TEXAS INSTITUTE FOR INTELLIGENT BIO-NANO MATERIALS AND STRUCTURES FOR AEROSPACE VEHICLES

A NASA University Research, Engineering and Technology Institute (URETI)

Reinforcement Learning for Active Length Control of Shape Memory Alloys

Kenton Kirkpatrick

John Valasek

Aerospace Engineering Department

Texas A&M University

Veritas Fugae

AIAA GNC Conference

21 August 2008

Honolulu, HI

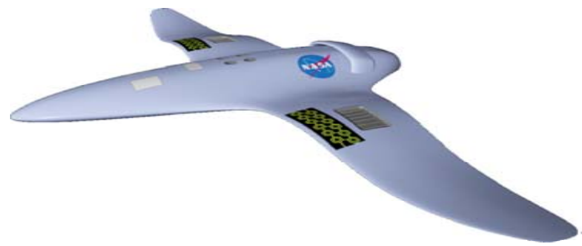
- Funded by the National Aeronautics and Space Administration
- Administered by the Texas Engineering Experiement Station
- A collaborative effort among: Prairie View A&M University | Rice University | Texas A&M University | | Texas Southern University | University of Houston | University of Texas - Arlington

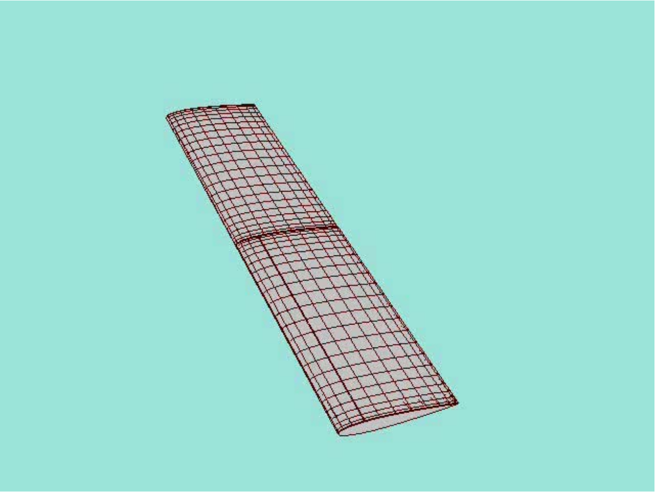
Basic Overview

- Motivation and Basics
- Learning
- Simulated Characterization Results
- Experimental Characterization Results
- Control Policy Learning Results
- Conclusions
- Challenges and Open Problems

Motivation & Basics

Morphing Aircraft





Why should it have morphing capabilities?

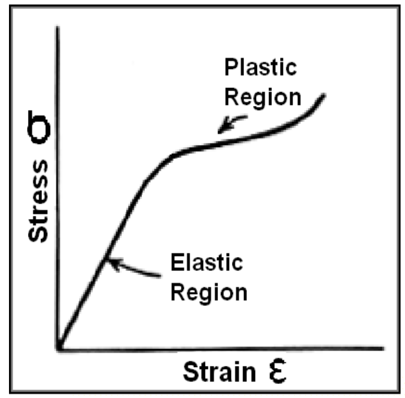
- Improved efficiency at multiple flight conditions
- Wider range of possible maneuvers

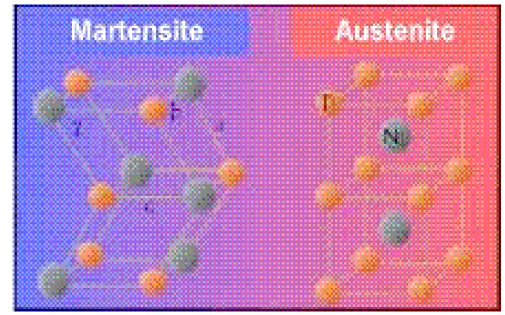
How would it physically achieve morphing?

- SMA Actuators

Shape Memory Alloys (SMA)

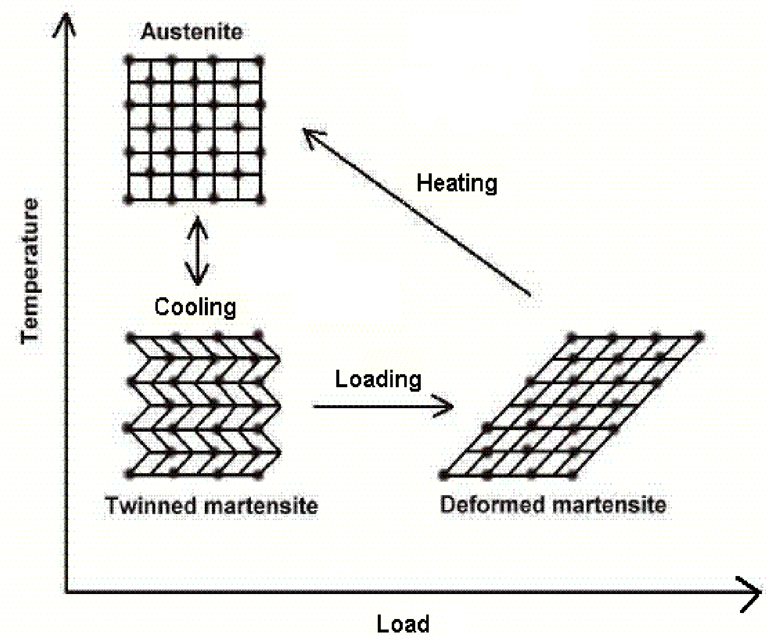
- Metallic Alloys used as actuators
- Shape Memory Effect – SMA can fully recover from a plastically deformed shape change by the addition of heat
- Electricity can be used to induce a cycle of heating, cooling, and deformation in an SMA in order to execute a dynamic task





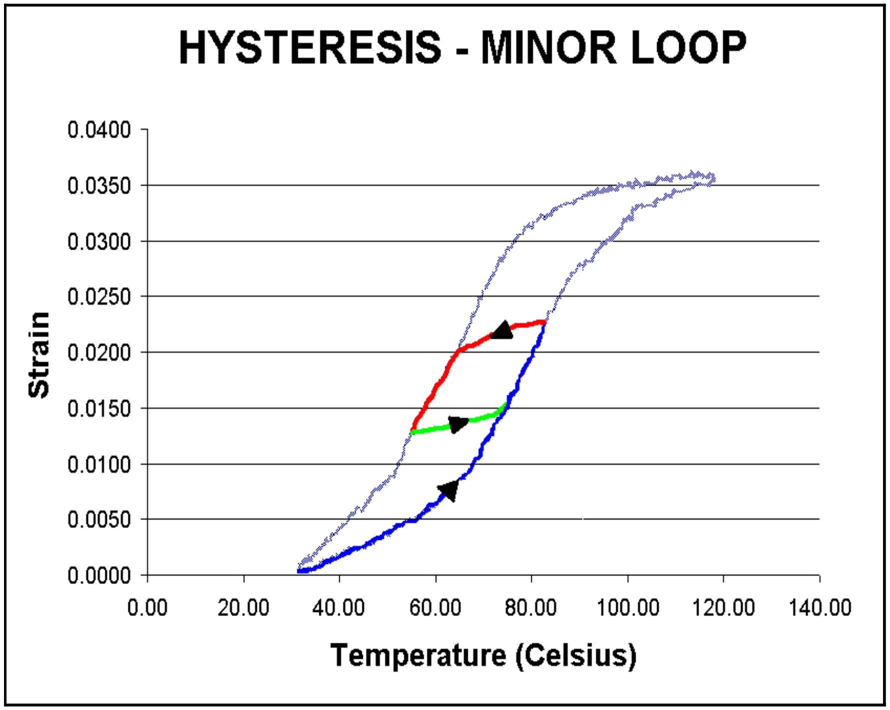
- PROBLEM:

  Efficiently characterizing and controlling SMA behavior



Why Is Characterizing SMAs Such A Challenge?

1. Uncertain Model Parameters
2. Temperature–Strain Relationship: *HYSTERESIS*
   - Behavior based on a <u>2-D FIELD</u>, not a 2-D PATH



SMA Characterization Methodologies

CURRENT METHOD: Experimentation and Testing to approximate constitutive model

- Mainly based on material property parameters OR system identification
  - Approximate Models:
    - Neglect hysteresis
    - Arrange antagonist SMAs to “cancel” the hysteretic effect
    - Modified plasticity model composed of averaged thermal effects
    - Disregard coupling of hysteretic and structural response
    - Other averaging techniques

What is Needed from SMA Learning?

