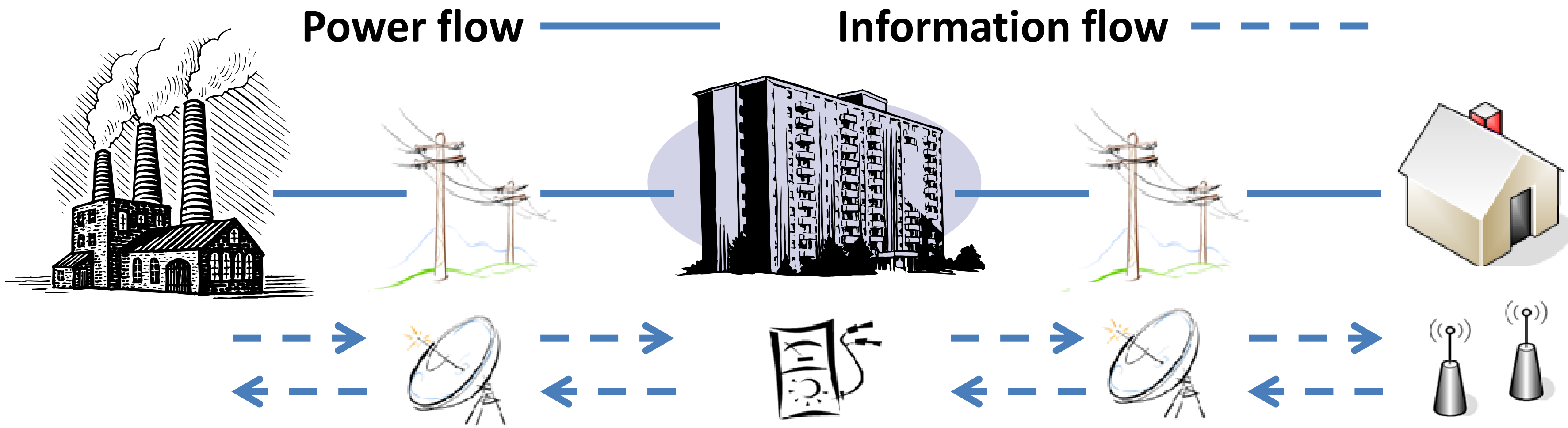
Learning-Based Energy Management Policy with Battery Depth-of-Discharge Considerations

Ting-Hsing Wang and Y.-W. Peter Hong

National Tsing Hua University

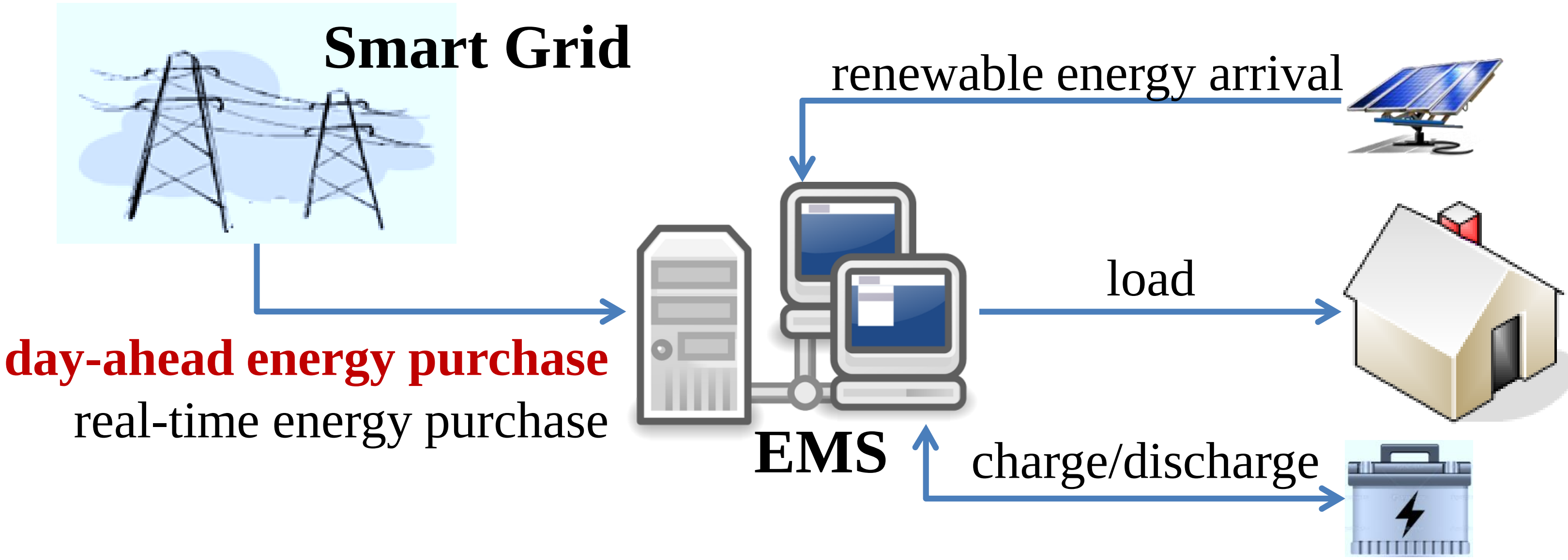
Date: 2015/12/16

Smart Grid



- Two-way flow of power and information, and more advanced sensory systems induce smartness into the power grid.
  - Intelligent control of generation and distribution.
  - Better integration of renewable energy.
  - Enhanced security and flexibility.
- More importantly, energy usage can be more efficient with *dynamic pricing* and **demand-side response** (i.e., load scheduling, battery control, and **energy purchase decisions**).

