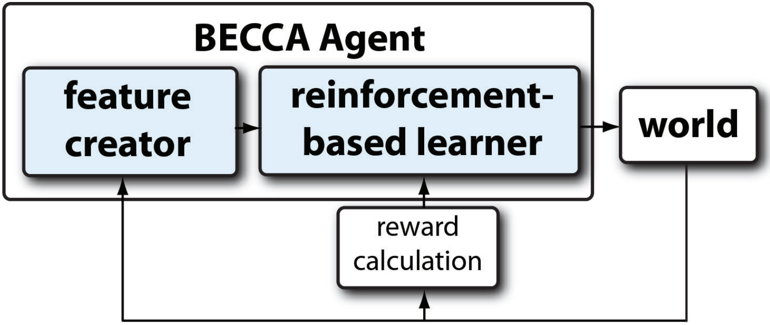
Figure: The WAM platform

PIC 1

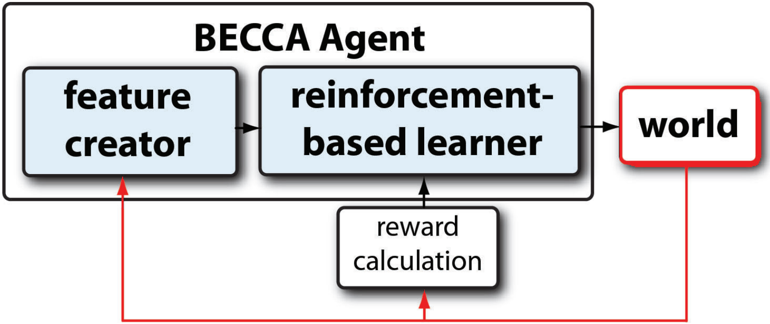
Figure: FIXME placeholder for the arch of the WAM

