UT Austin Learning Agents Research Group

THE UNIVERSITY OF TEXAS AT AUSTIN

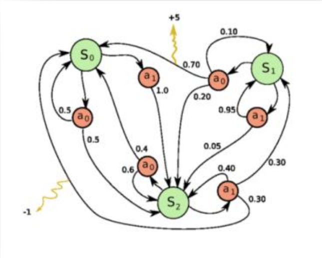
# A Learning Agent for Heat-Pump Thermostat Control

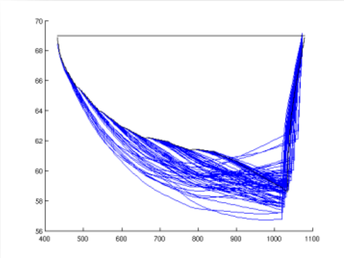

**Daniel Urieli and Peter Stone**

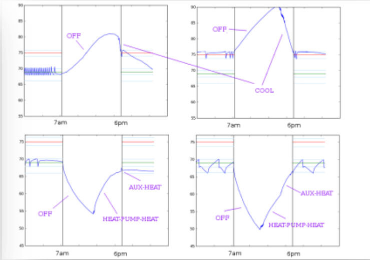
Department of Computer Science

**The University of Texas at Austin**

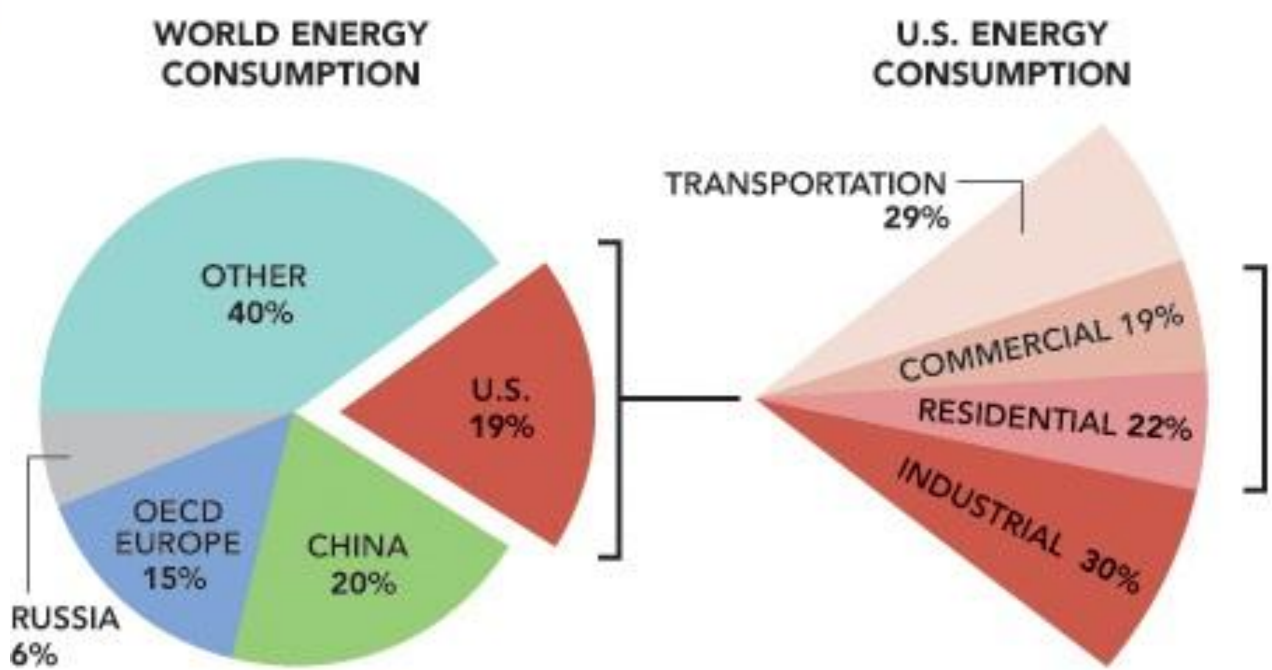
{urieli,pstone}@cs.utexas.edu





## Chapter 1: Buildings Sector



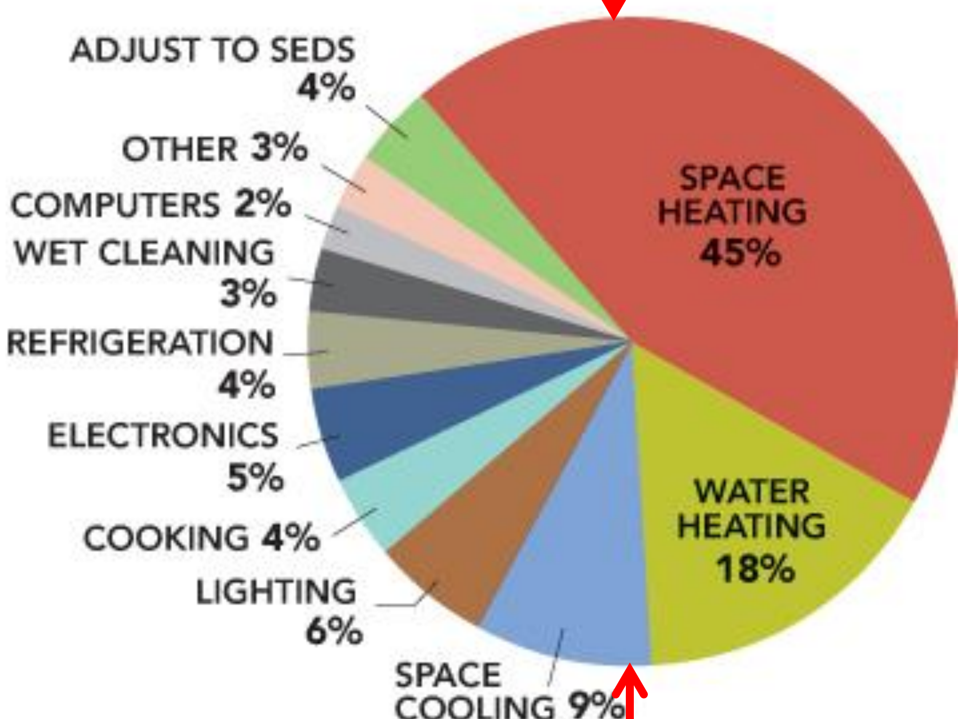
# BUILDINGS ENERGY DATA BOOK

U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy



## Chapter 2: Residential Sector

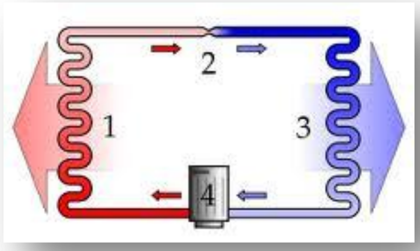
RESIDENTIAL SITE ENERGY CONSUMPTION BY END USE




Heating, Ventilation, and Air-conditioning (HVAC) systems

# Heat-Pump based HVAC System

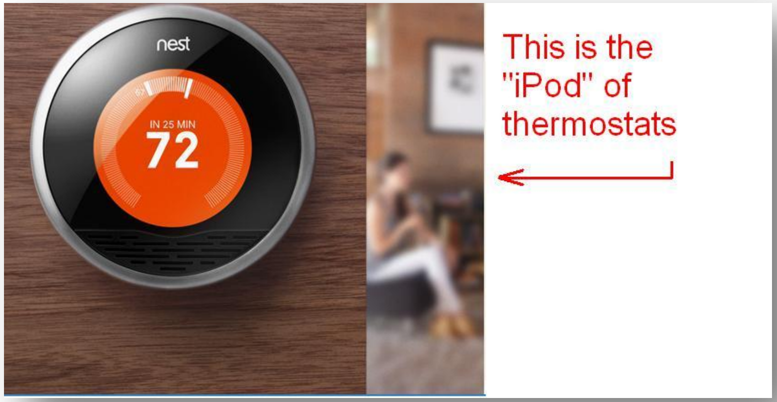
- Part of the efforts of moving to sustainable energy
- Heat-pump is widely used and highly efficient
  - Consumes renewable energy (electricity) rather than gas/oil
  - Its heat output is up to 3x-4x the energy it consumes
  - But: no longer effective in freezing outdoor temperatures
- Backed up by an auxiliary heater
  - Resistive heat coil
  - Unaffected by outdoor temperatures
  - But: consumes 2x the energy consumed by the heat-pump heater
- Heat pump is also used for cooling




+

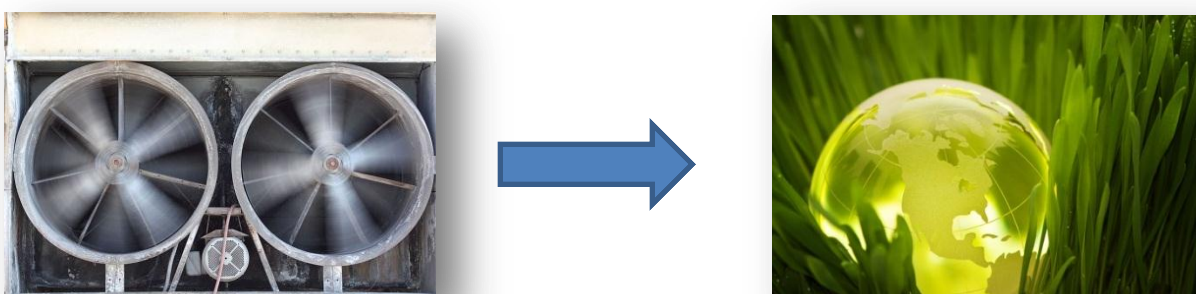
# Thermostat – an HVAC System's Decision Maker

- The thermostat :
  - Controls Comfort
  - Significantly affects energy consumption
- Current interest evident from companies like NEST, BuildingIQ



**<u>Goal</u>:**

Minimize energy consumption while satisfying comfort requirements



www.dot.gov

**Goal**:
Minimize energy consumption while satisfying comfort requirements

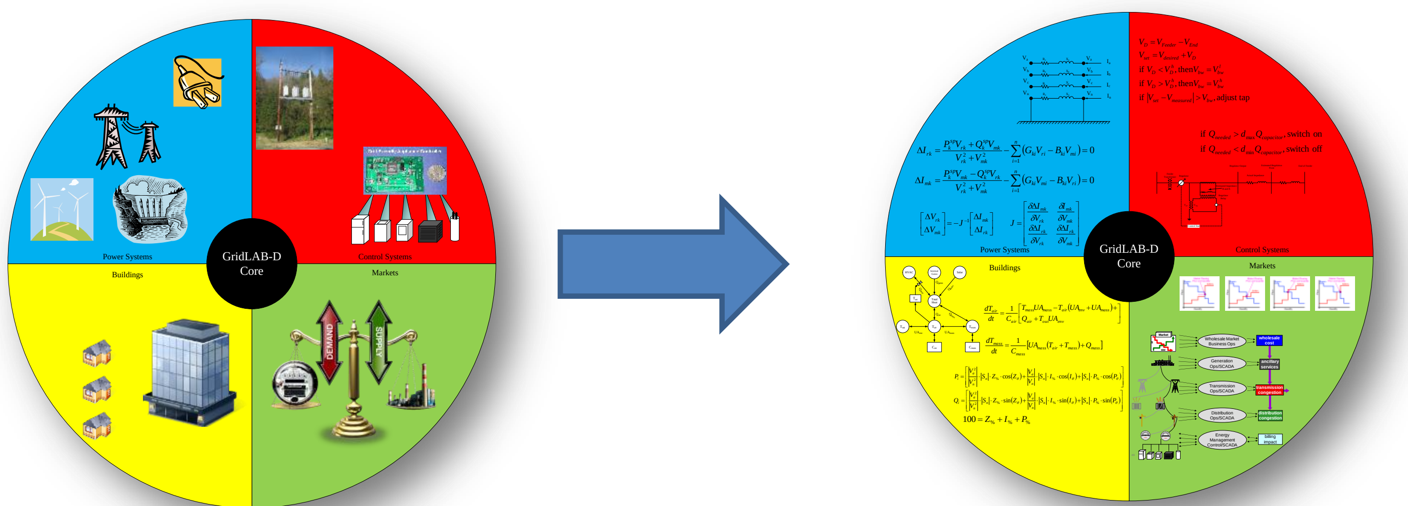
**Contributions**:
1. A complete reinforcement learning agent that learns and applies a new, adaptive control strategy for a heat-pump thermostat

