Adaptive Power Management for Energy Harvesting Sensor Nodes using Reinforcement Learning

Shaswot Shresthamali, Masaaki Kondo, Hiroshi Nakamura
The University of Tokyo

CONTEXT

Energy Harvesting Sensor Nodes

http://www.libelium.com/resources/images/content/products/plug-sense/details/solar_powered_photo.png

Sensor Node
(capable of varying the duty cycle)

+

Battery

+

Energy Harvesting Module

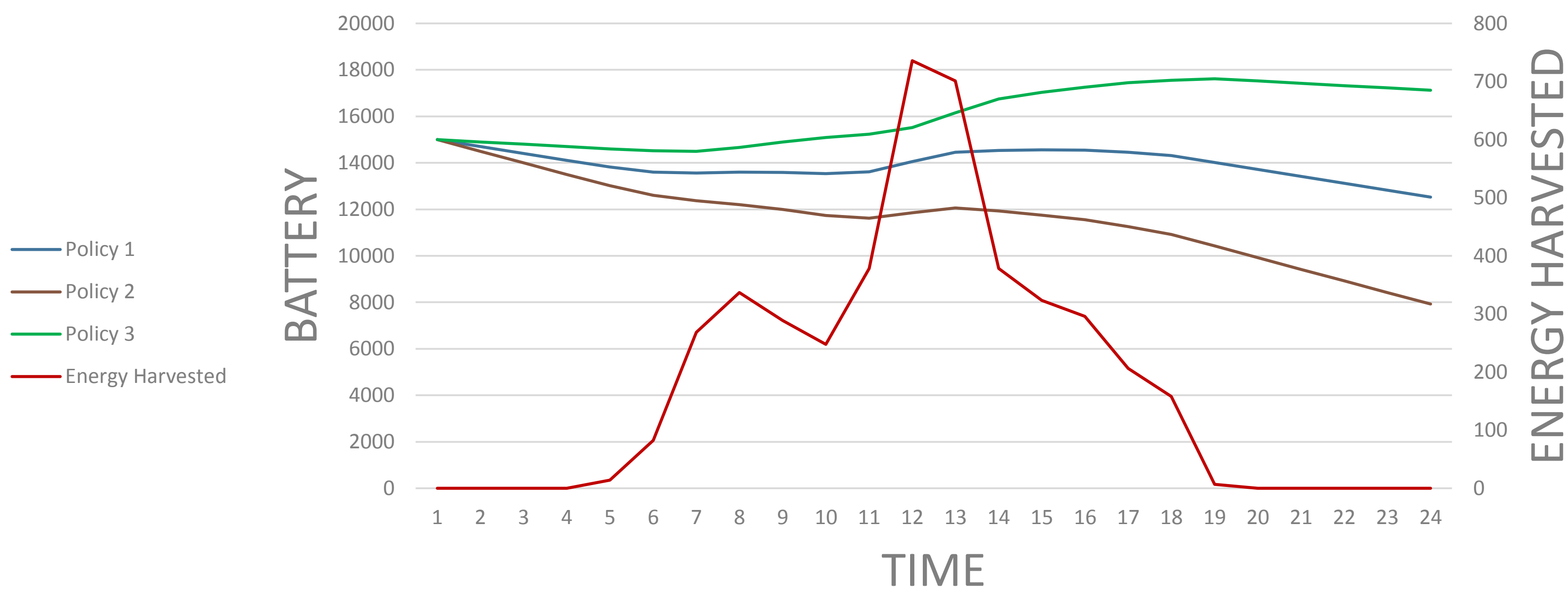
Theoretically capable of perpetual operation

Challenge I

Say your battery is at 75% and there is plenty of sunshine

Do you

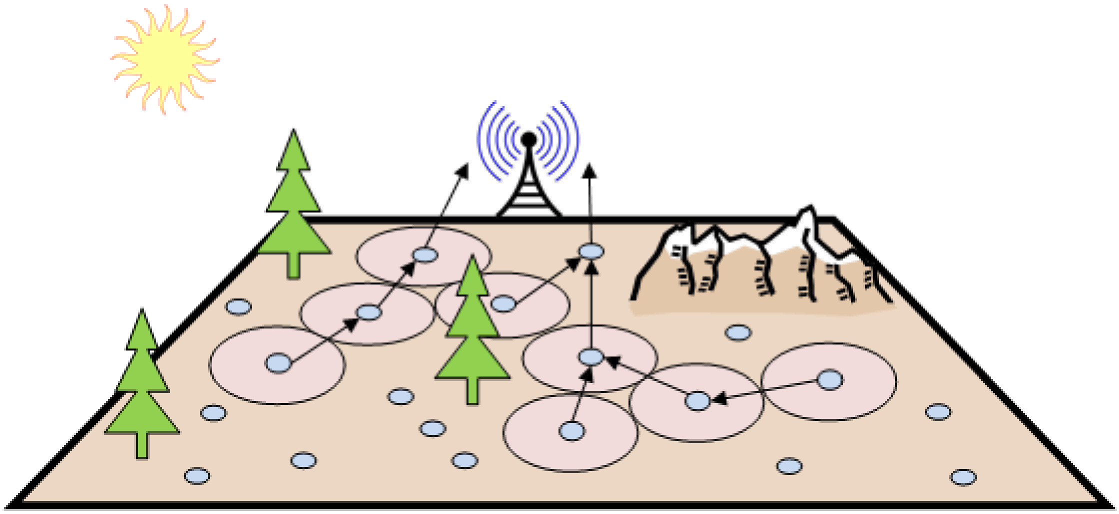
- Use the solar power to charge your battery only
- Use the solar power to charge your battery and drive the sensor node. If so, then with what proportion?