- 7-DoF robotic arm
- Cable driven
- Joint position encoders

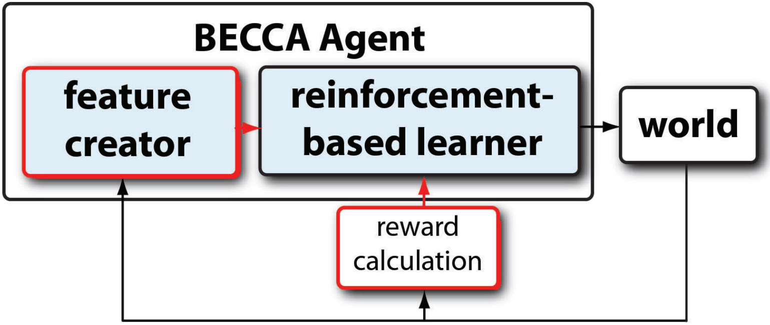
BECCA


BECCA Agent
feature creator
reinforcement-based learner
world
reward calculation

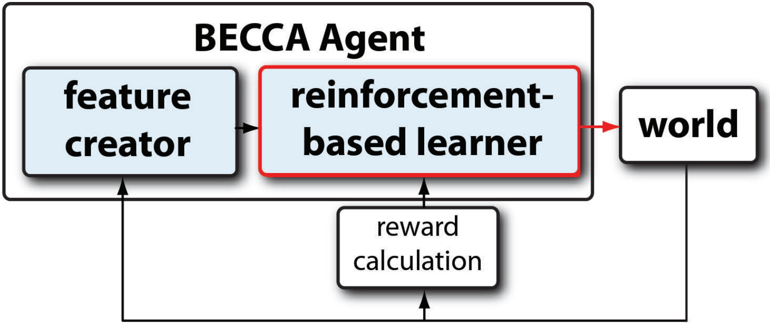
BECCA


BECCA Agent
feature creator
reinforcement-based learner
world
reward calculation

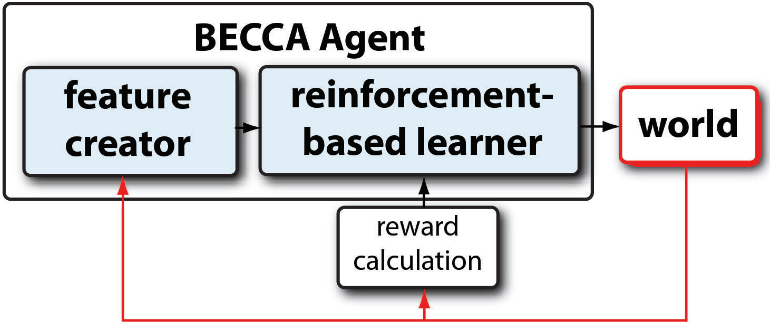
BECCA


BECCA Agent
feature creator
reinforcement-based learner
world
reward calculation

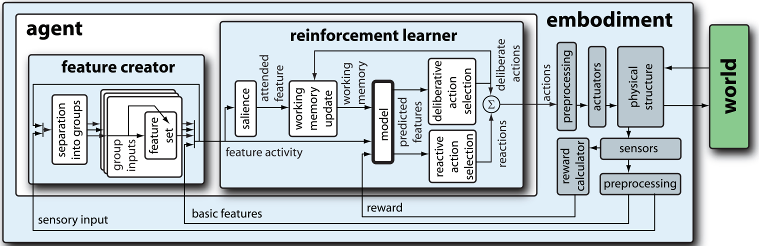
BECCA


BECCA Agent
feature creator
reinforcement-based learner
world
reward calculation

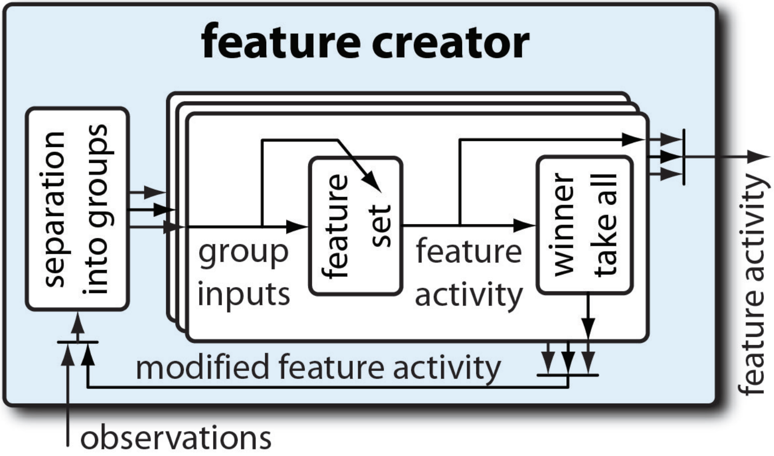
BECCA


BECCA Agent
feature creator
reinforcement-based learner
world
reward calculation

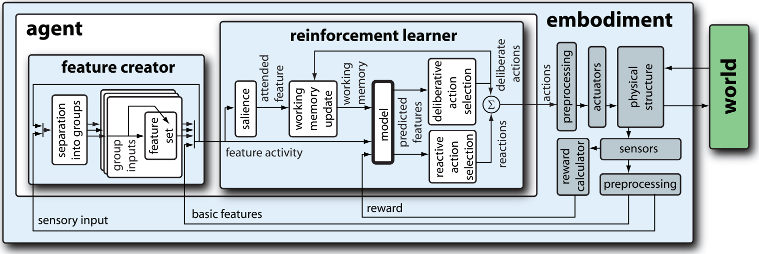
Feature Creator



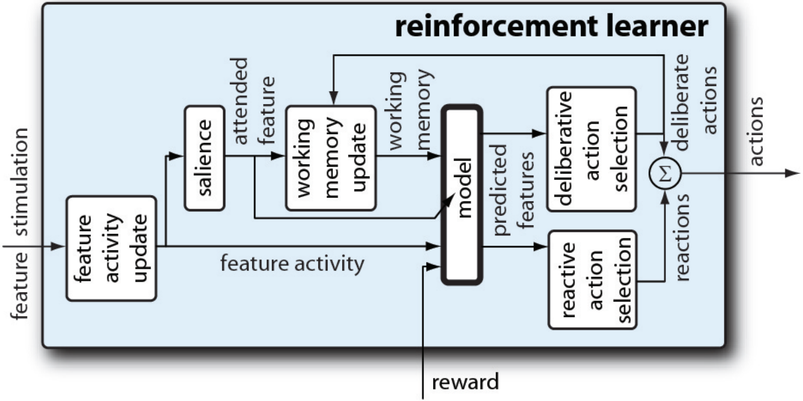
Feature Creator



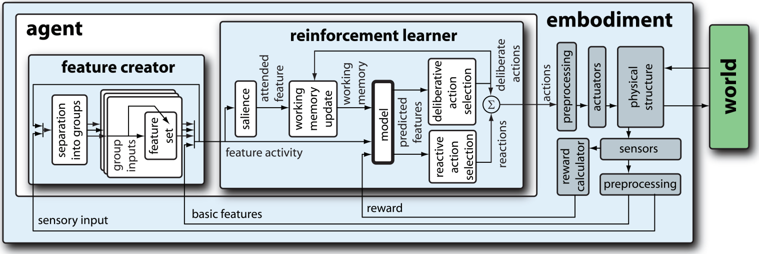
Reinforcement Learner



