Decision-Making Towards a Multi-Use Framework for Grid-Scale Energy Storage

Deep Dayaramani
Dept. of Civil and Environmental Engineering
Stanford University
Stanford, CA, USA
deepdaya@stanford.edu

Ed Han Xue
Dept. of Chemical Engineering
Stanford University
Stanford, CA, USA
edmondx@stanford.edu

Kevin Moy
Dept. of Energy Resources Engineering
Stanford University
Stanford, CA, USA
kmoy14@stanford.edu

***Abstract*—Energy storage systems (ESSs) on the electric grid participate in grid applications, for which their dispatch (charge and discharge) are financially compensated based on the value of that application. Furthermore, a single ESS is capable of participating in multiple grid applications, with the potential for multiple value streams for a single system, termed "value-stacking". This paper introduces a framework for decision making using reinforcement learning to analyze the financial advantage of value-stacking for ESSs, as applied to a single residential home with a single ESS. A policy is developed via Q-learning to dispatch the ESS between two grid applications: time-of-use (TOU) bill reduction and energy arbitrage on locational marginal price (LMP). The performance of the dispatch resulting from this learned policy is then compared to four other dispatch cases: a baseline of no dispatch, a naively-determined dispatch, and the optimal (highest revenue) dispatches for TOU and LMP separately. The TOU+LMP dispatch policy obtained via Q-learning led to the highest revenue and the lowest cost among all dispatch methods, successfully demonstrating the financial advantage of value-stacking.**

I. Introduction

Energy storage systems (ESSs) are a critical part of the renewable-fueled, sustainable energy grid of the future. Currently, ESSs are used in such renewable grid systems in grid applications, which support a variety of different stakeholders, including utilities, transmission operators, and utility customers (i.e. consumers of electricity) [1]. Generally, these services can be divided into "behind-the-meter" (BTM) services, which support customers, e.g. providing power during a blackout, and "front-of-meter" (FOM) services, which support the grid at large, e.g. regulating grid voltage. Currently, a single energy storage asset can only participate (provide and be compensated for) BTM or FOM services, but not both.

However, ESSs are still expensive, and many applications do not require energy storage dispatch at all times. From this, many operators of ESSs seek to use the ESSs in "value-stacking"; that is, using a singular energy storage resource for multiple different services/grid applications. Value-stacking is supported by federal energy policy, via FERC Order 841, which directed transmission grid operators to provide means for energy storage to participate in both BTM and FOM services [2]. It is also supported by California legislation, via the Multi-Use Application framework[1]. We hope to validate the premise of value-stacking: That by having the option to choose between multiple different grid applications, that the value is greater than participating in either grid application alone.

II. Methodology

A. Problem description

We can reduce the problem down into two grid applications: First, the application of time-of-use (TOU) bill reduction, or dispatching the energy storage to reduce the utility bill, by reducing the energy consumed by the load as measured through a utility meter, and second, the application of energy arbitrage, or "buying low/selling high" on the energy marketplace according to locational marginal prices (LMP). We assume a system that will allow for both such behavior, as shown in Figure 1.

We size the system to be equivalent to a Tesla Powerwall, rated at 5kW continuous power and 14kWh rated energy[2], and assume that the energy storage has no degradation and a dispatching efficiency of 100%, with SOC limits of 10% and 90% of rated energy. We also assume that the continuous power rating of 5 kW holds as the power limit for both charging and discharging.

The decision is whether to participate in TOU dispatch, LMP dispatch, or neither, for the goal of maximizing value, or the total TOU savings + LMP energy arbitrage revenue over the entire year. We can define this explicitly with the following equation and definitions in Table I to define the cumulative revenue V from the beginning of the year up until time period t:

$$V(t) = h \sum_{t'=0}^{t} m(t') \left[d_m(t') - c_m(t') \right] + u(t') \left[d_u(t') - l(t') - c_u(t') \right] \quad (1)$$

[1]https://www.utilitydive.com/news/california-regulators-first-to-allow-multiple-revenue-streams-for-energy-st/516927/