Challenge II

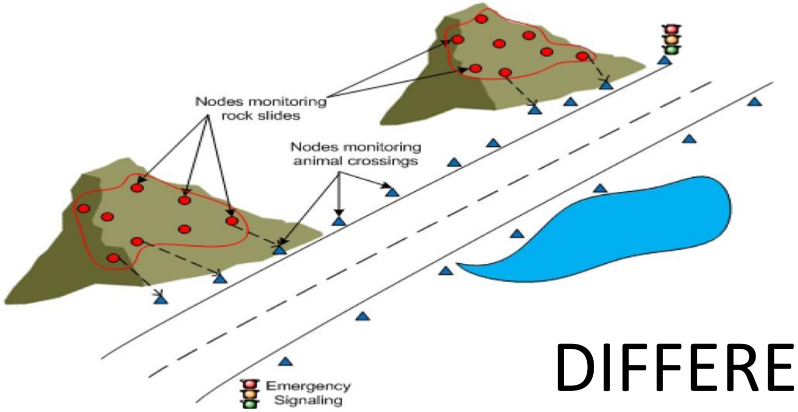
Environmental Sensor Networks – P.I. Corke et. al.

MOVING SENSORS



https://sites.google.com/site/sarmavrudhula/home/research/energy-management-of-wireless-sensor-networks

DIFFERENT ENVIRONMENTS



DIFFERENT SENSORS

http://www.mdpi.com/sensors/sensors-12-02175/article_deploy/html/images/sensors-12-02175f5-1024.png

Challenge III

BILLION AND TRILLIONS OF NODES

https://bendeetech.files.wordpress.com/2015/12/theinternetofthings2-540x334.jpg?w=350&h=200&crop=1

Challenges II and III

When dealing with *TRILLIONS* of sensor nodes,

Customizing each node is *impractical, impossible*

- Nodes should ***OPTIMIZE*** themselves.
- Nodes should ***ADAPT*** to their changing environments.

ENERGY HARVESTING NODES NEED TO BE

ADAPTABLE

SELF CALIBRATING

What this presentation is about

To demonstrate how to overcome the challenges by using Reinforcement Learning (RL)

- Brief introduction to Reinforcement Learning
- Our approach using RL
- How this strategy performs ***compares*** to other methods
- How this strategy ***adapts*** to changing environment

OBJECTIVES

Objectives

Energy Neutral Operation (ENO)