1. Characterization
   - Discovering how strain in a SMA wire is directly related to temperature during crystal phase transformation
     - Non-parametric

2. Control Policy
   - Input-output mapping between desired length and applied voltage
   - Essential element of a feedback control law

<u>Novelty of our approach</u>: ability to both Characterize and learn the Control Policy by the same method

Research Objective

- APPROACH
  - Using the Sarsa algorithm to characterize and control 1-D SMAs in real-time without human supervision
- KEY ISSUE
  - Need to characterize in temperature/strain space, but voltage needed for control
- SOLUTION
  - For control policy, learn directly in voltage/strain space
- BENEFIT
  - Higher accuracy and easier learning
  - Able to characterize and control in system interior and on boundary

Learning

Reinforcement Learning

- Does not require any prior knowledge.
  - Knowledge is based on experience and interaction with the environment, not on input-output data supplied by an external supervisor

- Achieves a specific **goal** by learning from interactions with the environment.
  - Considers **state** information (s)
  - Performs sequences of **actions**, (a), observing the consequences
  - Attempts to maximize **rewards** (r) over time
    - **These specify what is to be achieved,** ***not how to achieve it***

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots \sum_{i=0}^{\infty} \gamma^i r_{t+i}$$

  - Constructs a **state value function** (Q)
    - Learns an optimal control **policy**

$$\pi^* = \arg\max_a [r(s,a) + \gamma V^*(\delta(s,a))]$$

- Memory is contained in the state value function

RL Algorithm: Sarsa

$$Q(S,A,G) \leftarrow Q(S,A,G) + \alpha[R(S',A',G) + \gamma Q(S',A',G) - Q(S,A,G)]$$

3-D Control Policy Matrix

- States (*S*): Temperature and Strain Dependent
- Actions (*A*): Change Temperature (Voltage Application)
- Goal (*G*): Desired Strain
- α: Repetition Penalty
- γ: Future Policy Weight

Action Choice Method: ε-Greedy

- Explore or Exploit: Dependent upon ε (which varies with Episodes)

Why Not Q-Learning?

- Less chance of wire damage with Sarsa

SMA Characterization Results

a) Simulation

b) Experiment

Simulation Model

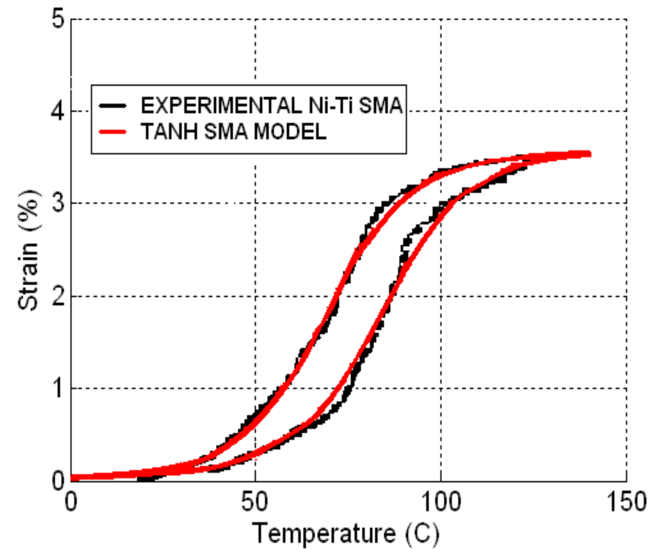
- **Temperature-Strain Relation:**
  - Hyperbolic Tangent Model

$$\varepsilon(i) = \frac{1}{2} h \tanh\left((T(i) - ctl)a\right) + s\left(T(i) - \frac{1}{2}ctl - \frac{1}{2}ctr\right) + \frac{1}{2}h + cs$$

- **Voltage-Temperature Relation:**

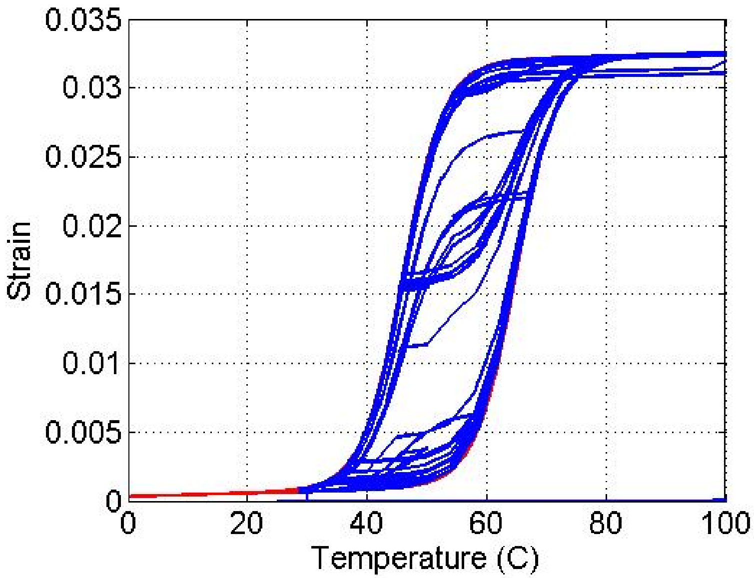
$$\frac{dT}{dt} = \frac{\frac{V^2}{R} - h\,\pi\,D\,L(T - T_\infty)}{m\,C_P}$$

- Learns input-output data (how to apply voltage to achieve a particular position state), not the constitutive model of an SMA

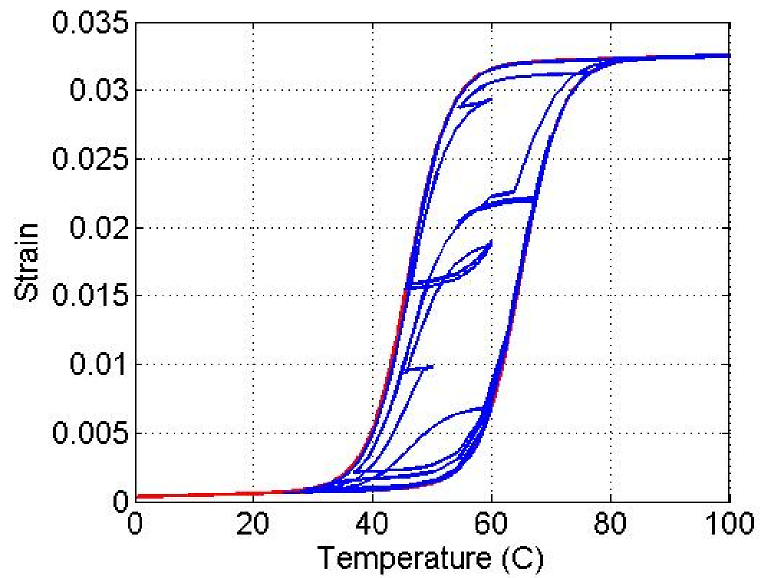


Simulation Results

After 10000 Episodes



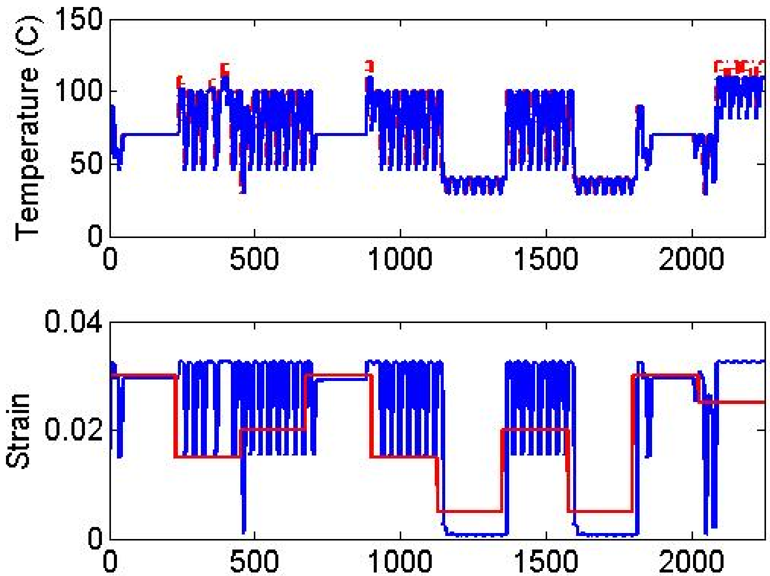
After 50000 Episodes



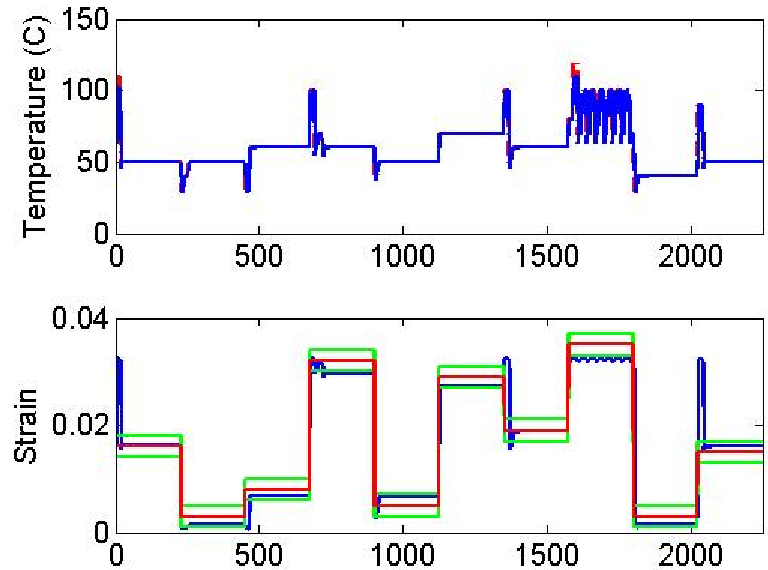
- **Goal States: Random Strain Goals**
- **Error range: ± 0.2% Strain**