Reinforcement Learner

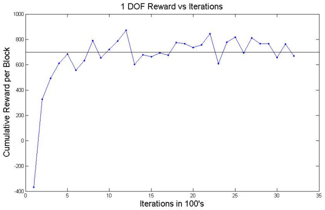


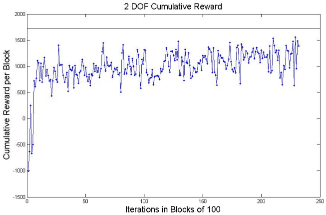
BECCA


agent
feature creator
separation into groups
group inputs
feature set
reinforcement learner
salience
attended feature
working memory update
working memory
model
predicted features
deliberative action selection
reactive action selection
deliberate actions
reactions
feature activity
embodiment
actions
preprocessing
actuators
physical structure
sensors
reward calculator
preprocessing
world
sensory input
basic features
reward

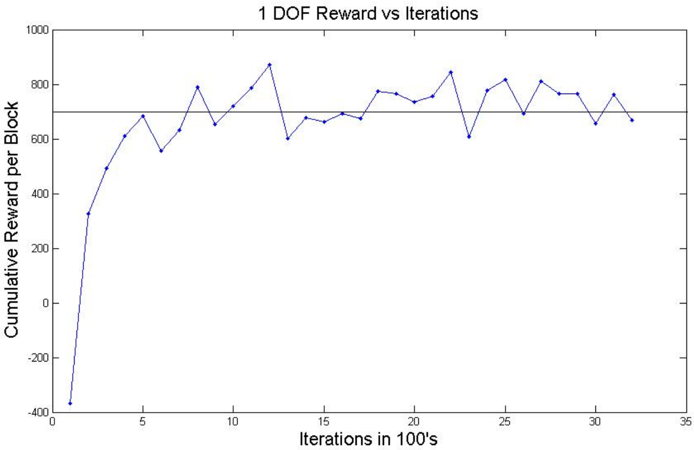
Baseline Tasks

- 1-DoF
  - 1 joint partitioned into 10 bins
  - Rewarded for being in the middle bin
- 2-DoF
  - 2 joints are partitioned into 10 bins each
  - Rewarded +10 for being in the middle bin of each joint
- 3-DoF
  - 3 joints are partitioned into 3 bins each
  - Rewarded +10 for being in the middle bin of each joint