2. Our agent achieves 7.0%-14.5% yearly energy savings

www.dot.gov

# Simulation Environment

- GridLAB-D: A realistic smart-grid simulator, simulates power generation, loads and markets
- Open-source software, developed for the U.S. DOE, simulates seconds to years
- Realistically models a residential home
  - Heat gains and losses, thermal mass, solar radiation and weather effects, uses real weather data recorded by NREL (www.nrel.gov)



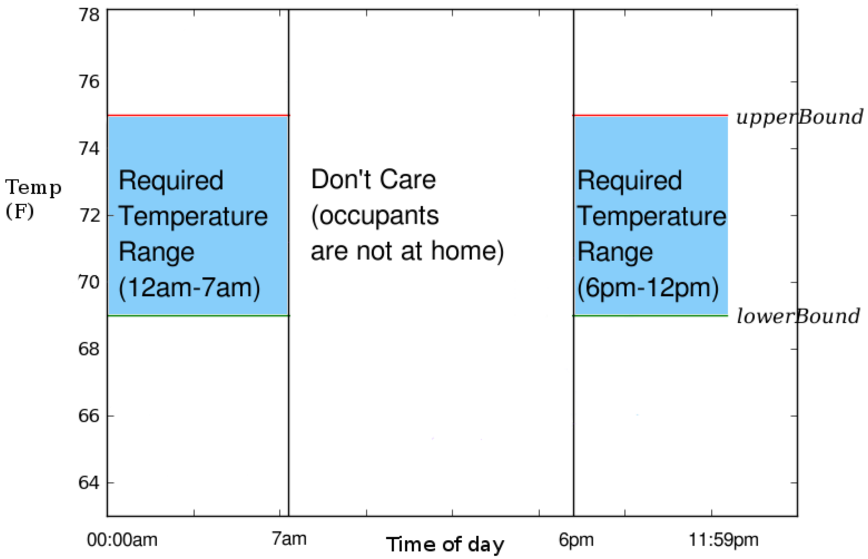
# Problem Setup

- Simulating a typical residential home

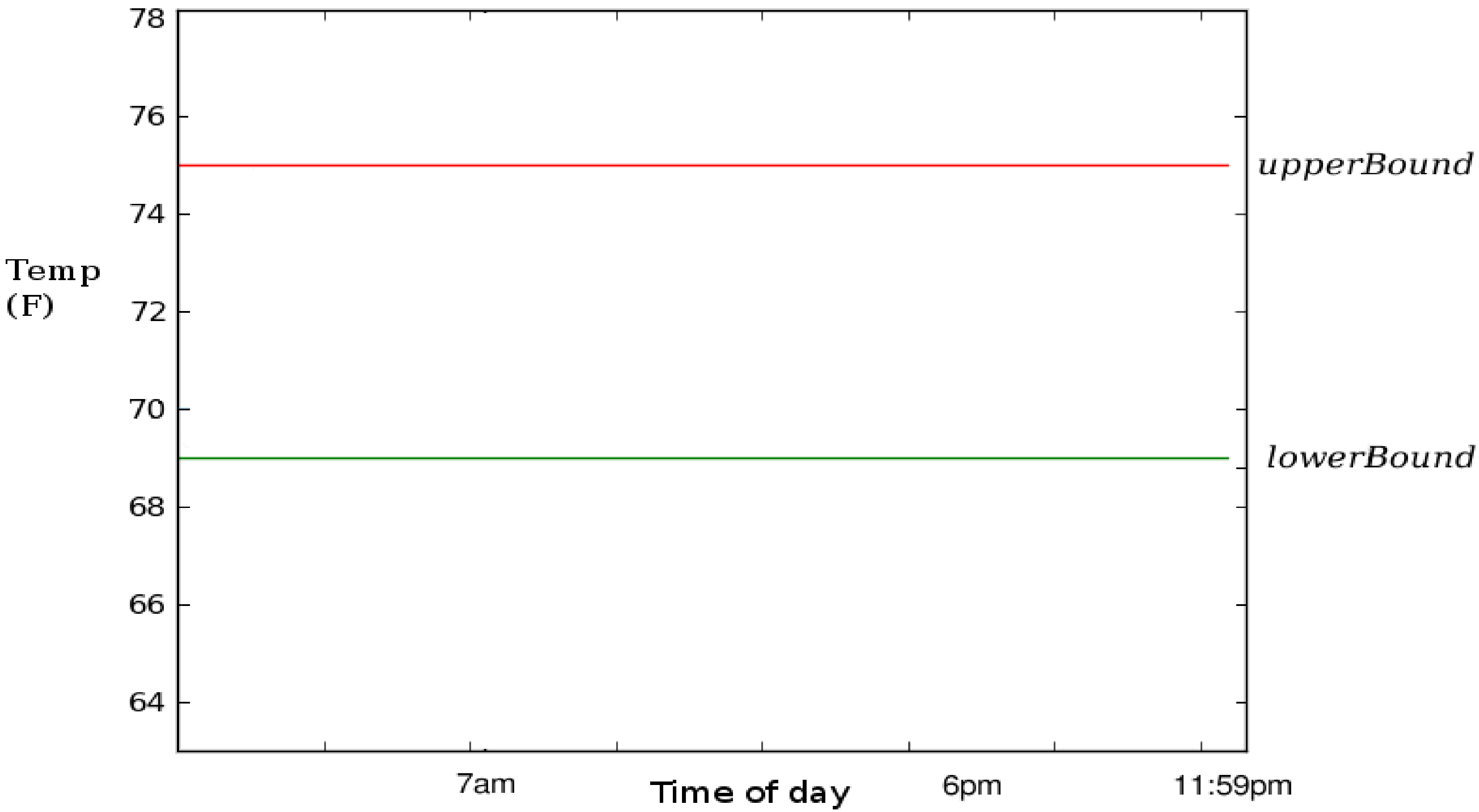
- **Goal:** minimize energy consumed by the HVAC, while satisfying the following comfort spec:

Occupants are

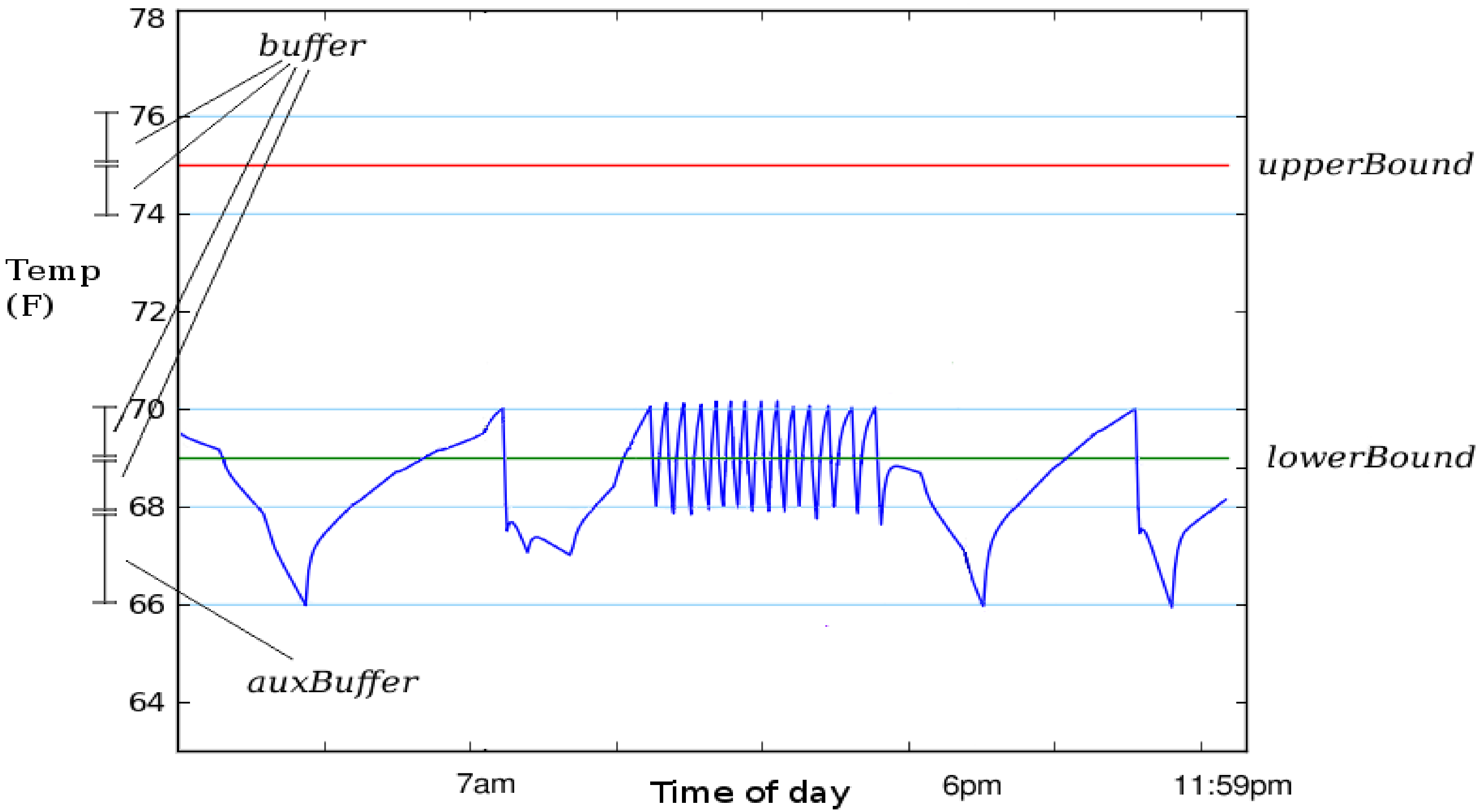
- 12am-7am: **At home.**
- 7am-6pm: **Not at home.** (the ***"don't care"*** period)
- 6pm-12am: **At home.**