- Energy consumed = Energy harvested

Maximize Performance

- Maximize Duty Cycle

Minimize Battery Downtime

- Battery should never drop to zero

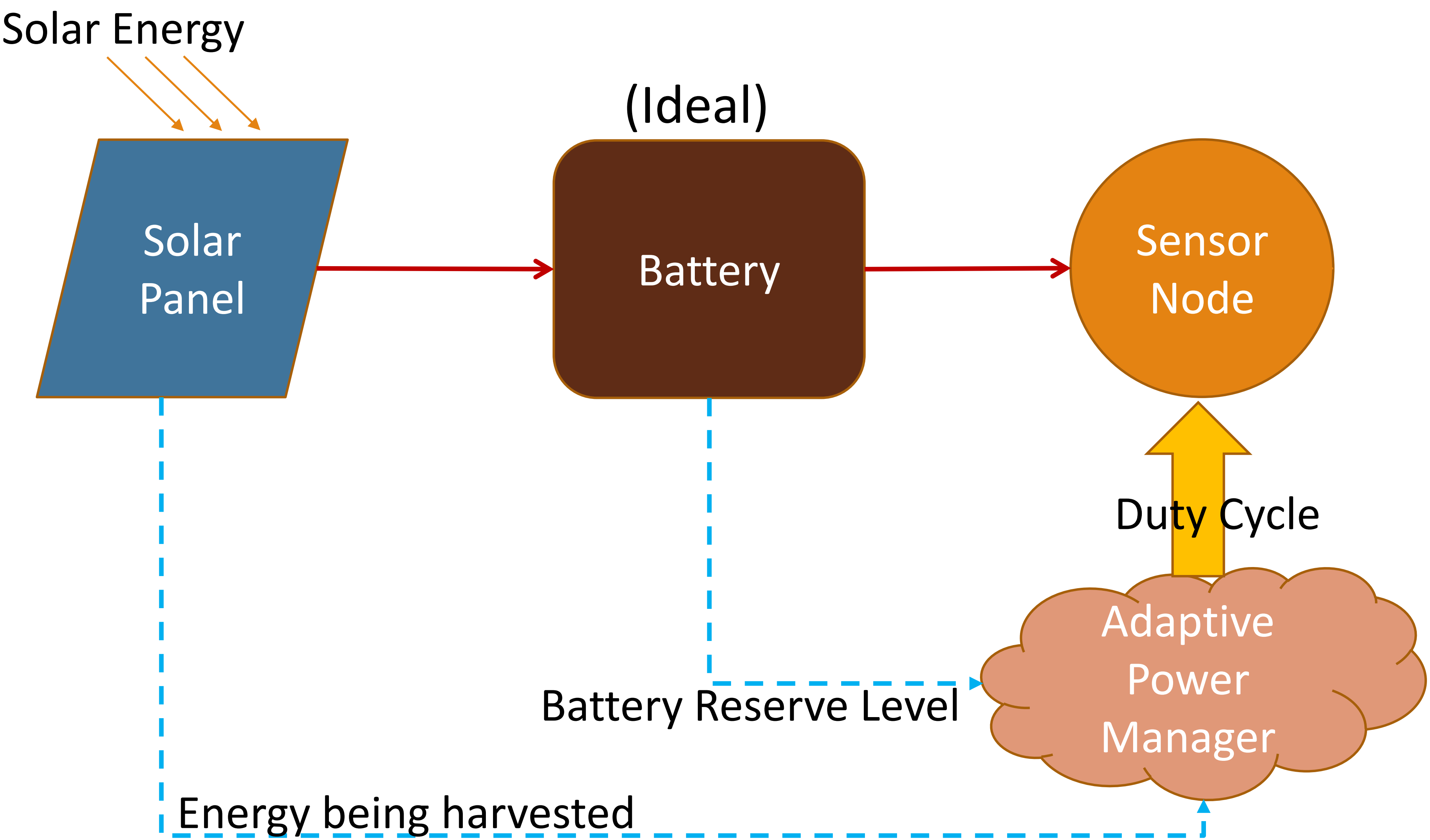
Minimize Energy Waste

- Battery should not overcharge

$$\text{Energy Waste} = \text{Energy Harvested} \\ -\text{Energy consumed by Node} \\ -\text{Energy to charge battery}$$

SYSTEM MODEL

System Model



REINFORCEMENT LEARNING (RL)

A BRIEF INTRODUCTION

What is RL

Type of Machine Learning - Learns by interacting with Environment

Suited for Sequential Decision Making Tasks

Map situations (states) into actions – receive as much reward as possible

Based on iterative process of trial and error – similar to how humans learn. (*Search and Memory*)

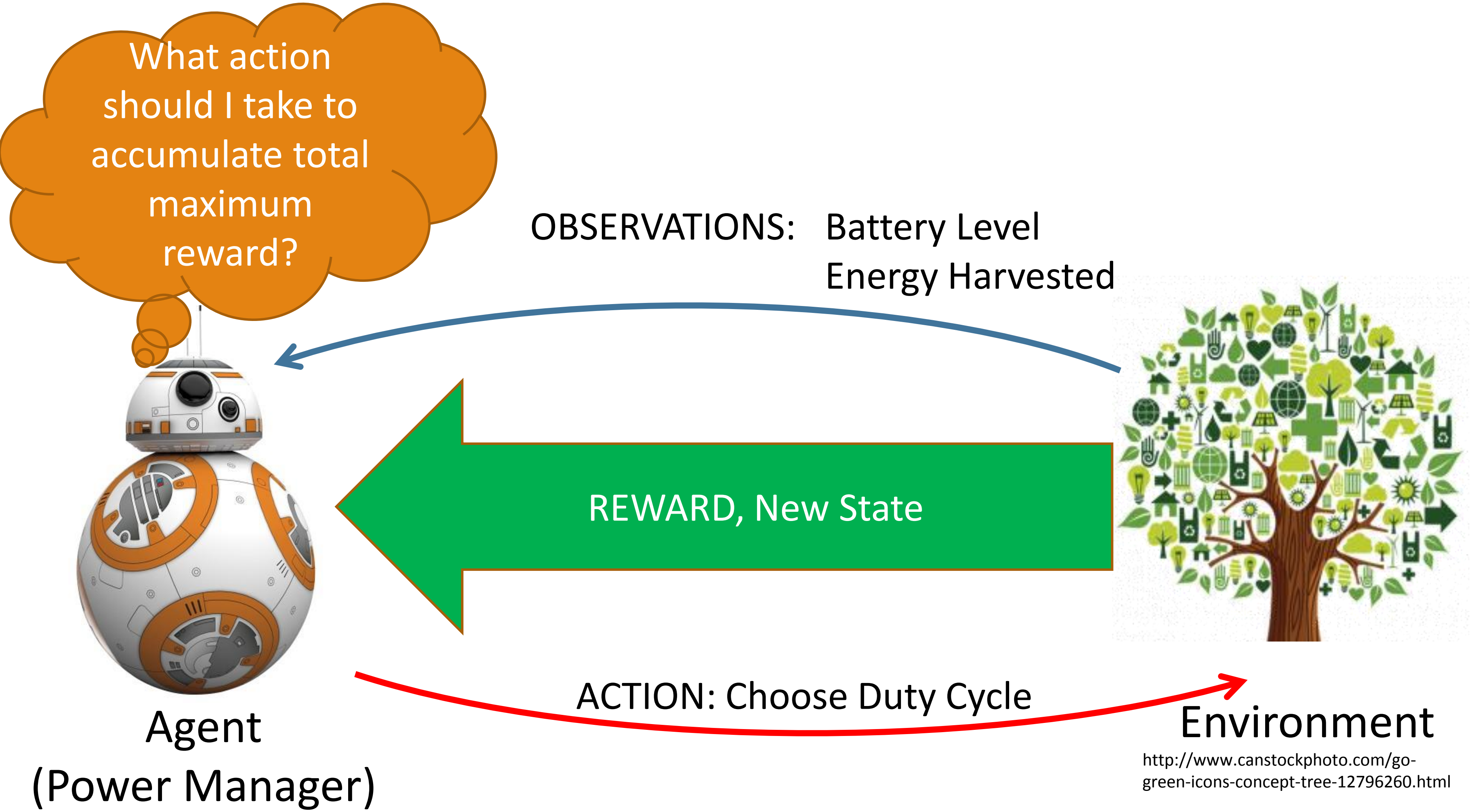
Why Reinforcement Learning

By using RL, it is possible

- To optimize nodes with raw high level data and minimal human input.

