

# Reinforcement Learning-Based Stackelberg Game for Optimizing EV Utilization and Shared Energy Storage System Charging Price

Yuito Ohno <sup>1)</sup> Shinkichi Inagaki <sup>2)</sup> Tatsuya Suzuki <sup>1)</sup> Hikari Nakano <sup>3)</sup>

<sup>1)</sup> Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8603, Japan

<sup>2)</sup> Faculty of Science and Engineering, Nanzan University, 18 Yamazato-cho, Showa-ku, Nagoya, Aichi 466-0824, Japan

<sup>3)</sup> NUVVE JAPAN Co., Ltd., 14F Hibiya Central Building, 1-2-9 Nishi-Shimbashi, Minato-ku, Tokyo 105-0003, Japan

**KEY WORDS:** EV and HV systems, Energy management, Energy control system, Shared Energy Storage System [A3]

This study investigates an energy management system that coordinates electric vehicles (EVs) and a shared battery within a local community for the effective utilization of surplus photovoltaic generation. We construct an operation method for the system and evaluate it from various perspectives, thereby discussing its effectiveness and remaining challenges toward practical implementation.

With the expanding installation of photovoltaic (PV) generation, surplus power that cannot be consumed within a community arises during the daytime, and using the batteries of EVs as demand-side flexibility is an effective way to absorb it. The authors have proposed an energy management system that coordinates a shared energy storage system (Shared ESS) owned by a commercial facility with a fleet of EVs, so that electricity is circulated within the community.

However, the EVs are owned by individual households, and the operator of the Shared ESS can command neither their travel nor their charging; the only variable it can manipulate is the charging price. Users decide whether and when to charge according to the time of day, their own state of charge and the posted price, while the number of EVs that can be connected simultaneously is limited. A low price concentrates the arrivals and produces waiting time once the ports are exceeded, whereas a high price relieves the congestion but drives the utility of visiting below the threshold of some users, who then lose the charging opportunity itself. Reducing the waiting time and securing charging opportunities are therefore in a trade-off that a single uniform price cannot escape.

Time-of-Use (ToU) tariffs switch the price between time slots, but their shape is determined by supply-side circumstances such as the surplus generation and does not reflect the congestion arising on the demand side. A Stackelberg game with the operator as the leader would anticipate the response of the users, but the follower decision here contains an integer variable representing whether to visit the facility, so replacing the lower-level problem by its KKT conditions is not valid. Reinforcement learning (RL) circumvents this difficulty: since the mapping from the price to the resulting cost is treated as a black box, the discrete visiting decision and the contention for the charging ports can be evaluated without approximation. Existing RL pricing studies, however, target a single charging station and do not couple the price with community-level dispatch.

In this study, we construct a simulator that integrates a discrete choice model of the EV user response, a first-come-first-served assignment to the charging ports, and the optimization of the power dispatch of the whole system, and determine the time-block charging price of the Shared ESS by RL. In the bi-level structure of Fig. 1, the leader divides the opening hours of the facility into equal blocks and posts the price vector once at the beginning of each day. Given this price sequence, its initial battery energy and its household demand, each follower EV solves its own optimization problem for the desired arrival step and the usage of the vehicle, the departure timing being governed by a binary logit utility of visiting that depends on the time, the price and the state of charge. The desired arrivals are then assigned to the finite number of ports, which fixes the waiting time of each EV and whether it could be charged before closing, and the power flows of the whole system are finally optimized.

The pricing problem is formulated as a Markov decision process whose step is one day and whose episode is seven consecutive days. The state comprises the battery energy of each EV and of the Shared ESS carried over from the previous day and the number of remaining days, the action is the vector of five block prices bounded within 5-20 JPY/kWh, and the reward is the negative weighted sum of the system cost, the total waiting time and the lost charging opportunities. The policy is trained by Soft Actor-Critic, with the mixed-integer follower response and dispatch replaced by surrogate models during training only and solved exactly with Gurobi at evaluation.

The method was validated with measured profiles of household demand and of the demand and PV generation of the facility, for ten households each owning a 60 kWh EV and one facility whose Shared ESS charges at most two EVs at a time at 36 kW between 10:00 and 20:00. The preference coefficients of the logit model were estimated from a choice experiment generated by a large language model, and the policy was trained over 70 days and evaluated over 21 days.

Table 1 lists the daily averages over the evaluation period. Against the best uniform price obtained by a grid search (BestFlat 8), the proposed method reduced the waiting time per charged EV from 27.6 to 4.6 minutes while increasing the number of EVs charged from 5.22 to 5.98 per day, and it lowered the bill of the whole community as well. The learned price is high from 10:00 to 12:00 and low from 14:00 to 16:00, which suppresses the arrivals immediately after opening and shifts them to the afternoon. Under BestFlat 8 all five visiting EVs requested to arrive at 10:00 and three waited up to one hour; ToU 8/16 reproduces the same concentration since it posts 8 JPY/kWh at the opening hour, and Flat 17.4 attracted no EV at all.

In conclusion, determining the charging price of a Shared ESS by RL resolves the trade-off between the waiting time and the charging opportunity that a uniform price cannot escape, while also reducing the electricity bill of the community. Future work will examine larger EV fleets and validate the preference coefficients against real user responses.

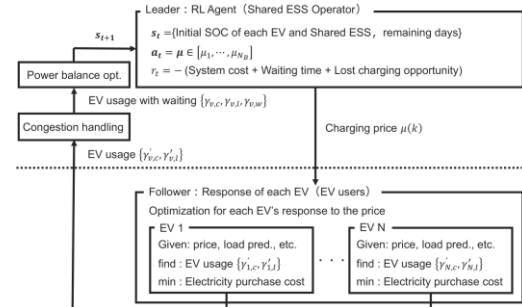


Fig. 1 Bi-level decision structure of the proposed framework.

Table 1 Comparison of the pricing schemes

|                                    | SAC (proposed) | BestFlat 8 | Flat 17.4 | ToU 8/16 |
|------------------------------------|----------------|------------|-----------|----------|
| Bill of the whole system [JPY/day] | 6675           | 6928       | 8666      | 6934     |
| Number of charged EVs [EV/day]     | 5.98           | 5.22       | 0.00      | 5.19     |
| Waiting time per charged EV [min]  | 4.6            | 27.6       | ---       | 27.2     |