Demand-Side Energy Management System



- Electricity market allows day-ahead and real-time purchases.
  - Day-ahead purchase for each hour of the next day at a discounted price.
  - Real-time purchase to compensate for any insufficient purchase.
- **Goal:** Derive a **day-ahead energy purchase policy** that takes into consideration the **cost of battery usage**.

Related Works and Main Contributions

- In the literature, energy management policies often assume …

knowledge of future demands [Qian 2013]

knowledge of renewable energy statistics [Jiang 2011]

knowledge of electricity price [Atzeni *et al.* 2013]

negligible battery cost [Koutsopoulos 2011] [Erseghe 2014]

Difficult to obtain accurately in practice.

Battery is NOT free.

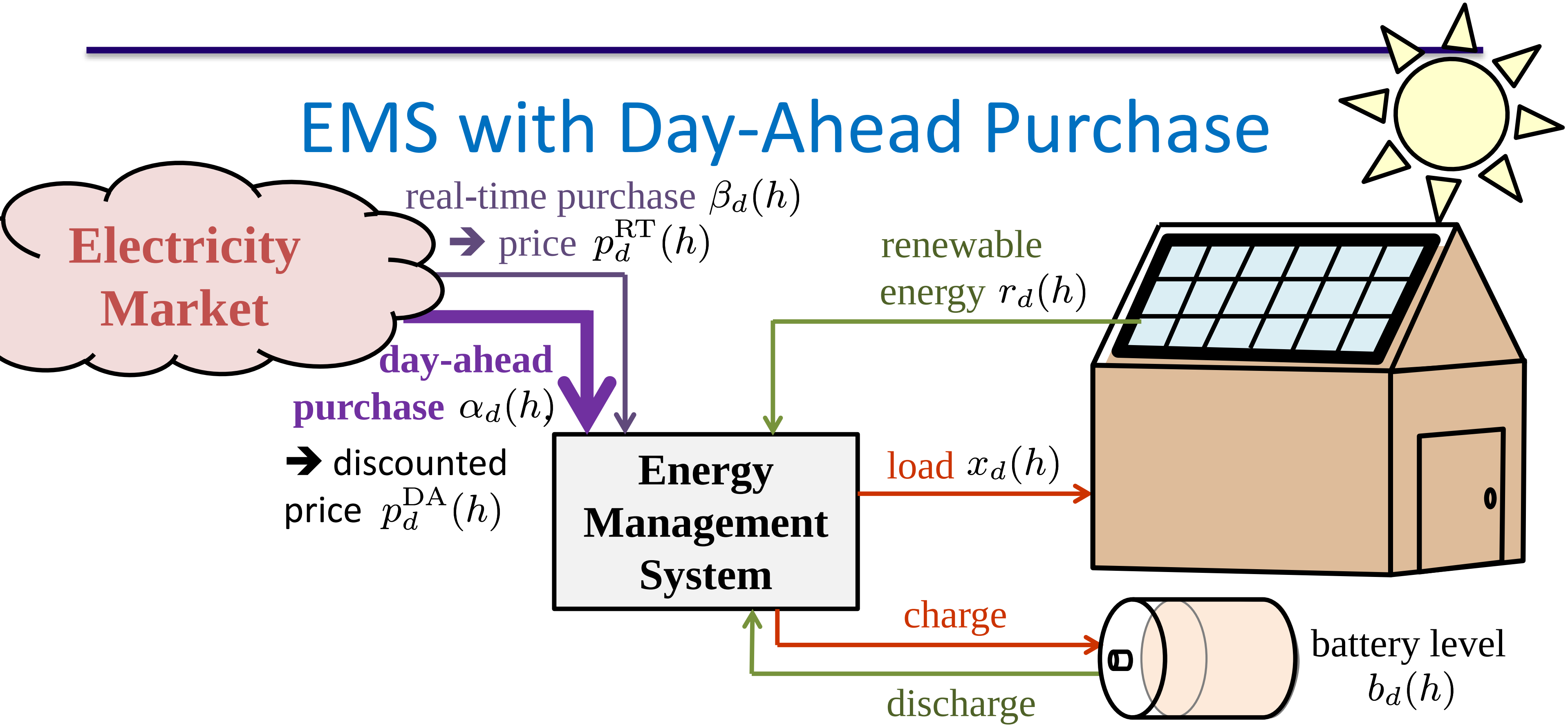
Main Contributions

- Propose a day-ahead energy purchase policy based solely on historical data utilize **reinforcement learning**.

  ➔NO need for any modelling of the load profile, renewable energy arrival, and pricing.

- Model battery cost based on its **depth-of-discharge (DoD)**.

EMS with Day-Ahead Purchase



- At day d, the EMS determines the **day-ahead energy purchase**

$$\boldsymbol{\alpha}_{d+1} = [\alpha_{d+1}(1), \ldots, \alpha_{d+1}(H)]$$

for the H hours of the next day, i.e., day $d+1$, based on the current-day load $\mathbf{x}_d = [x_d(1), \ldots, x_d(H)]$, renewable energy $\mathbf{r}_d = [r_d(1), \ldots, r_d(H)]$, and pricing $\mathbf{p}_d^{\mathrm{DA}} = [p_d^{\mathrm{DA}}(1), \ldots, p_d^{\mathrm{DA}}(H)]$ and $\mathbf{p}_d^{\mathrm{RT}} = [p_d^{\mathrm{RT}}(1), \ldots, p_d^{\mathrm{RT}}(H)]$.