- To adapt to changes in the environment parameters.

Reinforcement Learning



http://www.canstockphoto.com/go-green-icons-concept-tree-12796260.html

http://wedreamabout.com/product/bb-8-droid-the-coolest-star-wars-toy-ever

Reinforcement Learning

The question is:

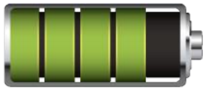
WHICH ACTION TO TAKE WHEN YOU ARE IN A GIVEN STATE?

EXAMPLE:

Lots of sunlight | Battery at 60%

Do you

[1] 

[2] 

- drive the sensor node at full strength without recharging?
- drive the sensor node at half strength with partial charging?

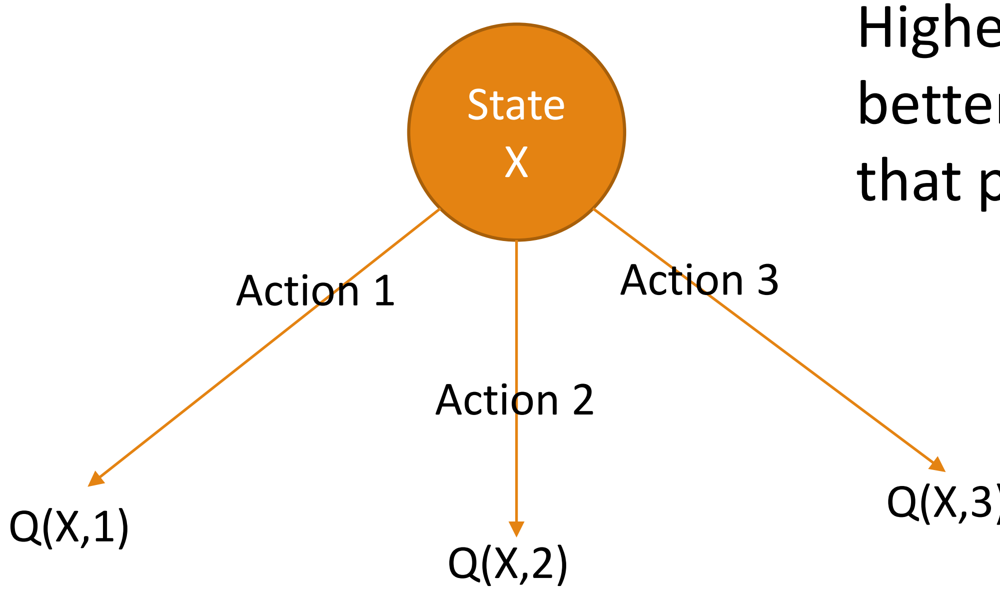
[1] https://spaceplace.nasa.gov/sun-corona/en/
[2] https://handyenergy.ru/

Q- Value

Assign every state-action pair → Q-Value, Q(s,a)

Q(s,a) means if the agent

- Starts from state *s*
- Takes action *a*
- Q(s,a) is the total reward it can expect in the best case scenario



Higher the Q-value, better the action for that particular state

Q-Learning

Challenge → Determining the Q-Values for all state-action pairs.

Q-table -> contains Q-Values of all possible state-action pairs

Accomplished by Q-Learning Algorithm

- Q-values are learned by interacting with environment.
- Iterative Process
- Bootstrapping approach

Q-Learning Algorithm

Q-Learning Algorithm

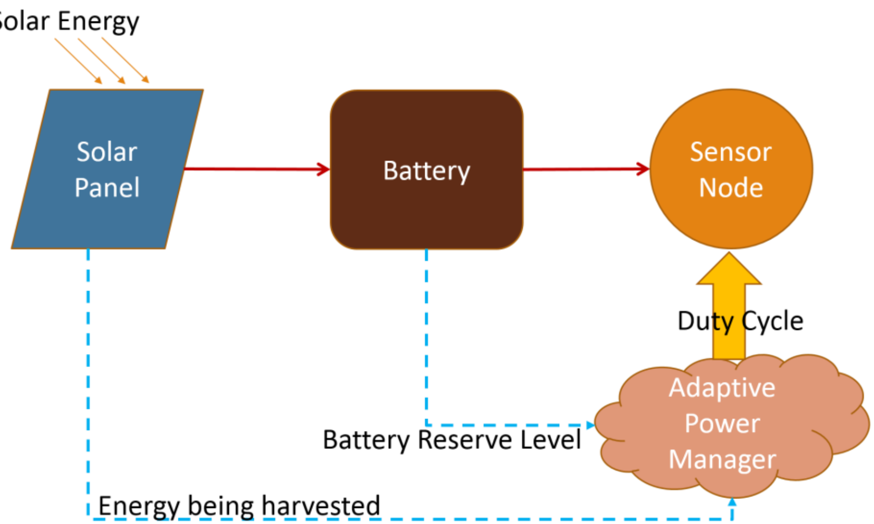
- Use arbitrary estimates for Q-values
- Use these estimates to decide on actions
- Update Q-table by using the rewards received
- Repeat until Q-value sufficiently converge