Simulation Results

After 10000 Episodes



After 50000 Episodes



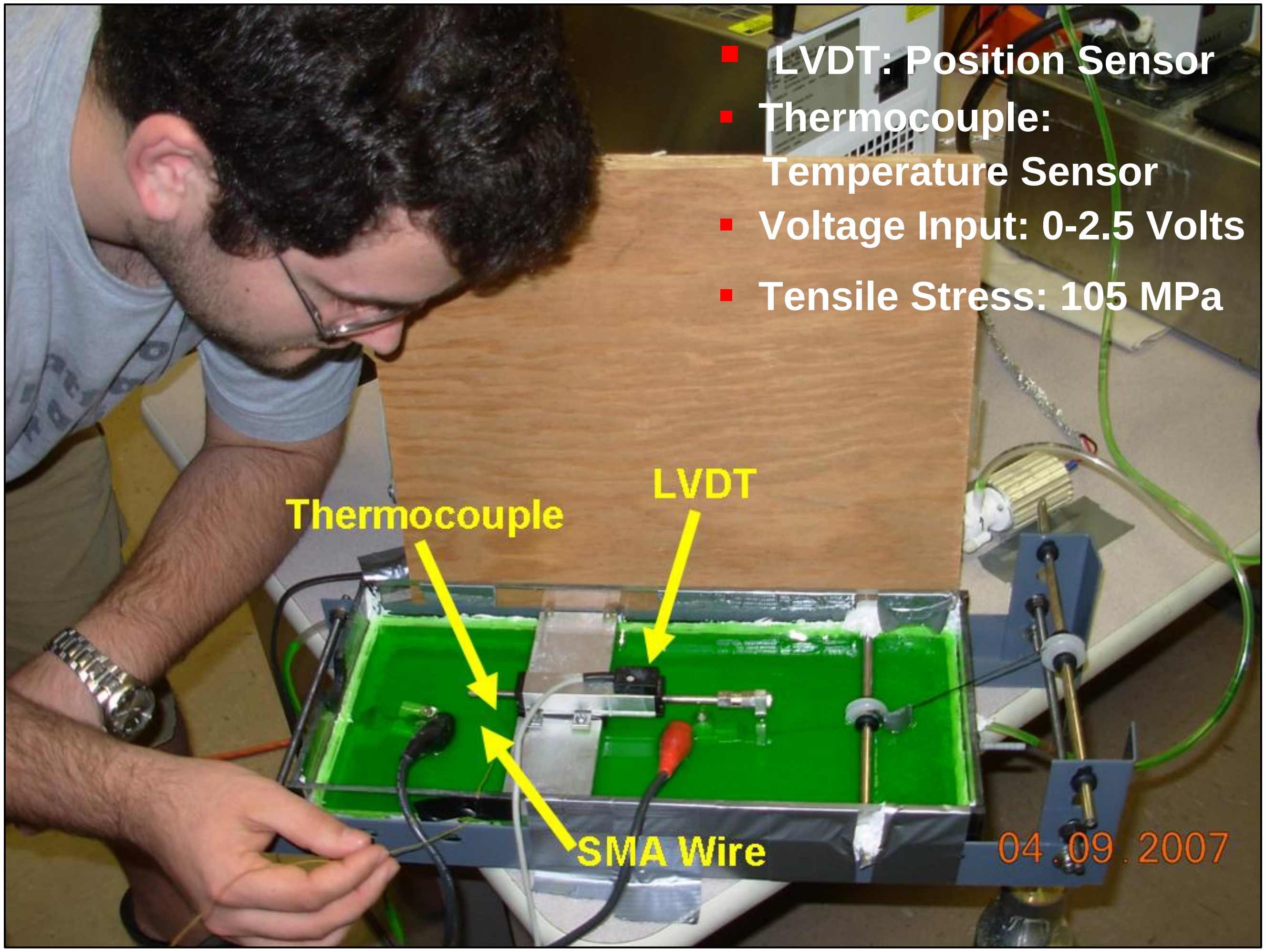
- **Goal Error Range: ± 0.2 % Strain**
- **Temp-Strain Mesh Size: 350 States**