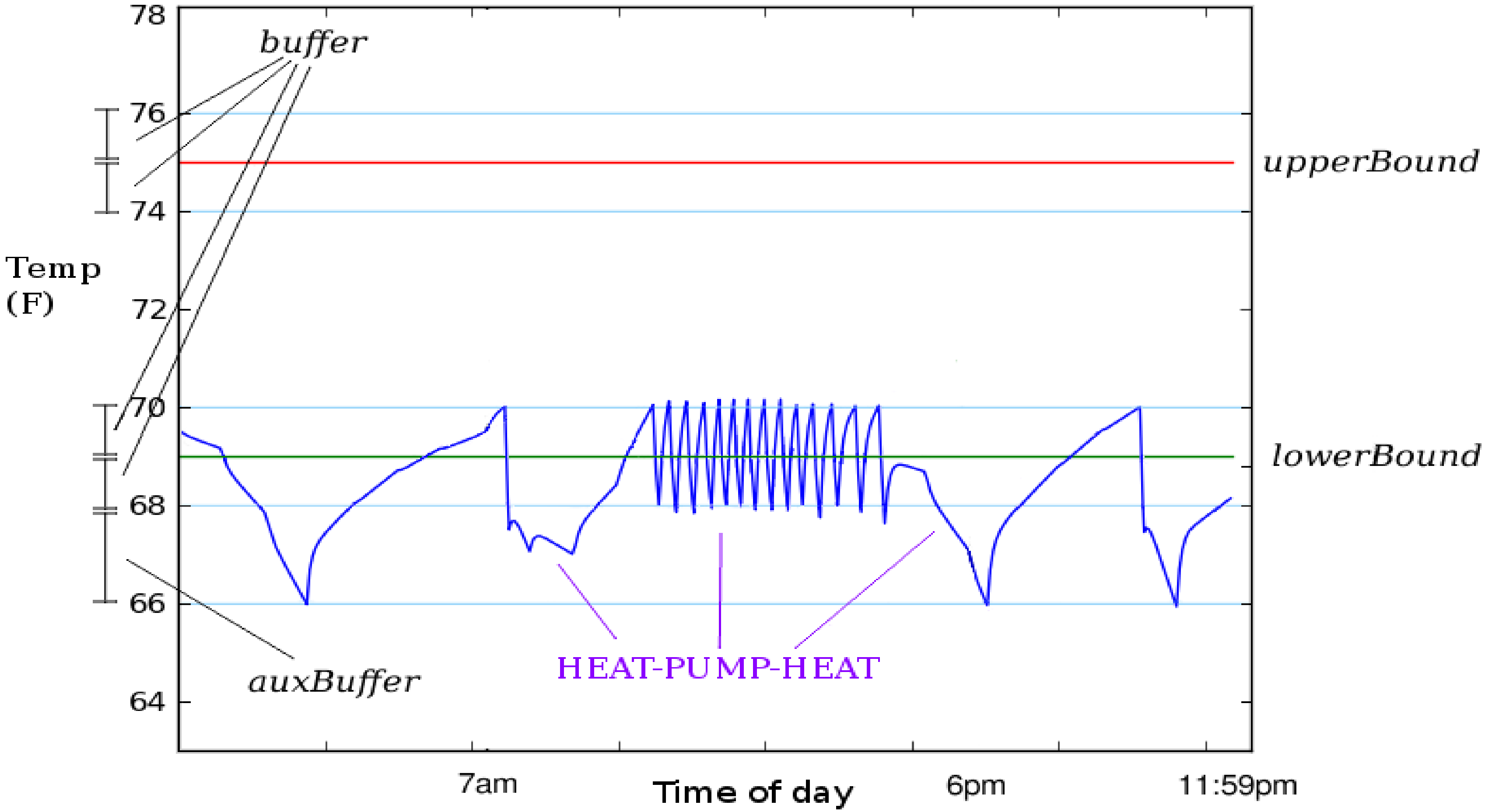
# The Default Thermostat



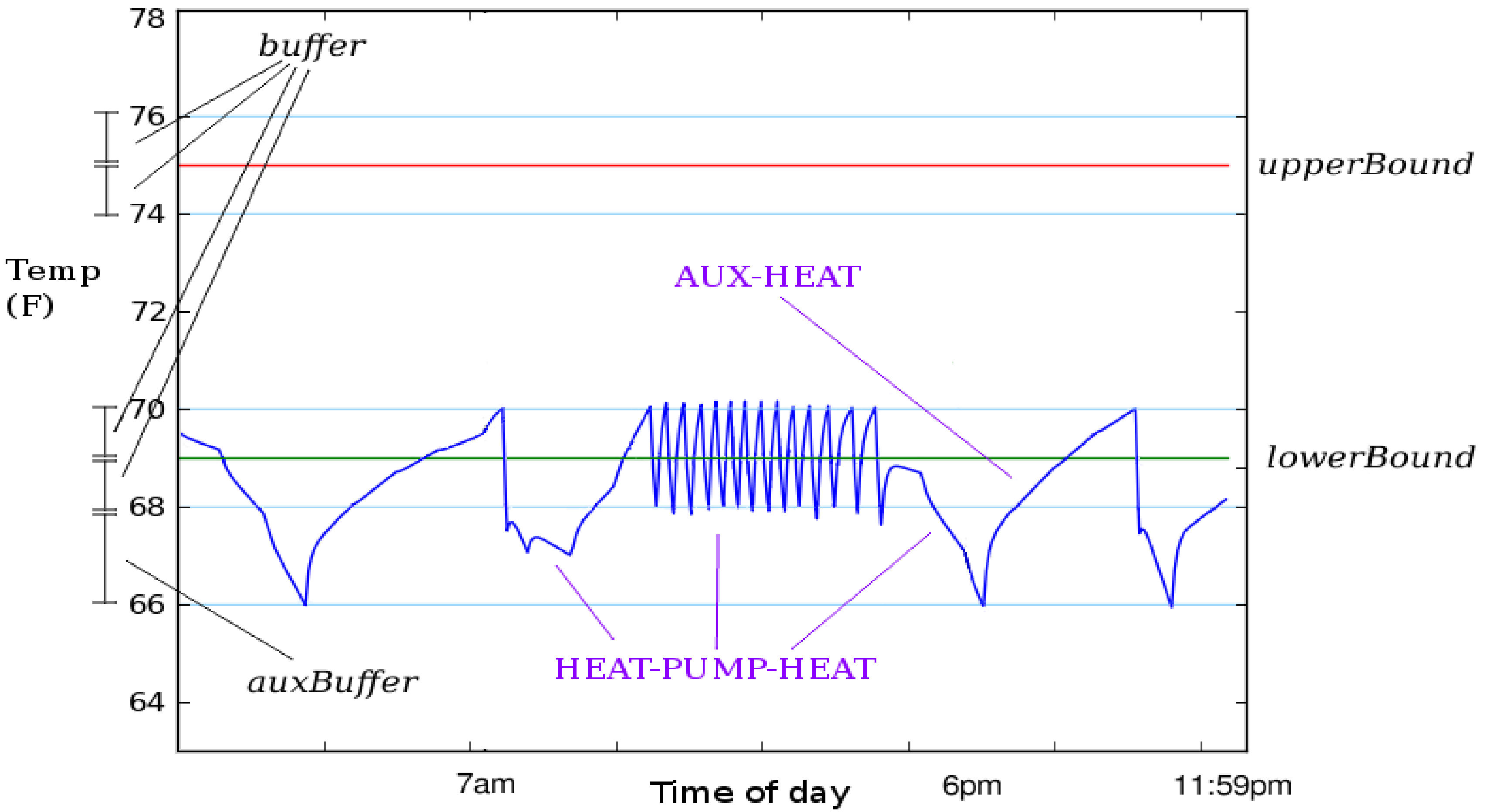
# The Default Thermostat



# The Default Thermostat



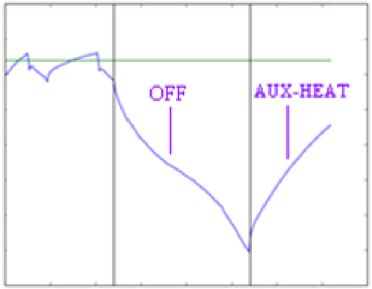
# The Default Thermostat



# Can We Just Shut-Down The Thermostat During "don't-care" Period?

- Effective way to save energy
  - Indoor temp. closer to outdoor ➡ heat dissipation slows down

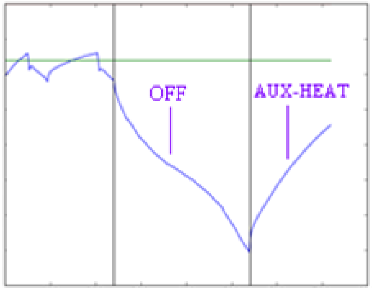
- Simulating it…




- In this case, the result is:
  - Increased energy consumption
  - Failure to satisfy the comfort spec

# Can We Shut-Down The Thermostat During "don't-care" Period?

- Effective way to save energy
  - Indoor temp. closer to outdoor ⇨ heat dissipation slows down
- Simulating it…



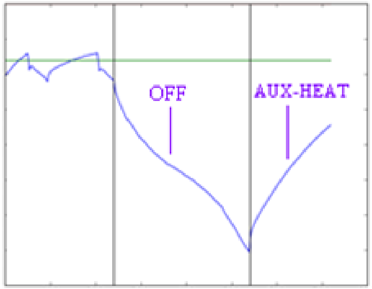

Therefore, people frequently prefer to leave the thermostat open all day

- In this case, the result is:
  - Increased energy consumption
  - Failure to satisfy the comfort spec

# Can We Shut-Down The Thermostat During "don't-care" Period?

- Effective way to save energy
  – Indoor temp. closer to outdoor ⇨ heat dissipation slows down

- Simulating it…

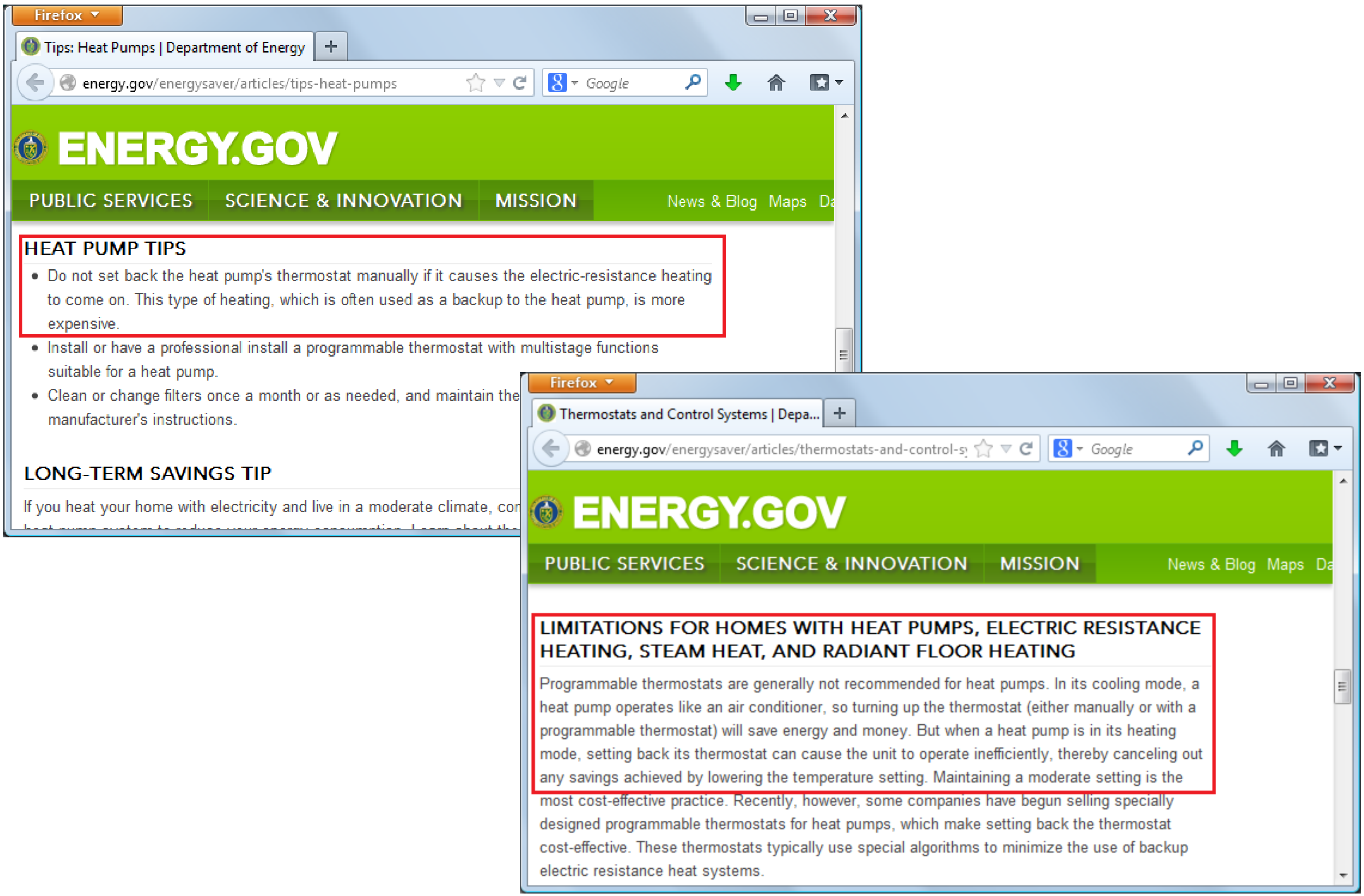



- In this case, the result is:
  – Increased energy consumption
  – Failure to satisfy the comfort spec