EXPERIMENTS ON ADAPTIVE POWER MANAGEMENT USING Q-LEARNING

State Space

State is defined by :

- amount of battery remaining
  - 200 possible levels
- amount of energy harvested
  - 5 possible levels



Total possible states: 200 x 5 = 1000

Action Space

Action: Choose duty cycle of the sensor node

$$A = a(t_k) \in \{10\%, 20\%, 30\% \ldots .100\%\}$$

$10\% \longrightarrow 50$ mW

$50\% \longrightarrow 250$ mW

$100\% \longrightarrow 500$ mW

Reward Function

The reward depends on:

- Distance from energy neutrality at time t_k

$$\Delta e_{neutral}(t_k) = e_{harvest}(t_k) - e_{node}(t_k)$$

- Amount of battery remaining

RESULTS

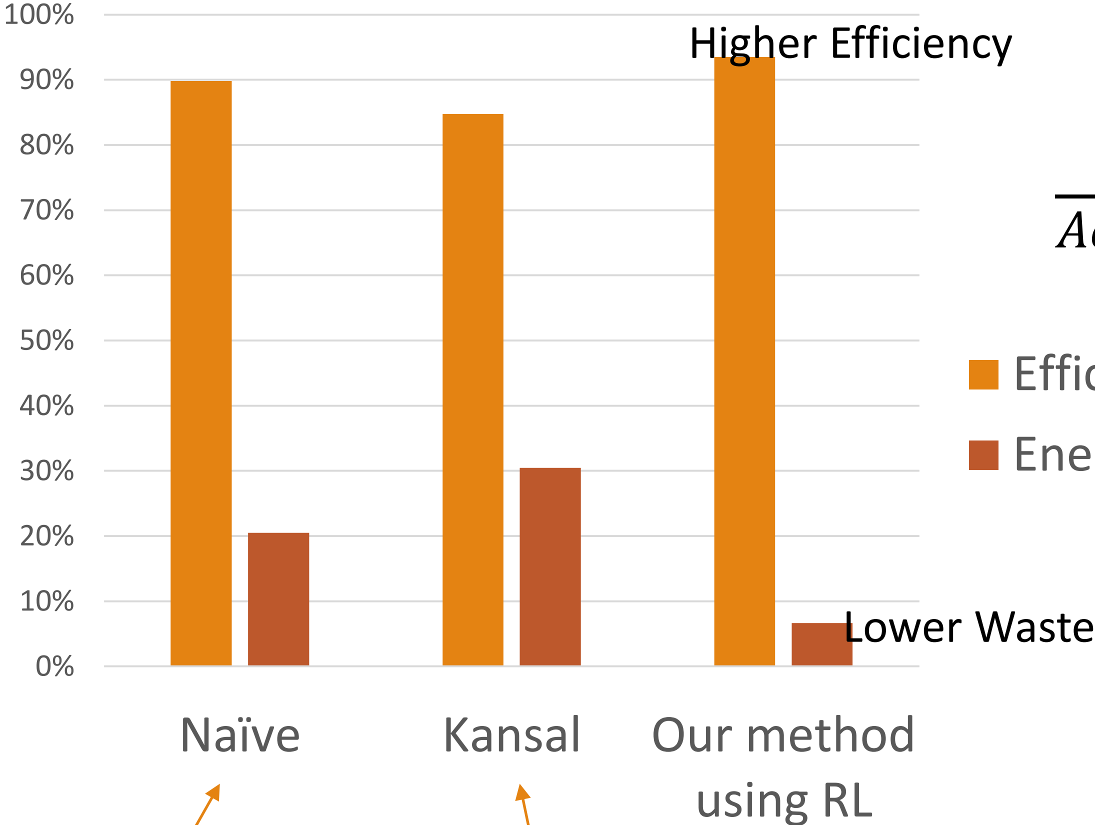
Training and Testing

Training: Tokyo (2000 to 2009)

Testing : Tokyo (2010)

- Compare it with other methods.
- Adaptation to diurnal and seasonal variations.
  - Greedy and ε-greedy Implementations

Comparison with other methods



Efficiency(%)

Energy Wasted(%)

