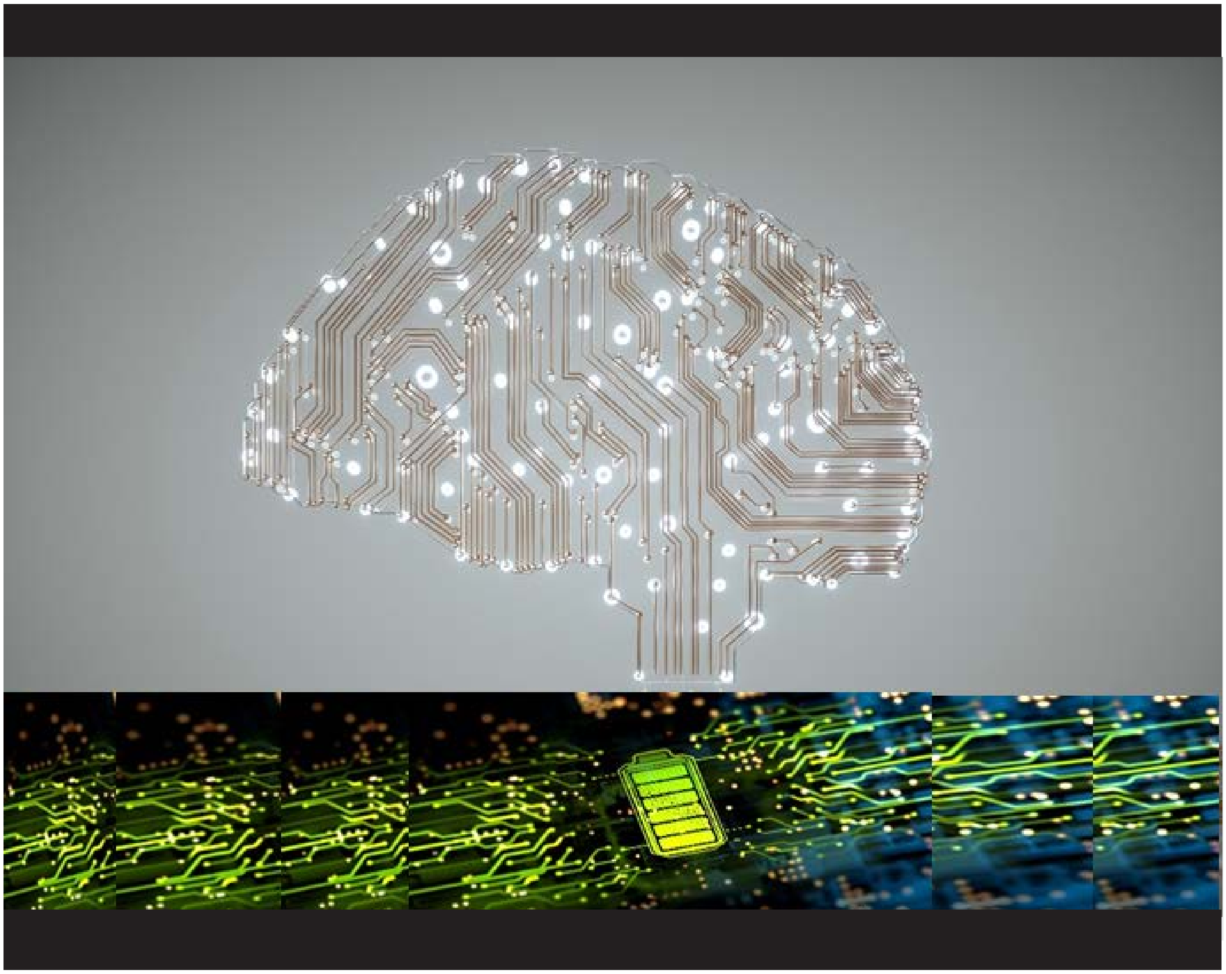
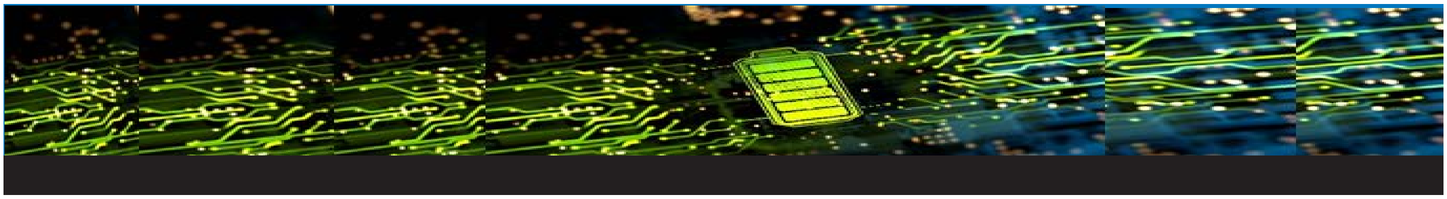
ARTIFICIAL INTELLIGENCE FOR ENERGY STORAGE OPERATION

Energy Arbitrage



January 2023



Introduction

Recent breakthroughs in Artificial Intelligence (AI) and Machine Learning (ML) algorithms have made powerful yet simple solutions possible for many challenges ranging from advanced automatic self-driving cars to the most basic web browser searches. These algorithms have also been applied to a number of modern-day electric power system challenges. This whitepaper focuses on energy storage applications in particular. Energy storage is an important distributed energy resources (DER) asset with the ability to make renewable generation a reliable power source. Further, energy storage can generate revenue by providing a range of market and ancillary services. In the near future, the payback period to own an energy storage system will be significantly shorter due to decreases in capital cost, increases in state and federal incentives and tax credits, and increase in avenues to generate economic value. It is therefore predicted that cumulative energy storage installations in the United States could grow from the current 25GW to 100GW by 2030.[1] With this expected growth potential , AI applications for this asset are inevitable and there are going to be more products in the near future.