[2]https://www.tesla.com/powerwall

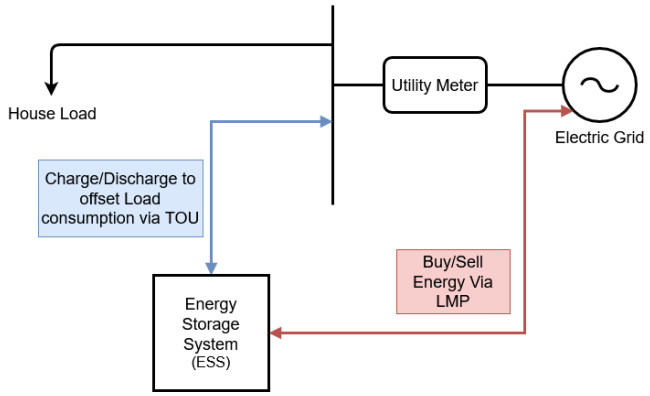

Fig. 1. Power flow diagram of system in this paper. Arrows denote direction of power flow. The ESS can charge and discharge via TOU to offset the energy consumed by the load from the grid, as measured via the utility meter. Likewise, it can also charge and discharge directly to the grid, buying and selling energy via LMP energy arbitrage.

The different terms in the summation represent different operation of the ESS. The first term represents direct power flow between the ESS and the grid, buying and selling energy on the LMP. The second term represents the power flow as measured by the utility meter, including the dispatch of the ESS for TOU. We restrict our ESS to prevent backfeeding, or the flow of energy back through the energy to the electric grid, which would result in a negative utility bill. Such behavior is not currently allowed under FERC Order 841.

TABLE I
LIST OF DEFINED VARIABLES IN THE PROBLEM DESCRIPTION.

Variable	Name	Units
$V(T)$	Total revenue up to time t	$
$m(t)$	LMP for time t	$/kWh
$u(t)$	TOU for time t	$/kWh
$l(t)$	Load at time t	kW
$E(t)$	SOC of energy storage at time t	kWh
$c_m(t)$	Charging power from LMP at time t	kW
$d_m(t)$	Discharging power to LMP at time t	kW
$c_u(t)$	Charging power for TOU at time t	kW
$d_u(t)$	Discharging power for TOU at time t	kW
h	Period, in hours, of the data (=0.25 for this paper)	unitless
C_{max}	Maximum charging power	kW
D_{max}	Maximum discharging power	kW
E_{min}	Minimum SOC	kWh
E_{max}	Maximum SOC	kWh

B. Data and Features

The data needed for BTM services problem is time-series data of: the building load, the TOU (price of energy bought from the grid), and for the FOM services problem, we use the LMP (price of energy on the energy marketplace). First, we obtained load data from the Pecan Street Dataport database[3] for a house in San Diego. From this, we selected the LMP pricing data from the COVID-EMDA datahub, using CAISO data for a node in the same location [3]. We also selected the relevant TOU pricing data (plan TOU-DR1) from the San Diego Gas & Electric tariff[4]. All data are obtained at 15-minute intervals for the dates of July 8th 00:00 to June 30th 23:45 of the next year, representing our year of data. For the LMP data, we collected data from July 8th 2018, 00:00 to 30 June 2019, 23:45 for the URBAN-N005 node in San Diego from CAISO's Open Access Same-Time Information System (OASIS) LMP DataBase [5] as the data for the 2014-2015 period wasn't available.

C. Q-learning implementation

We use the method of Q-learning to learn the best action for the ESS to take, given the TOU cost of energy, the LMP price of energy, the ESS state-of-charge (SOC), and the load profile.

1) State Space: Our state space includes four components, discretized as shown below:

1) TOU: There are only six unique prices, so this is already discretized to a space of 6.
2) SOC: The SOC is discretized into the following values: [1.4, 2.65, 3.9, 5.15, 6.4, 7.65, 8.9, 10.15, 11.4, 12.6]. This was determined using the Python function `arange(0.1*14, 0.9*14, 5*0.25)`, which ensures that the SOC is only a function of the maximum power rating (5 kW) of the ESS, given that we initialize the SOC $E(0) = 5.15$ kWh.
3) LMP: The LMP is a continuous cost, so we discretize it into 10 evenly spaced bins.
4) Load: Like LMP, the load is a continuous power, so we discretize it into 100 evenly spaced bins.

Each time step in the dataset is therefore represented with a tuple in $\mathbf{R}^4$.

2) Action Space: Our action space includes five components, discretized as shown below. Only one action is taken per time step (i.e. state tuple).