SMA Characterization Results

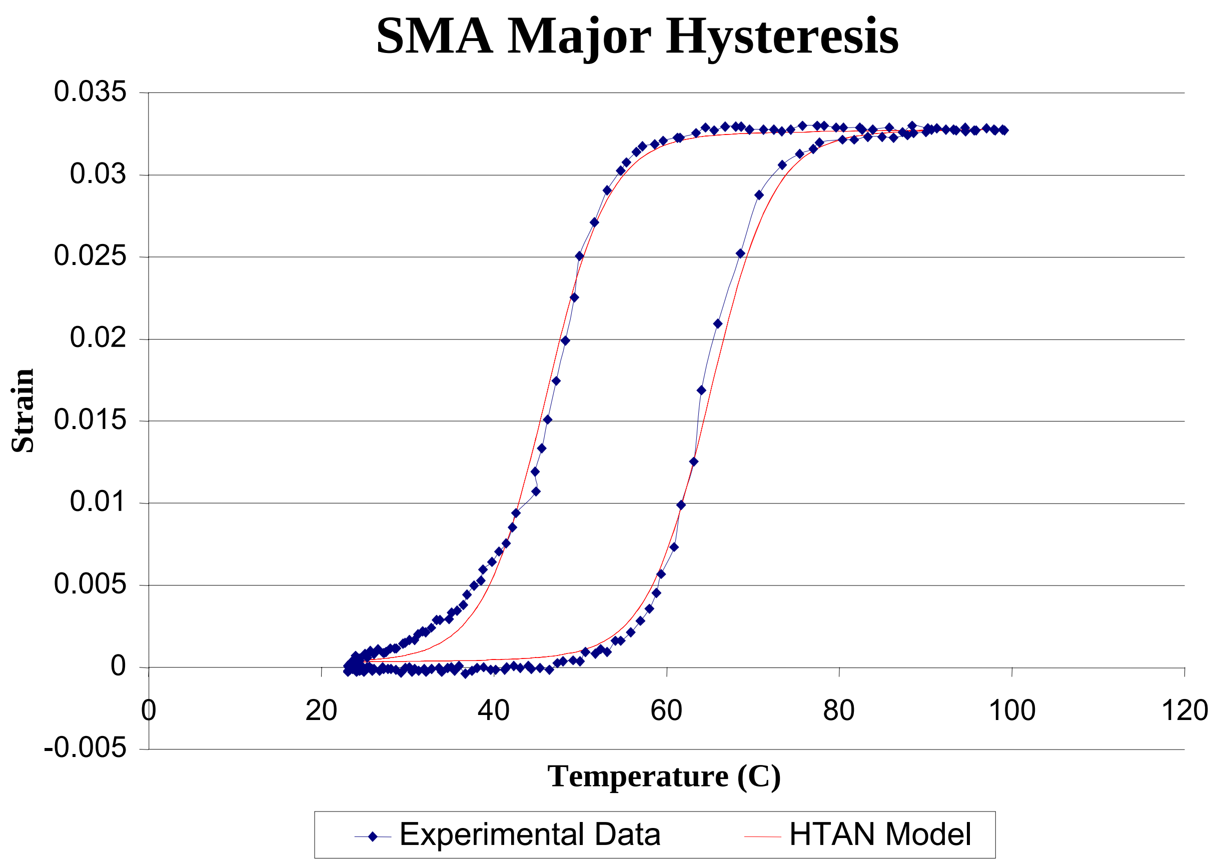
a) Simulation

b) Experiment

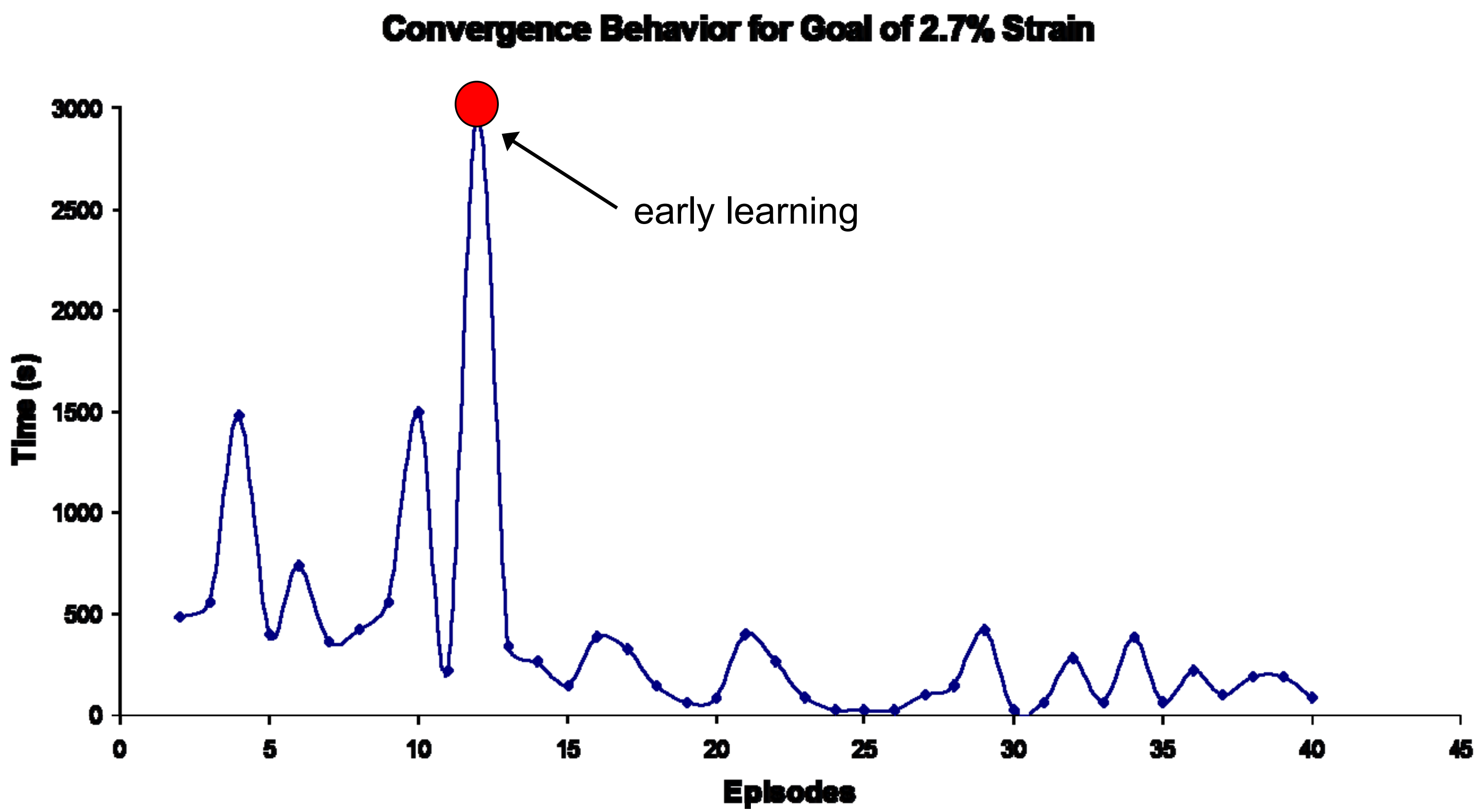
Experimental Setup



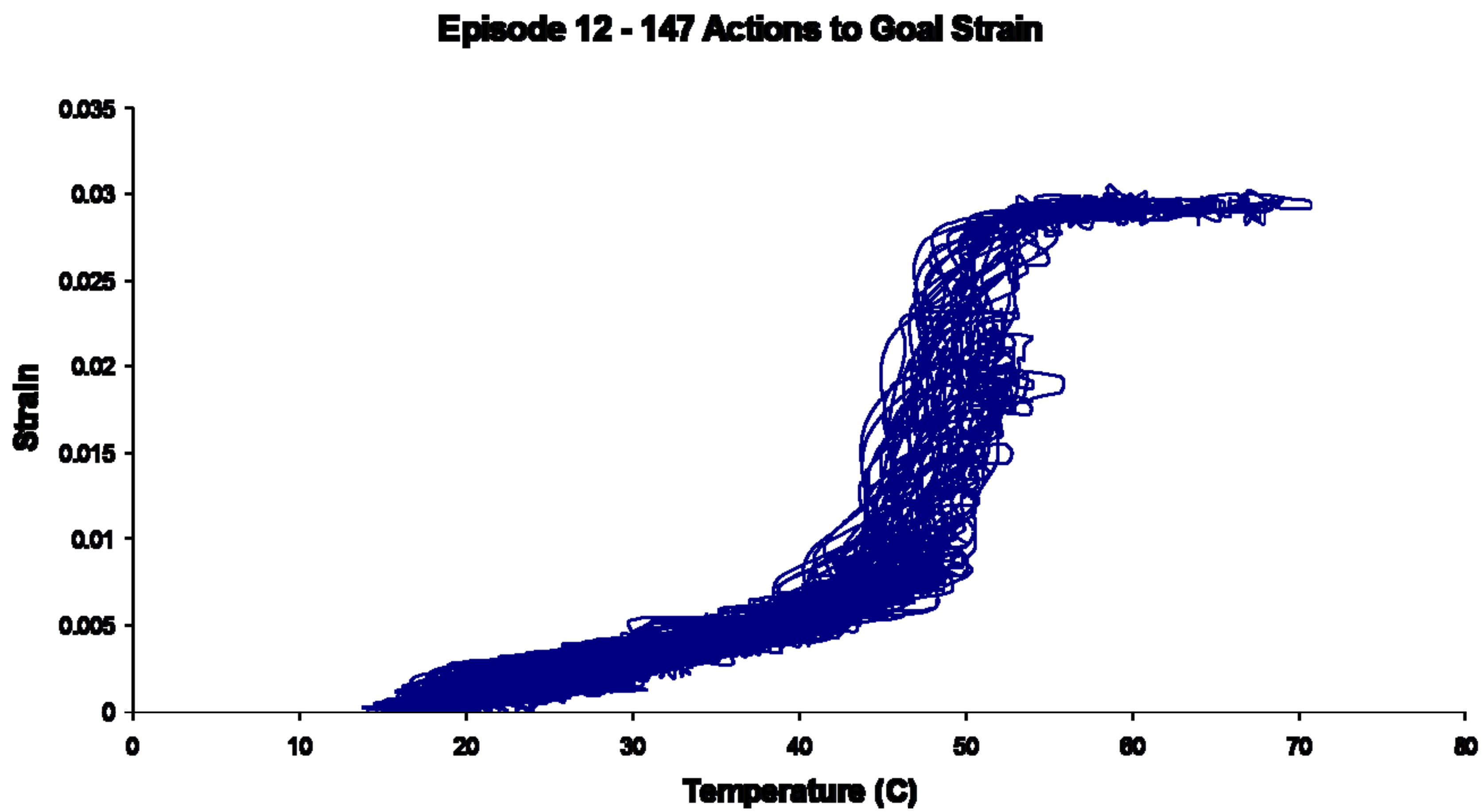
Experimental Forced Characterization: Validation