AI- and ML-based approaches have been applied for a wide range of energy storage applications such as energy storage dispatch optimization, solar plus storage operation, battery health monitoring, preventive health maintenance, and battery safety. There are many recent research projects and studies on these AI-based solutions for energy storage. The next section in this whitepaper presents a brief literature survey of some of the recent research articles. There are also several commercial products targeted for energy storage systems. For example, energy management solutions vendors such as BluWave-ai and Heila technologies claim to primarily use artificial intelligence-based methods. This whitepaper will serve as a primer to utilities that would like some background on how artificial intelligence and machine learning can be applied to energy storage system controls. It covers the key metrics and how to assess the operation of an AI-based product.

This whitepaper explores in detail one of the prevalent algorithms for the most common energy storage application; buying and selling energy in energy markets with prices established through day-ahead auctions. One of the major challenges with this application is the need for an accurate price forecasting tool and the ability to account for volatility in real time energy prices. Reinforcement learning, a data-driven, model-free, and self-adaptive approach that can learn from past data and decisions, is presented in this whitepaper. It has been shown to be simple and easy to implement and, more importantly, it eliminates the need for complex forecasting models.

The effectiveness of the algorithm is tested against DER-VET's perfect foresight optimization algorithm which, theoretically, produces the maximum possible revenue given the modeled conditions (energy market price profile). Furthermore, this whitepaper presents real time implementation of AI algorithms at an energy storage test facility located in Colorado.

AI Applications For Energy Storage Operation

AI-based approaches are used for a wide range of energy storage applications. This chapter presents an overview on both traditional approaches and AI-based algorithms for different energy storage applications.

Energy Arbitrage

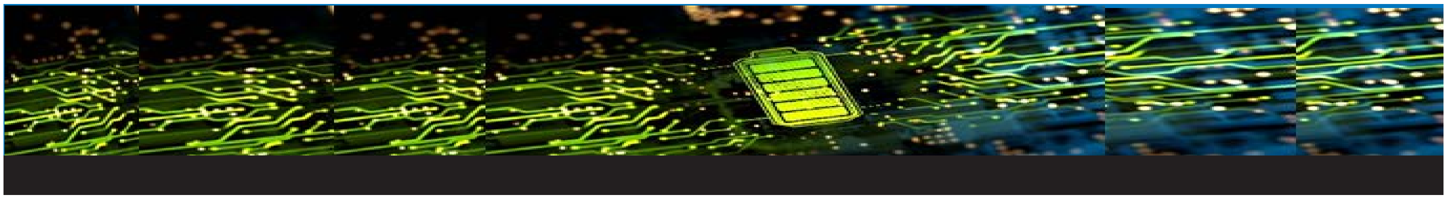
Large utility-scale energy storage systems can provide multiple value services, including energy arbitrage, based on day ahead ISO market signals, frequency regulation, spinning and non-spinning reserve, load following etc. Energy arbitrage is a major use case for energy storage systems. Since it is hard to predict price signal variations, most of the traditional approaches rely on sophisticated models to forecast price signals minutes to days in advance and control the energy storage accordingly. There are also few dynamic programming-based algorithms that circumvent the need to know the price signal beforehand.[2]

Given the challenges of price signal forecasting, data-driven model-free reinforcement learning (RL) approaches have recently emerged

Table of Contents

[1] BNEF 2021 Global Energy Storage Outlook (excludes pumped hydro).

[2] D. R. Jiang and W. B. Powell, "Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming," *INFORMS Journal on Computing*, vol. 27, no. 3, pp. 525-543, 2015.



as an attractive alternative for this problem. A reinforcement-based[3] algorithm may provide significant performance improvements and profit compared with existing traditional methods mentioned above. Following chapters will investigate these claims.

Customer Bill Reduction Services

Energy storage is increasingly being deployed by residential or commercial customer location as a behind the meter asset. It is primarily used for demand charge reduction and energy arbitrage functions, with a goal to lower the monthly utility bill for the customer. Traditionally, energy storage operations are implemented by either rule-based approaches or model predictive controllers. Reinforcement learning-based algorithms are used for customer bill reduction services[4] and it has been demonstrated that the monthly customer bill can be reduced up to 59% compared to a simple rule-based approach.

Energy Storage Degradation

Capacity degradation of an energy storage system is ***path dependent*** i.e., it depends on the energy storage system's operation and stress factors including temperature, average state-of-charge (SoC), charge, and discharge rates. A major challenge in predicting the life of energy storage assets is scarcity in availability of operational data. 1) Unfortunately rich Li-ion life testing data from electric vehicles are not generally useful for stationary applications. 2) Li-ion life evaluations in a laboratory setting for a 20-year/7,500 full cycle stationary application take about two years and are prohibitively expensive. Machine learning-based approaches enable rapid and precise prediction of Li-ion degradation and performance in complex and changing stationary use scenarios throughout a battery's entire life[5,6,7] Some of these algorithms are designed around predicting a cell's cycle life given data from its first 100 cycles.

[3] H. Wang and B. Zhang, "Energy storage arbitrage in real-time markets via reinforcement learning," in 2018 IEEE Power & Energy Society General Meeting (PESGM), 2018: IEEE, pp. 1-5.

[4] Guan, C., Wang, Y., Lin, X., Nazarian, S., & Pedram, M. (2015, January). Reinforcement learning-based control of residential energy storage systems for electric bill minimization. In 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC) (pp. 637-642). IEEE.