1) Charge from LMP ($c_m(t)$), buying the energy for charging the ESS at the current LMP price
2) Discharge to LMP ($d_m(t)$), selling the energy discharged from the ESS at the current LMP price
3) Do nothing, do not charge or discharge the ESS
4) Charge from TOU ($c_u(t)$), adding the energy for charging the ESS to the load energy consumption as billed at the TOU price
5) Discharge to TOU ($d_u(t)$), subtracting the energy discharged from the ESS from the load energy consumption as billed by the TOU price

This definition of the action space prevents the energy storage from attempting to charge and discharge at the same time, or charge/discharge from multiple sources.

In each action, the ESS will attempt to charge or discharge at the maximum allowable rate. In the case that the ESS does not run the risk of over/under charging, i.e. the charge/discharge does not cause $E(t)$ to go above or below E_{min} or E_{max},

[3]https://www.pecanstreet.org/dataport/

[4]https://www.sdge.com/whenmatters

[5]http://oasis.caiso.com/mrioasis/logon.do

then the dispatch will be at full rated power (5 kW), which will change the SOC by $(\pm 5kWh) * 0.25hr = \pm 1.25kWh$. Otherwise, the ESS will do nothing. This ensures that our SOC can be discretized as in the previous section.

Additionally, we ensure that the ESS will not cause the load to go negative (i.e. backfeed into the grid).

3) Reward function: We draw upon previous work for RL-based ESS energy arbitrage to determine the reward function as a moving average of recently observed prices [5].

This is kept separate from the overall performance of the policy, or the dollar cost paid for energy by the end of the dataset, that we are comparing between strategies.

At each timestep t (representing a single state), and possible action a, we construct the action function by using the moving average LMP $\overline{m(t)}$, but without the moving average on the TOU price, as the cost is dependent on the pre-determined and uncontrollable load $l(t)$, and therefore not subject to the same exploration benefit as for LMP energy arbitrage.

$$\begin{aligned} r(t,a) = h \Big[\Big(\overline{m(t')} - m(t) \Big) \left(c_m(t)|_{a_1} - d_m(t)|_{a_2} \right) \\ - u(t) \left(l(t) + c_u(t)|_{a_4} - d_u(t)|_{a_5} \right) \Big], \\ \overline{m(t)} = (1-\eta) m(t-1) + \eta m(t) \end{aligned}$$

We use the vertical line notation to refer to terms which are not zeroed out given a specific action. For example, action 1 (a_1) only preserves the $c_m(t)$ term, with all other $d_m(t), c_u(t), d_u(t) = 0$. This ensures that the reward directly follows from our action space definitions.

As the policy is developed and the actions are obtained, we can use Equation 1 to determine the cumulative revenue, V_{RL}.

4) Exploration: An ϵ-greedy approach to exploration was applied with the value of $\epsilon = 0.65$. The algorithm helps balance exploration and exploitation of what we know. It chooses a random valid action with 65% probability given the SOC and chooses the best action with the other 35% probability.

5) Q-Learning Model: For the Q-learning model, we developed a python class, Residential, that would take care of the hyperparameters and parameters of the model and exploration. This class takes care of action maps, generation of the initial Q, S, and policy. A high level pseudocode of this program is given in Algorithm 1.[6]

Algorithm 1: Overall Residential Class- Q-Learning Algorithm

Initialize Q, S, ϵ, A, π, constants;
get_allowed_actions(state)
epsilonGreedyPolicy(Q, ϵ, a_n)
Q_learning()
get_next_state_reward(s, a)
calc_revenue(): Use Equation 1 to calculate cumulative reward & Store best policy π

D. Comparison to other dispatch methods

We also compare the dispatch behavior from Q-learning to other dispatch methods, using Equation 1 as a comparison. These dispatch methods are shown below.