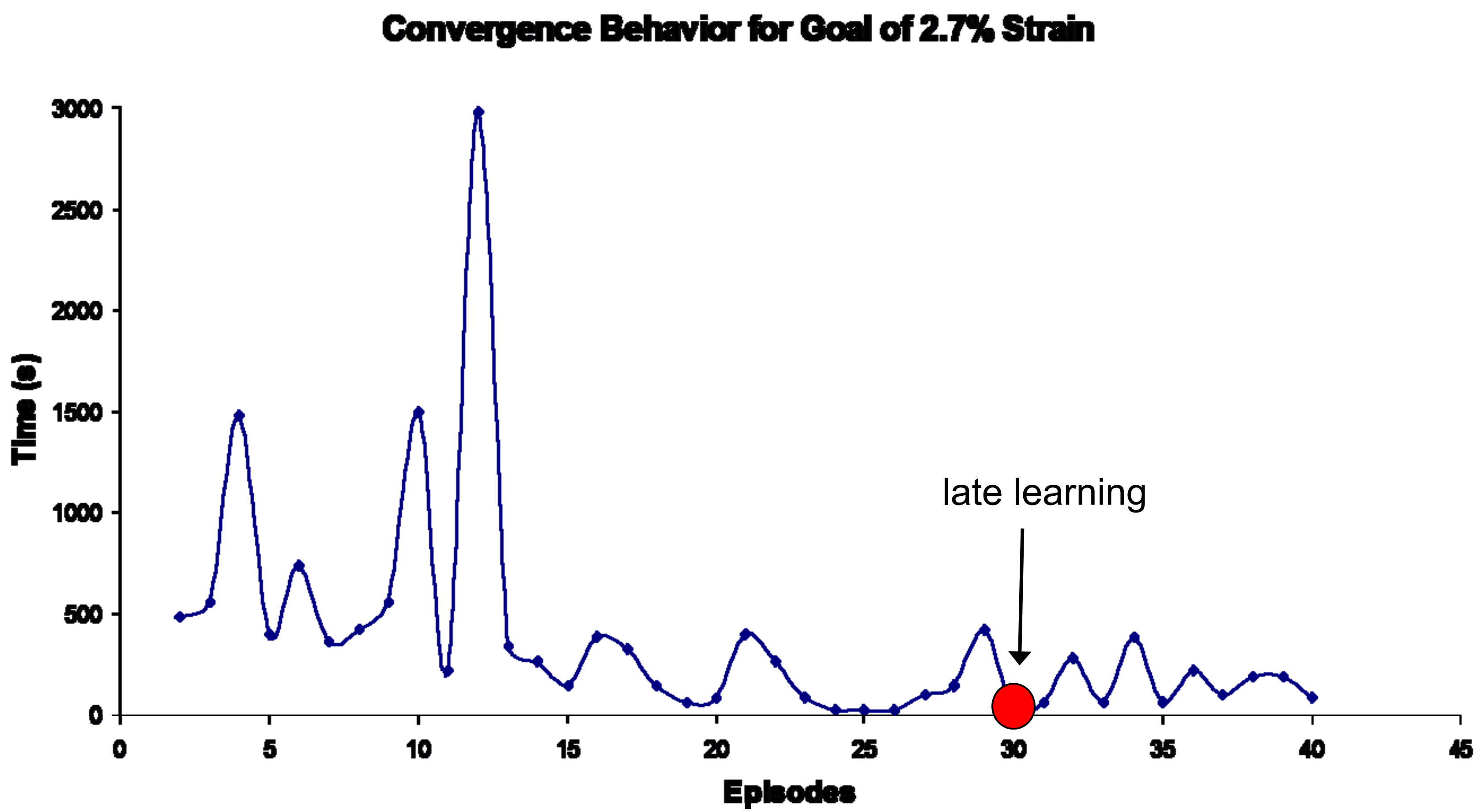
Experiment – Goals v. Time



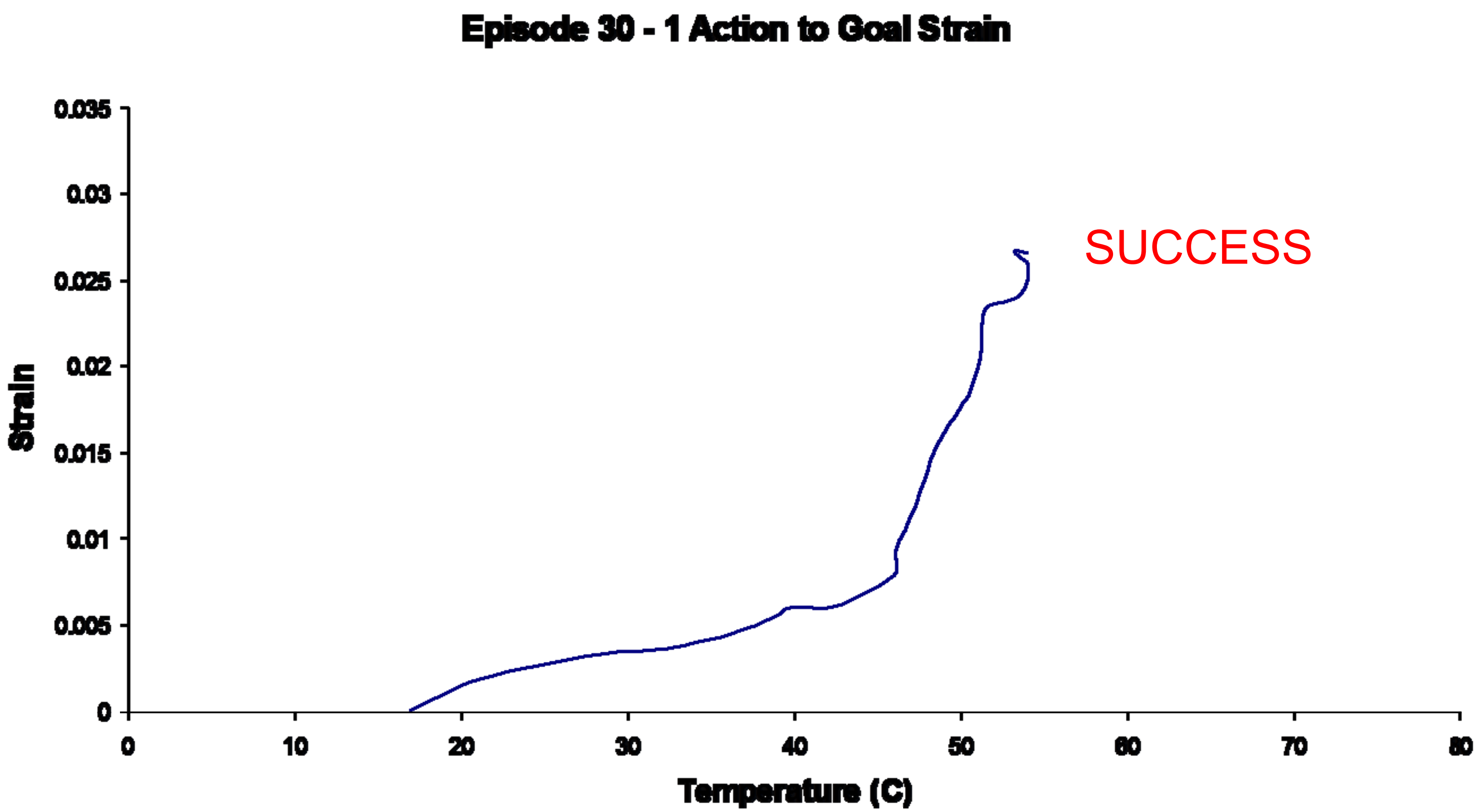
Experiment – Early Learning Refinement



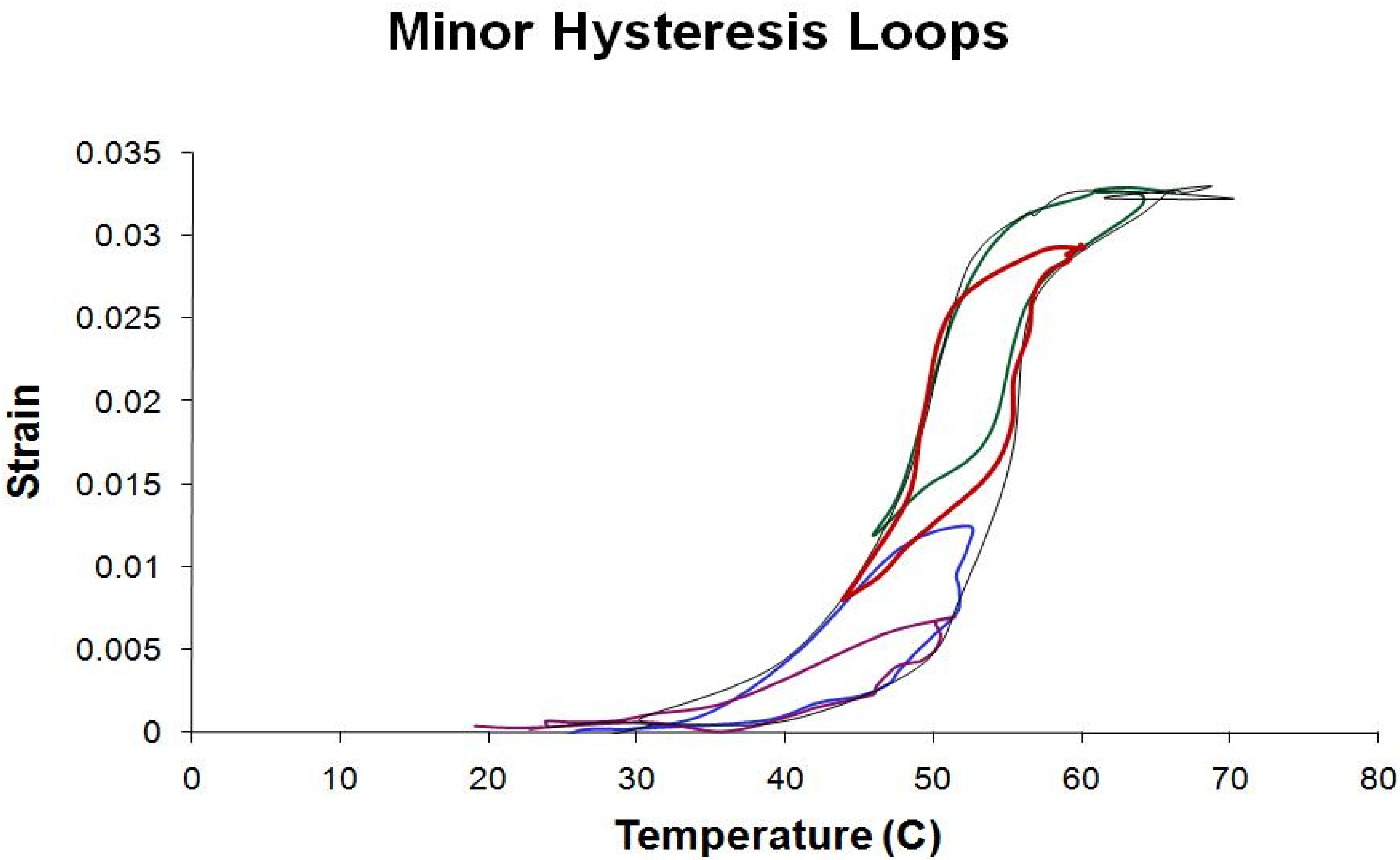
Experiment – Goals v. Time



Experiment – Late Learning Refinement



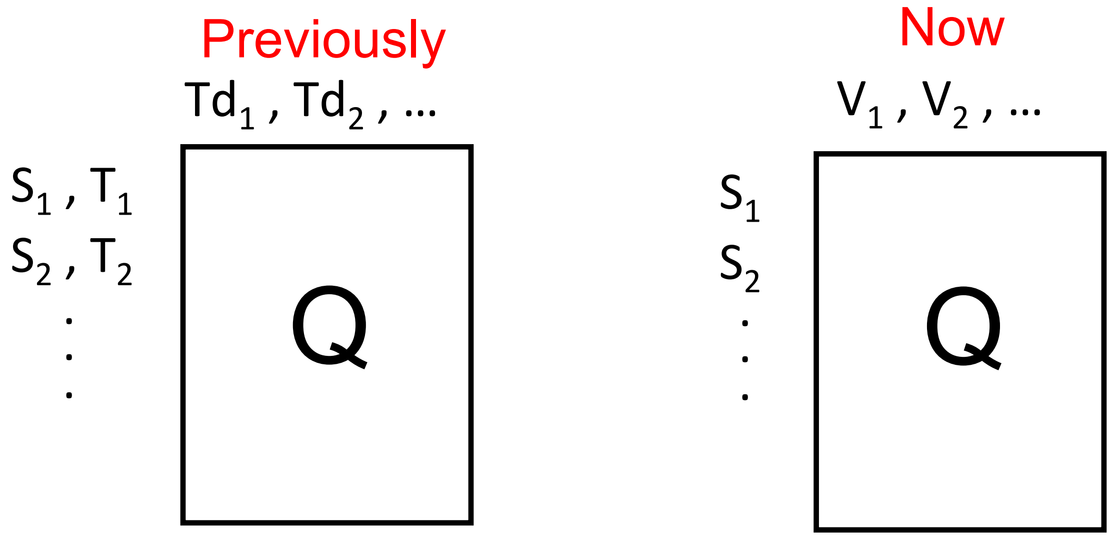
Experiment – Minor Hysteresis



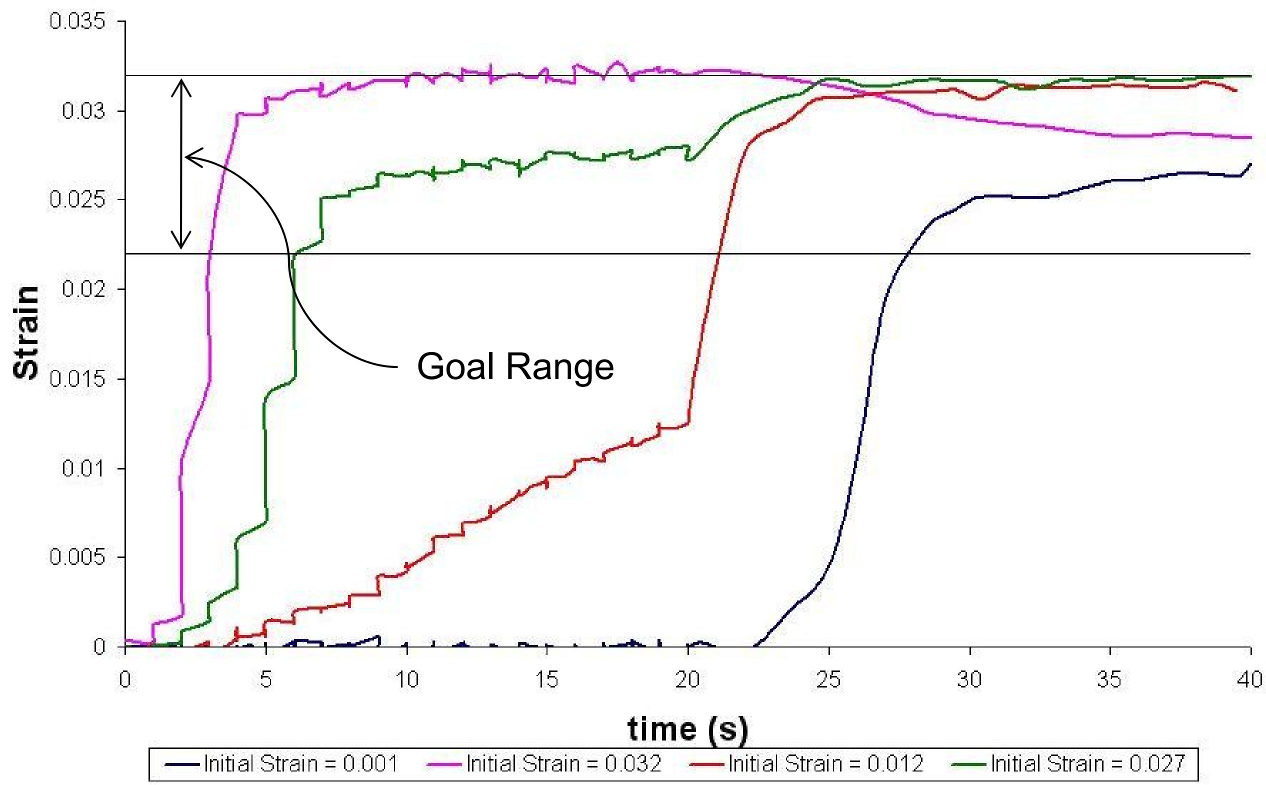
SMA Control
Experimental Results

Temp/Strain vs. Voltage/Position

- Getting voltage/position relationship (or voltage/strain) is better from a control standpoint
  - Voltage outputs far more accurate than temperature measurements
    - Thermocouple noise: ± 5-10 degrees Celsius
    - Voltage supply noise: < 0.01 Volts
  - Less states to explore (350 states versus 35 states)