Cost of (Grid) Energy Purchase

- **Residual (or excess) energy** at time h is

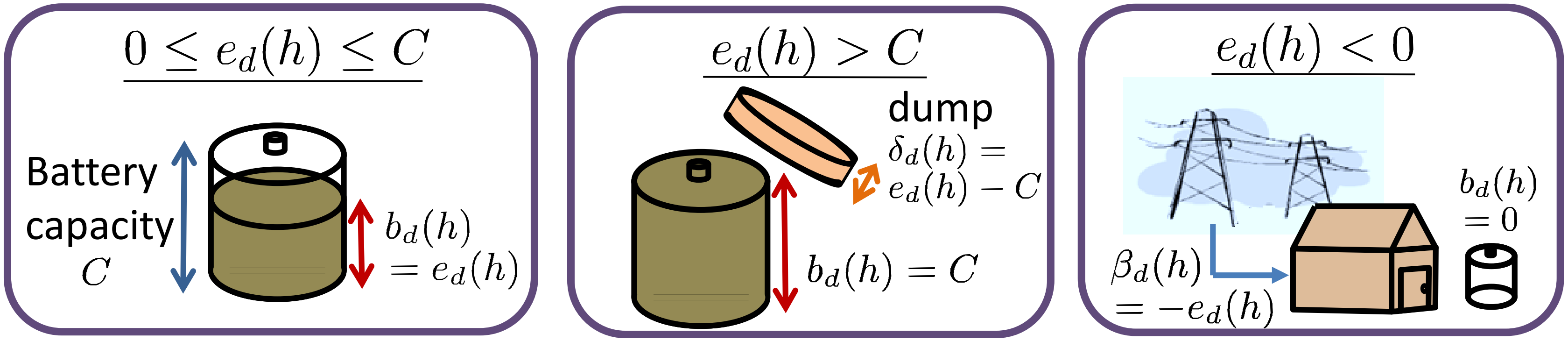
$$e_d(h) = b_d(h-1) + r_d(h) + \alpha_d(h) - x_d(h)$$

Day-ahead purchase for time h

Residual battery at time $h-1$

Renewable energy at time h

Load at time h



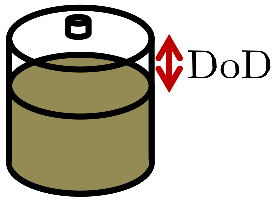
- The **cost of (grid) energy purchase** is

$$\kappa_d^{\text{grid}}(h) = p_d^{\text{DA}}(h)\alpha_d(h) + p_d^{\text{RT}}(h)\beta_d(h) + p_d^{\text{DP}}(h)\delta_d(h)$$

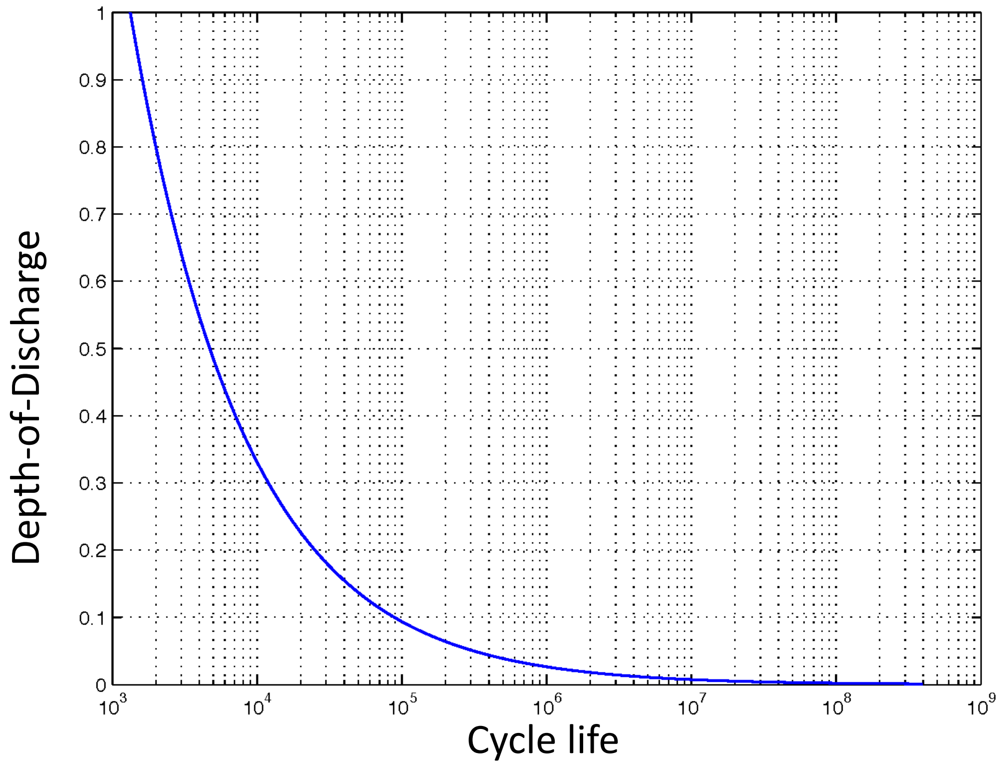
where $p_d^{\text{DP}}(h)$ is the dumping cost.

Battery Depth-of-Discharge (DoD)

- **Depth-of-Discharge (DoD)**: ratio of battery discharge over the maximum capacity C.