[5] Chen, B.R., M.R. Kunz, T.R. Tanim, E.J. Dufek, "A machine learning framework for early detection of lithium plating combining multiple physics-based electrochemical signatures" *Cell Reports Physical Science,* 2(3), 100352 (2021).

[6] Kunz, M.R., E.J. Dufek, Z. Yi, K.L. Gering, M.G. Shirk, K. Smith, B.R. Chen, Q. Wang, P. Gasper, R.L. Bewley, T.R. Tanim, "Early battery performance prediction for mixed use charging profiles using hierarchal machine learning" *Batteries & Supercaps*, 4, 1186-1196 (2021)

[7] Severson, K.A., et al, "Data-driven prediction of battery cycle life before capacity degradation", *Nature Energy,* 4, 383-391 (2019).

Secondary battery utilization strategy requires understanding the state of health of retired battery cells. A machine learning-based approach is used to screen and develop a reuse strategy, which has the potential to solve the problem of battery recycling in the future.[8]

Battery Safety

Temperature is a crucial state for guaranteeing the reliability and safety of a battery during operation. The ability to estimate battery temperature, especially the internal temperature, is of paramount importance to the battery management system for monitoring and thermal control purposes. A data-driven approach combining neural network (NN) and the extended Kalman filter (EKF) is proposed to estimate the internal temperature for lithium-ion battery thermal management.[9] An advantage of a data-driven approach is that it eliminates the need for any battery background knowledge, so the developed hybrid approach is applicable for any type of battery, irrespective of chemistry, and it is convenient to achieve reasonable thermal management.

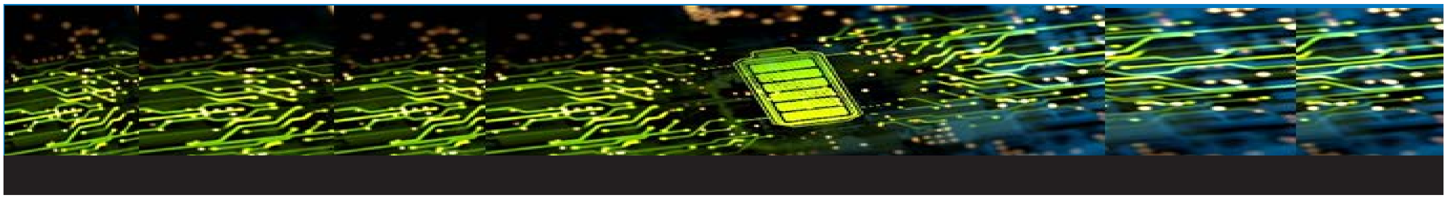
Anomaly Detection

A battery's state-of-power (SoP) refers to the maximum power that can be charged or discharged steadily by the battery within a predetermined time interval. There is a need to study the effect of SoP on battery health and to identify any anomaly ahead of time. This is especially important with the increase in fast-charging infrastructure for electric vehicles. Two issues should be considered: (1) the influence of both the ambient temperature and the rise in temperature caused by high power, and (2) the influence of changes in the state of charge (SoC). A model-based extreme learning machine is demonstrated to predict future battery voltage, power, and surface temperature for any given load current.[10]

[8] Garg, A., Yun, L., Gao, L., & Putungan, D. B. (2020). Development of recycling strategy for large stacked systems: Experimental and machine learning approach to form reuse battery packs for secondary applications. *Journal of Cleaner Production,* 275, 124152.

[9] Liu, Kailong, Kang Li, Qiao Peng, Yuanjun Guo, and Li Zhang. "Data-driven hybrid internal temperature estimation approach for battery thermal management." *Complexity* 2018 (2018).

[10] Lin, Peng, Peng Jin, Jichao Hong, and Zhenpo Wang. "Battery voltage and state of power prediction based on an improved novel polarization voltage model." *Energy Reports* 6 (2020): 2299-2308.



Solar PV Power Smoothing

Battery Energy Storage Systems (BESS) coupled with solar generation can strengthen the renewable power generation source and smooth power output, thereby avoiding any fluctuations that can appear in voltage and power at the point of interconnection. Smoothing filters such as Low Pass Filters (LPFs) are integrated along with the BESS for optimal functioning and cost reduction. It has been established that the time constant of an LPF directly impacts the degree of solar PV smoothing. The machine learning-based method has been demonstrated to intelligently control the LPF time constant to efficiently rid the PV profile from fluctuations while operating under practical constraints.[11]

Other Miscellaneous Applications

Energy storage coupled with solar PV has multiple applications. Machine learning and reinforcement learning-based algorithms have been proposed for energy management and bidding in the capacity market.[12,13] In addition, there are some algorithms for optimal real-time operation of the asset.[14]

Reinforcement Learning-Based Approach for Energy Arbitrage

EPRI has developed the DER-VET software as a planning tool to estimate the revenue a system can generate by performing various services, like energy arbitrage. The optimization-based planning tool gives an upper limit on how much revenue the storage system can generate under a set of conditions. The tool relies on explicit price data as an input for its perfect foresight modeling. Contrarily, a properly implemented reinforcement learning control algorithm may enable profitable operation of an energy storage system without a priori knowledge or forecasts of future prices, but this will obviously underperform relative to DER-VETs perfect foresight model. This whitepaper assesses the feasibility of RL-based algorithm for energy arbitrage application.

Reinforcement Learning Algorithm

A reinforcement learning policy behaves similarly to a controller in a control system in terms of how it observes the environment and generates actions to carry out a task as efficiently as possible. Numerous control problems, e.g., robotics and automated driving, need sophisticated, nonlinear control structures. These issues can be solved using methods like gain scheduling, robust control, and nonlinear model predictive control (MPC), but doing so frequently necessitates a control engineer with a high level of domain knowledge. The complexity of a nonlinear MPC's computational requirements can make the resulting controllers difficult to implement.