Therefore, people frequently prefer to leave the thermostat open all day

However, a smarter shut-down should still be able to save energy

# From the US Dept. of Energy's website

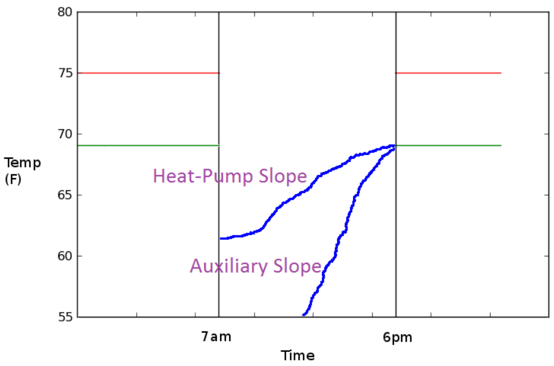


# Challenges

<u>Desired behavior:</u>

- Maximize shut-down time while staying above the heat-pump slope
- Similarly for cooling (no AUX)

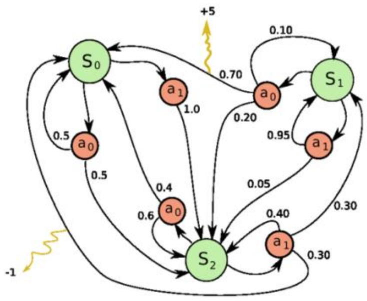
<u>Challenges:</u>




- The heat-pump slope:
  - Is unknown in advance
  - Changes every day
  - Depends on future weather
  - Depends on specific house characteristics
- Action effects are:
  - Drifting rather than constant: since heat is being <u>moved</u> rather than <u>generated</u>, heat output strongly depends on the temperatures indoors, outdoors and along the heat path
  - Noisy due to hidden physical conditions
  - Delayed due to heat capacitors like walls and furniture
- Also, in a realistic deployment:
  - Exploration cannot be too long or too aggressive
  - Customer acceptance will probably depend on worst-case behavior
- Making decisions in continuous, high dimensional space

# Our Problem as a Markov Decision Process (MDP)

- **States:**
- **Actions:**
- **Transition:**
- **Reward:**
- **Terminal States:**

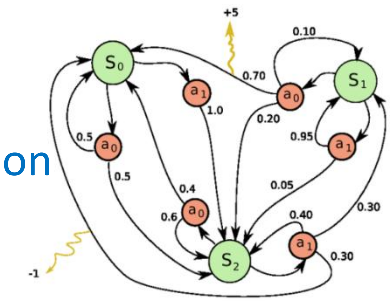


- Action is taken every 6 minutes
  - Modeling a realistic lockout of the system

# Our Problem as a Markov Decision Process (MDP)

- **States:**
- **Actions:** {COOL, OFF, HEAT, AUX}

  1 : 0 : 2 : 4 ⟸ consumption ($e_a$) proportion



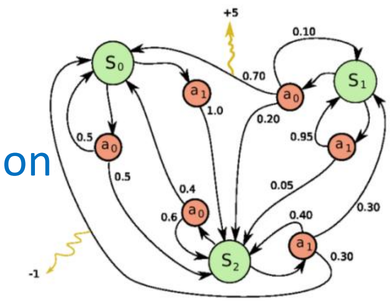

- **Transition:**
- **Reward:**
- **Terminal States:**

---

- Action is taken every 6 minutes
  - Modeling a realistic lockout of the system

# Our Problem as a Markov Decision Process (MDP)

- **States:**
- **Actions:** {COOL, OFF, HEAT, AUX}
  1 : 0 : 2 : 4 ⟸ consumption ($e_a$) proportion




- **Transition:**
- **Reward:** $-\,e_a - 100000\,\Delta^2_{6pm}$ *where:*
  $\Delta^2_{6pm}$ := (indoor_temp_at_6pm – required_indoor_temp_at_6pm)
- **Terminal States:**

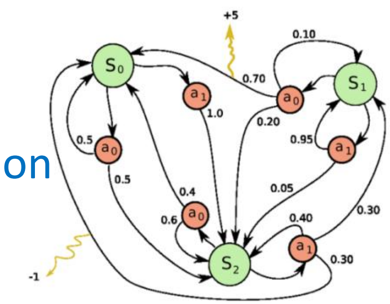
---

- Action is taken every 6 minutes
  - Modeling a realistic lockout of the system

# Our Problem as a Markov Decision Process (MDP)

- **States:** **???**
- **Actions:** {COOL, OFF, HEAT, AUX}

  1 : 0 : 2 : 4 consumption ($e_a$) proportion




- **Transition:**
- **Reward:** $-\,e_a - 100000\,\Delta^2_{6pm}$ *where:*

  $\Delta^2_{6pm}$ := (indoor_temp_at_6pm – required_indoor_temp_at_6pm)
- **Terminal States:**

---

- Action is taken every 6 minutes
  - Modeling a realistic lockout of the system

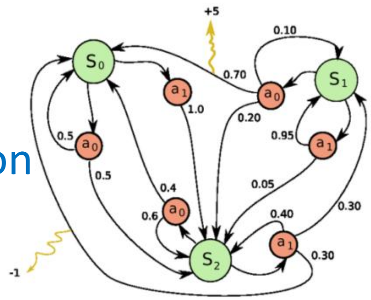
# How Should We Model State?

- Choosing a state representation is an important design decision. A state variable:
  - captures what we need to know about the system at a given moment
  - is the variable around which we construct value function approximations
    [Powell 2011]

- Definition 5.4.1 from [Powell 2011]:
  - ***A state variable** is the minimally dimensioned function of history that is necessary and sufficient to compute the decision function, the transition function, and the contribution function.*

# Our Problem as a Markov Decision Process (MDP)

- **States:** $\langle T_{in}, \text{Time}, e_a \rangle$
- **Actions:** {COOL, OFF, HEAT, AUX}
  1 : 0 : 2 : 4 ⟸ consumption ($e_a$) proportion




- **Transition:**
- **Reward:** $-e_a - 100000\,\Delta^2_{6pm}$ *where:*
  $\Delta^2_{6pm}$ := (indoor_temp_at_6pm – required_indoor_temp_at_6pm)
- **Terminal States:**

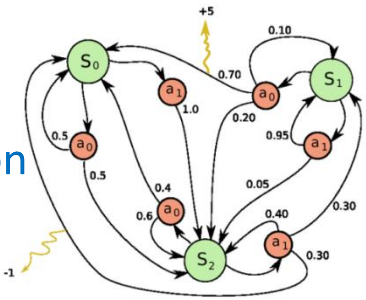
---

- Action is taken every 6 minutes
  - Modeling a realistic lockout of the system

# Our Problem as a Markov Decision Process (MDP)

- **States:** $<T_{in}, \text{Time}, e_a>$
- **Actions:** {COOL, OFF, HEAT, AUX}

  1 : 0 : 2 : 4 ⟸ consumption ($e_a$) proportion



- **Transition:**
- **Reward:** $-\,e_a - 100000\,\Delta^2_{6pm}$ *where:*

  $\Delta^2_{6pm}$ := (indoor_temp_at_6pm – required_indoor_temp_at_6pm)

- **Terminal States:**

---

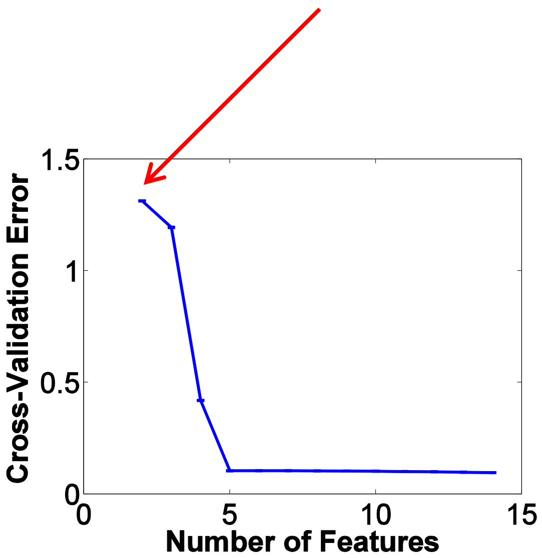
- Action is taken every 6 minutes
  - Modeling a realistic lockout of the system

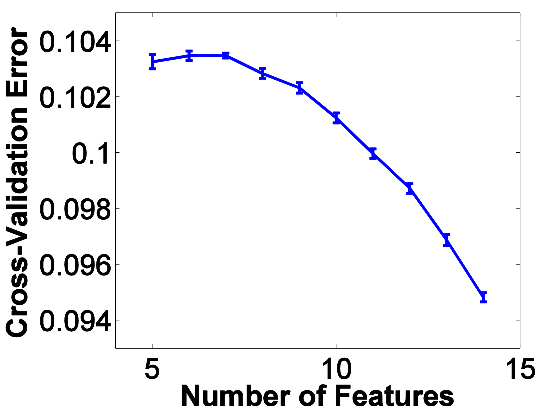
# Expanding State to Compute the Transition Function

- Can we predict action effects for each of the state variables?
- Current state representation: $\langle T_{in}, \mathrm{Time}, e_a \rangle$
- Need to be able to predict $T_{in}$ and $e_a$
- Method: generate simulated data, use cross-validation to test for regression prediction accuracy

# Predicting $T_{in}$

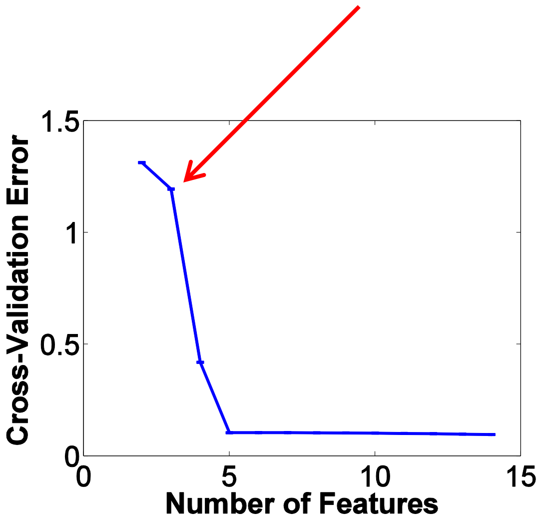
- Prediction error is unacceptably high – state <$T_{in}$, Time, $e_a$> doesn't capture enough information

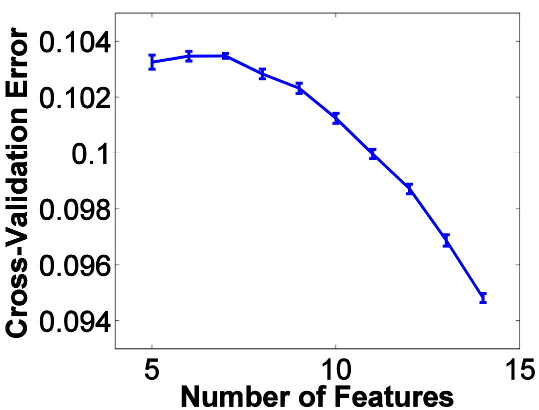


# Predicting $T_{in}$

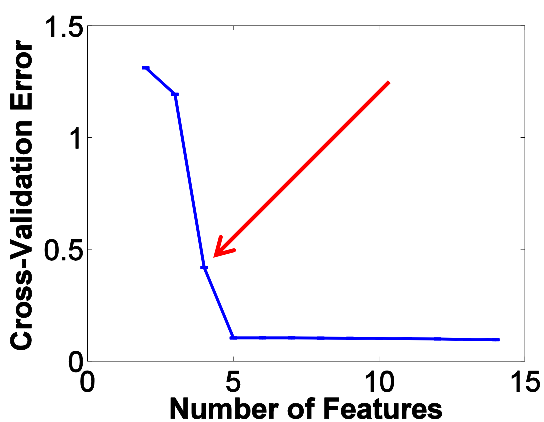
- Prediction error is unacceptably high – state $\langle T_{in}$, Time, $e_a \rangle$ doesn't capture enough information
- Add $T_{out}$ – directly affects $T_{in}$ . Prediction error still unacceptably high