- The no. of charges/discharges in a lifetime (i.e., its **cycle life**) is related to the DoD by $N_{\text{cycle}} = c_1 \cdot \text{DoD}^{-c_2}$. [David *et al.* 2013]



– Li-ion battery (2012): $c_1 = 1331, c_2 = 1.825$

Cost of Battery Usage

- For a battery of price p_{batt} (USD/kWh), the cost of battery discharge up to a certain DoD is

$$\frac{p_{\text{batt}} \cdot C}{N_{\text{cycle}}} = \frac{p_{\text{batt}} \cdot C}{c_1 \cdot \text{DoD}^{-c_2}}.$$

- Define the **current depth-of-discharge (cDoD)** at time h as the DoD relative to last charge, which is given by

$$\text{cDoD}_d(h) = \begin{cases} \text{cDoD}_d(h-1) + \frac{b_d(h-1) - b_d(h)}{C}, & \text{if } b_d(h-1) \geq b_d(h) \\ 0, & \text{otherwise.} \end{cases}$$

- Define **cost of battery usage** at time h as the marginal cost of increasing the DoD after the current usage, i.e.,

$$\kappa_d^{\text{batt}}(h) = \max\left\{0, \frac{p_{\text{batt}} \cdot C}{c_1 \cdot \text{cDoD}_d(h)^{-c_2}} - \frac{p_{\text{batt}} \cdot C}{c_1 \cdot \text{cDoD}_d(h-1)^{-c_2}}\right\}$$

EMS based on Reinforcement Learning

- **Goal:** Develop a day-ahead energy purchase policy based only on historical data using reinforcement learning.
- Reinforcement learning determines an action for each state (i.e., a policy) so as to maximize the future expected reward.
  - E.g., determine the energy purchase (action) for the current day so as to minimize the expected future energy cost.
- The policy is learned by observing the rewards obtained from the actions taken in past states (i.e., historical data).
  - E.g., by observing the energy costs obtained from the energy purchase decisions (actions) made in previous days.

State, Action, and Reward of an EMS

- In our problem, the **state** of day d is

$$\mathbf{s}_d = (\mathbf{r}_d, \mathbf{x}_d, \mathbf{p}_{d+1}^{\mathrm{DA}}, \mathbf{p}_d^{\mathrm{RT}}, \mathbf{b}_d)$$

Renewable energy (over H hrs) | Load profile | Day-ahead price | Real-time price | Battery level

- The **action** is the energy purchase over H hours of day $d+1$, i.e., $\boldsymbol{\alpha}_{d+1} = [\alpha_{d+1}(1), \ldots, \alpha_{d+1}(H)]$
- The **reward** is the negative of the total energy cost, i.e.,

$$\mathcal{R}(\mathbf{s}_d, \boldsymbol{\alpha}_d) = -\sum_{h=1}^{H} \left[\underbrace{\kappa_{d+1}^{\mathrm{grid}}(h)}_{\text{= grid cost}} + \underbrace{\kappa_{d+1}^{\mathrm{batt}}(h)}_{\text{= battery cost}} \right]$$

= grid cost

= battery cost (measured by marginal cost of DoD increase)

Least-Square Policy Iteration (LSPI)

- **Goal of RL:** Find an action $\boldsymbol{\alpha}$ for each state $\mathbf{s}_d$ that maximizes the **state-action value function**

discount factor

$$Q^{\pi}(\mathbf{s}_d, \boldsymbol{\alpha}) = \mathcal{R}(\mathbf{s}_d, \boldsymbol{\alpha}) + \gamma \mathrm{E}\left[Q^{\pi}(\mathbf{s}_{d+1}, \pi(\mathbf{s}_{d+1})) \middle| \mathbf{s}_d, \boldsymbol{\alpha}\right]$$

i.e., the expected future reward.

(1) **Linear approximation** (for dimensionality reduction):

$$\widehat{Q}^{\pi}\left(\mathbf{s}_d, \boldsymbol{\alpha}_{d+1}; \{w_j\}_{j=1}^{k}\right) = \sum_{j=1}^{k} \phi_j(\mathbf{s}_d, \boldsymbol{\alpha}_{d+1}) w_j$$

problem-specific basis functions

(2) **Policy Iteration:**

Learns weights $\{w_j\}_{j=1}^{k}$ for policy π based on historical data.

$\{w_j\}_{j=1}^{k}$

Update π such that

$$\pi(\mathbf{s}) \leftarrow \arg\max_{\boldsymbol{\alpha}} \widehat{Q}^{\pi}\left(\mathbf{s}, \boldsymbol{\alpha}; \{w_j\}_{j=1}^{k}\right)$$

π

➔ Parameter-by-parameter optimization

Basis Selection

- Suppose that $H = 24$.

$$
\begin{aligned}
\phi_1(\mathbf{s}_d, \boldsymbol{\alpha}) &= -\sum_{h=1}^{24} \hat{\kappa}_{d+1}^{\text{grid}}(h) && \rightarrow \text{estimated grid cost} \\
\phi_2(\mathbf{s}_d, \boldsymbol{\alpha}) &= -\sum_{h=1}^{24} \hat{\kappa}_{d+1}^{\text{batt}}(h) && \rightarrow \text{estimated battery cost} \\
\phi_3(\mathbf{s}_d, \boldsymbol{\alpha}) &= \hat{\mathbf{b}}_{d+1}(24) && \rightarrow \text{estimated end day battery level} \\
\phi_4(\mathbf{s}_d, \boldsymbol{\alpha}) &= \sum_{h=1}^{4} \mathbf{p}_d^{\text{RT}}(h) \\
\phi_5(\mathbf{s}_d, \boldsymbol{\alpha}) &= \sum_{h=5}^{8} \mathbf{p}_d^{\text{RT}}(h) \\
&\vdots \\
\phi_9(\mathbf{s}_d, \boldsymbol{\alpha}) &= \sum_{h=21}^{24} \mathbf{p}_d^{\text{RT}}(h)
\end{aligned}
$$