Reinforcement learning uses a general framework where current actions depend on the present system states (state space). In this implementation, there are only three possible actions possible for energy storage – charge or discharge at full power and stay idle. For every action, there is a reward generated to receive feedback on the activity. The rewards from the past actions are used to update a matrix referred to as the Q matrix. All the components are explained in detail in the following sections.

State Space

In reinforcement learning, the energy arbitrage for energy storage is designed as a state space which is a finite number of discrete states. To be specific, the system's state is represented by the current price p_t and the ability of the energy storage to further charge or discharge, which depends on the energy storage system's current state of charge (SoC). The state space for this problem is represented as follows,

$$S=\{1,\dots,M\}\times\{0,1\}\times\{0,1\}$$

where {1,…., M} represents M even price intervals from the lowest to the highest, and the other two binary states correspond to 'allowed to charge' and 'allowed to discharge.'

Action Space

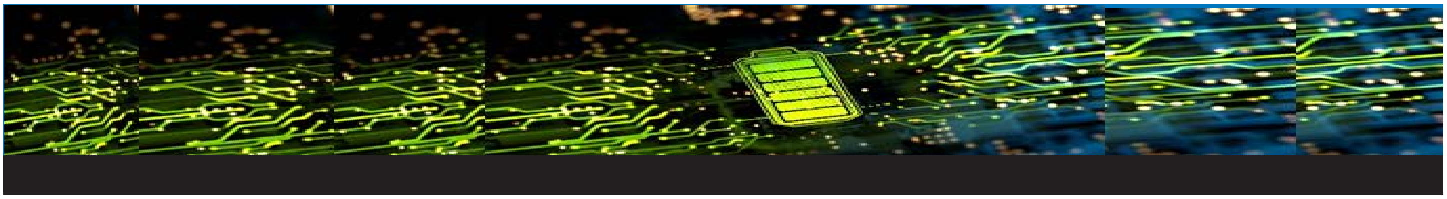
In practice, there are infinite number of possible energy storage operations. In this formulation, the action space is simplified into three possibilities by assuming that the energy storage can either stay idle or charge/discharge at rated power capacity, as long as

[11] Syed, M. A., & Khalid, M. (2021, January). Machine learning based controlled filtering for solar PV variability reduction with BESS. In 2021 *International Conference on Sustainable Energy and Future Electric Transportation (SEFET).* (pp. 1-5). IEEE.

[12] Huang, B., & Wang, J. (2020). Deep-reinforcement-learning-based capacity scheduling for PV-battery storage system. *IEEE Transactions on Smart Grid*, 12(3), 2272-2283.

[13] Keerthisinghe, Chanaka, Archie C. Chapman, and Gregor Verbič. "Energy management of PV-storage systems: Policy approximations using machine learning." *IEEE Transactions on Industrial Informatics* 15, no. 1 (2018): 257-265.

[14] Henri, Gonzague, Ning Lu, and Carlos Carreio. "A machine learning approach for real-time battery optimal operation mode prediction and control." In 2*018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pp. 1-9. IEEE, 2018.



it doesn't exceed the rated energy capacity. The allowable charge/discharge rates are $\widetilde{D}_{max} = \min\{D_{max}, E_{t-1} - E_{min}\}$ and $\tilde{C}_{max} = \min\{C_{max}, E_{max} - E_{t-1}\}$.

$$A = \{-\widetilde{D}_{max}, 0, \tilde{C}_{max}\}$$

Where, $a = -\widetilde{D}_{max}$, denotes discharge either at maximum rate D_{max} or until the storage hits the minimum limit of allowed stored energy E_{min}. Similarly, action $a = \tilde{C}_{max}$ denotes charge at maximum rate C_{max} or until the storage reaches the minimum allowed limit on energy.

Reward

Reward is a feedback mechanism for the reinforcement algorithm to update its policy. It is calculated based on the interactions in a specific environment, which in this case is the energy market and the day-ahead price signals. The reward function is calculated as shown below based on the average prices,

$$r_t = \begin{cases} (\bar{p}_t - p_t)\tilde{C}_{max} & \text{if charge} \\ 0 & \text{if standstill} \\ (p_t - \bar{p}_t)\widetilde{D}_{max} & \text{if discharge} \end{cases}$$

where the average price $\bar{p}_t$ is calculated by,

$$\bar{p}_t = (1-\eta)\bar{p}_{t-1} + \eta p_t$$

in which η is the smoothing parameter. When the energy storage charges at a price lower than the average price (i.e., $p_t < \bar{p}_t$), it will get a positive reward, otherwise it will receive a loss. Similarly, the reward function encourages the energy storage to discharge at a high price.

Q Learning Algorithm

With the state, action, and reward defined, the real-time charge and discharge policy is updated based on Q-learning, a popular subclass of reinforcement algorithms. A relatively simple Reinforcement Learning algorithm (Q-Learning) was utilized as described in the paper "Energy Storage Arbitrage in Real-Time Markets via Reinforcement Learning."[15] The energy storage maintains a state-action value matrix Q, where each entry Q(s,a) is defined for each pair of state s and action a. When the energy storage takes a charge/discharge action, the value matrix is updated as follows:

$$Q(s,a)_t = (1-\alpha)Q(s,a)_{t-1} + \alpha\left[r_t + \gamma \max_{a'} Q(s', a')\right]$$

where the parameter

α ε *(0,1]* is the learning rate weighting the past value and new reward.