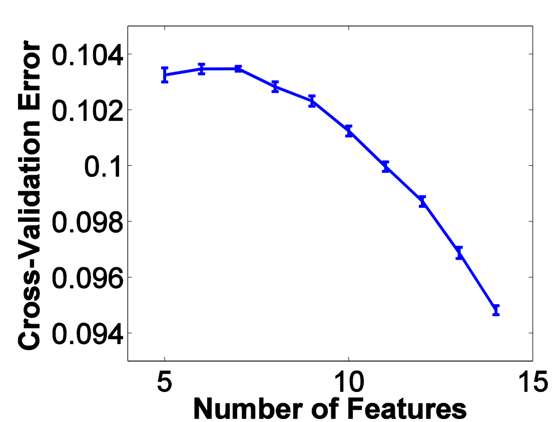


# Predicting $T_{in}$

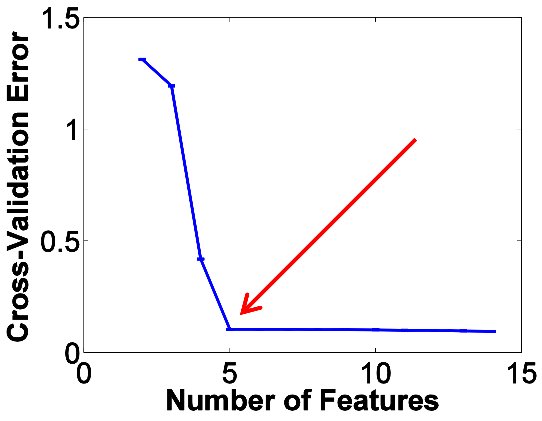
- Prediction error is unacceptably high – state $<T_{in}, \text{Time}, e_a>$ doesn't capture enough information
- Add $T_{out}$ – directly affects $T_{in}$ . Prediction error still unacceptably high
- Noise explained as hidden home state ➡ add history of observable information
  - Previous action

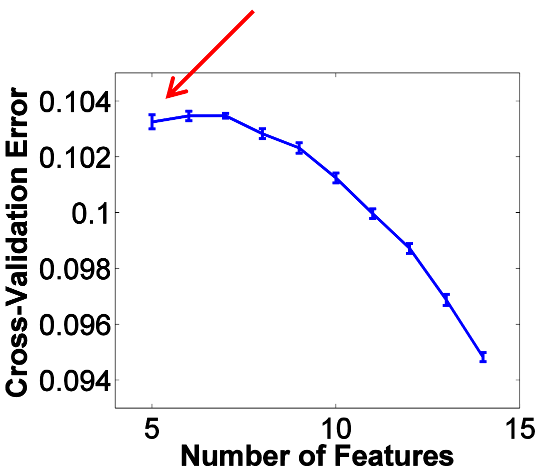


# Predicting $T_{in}$

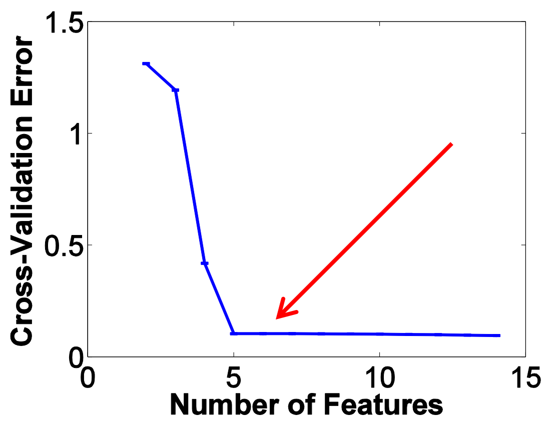
- Prediction error is unacceptably high – state <$T_{in}$, Time, $e_a$> doesn't capture enough information
- Add $T_{out}$ – directly affects $T_{in}$ . Prediction error still unacceptably high
- Noise explained as hidden home state ⇨ add history of observable information
  - Previous action
  - Measured $T_{in}$ history of 10 temperatures: <$t_0$>

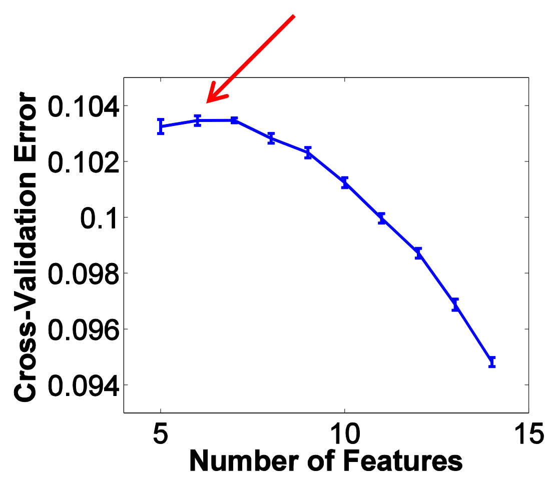


# Predicting $T_{in}$

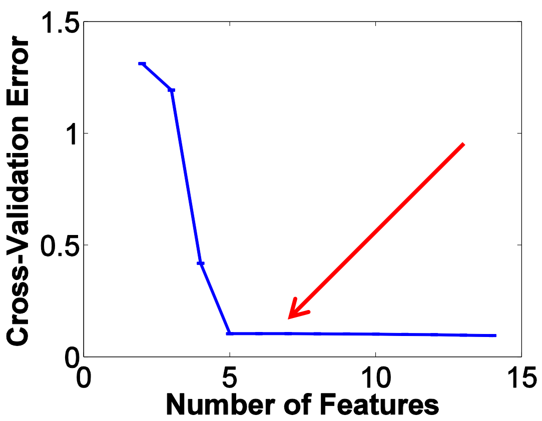
- Prediction error is unacceptably high – state $<T_{in}$, Time, $e_a>$ doesn't capture enough information
- Add $T_{out}$ – directly affects $T_{in}$ . Prediction error still unacceptably high
- Noise explained as hidden home state ⇨ add history of observable information
  - Previous action
  - Measured $T_{in}$ history of 10 temperatures: $<t_0, t_1>$

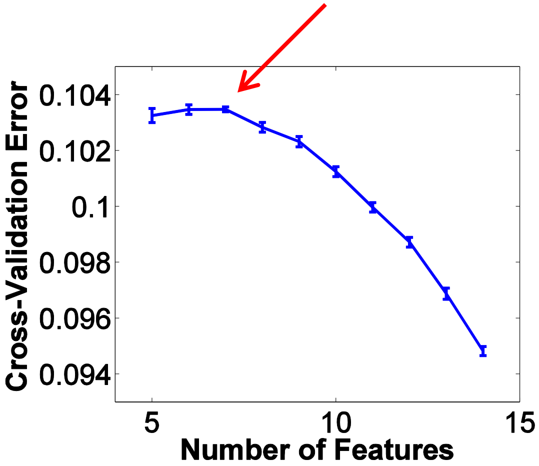


# Predicting $T_{in}$

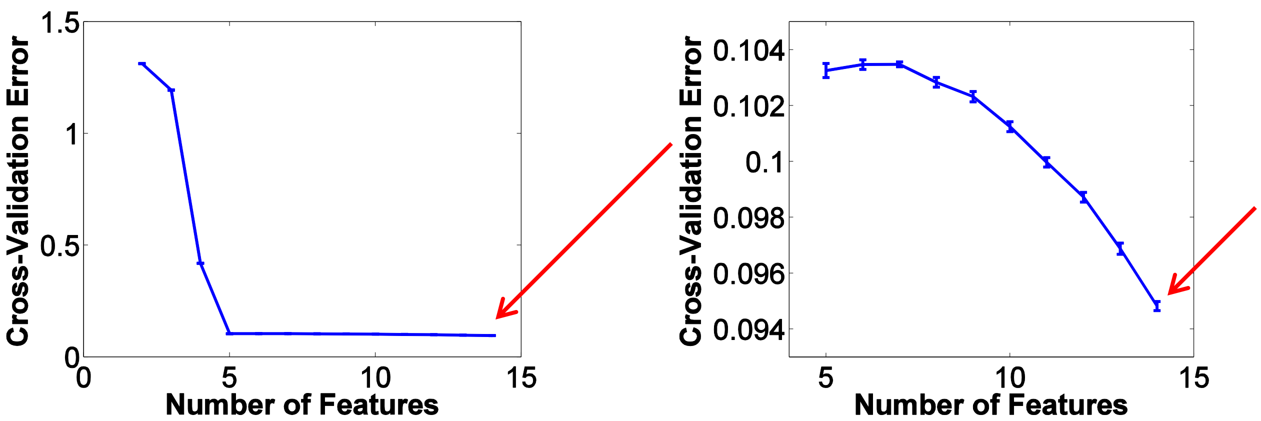
- Prediction error is unacceptably high – state $<T_{in}$, Time, $e_a>$ doesn't capture enough information
- Add $T_{out}$ – directly affects $T_{in}$ . Prediction error still unacceptably high
- Noise explained as hidden home state ⇨ add history of observable information
  - Previous action
  - Measured $T_{in}$ history of 10 temperatures: $<t_0, t_1, t_2>$



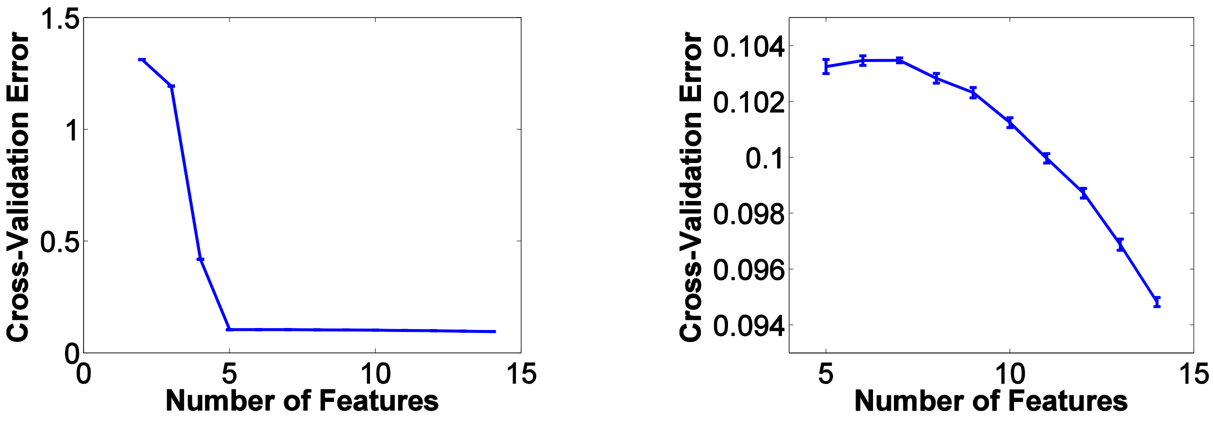
# Predicting $T_{in}$

- Prediction error is unacceptably high – state <$T_{in}$, Time, $e_a$> doesn't capture enough information
- Add $T_{out}$ – directly affects $T_{in}$ . Prediction error still unacceptably high
- Noise explained as hidden home state ⇨ add history of observable information
  - Previous action
  - Measured $T_{in}$ history of 10 temperatures: <$t_0$, $t_1$, $t_2$, …, $t_9$>