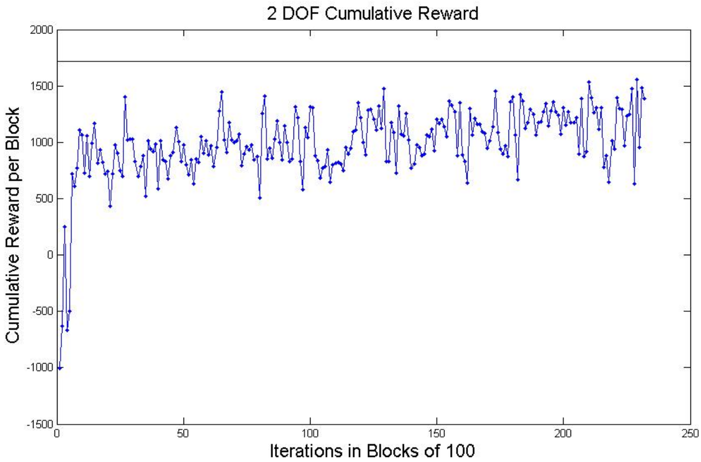




Baseline Tasks



Baseline Tasks



Baseline Tasks

3-DoF graph goes here

PRM

- Configuration Space (C-Space)
  - n DoFs results in n axes
  - A configuration is a point in the n dimensional space
- Roadmap Construction
  - Randomly sample configurations (vertices) in C-Space
  - Connect pairs of configurations (to form edges)
  - Roadmaps approximate the topology of C-Space

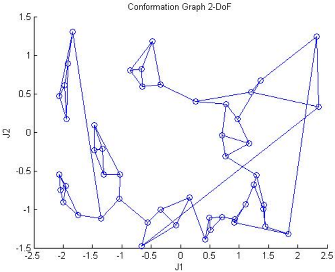
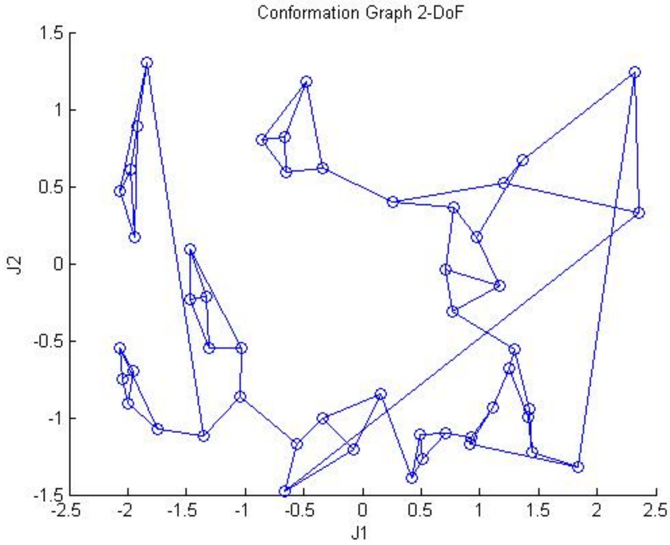
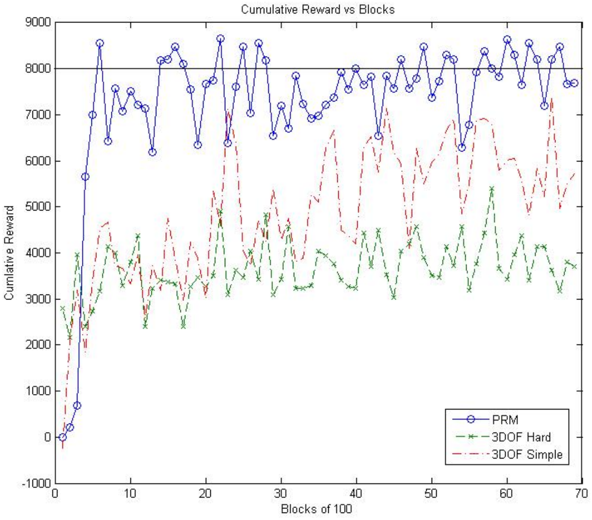