$$\frac{Actual\ Duty\ Cycle}{Achievable\ Maximum\ Duty\ Cycle}$$

$$\frac{Total\ Energy\ Wasted}{Total\ Energy\ Harvested}$$

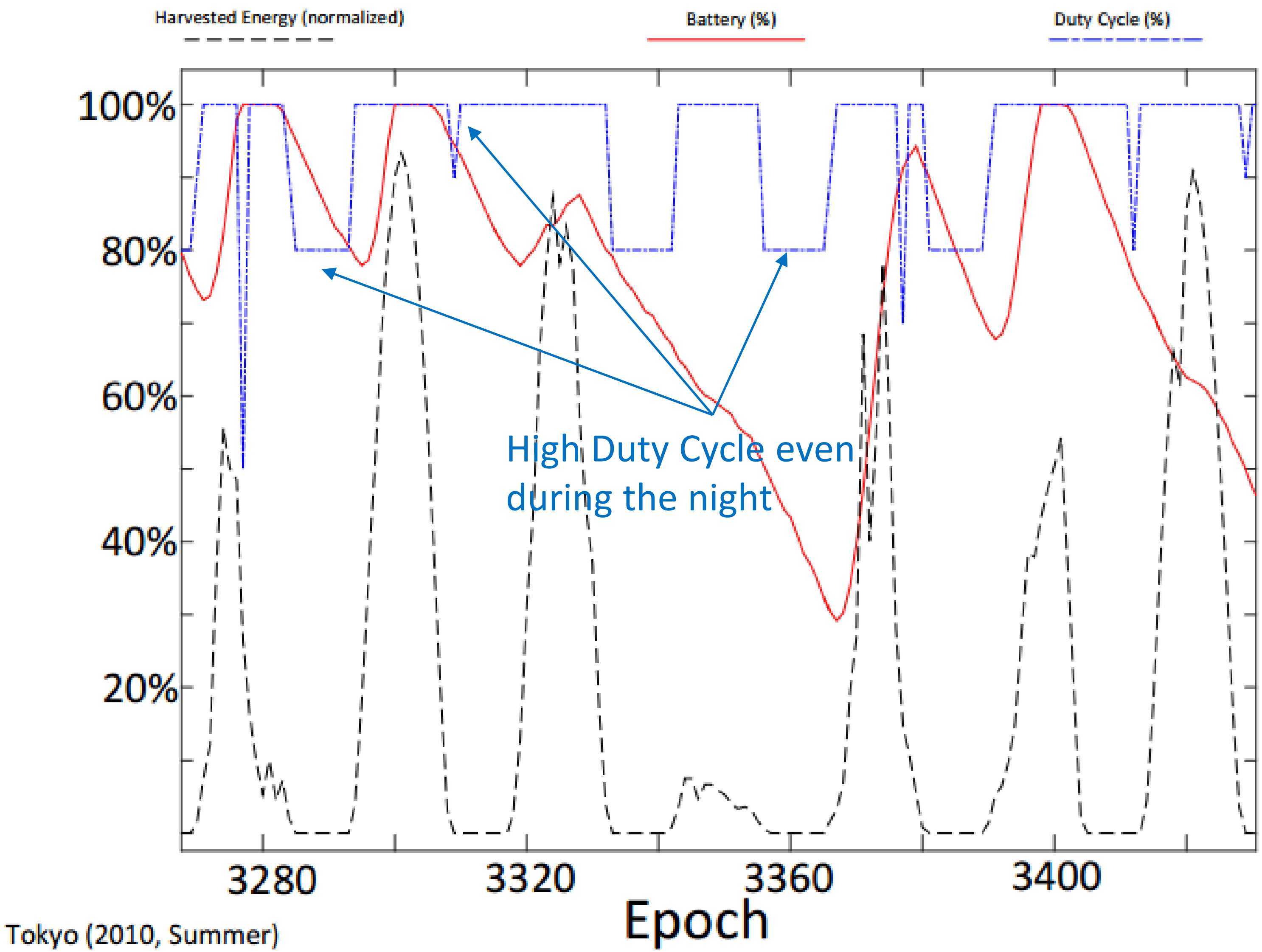
Duty Cycle is proportional to battery level

Fix duty cycle for present day by predicting total energy for next day

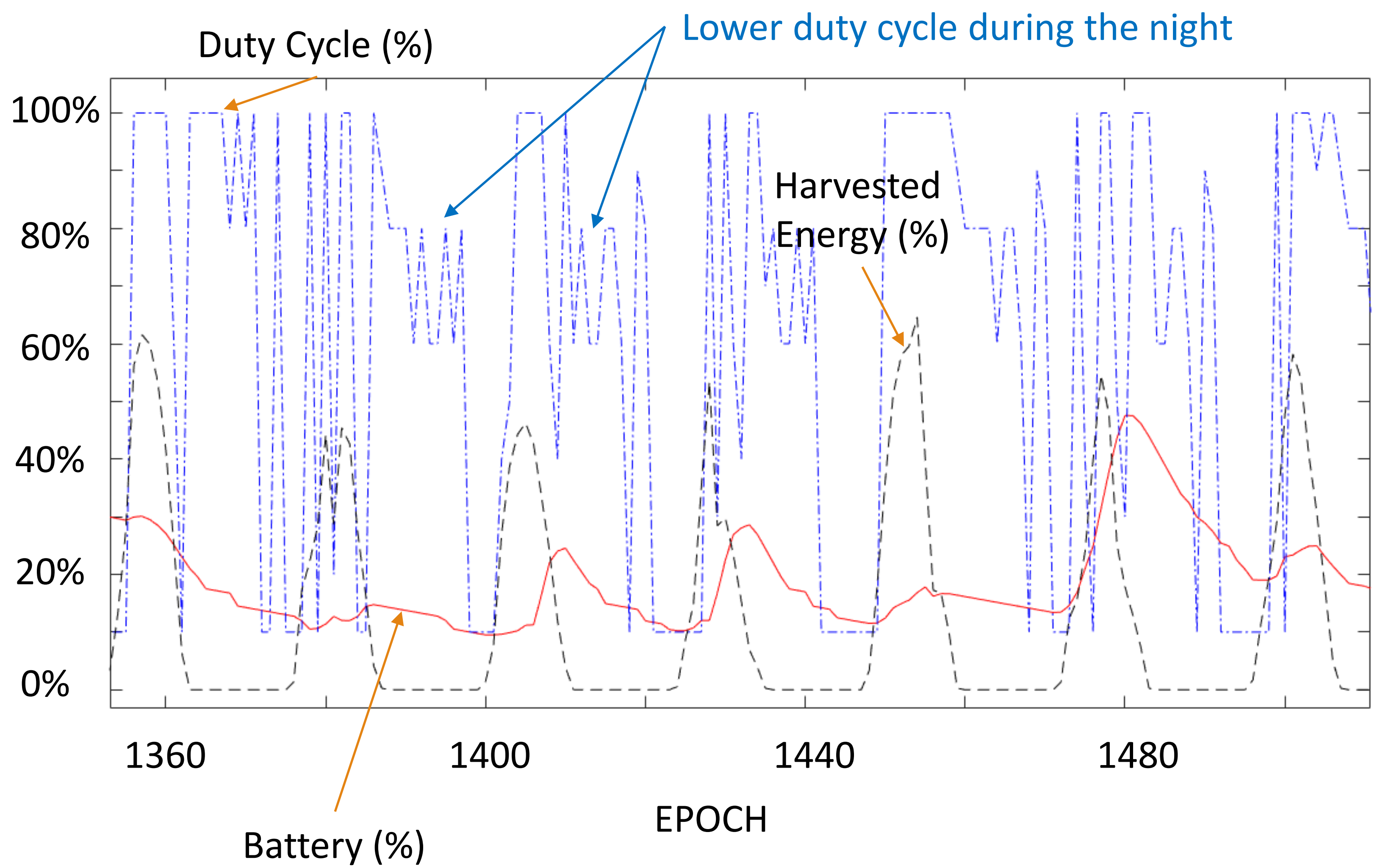
$$\begin{aligned} &Energy\ Waste \\ &= Energy\ Harvested \\ &\qquad - Node\ Energy \\ &\qquad - Charging\ Energy \end{aligned}$$