# Predicting $T_{in}$

- Prediction error is unacceptably high – state $\langle T_{in}, Time, e_a \rangle$ doesn't capture enough information
- Add $T_{out}$ – directly affects $T_{in}$ . Prediction error still unacceptably high
- Noise explained as hidden home state ⇨ add history of observable information
  - Previous action
  - Measured $T_{in}$ history of 10 temperatures: $\langle t_0, t_1, t_2, \ldots, t_9 \rangle$
  - Resulting state: $\langle T_{in}, T_{out}, Time, e_a, prevAction, t_0, \ldots, t_9 \rangle$



# Completing the state definition

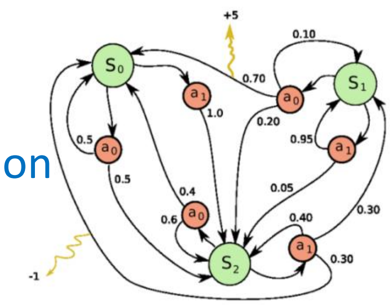
- Resulting state: $<T_{in}, T_{out}, \text{Time}, e_a, \text{prevAction}, t_0, \ldots, t_9 >$
- Can we predict the newly added variables?
- Trivially, except for $T_{out}$
- Therefore, add weatherForecast to state
- weatherForecast doesn't need to be predicted in our transition function
- This completes our state definition
- The final resulting state is:

  $<T_{in}, T_{out}, \text{Time}, e_a, \text{prevAction}, t_0, \ldots, t_{9,}\ \text{weatherForecast}>$

# Our Problem as a Markov Decision Process (MDP)

- **States:** <$T_{in}$, $T_{out}$, Time, $e_a$, prevAction, $t_0$, …,$t_9$, weatherForecast>
- **Actions:** {COOL, OFF, HEAT, AUX}

  1 : 0 : 2 : 4 ⟸ consumption ($e_a$) proportion



- **Transition:** unknown in advance ⟹ learned
- **Reward:** $-\, e_a - 100000\, \Delta^2_{6pm}$ *where:*

  $\Delta^2_{6pm}$ := (indoor_temp_at_6pm – required_indoor_temp_at_6pm)
- **Terminal States:** {s | s.time = 11:59pm}

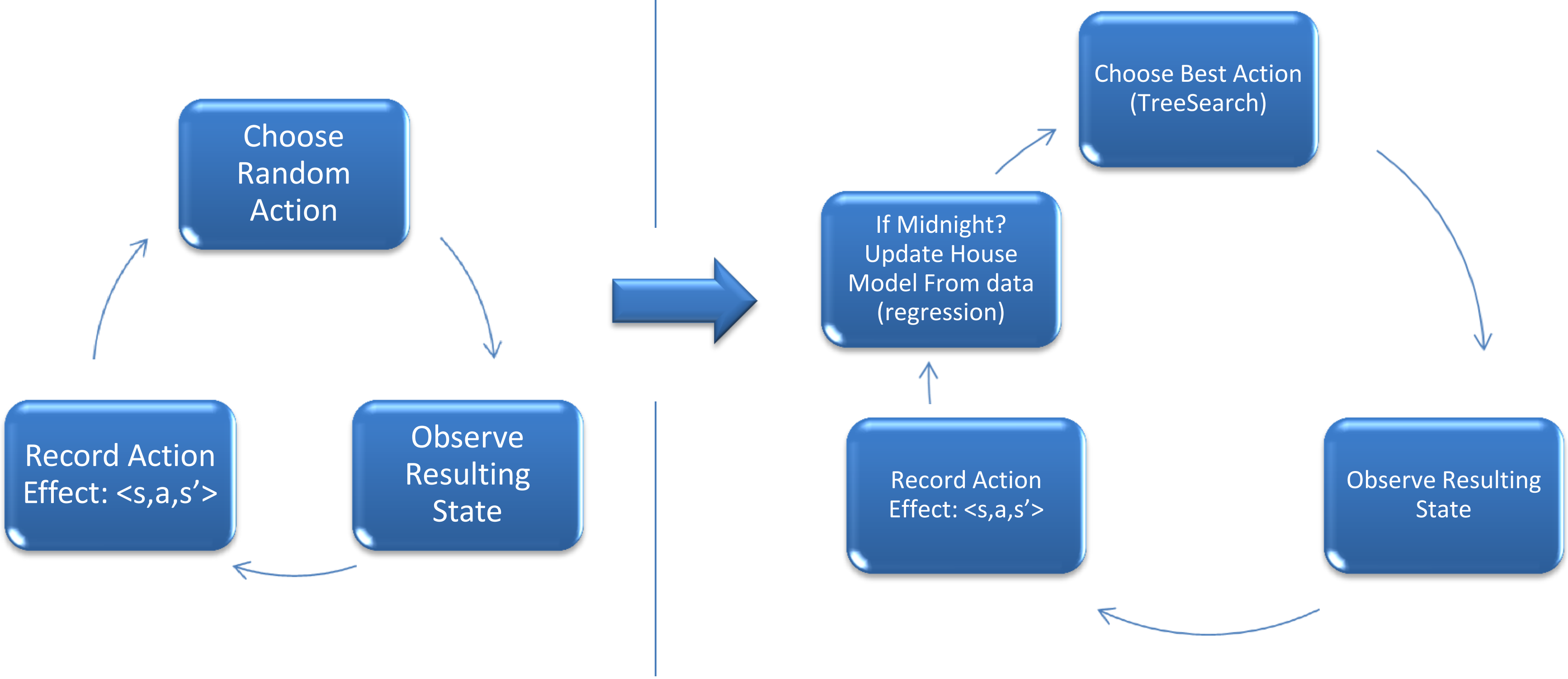
---

- Action taken every 6 minutes
  - Modeling a realistic lockout of the system
- State space is continuous and high dimensional

# Agent Operation

First 3 days:
exploration

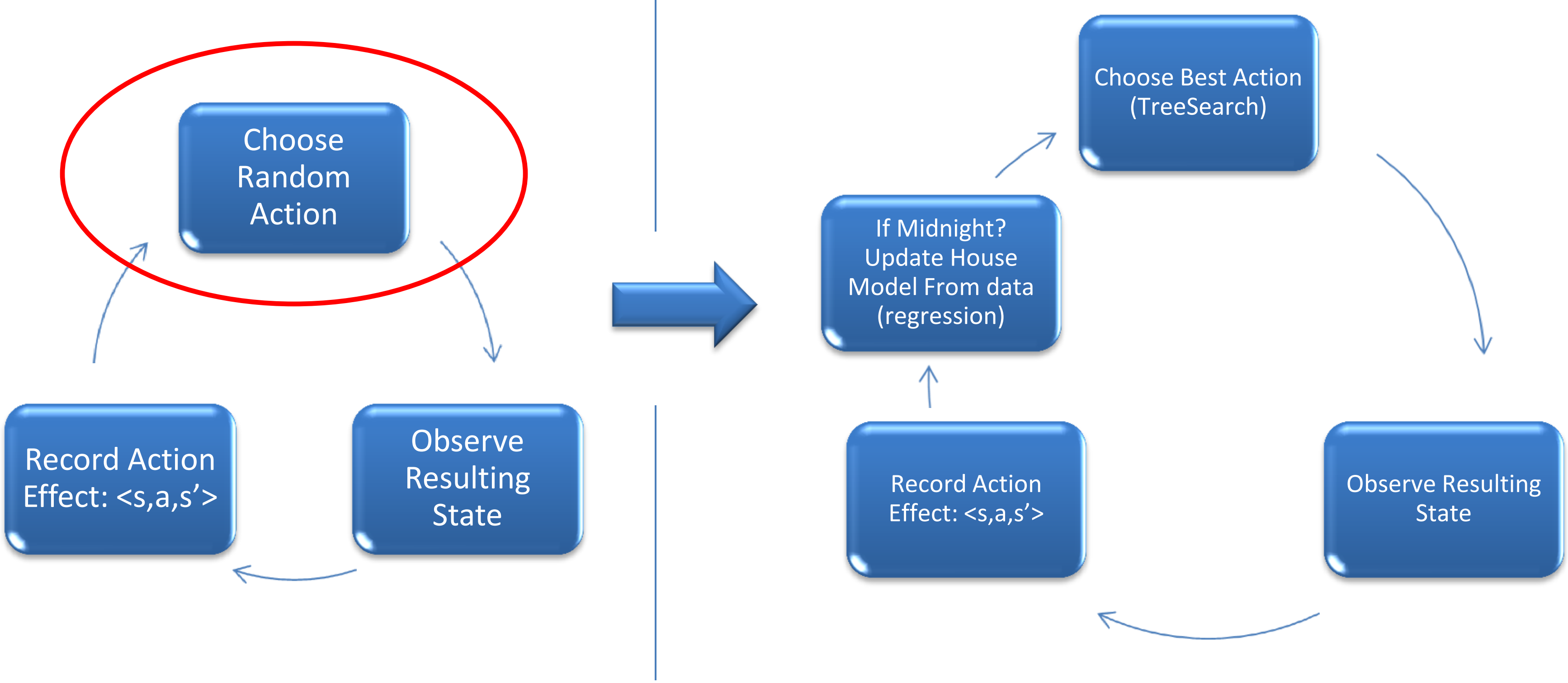
Starting day 4:
energy-saving
setback policy



# Agent Operation

First 3 days:
exploration

Starting day 4:
energy-saving
setback policy



# Exploration

- Random actions for 3 days
- Could use more advanced exploration policy
- However, this is still a realistic setup

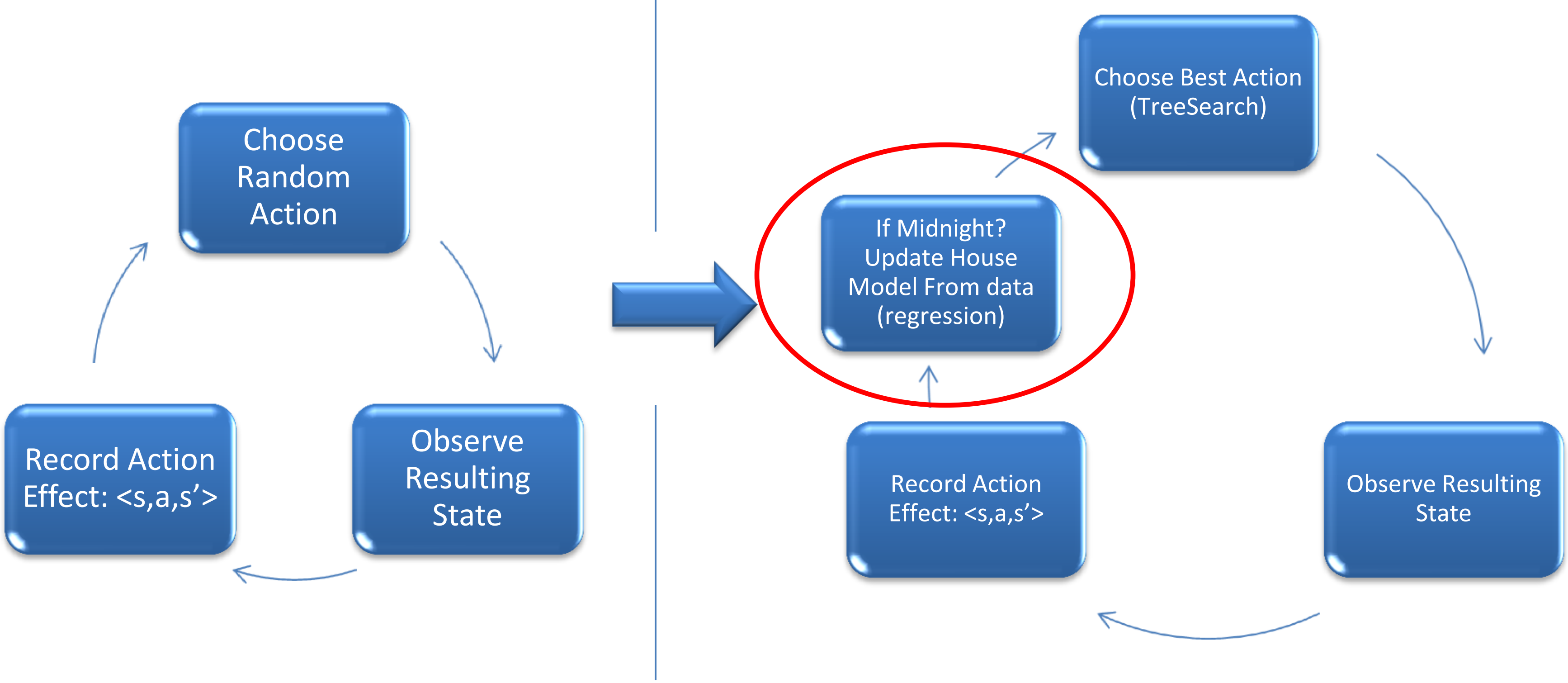
# Exploration

- Random actions for 3 days
- Could use more advanced exploration policy
- However, this is still a realistic setup
  - For instance when occupants are traveling during the weekend