Figure: 2-DoF Roadmap

PRM


Conformation Graph 2-DoF
1.5
1
0.5
0
-0.5
-1
-1.5
J2
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
J1

PRM Tasks


Cumulative Reward vs Blocks
Cumlative Reward
Blocks of 100
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
0
10
20
30
40
50
60
70
PRM
3DOF Hard
3DOF Simple

PRM Advantage

State Space

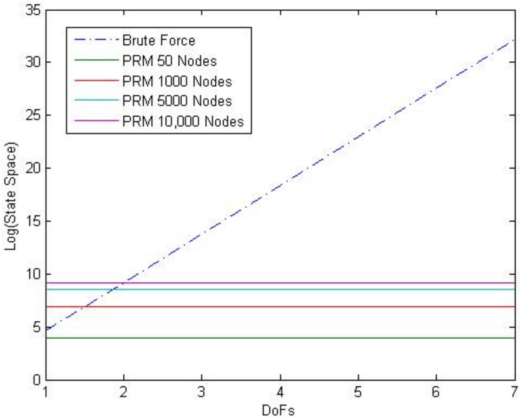
$$|States| \ * \ |Actions|$$

Task	States	Actions	State Space
1-DoF	10	10	100
2-DoF	100	100	10000
3-DoF	1000	1000	1000000
*PRM**	50	4	200

* for the experiments

PRM Advantage

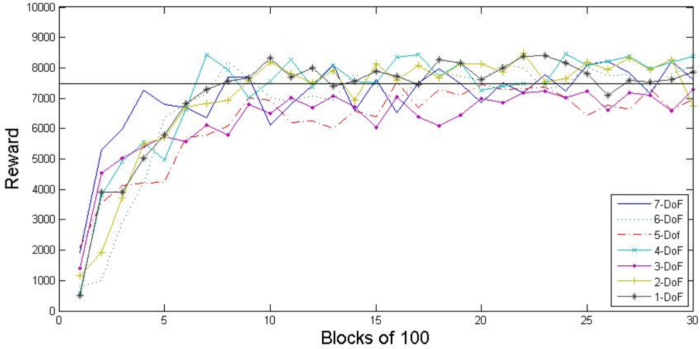
PRMs approximate the topology of the high dimensional state space.



note: k = 4 for all PRMs shown

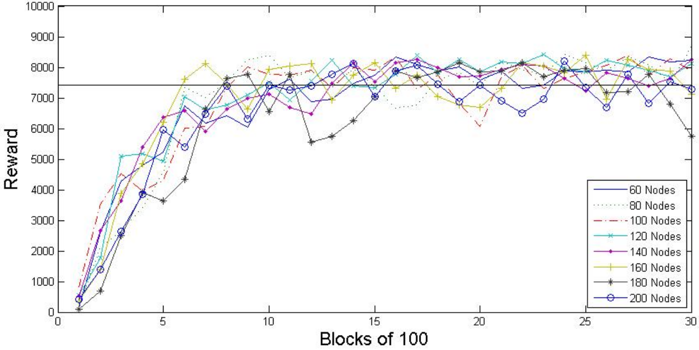
Scalability

Fixed roadmap size with variable DoFs



Roadmap Size

Fixed DoF's with variable roadmap size



Conclusion

- Trained in simulation
- Run on real robots after training
- PRM BECCA is much more scalable than a brute force approach
- PRM's approximate the topology of the state space
- Scalability is one of the hardest challenges for Reinforcement Learning, and machine learning in general

Acknowledgments

This work was supported by Sandia National Laboratories PO# 1074659. Tapia supported in part by the National Institutes of Health (NIH) Grant P20RR018754 to the Center for Evolutionary and Theoretical Immunology. We also thank Dr. Dave Vick for his help with the robotic hardware setup.

UNM

Sandia National Laboratories