γ ε *[0,1]* is the discount rate determining the importance of future rewards. After taking an action a, the state transits from *s* to *s´*, and the energy storage updates the value matrix incorporating the instant reward r_t and the future value $\max_{a'} Q(s', a')$ in state *s´*. Over time, the energy storage can learn the value each action in all states. When *Q(s,a)* converges to the optimal state-action values, we obtain the optimal arbitrage policy.

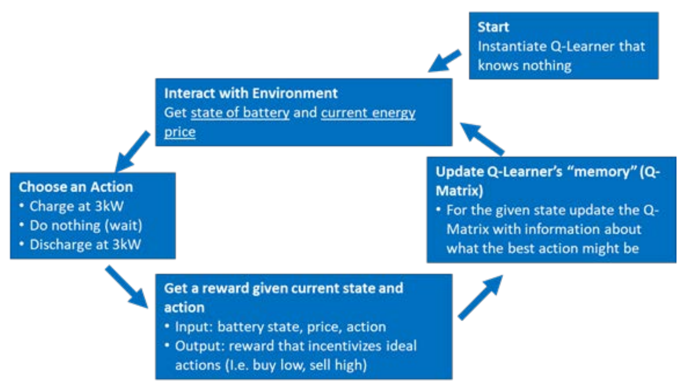


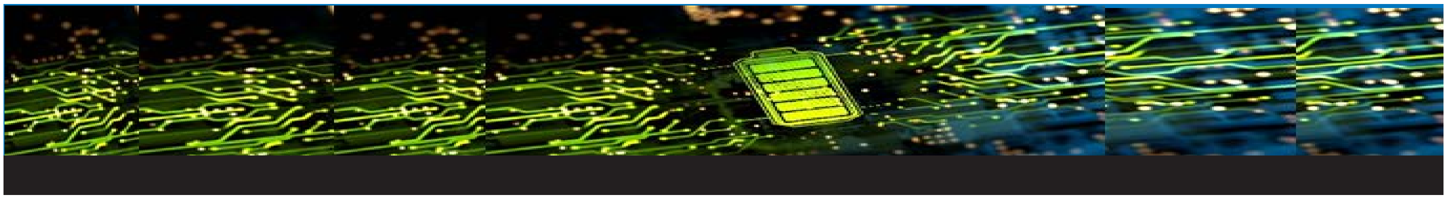
Figure 1 Simple Reinforcement Learning Algorithm Flow Chart

The step-by-step Q-learning algorithm for energy arbitrage is presented in the above Algorithm. To avoid the learning algorithm getting stuck at sub-optimal solutions, ε-greedy approach is used. The algorithm not only exploits the best action but also explores other actions, which could potentially be better. Specifically, using ε-greedy, the algorithm will randomly choose actions with probability ε ϵ [0,1] and choose the best action with probability *1*-ε.

Energy Arbitrage Revenue Results

To test the algorithm, the total revenue that can be generated from historical market data is calculated. NYISO 2021 market price signals ($/kWh) (top of Figure 2) were used for this analysis. Energy

[15] H. Wang and B. Zhang, "Energy storage arbitrage in real-time markets via reinforcement learning" in 2018 IEEE Power & Energy Society General Meeting (PESGM), 2018: IEEE, pp. 1-5.



storage size of 3kWh and 9.8kWh was used for the analysis. This sizing was chosen to match the real battery system with which this algorithm would be tested in Colorado.

The solution performance and reliability in the RL-based inference models used for decision-making can be significantly impacted by choosing the appropriate hyperparameters i.e., α-learning rate, γ-discount rate and ε-exploration probability. Finding the optimal settings is a computationally expensive process, called a hyperparameter search. The grid search method is used to find the best hyperparameters that generate the maximum revenue for the historical price signal.

Figure 2 shows the grid matrix of cumulative revenue for all possible combinations of alpha, gamma, and epsilon. Epsilon is the probability number that determines if the system has to do exploration or use the Q matrix value. A high epsilon means ignore the Q values and use only the random exploration. It is obvious that the high revenue is possible when epsilon is low. For a fixed epsilon, the variation in alpha and beta did not have much impact on the cumulative revenue i.e., it is between \$10.10 and \$10.60.

The best alpha, gamma, and epsilon values are 0.55, 0.88 and 0.35 respectively. These values will be used for comparison purposes throughout this whitepaper. The dispatch results corresponding to these hypermeter values are shown in Figure 3.

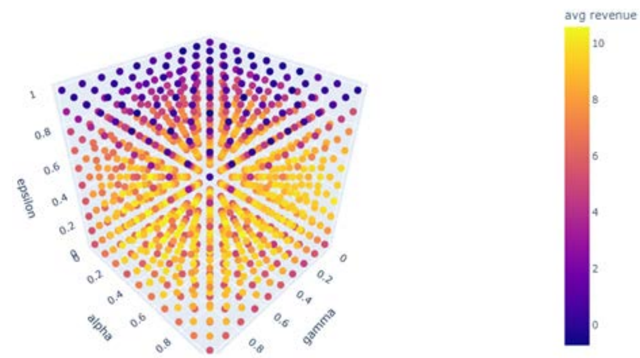


Figure 2 Average revenue (in \$) from a sensitivity analysis to discover the best alpha, gamma, and epsilon values for this Q-learning instance given the price signal.

The top sub figure in Figure 3 shows the historical NYISO price signal (\$/kWh) that will be used for the yearly simulation. It is 15-min resolution data and the actions are calculated for every 15-minute interval. The second subfigure shows the action state of the energy storage, which is either buy, sell, or stay idle. The third subfigure is the state of charge, as a result of the charge and discharge actions throughout the year. The SoC is restricted to vary between 20-80% only. The fourth subfigure is the instantaneous reward that is generated according to the defined function, which is based on the average price. In this implementation, a moving average for two days is used. Finally, the last subfigure is the sum of all instantaneous rewards over time to calculate the cumulative reward.