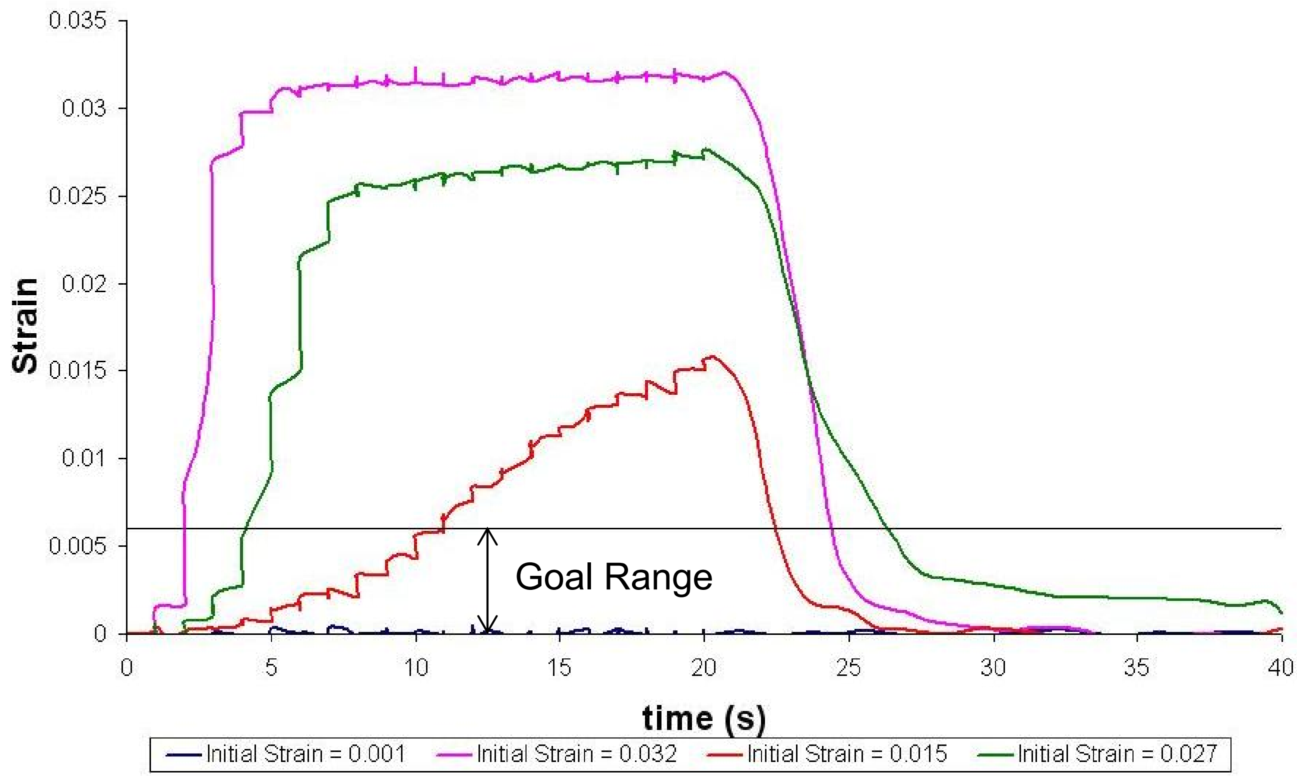


Control Policy Test for Goal = 2.7% Strain after 50 Episodes



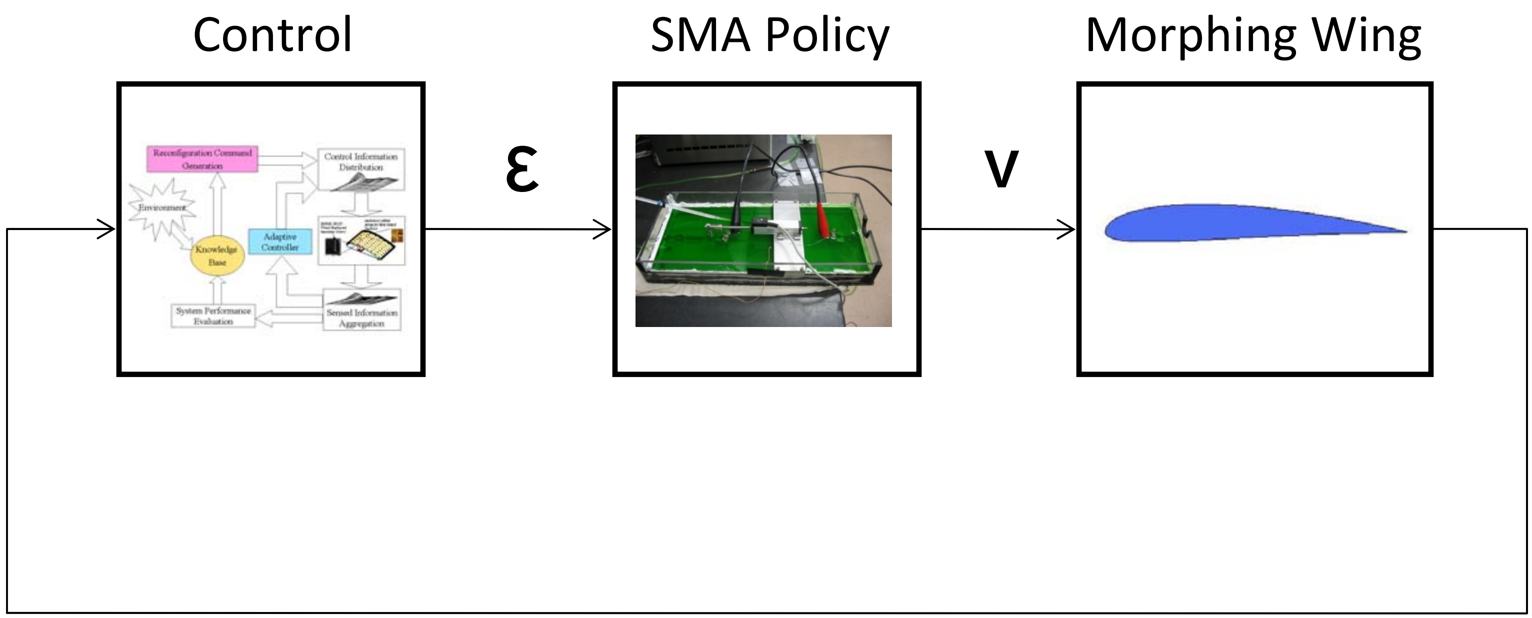
- Control Policy exploitation begins at t = 21 seconds
- Successful goal achievement from each initial state

Control Policy Test for Goal = 0.1% Strain after 50 Episodes



- Control Policy exploitation begins at t = 21 seconds
- Successful goal achievement from each initial state

Role in Morphing



- Controller sends desired strain to the SMA control policy
- SMA Control Policy send required voltage to wing
- Wing applies required voltage to actuators and morphs

Conclusions

- Reinforcement Learning successfully learns (without human supervision):
  - SMA hysteresis behavior in Temperature/Strain space
    - Major loop
    - Minor loops
  - Control Policy directly in Voltage/Strain space
    - Proven for both interior and boundary of system
    - Accomplished with goal range of ± 0.5% Strain (± 0.64 mm)
- Learning in voltage/strain space is easier and more useful for control than temperature/strain space
  - Voltage noise far less than thermocouple noise
  - Voltage/strain policy directly useful in feedback control law

Goals Achieved

- ✓ Investigate Machine Learning algorithms to characterize and control SMA temperature-strain behavior

- ✓ Understand control of SMA wire length after validating the model in the temperature-strain scope

- ✓ Determine optimal control policy for SMA wire length control (1-D)

Challenges and Open Problems

- Open Problems
  - Validation of simulation by comparing to Preisach model of SMA hysteresis needed to provide a good conventional comparison.
  - Completion of control policy by learning for all other goal states needed to use the policy in a feedback control law.
  - Expanding scope from wire (one spatial dimension displacement) to surface (two and three-dimensional displacements) would expand these methods to greater generalities.
  - Learning how to control arrays of 1-D wires needed to make 1-D SMA actuators useable for morphing.

- Challenges
  - Extension to higher dimensions produces much longer learning times due to spatial complexity.
  - The ability to heat 2-D and 3-D SMA objects is difficult to accomplish through electrical resistance.