ϕ_4–ϕ_9 → real-time electricity price (for every 4 time slots)

$$
\begin{aligned}
\phi_{10}(\mathbf{s}_d, \boldsymbol{\alpha}) &= \sum_{h=1}^{4} (\mathbf{r}_d(h) - \mathbf{x}_d(h)) \\
\phi_{11}(\mathbf{s}_d, \boldsymbol{\alpha}) &= \sum_{h=5}^{8} (\mathbf{r}_d(h) - \mathbf{x}_d(h)) \\
&\vdots \\
\phi_{15}(\mathbf{s}_d, \boldsymbol{\alpha}) &= \sum_{h=21}^{24} (\mathbf{r}_d(h) - \mathbf{x}_d(h))
\end{aligned}
$$

ϕ_{10}–ϕ_{15} → residual renewable energy

$$
\begin{aligned}
\phi_{16}(\mathbf{s}_d, \boldsymbol{\alpha}) &= \mathbf{b}_d(24) && \rightarrow \text{end day battery level} \\
\phi_{17}(\mathbf{s}_d, \boldsymbol{\alpha}) &= 1
\end{aligned}
$$

Prediction of Future States

- Bases functions ϕ_1, ϕ_2, and ϕ_3 require estimates (i.e., predictions) of the next states $\hat{\mathbf{r}}_{d+1}$, $\hat{\mathbf{x}}_{d+1}$, and $\hat{\mathbf{p}}_{d+1}^{\mathrm{RT}}(h)$.

- Linear least-square predictions

$$\hat{\mathbf{r}}_{d+1} = \mathbf{F}_{\mathbf{r}}\mathbf{r}_d, \quad \hat{\mathbf{x}}_{d+1} = \mathbf{F}_{\mathbf{x}}\mathbf{x}_d, \quad \hat{\mathbf{p}}_{d+1}^{\mathrm{RT}} = \mathbf{F}_{\mathbf{p}}\mathbf{p}_d^{\mathrm{RT}}.$$

  where $\mathbf{F}_{\mathbf{r}}$, $\mathbf{F}_{\mathbf{x}}$, and $\mathbf{F}_{\mathbf{p}}$ are least square estimators.

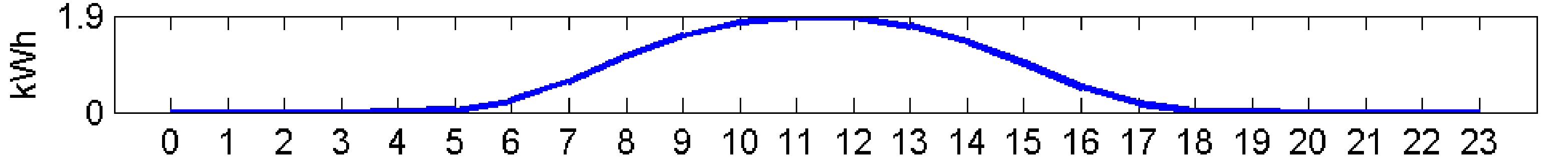
Remark: Data preprocessing is required to deal with the non-stationarity of the data and the effect of outliers.

✓ Cancellation of the mean over the most recent $D = 30$ days.

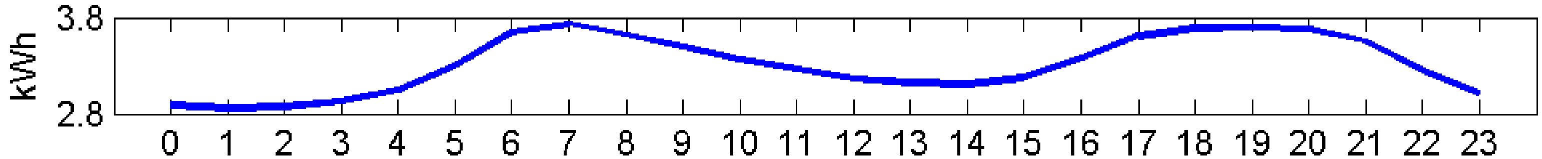
✓ Negligence of the data that lies outside of 3 standard deviations.

Real-Life Data

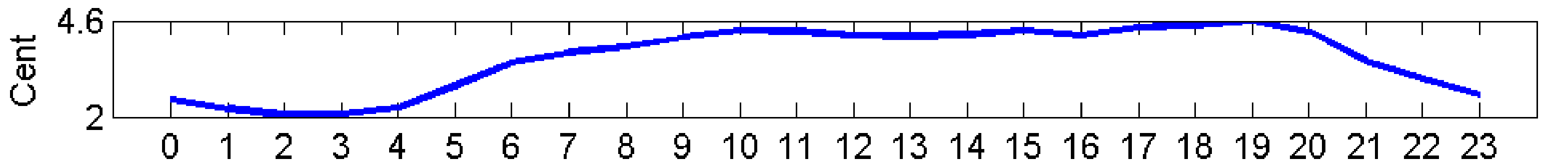
- Number of time slot in a day: $H = 24$.
- Solar energy (Chicago, PV panel size 25 m^2) [NREL PVWatts Calculator]



- Residential load (average of a class of residential single families with electric space heat delivery, 2010-2014) [ComEd]



- Day-ahead electricity price (2010-2014) [ComEd RRTP program]