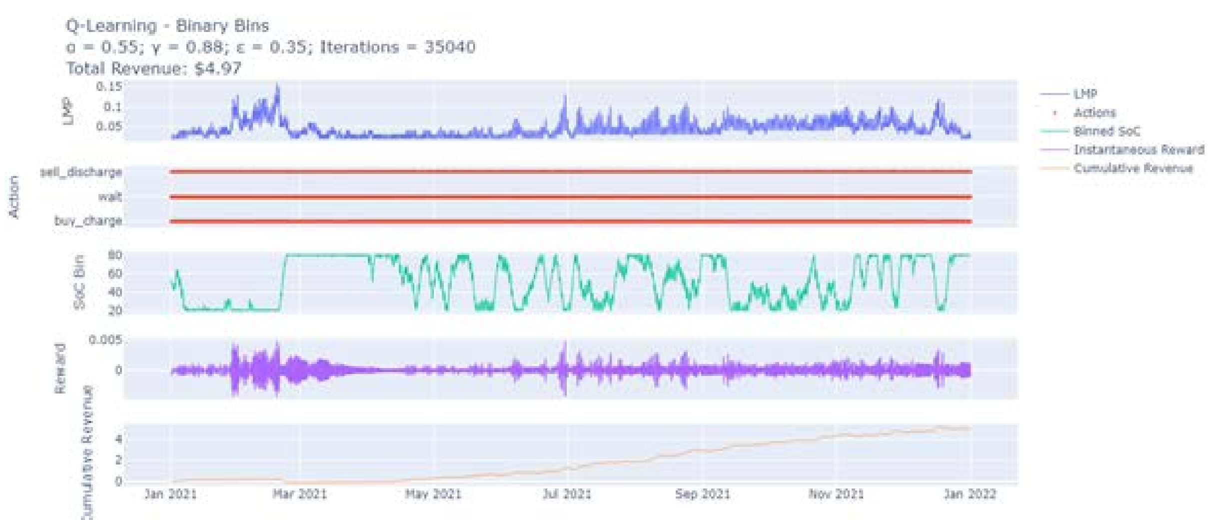
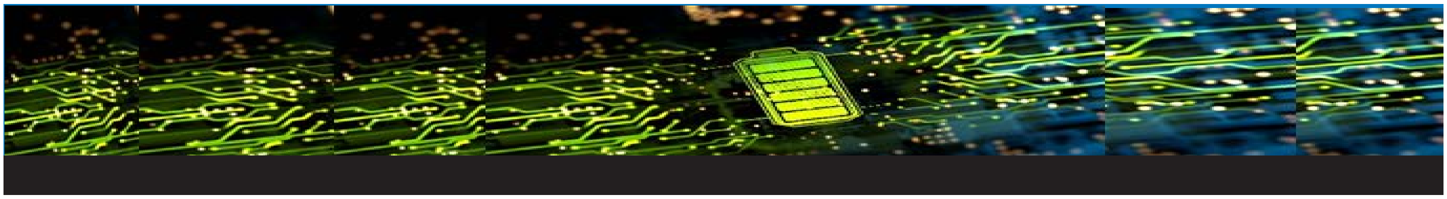


Figure 3 Results from a simulated one-year run of the Q-learning algorithm using historical NYISO data. Subplot descriptions (top to bottom) a. 15-minute price signal being fed to the Q-learner, b. The action taken by the Q-learner at each 15-minute interval, c. The SoC of the system, d. Instantaneous reward from the reward function given the state and chosen action, e. The cumulative revenue of the Q-learner.



Comparison with DER-VET

EPRI has developed DER-VET software, a planning tool to estimate the revenue that the energy storage system can generate from energy arbitrage and other services. The optimization-based planning tool gives an upper estimate on what the energy storage can generate. DER-VET is a mixed integer linear programming-based optimization tool with a monthly optimization window for this implementation.

Table 1 shows the comparison of the revenue generated from DER-VET and Q-learning for the same modeled energy storage and market price signal. As expected, results show the maximum revenue from DER-VET is substantially higher than the RL model-based results. This is because the perfect foresight-based algorithm has the ability to look ahead and plan its dispatches based on perfect knowledge of the price in a monthly optimization window. Reinforcement learning dispatches storage using a policy that begins with no knowledge and is updated based on each decision and corresponding reward with no ability to look ahead.

Table 1. Comparison of profit for simulation

	DER-VET ($) (15 min)	Q-Learning ($) (15 min) (All past data)
NY-ISO	14	4.95

Figure 4 shows the dispatch results from the reinforcement algorithm (green) compared to DER-VET (blue). It can be observed that both the algorithms have similar trends in charging and discharging the energy storage. For example, at those instances indicated with a red arrow, both algorithms discharge energy storage to its fullest. At the same time, instances indicated with a blue arrow show that the RL algorithm is conservative, and did not discharge the battery to the lowest SoC level. This is because DER-VET knows ahead of time that the price is not going up any further and it can discharge now at this given price..

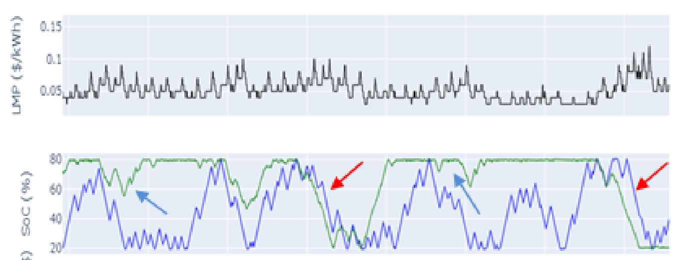


Figure 4 Comparing the optimal operation from DER-VET to the simulated operation from the Q-learning algorithm. At some points the Q-learner performs an action very similar to the optimal strategy (red arrows) at other times the Q-learner acts more conservatively and chooses not to fully discharge (blue arrows).