# Agent Operation

First 3 days:
exploration

Starting day 4:
energy-saving
setback policy



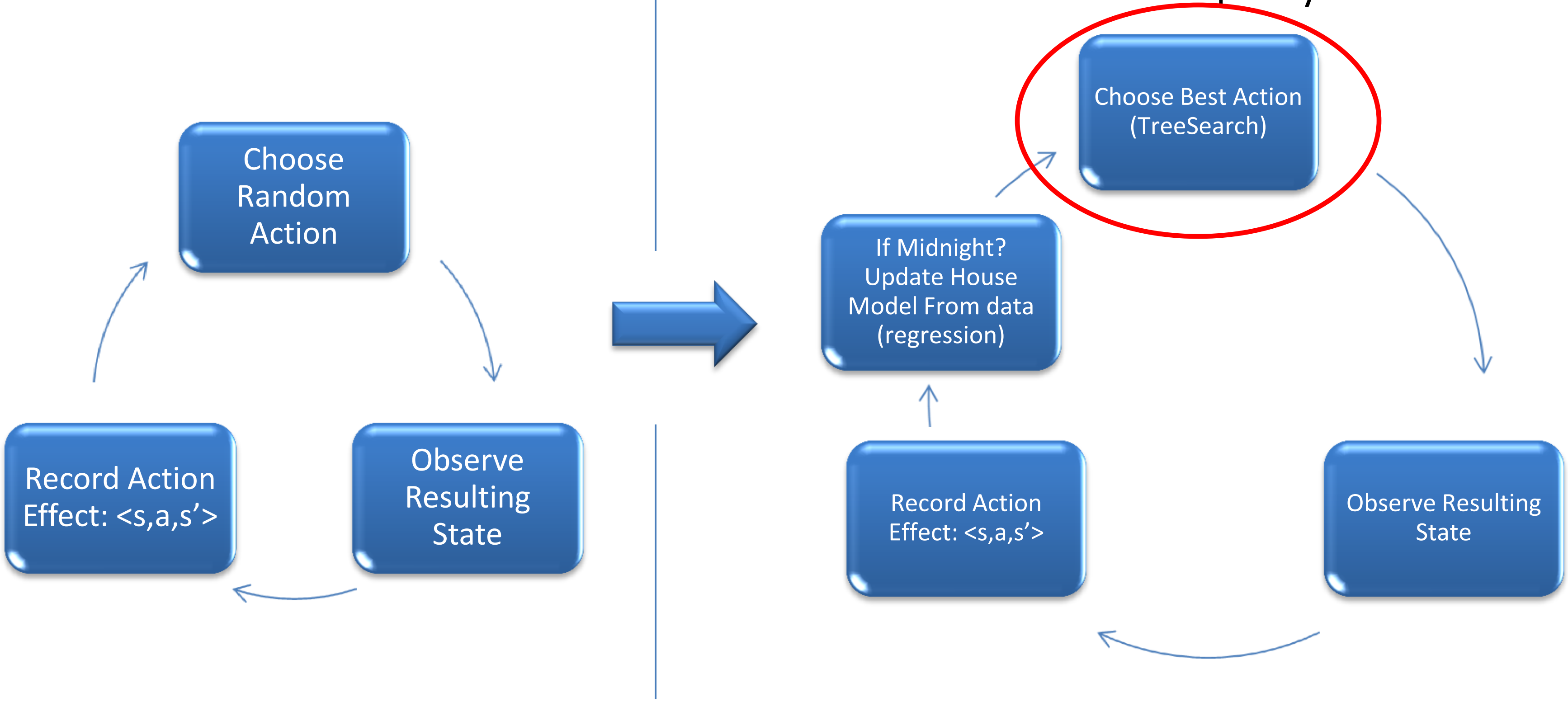
# Update House Model from Data

- Every midnight, use all the recorded data $<s, a, s'>$ to estimate the house's transition function
- Linear Regression to estimate $<s,a> \longrightarrow s'$

# Agent Operation

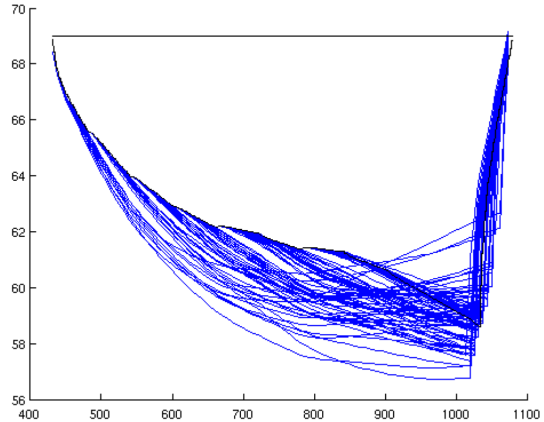
First 3 days:
exploration

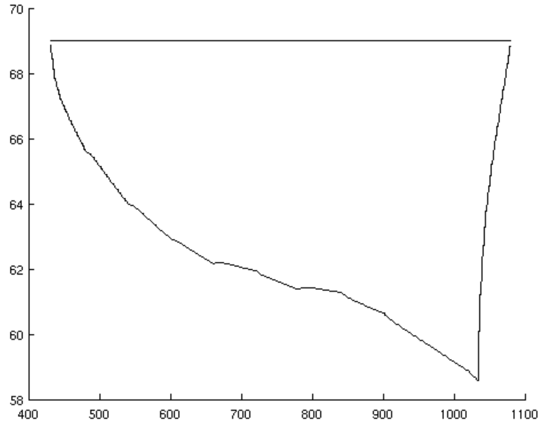
Starting day 4:
energy-saving
setback policy



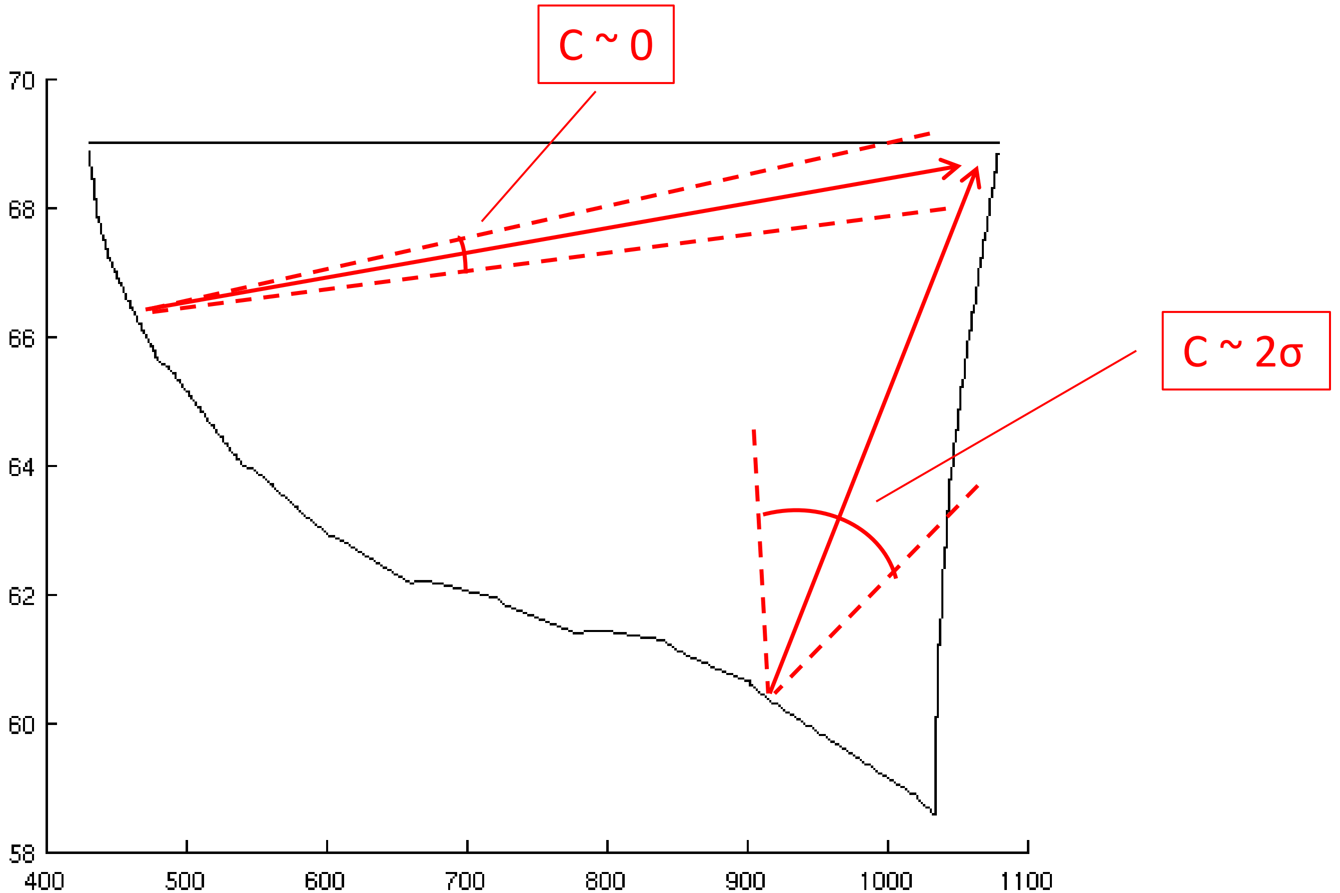
# Choosing the Best Action

- Dealing with continuous high-dimensional state
- Impractical to compute a value function
- ⇨ Run a tree search at every step
- Choose the first action of the best search as the next action



# Safety Buffer in a Tree Search



# Results

- Simulate 1 year under different weather conditions
- 21 residential homes of sizes 1000-4000 $ft^2$
- Using real data weather recorded in