Acknowledgements

- NASA Headquarters as part of the University Research & Education Training Institute (URETI) program

  Texas Institute for Intelligent

  Bio-Nanomaterials and Structures for Aerospace

  Vehicles

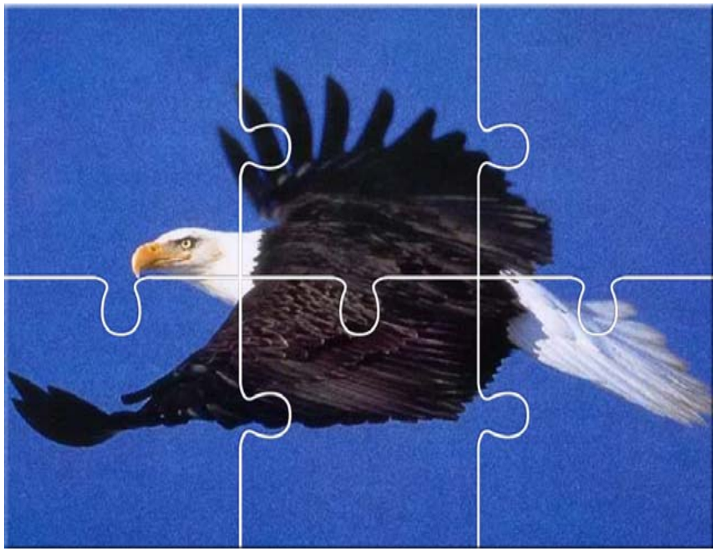
- National Science Foundation ***Graduate Research Fellowship Program***

- Texas A&M University ***Undergraduate Summer Research Grant*** (USRG)

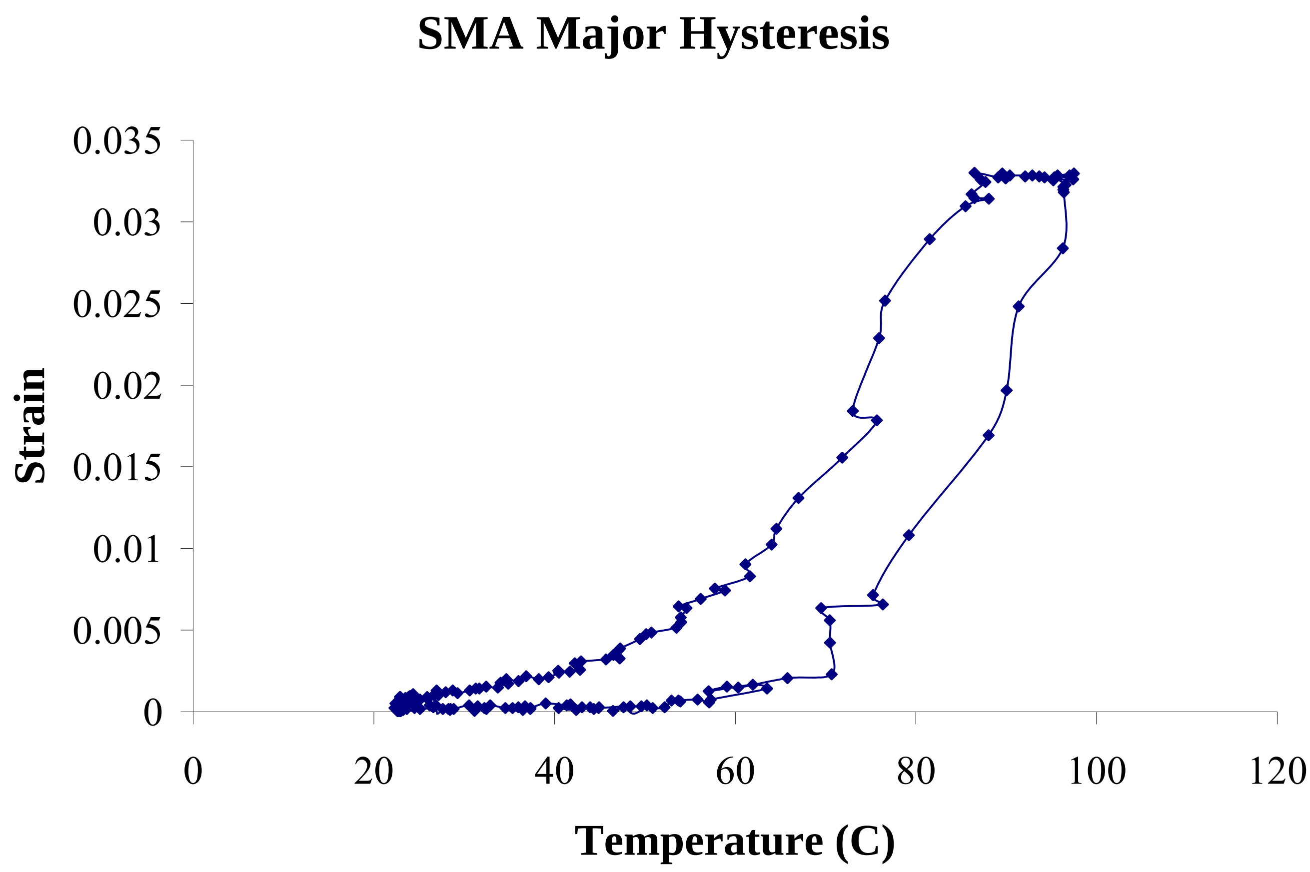
- This work was sponsored (in part) by the Air Force Office of Scientific Research, USAF, under grant/contract number FA9550-08-1-0038. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

This support is gratefully acknowledged by the authors

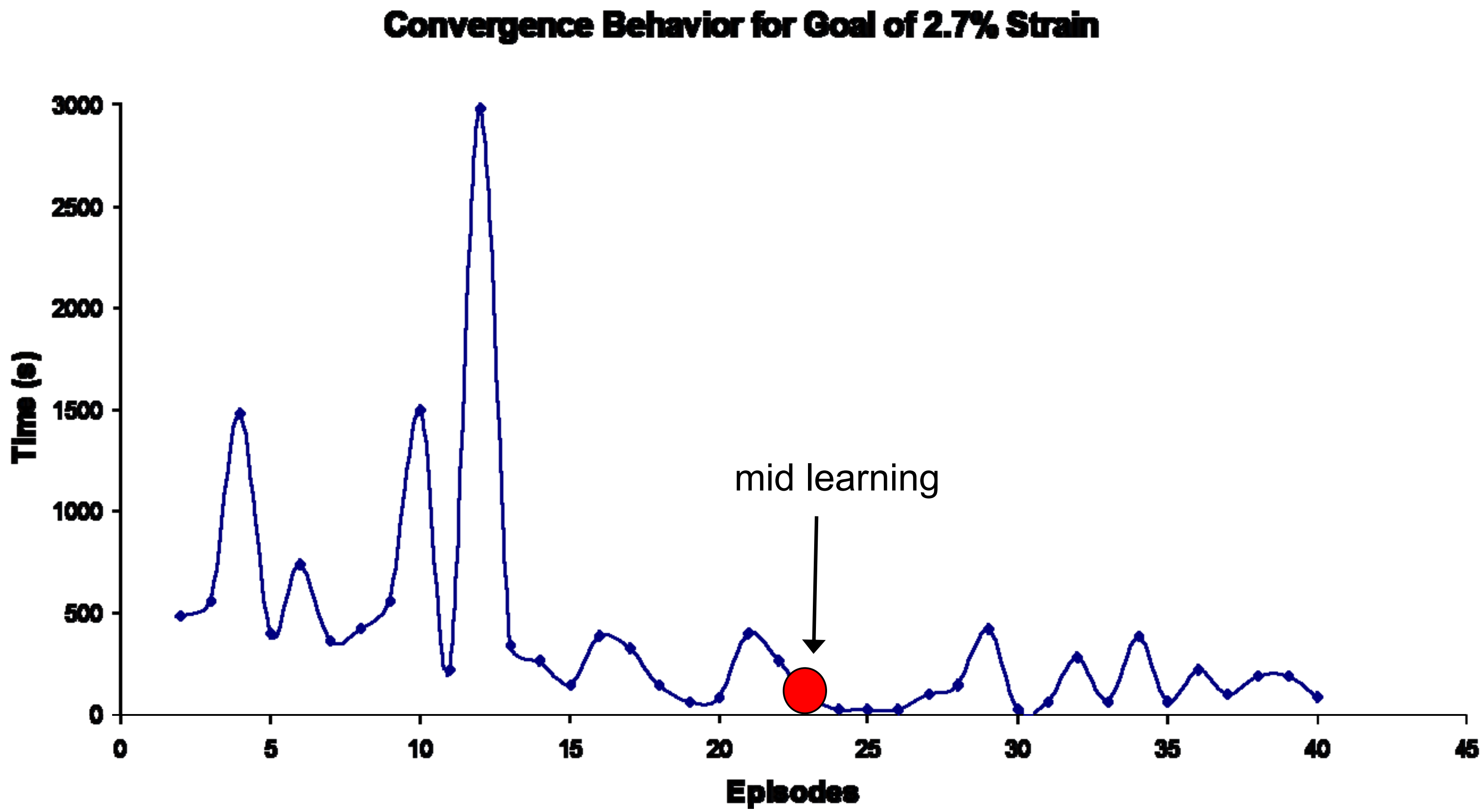
Questions?



Experimental Forced Characterization: Water



Experiment – Goals v. Time



Experiment – Mid Learning Refinement