a) Baseline case: In this case, V is calculated with no dispatch from the ESS. Therefore, the cumulative revenue takes on a simple expression:

$$V_{baseline}(t) = h \sum_{t'=0}^{t} u(t') \left[-l(t') \right] \tag{2}$$

[6]The full codebase can be found on Github: https://github.com/kmoy14-stanford/aa-228-final-project

b) Naive TOU dispatch case: For the most naive use of TOU pricing, the ESS charges at the full rated power only at the lowest tariff price, and discharges only at the highest tariff price. If the time periods of the lowest tariff price form the set T_{low}, and the highest tariff price T_{high}, then the cumulative revenue takes on the following expression:

$$\begin{aligned} V(t) &= h \sum_{t'=0}^{t} u(t') \left[d_u(t') - l(t') - c_u(t') \right] \\ & c_u(t') \\ &= \begin{cases} C_{max} & t' \in T_{low} \\ 0 & \text{o.w.} \end{cases} \\ & d_u(t') \\ &= \begin{cases} D_{max} & t' \in T_{high} \\ 0 & \text{o.w.} \end{cases} \end{aligned} \tag{3}$$

c) Optimal TOU dispatch case: We solve the well-studied LP optimization problem below for TOU pricing, with $g(t)$ as the total power drawn from the grid, $l_g(t)$ the load power supplied by the grid, and $E(t)$ initialized to some value $E(0) = E_{init}$ [4]:

$$\begin{aligned} \text{min.} : & \; h \sum_{t=0}^{T_{max}} g(t) * u(t) \\ \text{s. t.} : & \begin{cases} g(t) = c_u(t) + l_g(t) \\ l(t) = d_u(t) + l_g(t) \\ c_u(t) \le C_{max} \\ d_u(t) \le D_{max} \\ E_{min} \le E(t) \le E_{max} \\ E(t) \ge h * d_u(t) \\ g(t), l_g(t), c_u(t), d_u(t), E(t) \ge 0 \end{cases} \quad , \forall t \in \{0, T_{max}\} \\ & E(t) = E(t-1) + h \left[c_u(t-1) - d_u(t-1) \right], \\ & \forall t \in \{1, T_{max}\} \end{aligned}$$

Once the optimal dispatch c_u^* and d_u^* are achieved, we can use Equation 1 with $c_m = d_m = 0$ to determine the cumulative revenue, V_{TOU}^*.

d) Optimal LMP dispatch case: We solve the well-studied energy arbitrage LP problem as shown below for LMP pricing, with $s(t)$ as the dispatch in for timestep t, i.e. $s(t) = d_m(t) - c_m(t)$, $p(t)$ the LMP at timestep t, and E(t) initialized to some randomized value $E(0) = E_{init}$ between E_{max} and E_{min} [5], [6]:

$$\max. : h \sum_{t=1}^{T_{max}} p(t) * s(t)$$

$$\text{s. t. : } \begin{cases} -C_{max} \leq s(t) \leq D_{max} \\ E_{min} \leq E(t) \leq E_{max} \end{cases} , \forall t \in \{0, T_{max}\}$$

$$E(t) = E(t-1) + hr(t-1),$$

$$\forall t \in \{1, T_{max}\}$$

Here, the rate is negative in the case that the ESS is buying energy from the grid at LMP, thus converting the revenue into a cost. Once the optimal dispatch $s(t)$ is achieved, we can use Equation 1 with $c_u = d_u = 0$ and by setting $c_m(t)$ and $d_m(t)$ as the negative and positive components of $s(t)$, i.e $c_m(t) = -\min(0, s(t))$ and $d_m(t) = \max(0, s(t))$, to determine the cumulative revenue, V^*_{LMP}.

III. Results

We learn a policy for the combination of TOU and LMP dispatch using Q-learning as in Section II-C, and apply it to our dataset. The resulting cumulative cost of dispatching according to the Q-learning policy, as well as all other dispatch methods, is shown in Figure 2. We find that the Q-learning policy resulted in actions that reduced the total energy cost the most from the baseline case.

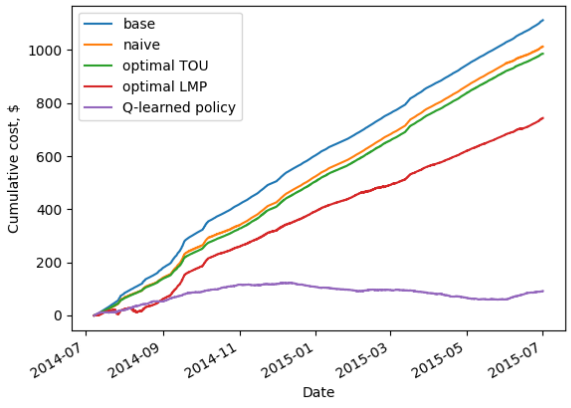

Fig. 2. Comparison of cumulative cost for the dataset for each dispatch method. Note that the cost is negative revenue; all methods resulted in net system costs.