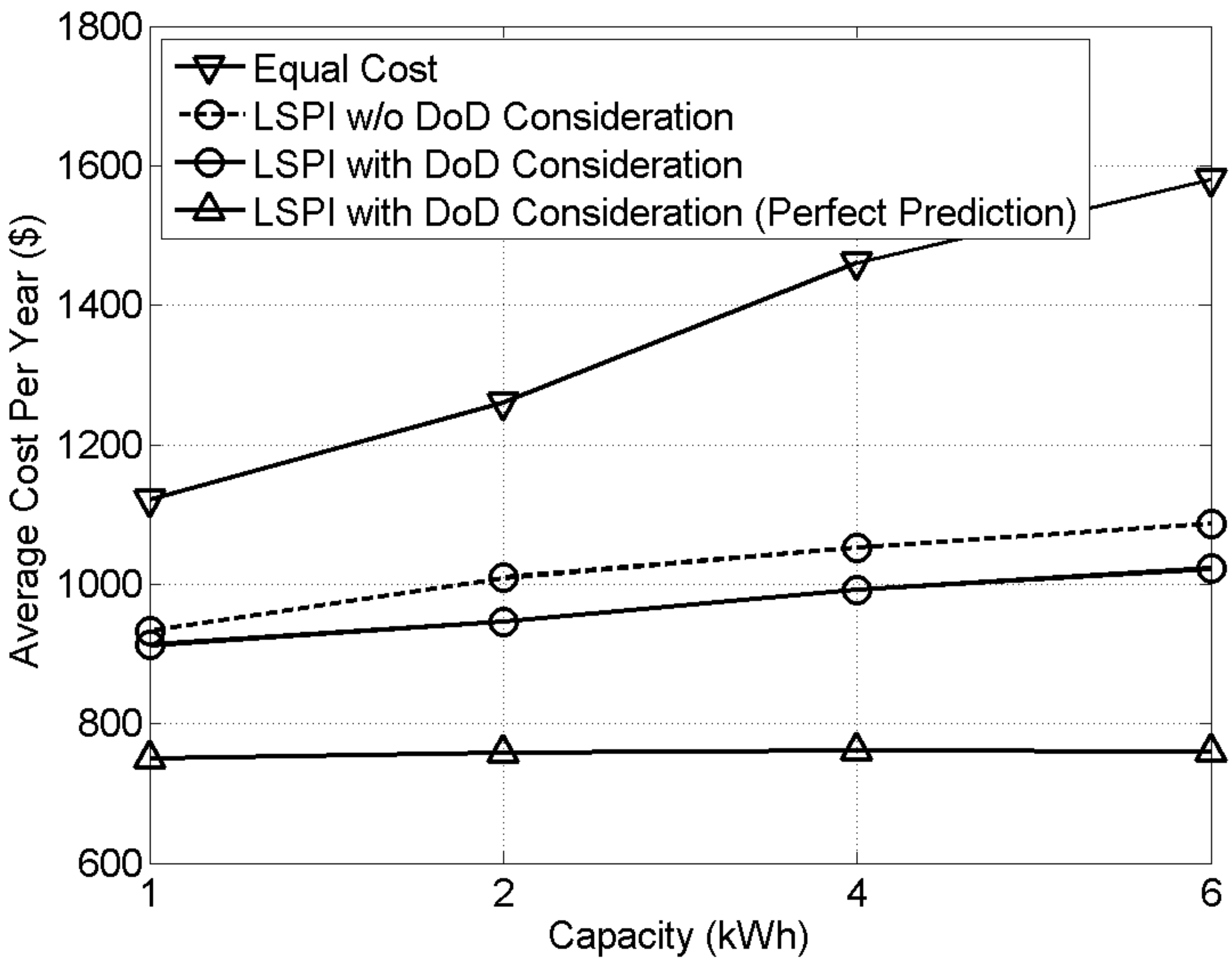
Simulation Settings

- Battery price: $p_{batt} = 500$ (USD/kWh) [Tesla 2015]
- Dumping cost: $p_d^{\mathrm{DP}}(h) = 0.004$ (USD/kWh) [Ramavajjala *et al.* 2012]
- Discount factor: $\gamma = 0.9$
- Stopping criterion: $\epsilon = 10^{-3}$
- Baseline (Equal cost policy):
  The action at time h of day $d+1$ is chosen as