ADAPTATION TO SEASONAL CHANGES

Performance in Summer



Performance in Winter



ADAPTATION TO CHANGE IN LOCATION

Implementation: ε-greedy approach

Perfect Q-convergence takes too long.

Instead, use ε-greedy approach with non-converged Q-table.

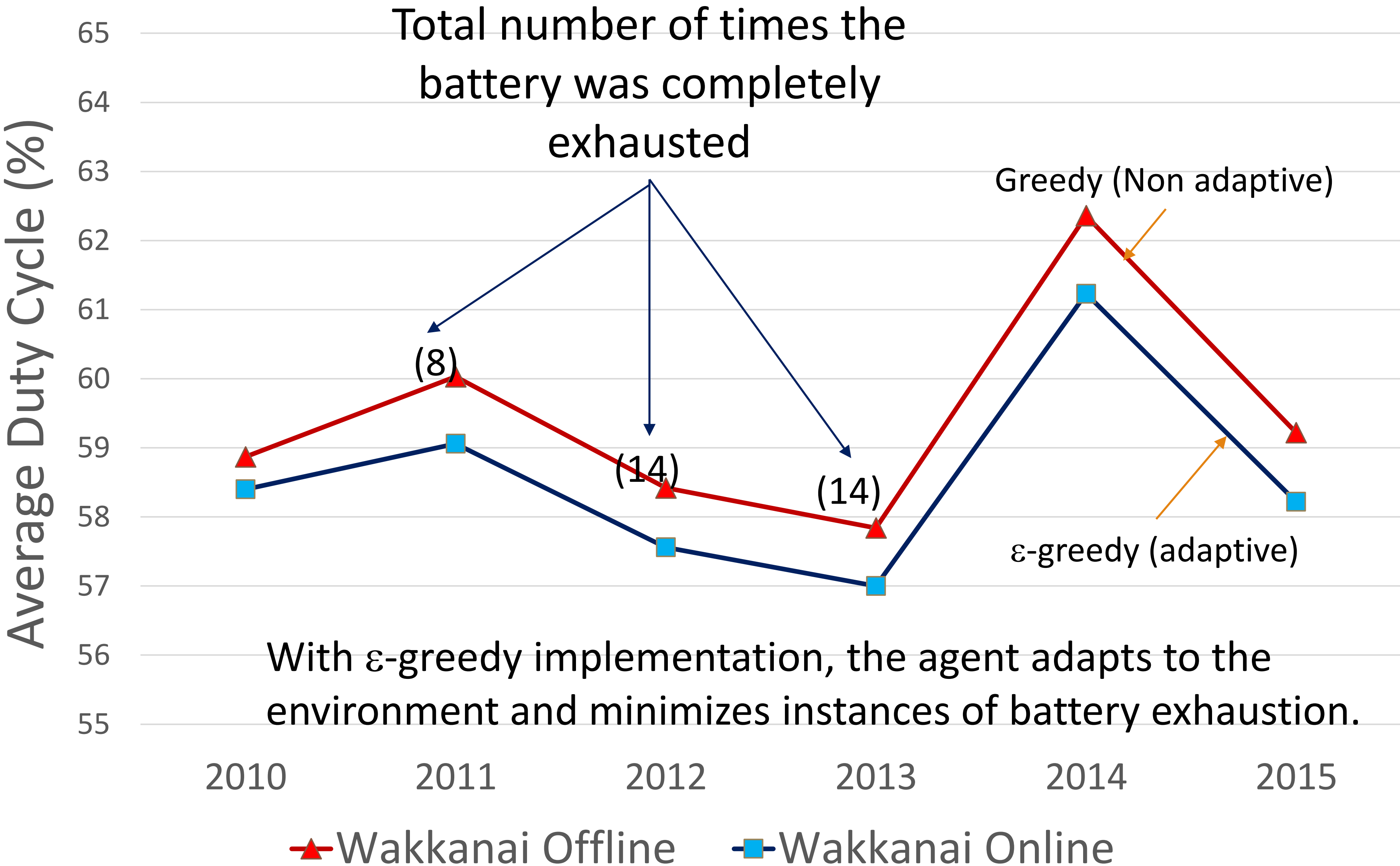
ε-greedy approach:

- Take the best action by default.
- Take a random action with probability ε.
- Increasing ε → Exploration
- Decreasing ε → Exploitation

Adaptation to change in climate

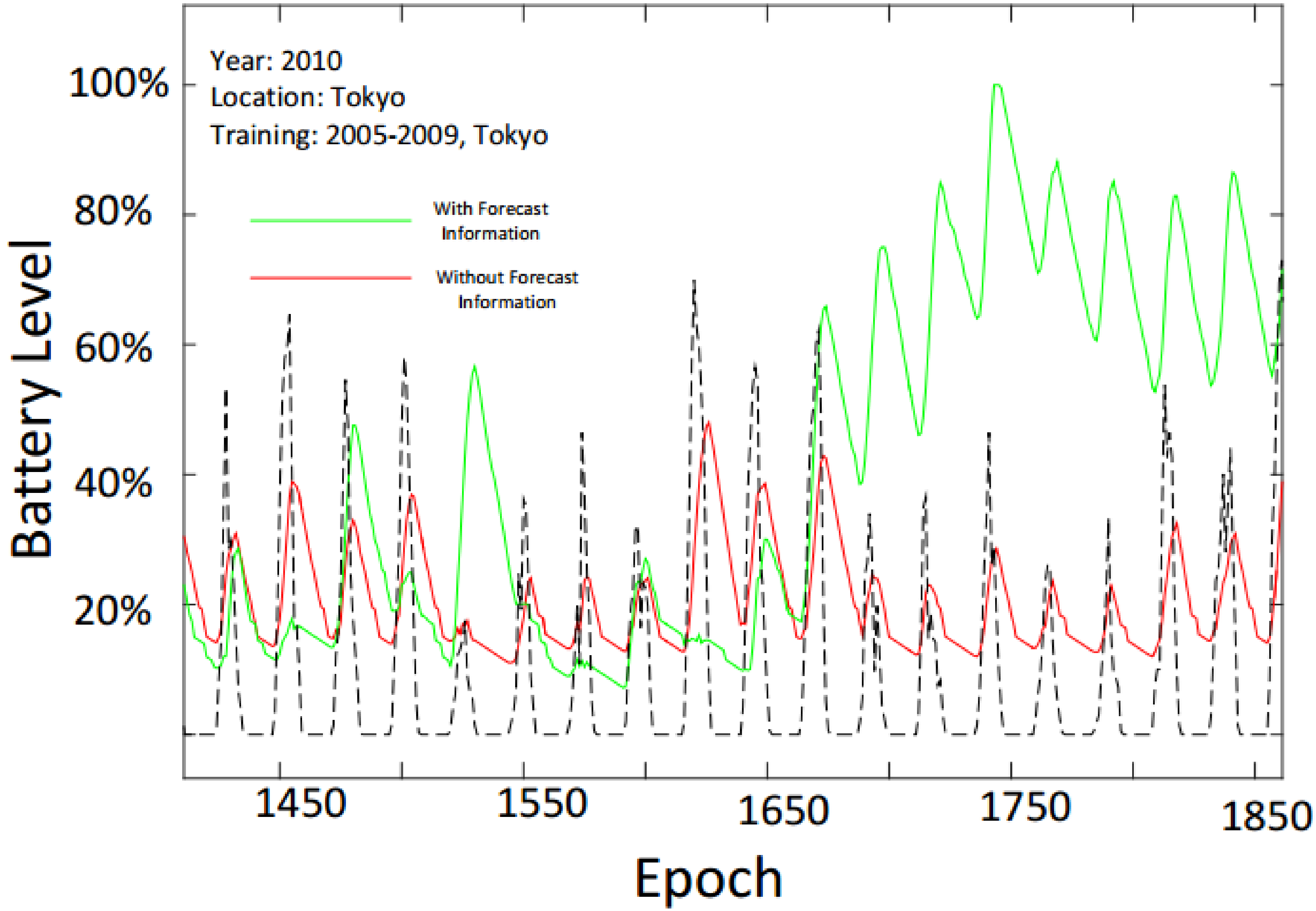
- Wakkanai (very little sunshine)
- Compare between
  - a greedy approach (Offline) and
  - an ε-greedy approach (Online).
- Training: 2000-2009 Tokyo
- Testing: 2010 Wakkanai

Adaptation to change in location



With ε-greedy implementation, the agent adapts to the environment and minimizes instances of battery exhaustion.

With and Without Forecast Information



CONCLUSION

CONCLUSION

- Proposed system is able to meet objectives of
  - Energy neutrality
  - Maximizing performance

- Exceeds the performance of other schemes

- Capable of adaptation

- Inclusion of weather forecast results in smarter operation

THANK YOU FOR LISTENING

ANY COMMENTS OR QUESTIONS ARE WELCOME