To explain this, we can also examine the distribution of Q-learning policy actions throughout the year. This is shown below in Figure 3. We can see that this policy actually chose a majority of the time to discharge the ESS for TOU, offsetting the TOU cost of energy for the load, and spends the least time selling energy at LMP. Therefore, the policy is choosing to minimize the negative reward over exploiting the potential revenue on the LMP market. However, we also highlight that the cost savings is far greater than that for TOU alone. First, while relatively infrequent, the policy is able to generate some revenue by selling into the LMP market. More importantly, the policy is taking advantage of two sources of energy, charging from either TOU or LMP, and therefore able to choose the cheaper of the two, further reducing the cost of energy.

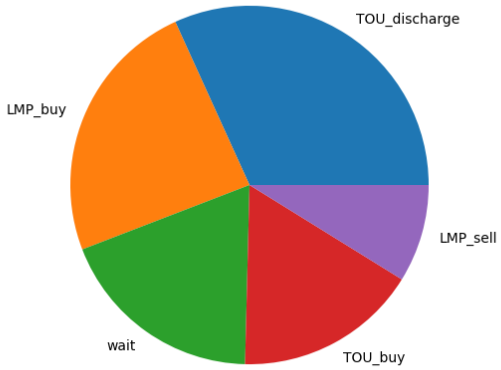

Fig. 3. Distribution of actions throughout the year for the Q-learning policy.

IV. Conclusions

We presented an application of Q-learning to the problem of an ESS participating in multiple different grid applications, with the aim of maximizing the revenue and reducing the cost of energy for a given system. We successfully demonstrated that for the combination of TOU dispatch and LMP energy arbitrage, that following a policy obtained via Q-learning lead to the greatest reduction in energy cost, by taking advantage of multiple sources of energy at different costs. Therefore, we also successfully validated the premise of value-stacking, in that the cost savings combining two different grid applications was greater than either application alone, and even reduced the cost more than both separate grid applications combined.

For future work, we could study whether this policy has been overtrained to our particular dataset, and whether we can extend or generalize the policy to apply to multiple different systems. We can also try to add other grid applications, such as frequency regulation, to study whether there could be further reductions in cost, or even a net positive revenue. The system we chose also originally had solar power, which actually produced net positive power for some time periods. The inclusion of solar power would add an additional power source and potential for even greater revenue, but will add complexity to our action space (e.g. using solar power to supply the load, charge the ESS, or sell at LMP). Lastly, we can perform more hyperparameter tuning, or try different

reinforcement learning approaches, such as Deep Q-Networks, SARSA, or the addition of eligibility traces, to see whether improvements could be made towards learning the dispatch policy.

V. Contributions

Kevin:

- Problem and RL formulation
- Optimal TOU case
- Report writing, all figures

Ed:

- Base Case, Naive TOU Case
- coding epsilon-greedy Q-learning class/algorithm
- generating RL policy and associated data

Deep:

- Optimal LMP Case
- Report Writing
- State Space Discretization

References

[1] Rocky Mountain Institute, "The Economics of Battery Energy Storage", Rocky Mountain Institute, Tech. Rep., 2015, Available: https://rmi.org/insight/economics-battery-energy-storage/.

[2] Federal Energy Regulatory Commission, Order 841, Effective Date 06/04/2018. Available: https://www.federalregister.gov/d/2018-03708

[3] G. Ruan, D. Wu, X. Zheng, H. Zhong, C. Kang, M. A. Dahleh, S. Sivaranjani, and L. Xie, "A Cross-Domain Approach to Analyzing the Short-Run Impact of COVID-19 on the U.S. Electricity Sector," Joule, 2020. (Accepted)

[4] M. Dabbagh, A. Rayes, B. Hamdaoui and M. Guizani, "Peak shaving through optimal energy storage control for data centers," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-6, doi: 10.1109/ICC.2016.7511242.

[5] H. Wang and B. Zhang, "Energy Storage Arbitrage in Real-Time Markets via Reinforcement Learning," 2018 IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, 2018, pp. 1-5, doi: 10.1109/PESGM.2018.8586321.

[6] Ariss, R., Buard, J., Capelo, M., Duverneuil, B., Hatchuel, A., May (2016). Cost-Optimization of Battery Sizing and Operation.