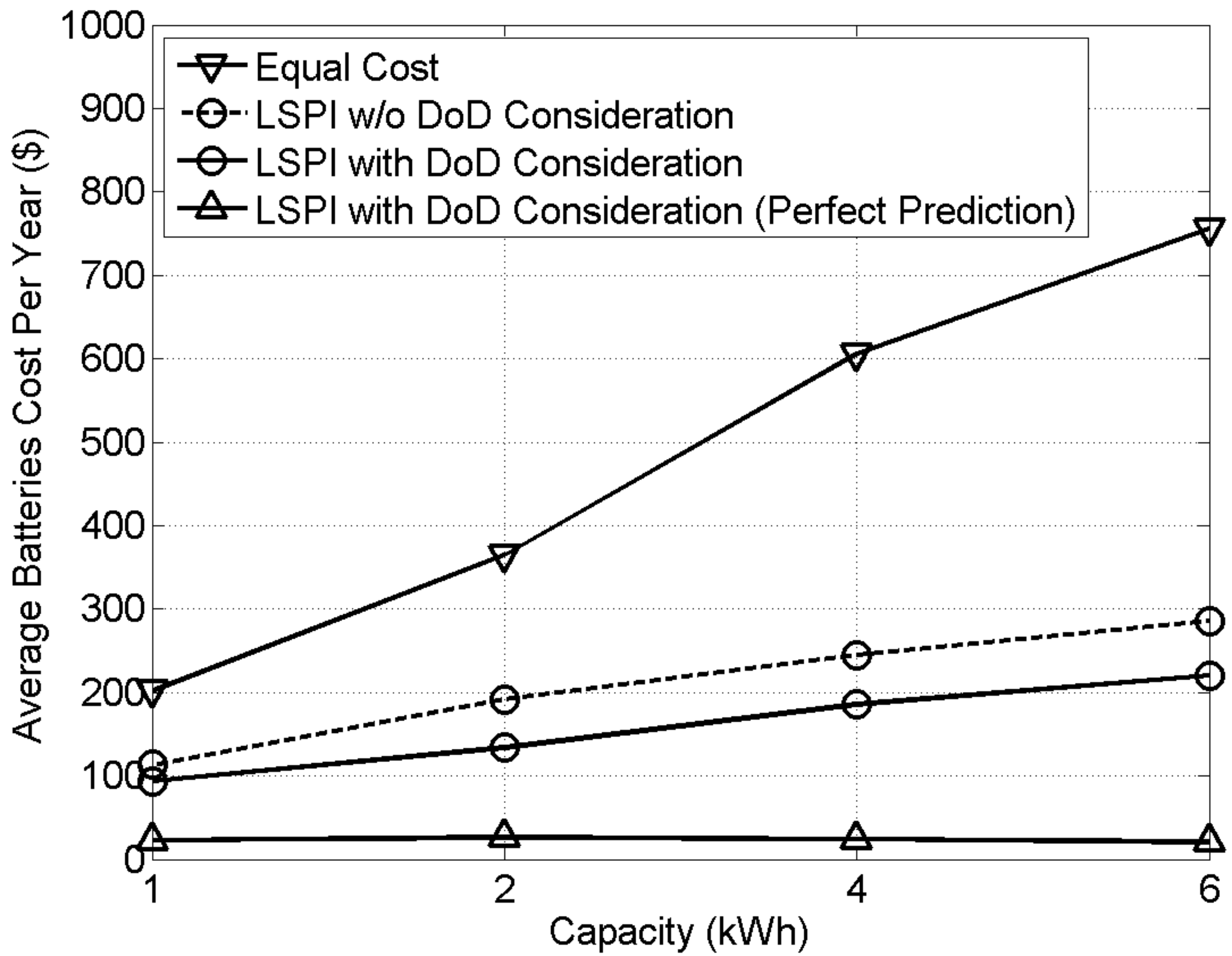
$$\alpha_{d+1}(h) = \sum_{h_1=1}^{24} \left(\hat{x}_{d+1}(h_1) - \hat{r}_{d+1}(h_1)\right) \frac{1/\hat{p}_{d+1}^{\mathrm{DA}}(h)}{\sum_{h_2=1}^{24} 1/\hat{p}_{d+1}^{\mathrm{DA}}(h_2)}$$

so that the anticipated cost is equal in each time slot.