Testing on a Real System

This chapter explores the possibility of implementing a reinforcement learning policy for energy arbitrage on a real energy storage system using a Raspberry Pi 4B single-board computer, a small, inexpensive computing device. The Raspberry Pi in this project enables autonomous system control, remote monitoring, and data collection.

The Test Facility

The reinforcement algorithm was tested on a real energy storage system located at the Solar Technology Acceleration Center (SolarTAC), a test facility near Denver, Colorado (Figure 5). SolarTAC offers plug and play microgrid testing capabilities with over 2MW of local renewable generation capacity. The next section provides details of the hardware and software used for testing.

Figure 5 An image of the Solar Technology Acceleration Center (SolarTAC) in Watkins, CO where the battery was located and tested.

Hardware

The hardware installed at SolarTAC to make this testing possible includes:

- A previously installed 7.6 kW Solar Edge inverter with a 5kW-9.8kWh LG Chem battery that is set up on one of the SolarTAC test pads. For this testing, the Q-learner will only be able to use 3kW of the 5kW available and must keep the SoC between 20%-80%.
- Raspberry Pi 4B with 4GB of RAM to facilitate remote and autonomous control and data collection from the Solar Edge system.
- Cradlepoint IBR600C cellular modem to provide remote access to the equipment on site.

- Off-site EPRI server running a PostgreSQL database for real-time data collection and remote analysis.

Software

The most important software used in this project is the Volttron, an open-source, Python-based tool developed by Pacific Northwest National Laboratory that provides an environment where "data and devices can connect seamlessly and securely to make decisions based on a user's needs and preferences." It is often used to orchestrate relatively complex controls and optimizations within smart buildings that contain many different internet of things (IoT) devices that would not otherwise communicate in a desired fashion. It is designed to run on low resource Linux machines, like the Raspberry Pi, and easily supports custom development to fit specific project needs.

Volttron is a platform where Volttron agents, each with its own dedicated role to perform, communicate with each other over a ZeroMQ message bus. These agents can be combined in many ways to suit a specific project's needs.

For our use case, Volttron, and the underlying agents, are used to accomplish the following:

- **Platform, device, and actuator agents**: Manage modbus communication (read and write) between Volttron agents and the Solar Edge energy management system (EMS).
- **SQL historian agent**: Collect data every 10 seconds from the inverter, battery, and other Volttron agents. Write these data to the remote PostgreSQL database.
- **Custom developed agents**:
  - **Price signal agent**: Reads in a price signal from a CSV file and iterates through the rows broadcasting a synthetic price signal to the other agents. In the future, this could be replaced with actual ISO signals.
  - **Q-learning agent**: Contains the q-learning code and acts on new price signals to decide the best course of action given the current state (battery SoC and energy price).

Figure 6 shows the test setup that interfaces with computing facilities from different geographic locations. At SolarTAC in Denver, CO, there is a Solar Edge energy storage and inverter system and Raspberry Pi 4B. The Raspberry Pi hosts the Volttron platform with agents that facilitate remote and autonomous control and data acquisition. The Q-Learning Controller code runs as one of the Volttron "agents". The computing facility at Palo Alto, CA, hosts the SQL database. It also offers resources for post processing for the data visualization, and reporting.

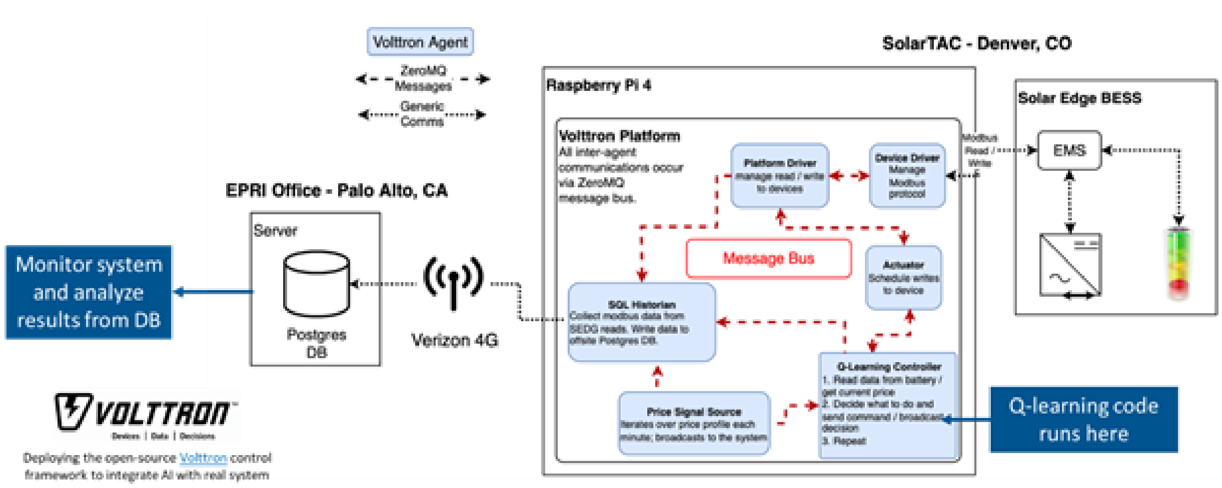
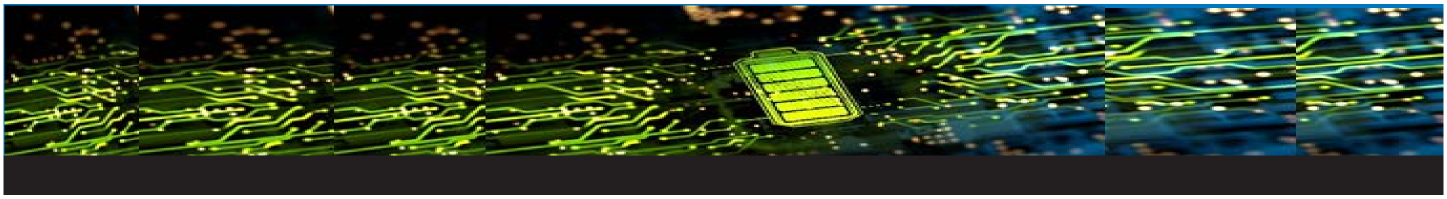