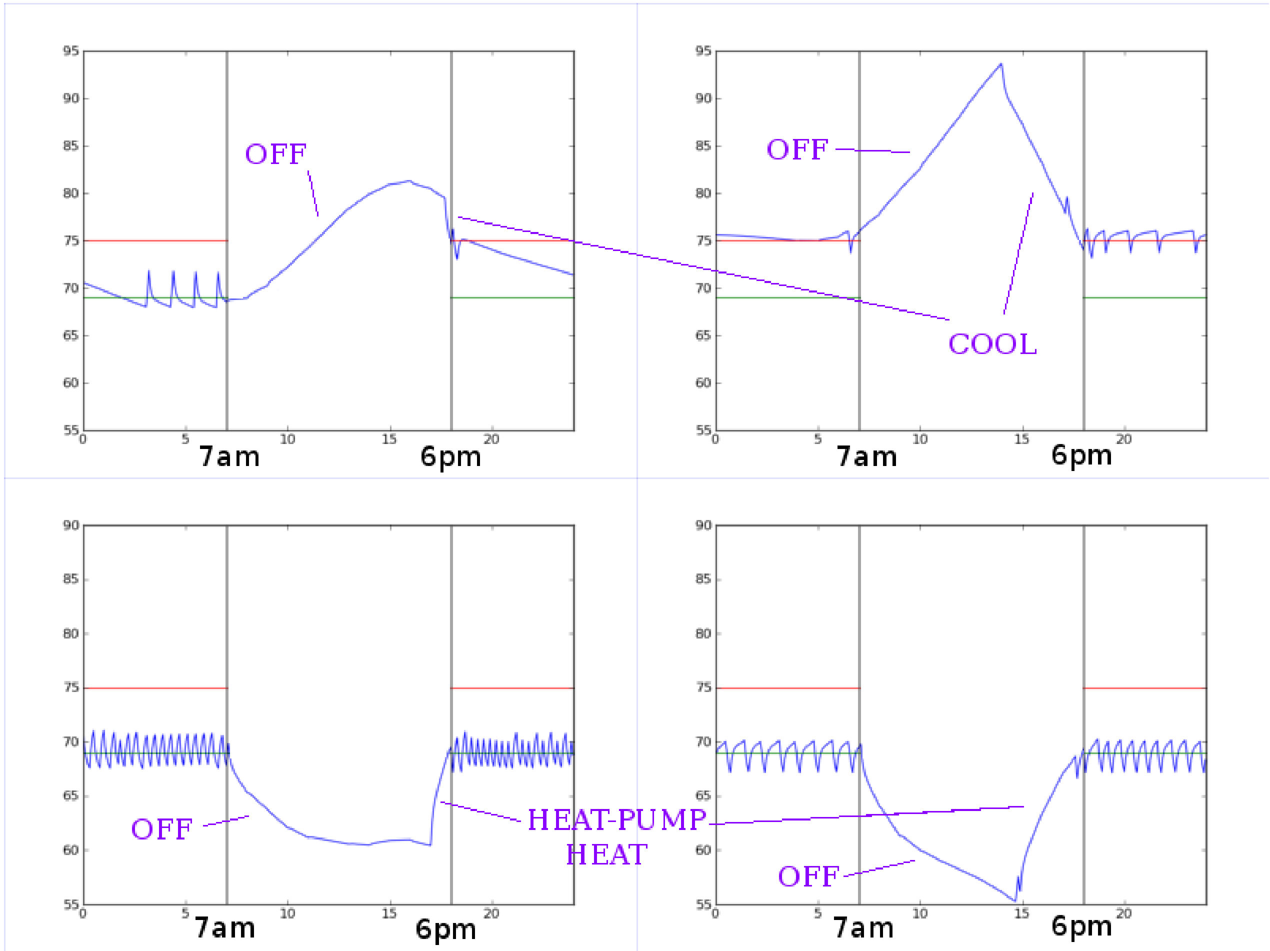
NYC 

Boston 

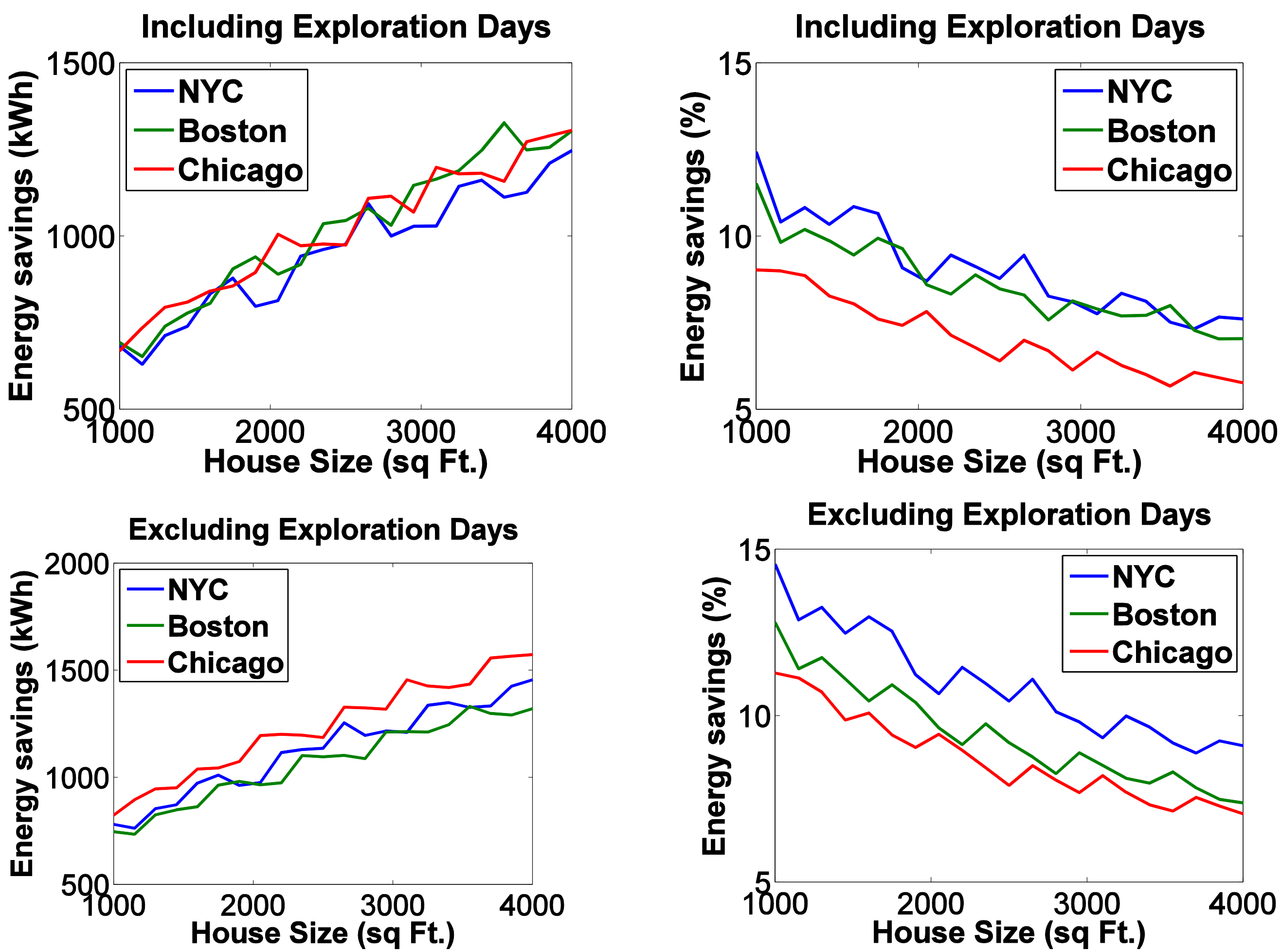
Chicago 

- Why cold cities? Since heating consumes 2x-4x more energy

# Temperature Graphs – Learned Setback Policy

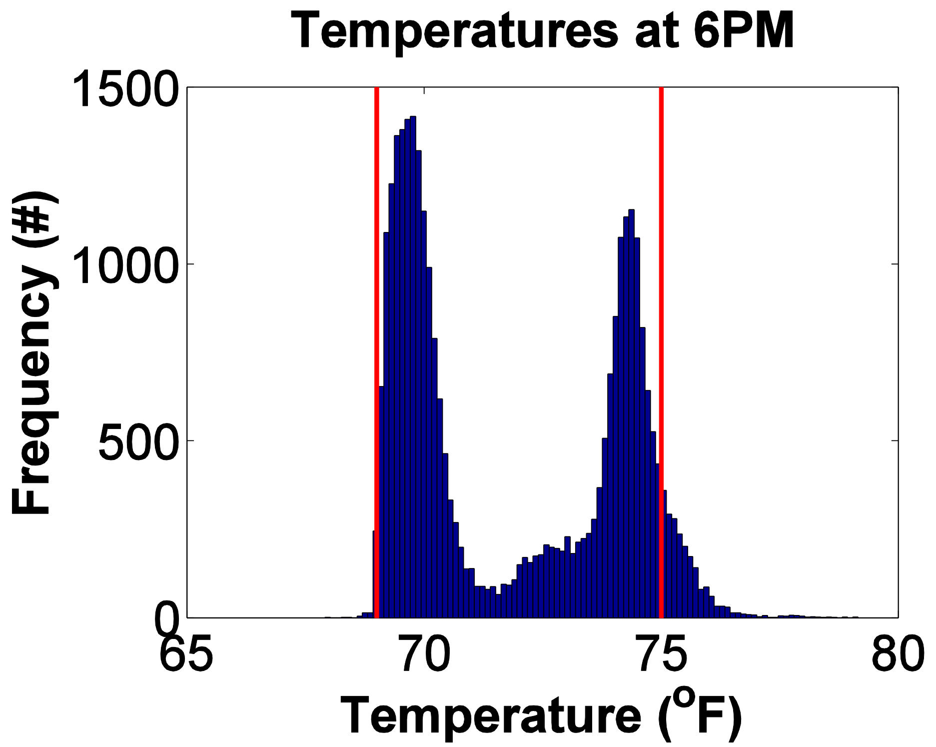


# Energy Savings



# Comfort Performance

- In more than 22,000 simulated days



# Ablation Analysis

Table 1: Ablation Analysis

| Analysis Type | | Energy Consumption (kWh) | Comfort Violations (#) | Range of 6pm Temp. |
|---|---|---|---|---|
| Removed Feature | prevAct+hist+ conf | 1112(+9.5%) | 232 | 60.1-84.4 |
| | prevAct+hist | 1070(+5.4%) | 193 | 60.8-80.9 |
| | conf | 1024(+0.8%) | 138 | 67.5-78.3 |
| | hist | 1016(+0.0%) | 133 | 67.1-77.7 |
| | prevAct | 1015(+0.0%) | 65 | 67.8-76.5 |
| Other conf. bounds | $2\sigma$ | 1090(+7.3%) | 29 | 69.0-78.5 |
| | $c = 2$ | 1039(+2.3%) | 27 | 69.0-77.8 |
| **Final Agent** | | **1015** | **23** | **68.8-76.6** |

- Removing features and their combinations
  - State features:
    - prevAct: previousAction
    - Hist: temperature history $t_0, \ldots, t_9$
  - conf: confidence buffer
- Setting other values to the confidence bound

# Related Work

- [Rogers et al. 2011] – adaptive thermostat that tries to minimize price & peak demand rather than the total amount of energy.

- [Hafner and Riedmiller 2011; Kretchmar 2000] – use RL inside an HVAC systems, but for tuning the system itself.

- [T. Peffer et al. 2011] – How people use thermostats in homes

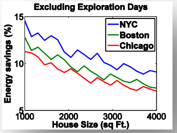
- [Powell 2011] – Approximate Dynamic Programming

- Learning thermostats
  - Commercial companies: NEST, more…
  - Do not publish the algorithms used

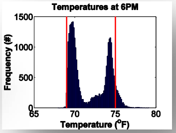
# Summary

- A complete, adaptive, RL agent
  for controlling a heat-pump thermostat

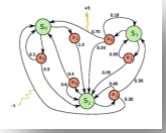
- Achieves 7%-14.5% yearly energy savings in simulation, while satisfying comfort requirements



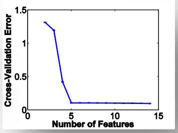

+



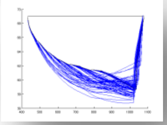

- Techniques:
  - Carefully defined the problem as an MDP
  - Carefully chose a state representation
  - Using an efficient, specialized tree-search



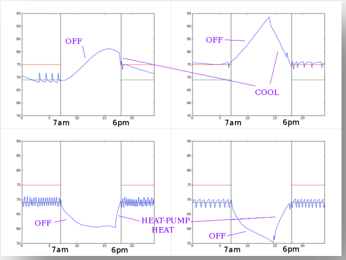
+




+



- Experiments run on a range of homes and weather conditions




Thank you!