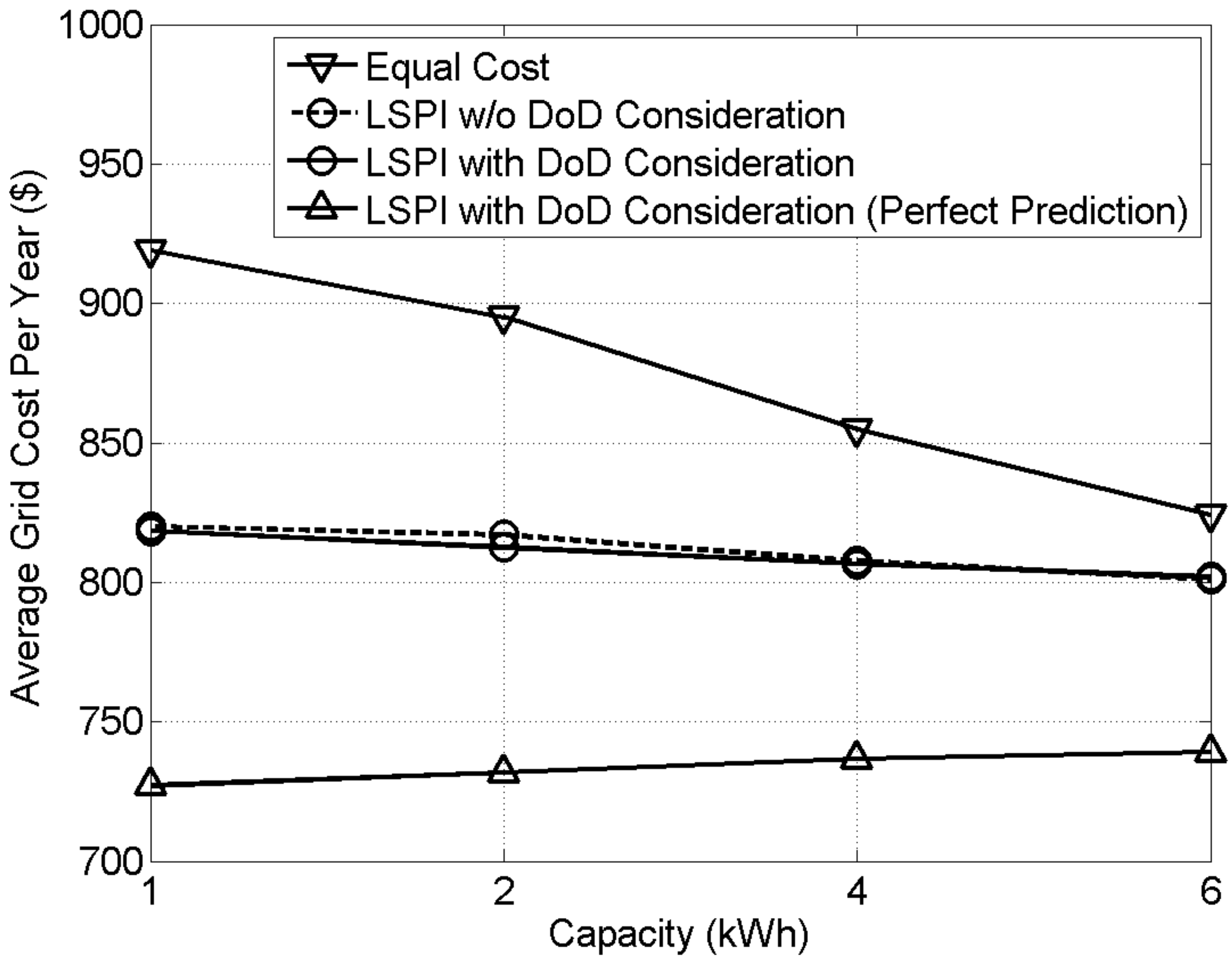
Total Cost Per Year vs. Battery Size



Battery Cost Per Year vs. Battery Size

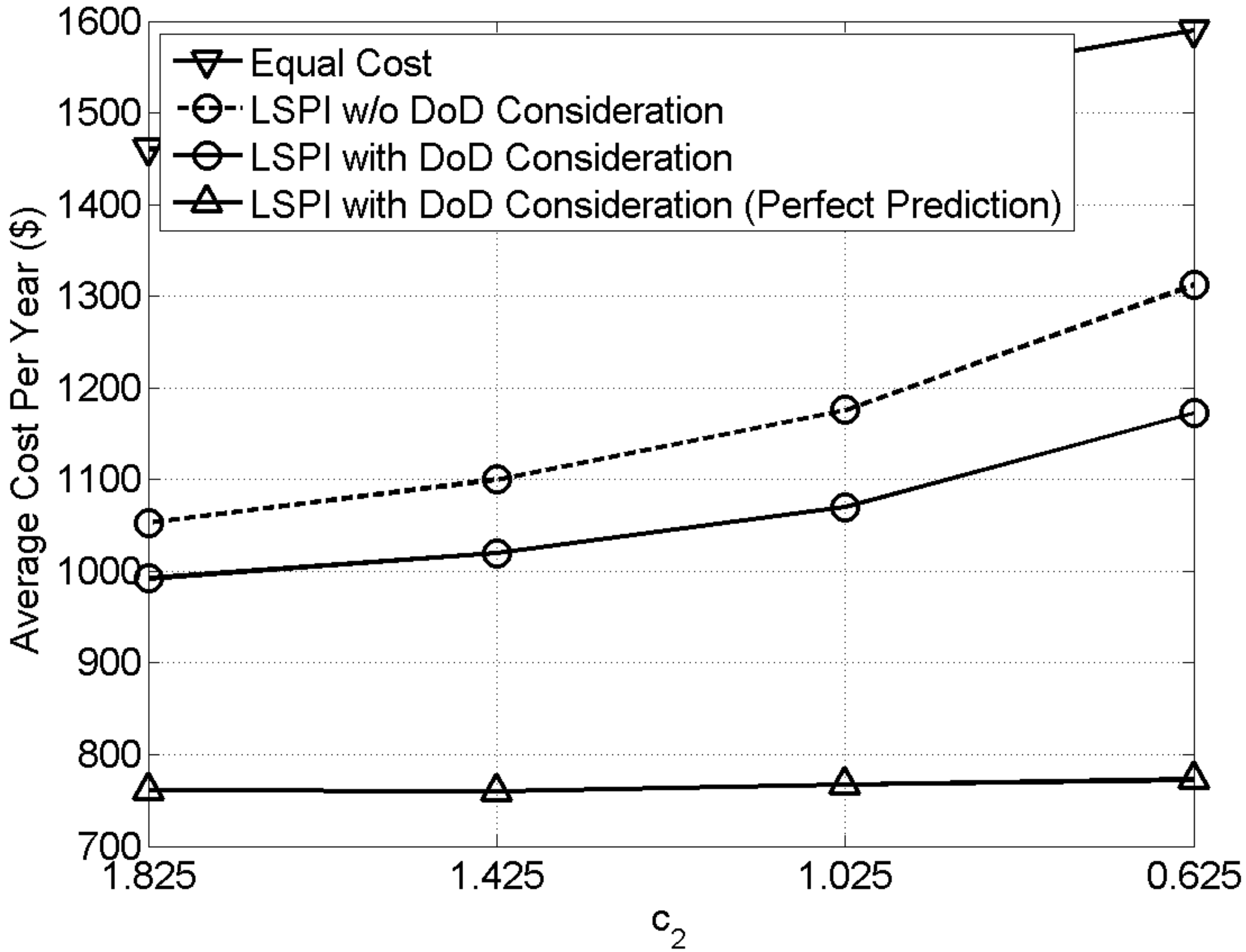


Grid Cost Per Year vs. Battery Size



Average Cost Per Year vs. c_2

➔ Recall that $N_{\text{cycle}} = c_1 \cdot \text{DoD}^{-c_2}$.



Conclusion

- Proposed a novel method for evaluating the cost of battery usage based on the **depth-of-discharge (DoD)**.
- Utilized **reinforcement learning** to develop an energy management policy that relies only on historical data (instead of assuming statistical models).
- *Battery cost is not negligible* and is essential to finding a practical energy management policy.