Figure 6 Diagram showing the hardware and software configuration used for real system testing of the Q-learning algorithm.



Results

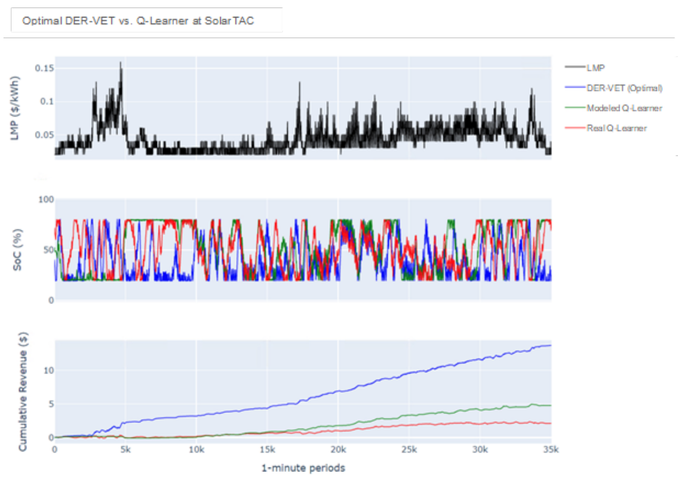


Figure 7 A comparison showing the differing operation and revenue for 3 cases; DER-VET's perfect foresight model (blue), the simulated Q-learning operation (green), and the same Q-learning setup applied to the real battery system at SolarTAC.

The Q-learning policy for energy arbitrage was successfully tested at the SolarTAC test facility. For comparison, the same one-year price signal as used in the modeling phase is also used for real time implementation. Note that this one-year, 15-minute price profile had to be condensed to a 1-minute interval for real system testing because there was not enough time to test the system and collect results for a full year. Figure 7 shows the comparison of the state of charge and cumulative revenue due to the charge and discharge actions on the energy storage. As discussed earlier, the DER-VET results set the upper limit on the optimal operation of the energy storage asset. In this case it is $15 for the 3kW, 9.8kWh energy storage system. The simulated Q learner is shown to provide a cumulative revenue of $5, while the real-time implementation was slightly less than the simulated value – $2.50. Ideally, the revenue from the simulated and real implementations should not have been very different. But this is likely due to the indeterministic nature of this reinforcement learning algorithm. Although separate instances of the Q-learner should eventually converge on the same policy, no two runs of the algorithm are exactly alike. This is because, especially early on in the run when the algorithm is still exploring, there is an inherent aspect of randomness to the decisions. So, it is possible that the simulated Q-learner shown here happened to "pick up" on the proper charge/discharge strategy earlier than the real system Q-learner.

Conclusion

There are a growing number of applications of AI for energy storage. There are also a number of AI algorithms and policies for every application. This whitepaper looked at one of the reinforcement learning approaches for energy arbitrage application. The results from the algorithm were compared against optimization results with perfect foresight on price signals. It is clear that there is still room for improvement. Following are some of the avenues that could help improve the algorithm's performance:

- State space: This implementation considered 10 discrete price bins and 2 other binary states based on whether it is possible to charge or discharge. Considering additional state spaces such as the time of the day, seasons, and other parameters that could add more insights into the price variations can help develop a robust tool.
- Reward: This implementation uses average price as a reward parameter. The reward signal could also be sophisticated to improve the performance of the algorithm.
- Reinforcement learning (RL) can use deep neural networks to learn an effective control strategy in very complicated environments. With proper design, this feature offers tremendous potential for RL to develop a battery charging/discharging policy for energy arbitrage.

On the positive side, this whitepaper has shown the possibility of using reinforcement learning without any prior knowledge of the market, or complex engineering, to provide energy arbitrage functionality. The implementation has shown the algorithm can learn from past historical dispatches and adapt to reach an optimal dispatch policy.

Future work should investigate how this control technique compares to contemporary control algorithms to assess if the purported benefits of reinforcement learning are justified.

Note

For further information about EPRI, call the EPRI Customer Assistance Center at 800.313.3774 or e-mail askepri@epri.com.

About EPRI

Founded in 1972, EPRI is the world's preeminent independent, non-profit energy research and development organization, with offices around the world. EPRI's trusted experts collaborate with more than 450 companies in 45 countries, driving innovation to ensure the public has clean, safe, reliable, affordable, and equitable access to electricity across the globe. Together, we are shaping the future of energy.

EPRI RESOURCES

Suma Jothibasu, *Technical Leader*
650.855.7933, sjothibasu@epri.com

Joseph Thompson, *Engineer/Scientist III*
650.855.2587, jthompson@epri.com

Fouad Hasan, *Student Employee*

Saumil Shah, *Student Employee*

AI Board Initiative

3002023924 January 2023

EPRI
3420 Hillview Avenue, Palo Alto, California 94304-1338 • USA • 800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com