Enhancing Flexibility in V2B Applications with Renewable Energy Resources

Maximiliano Trimboli, Nicolás Antonelli and Luis Avila

Laboratorio de Sistemas Inteligentes, CONICET-UNSL
mdtrimboli@unsl.edu.ar, nantonelli@unsl.edu.ar and loavila@unsl.edu.ar

Abstract. The incorporation of EV parking within vehicle-to-building (V2B) frameworks signifies not only a technological evolution but also a pivotal step towards constructing smarter and environmentally friendly urban environments. This initiative actively contributes to the optimization of system resources while also enabling the incorporation of renewable energy resources. In this study, we propose the development of reinforcement learning (RL) algorithms for the management of smart parking lots, aiming to minimize building energy purchases from the grid while ensuring efficient charging of EVs. The proposed methods obtained a 15% to 17% improvement in the evaluation reward in comparison with rule based method as a benchmark. In the realm of grid energy, they saved 9 to 11% in average purchase cost. In essence, these algorithms, after training, make more efficient decisions than more traditional control methods while ensuring electric vehicle (EV) charging.

Keywords: Electric vehicles · Smart Charging · Renewable Energy · Reinforcement Learning.

1 Introduction

The widespread adoption of EVs brings significant benefits, with a notable reduction in harmful emissions being paramount. As e-mobility continues to surge, investments in battery technologies and charge point infrastructures are also on the rise. Among the numerous challenges to address, those linked to changes in energy consumption behaviors associated with evolving charging processes are particularly crucial. Challenges also include difficulties in identifying charging locations during daylight hours, rising infrastructure and equipment costs, and potential overloads during peak periods. All of these constitute critical considerations for advancing EV charging infrastructure. As our society aspires to inhabit increasingly interconnected cities, the development of smart infrastructure must span various domains, including smart buildings, mobility solutions, energy systems, and monitoring of water and air quality [1]. In such scenarios, the implementation of smart parking can mitigate harmful emissions by incorporating renewable energy sources, predominantly through the integration of solar arrays [2]. Furthermore, the large-scale integration of EVs is expected to enhance sustainability by providing energy storage and creating new revenue sources from the EVs' batteries [3].

The integration of EV technologies not only demands policies fostering sustainability and grid stability but also initiatives encouraging user participation. For instance, prolonged waiting times at charging stations can significantly delay the recharging process, potentially dissuading consumers from opting for environmentally conscious mobility. Moreover, uncoordinated charging strategies within a limited charging infrastructure may contribute to increased demand peaks on the grid. Therefore, charge management algorithms will play a crucial role in maximizing the potential of smart structures [4]. Ultimately, the substantial energy demands at charging terminals could strain electrical systems, potentially diminishing service quality. In a study by [5], Moghaddam et al. (2017) designed a charging strategy to address these concerns by providing multiple charging options, modeling the optimal charging station as a multi-objective optimization problem. Additionally, [6] Bose et al. (2023) utilized classical topology results to formulate a new charging station placement algorithm in the context of smart cities.

Considering that EVs can offer flexibility to support the operation of electrical systems in smart buildings, the concept of utilizing parked EVs as energy storage devices is exceedingly attractive [7, 8]. However, the introduction of EVs also introduces uncertainty into the grid, as, with the Vehicle-to-Building (V2B) functionality [9], they can provide energy to the building loads by discharging the battery. In this context, the works of [7] and [10] assess the utilization of parked EVs to shave the peak load in building-integrated microgrids. The former develops an optimization framework to control the microgrid's operation and manage power flow exchanges, ensuring a high quality of service to EV owners. The latter predicts the day-ahead building electricity demand profile and identifies the optimal schedule for charging and discharging EVs to minimize electricity peak demand.

Due to the ability of EVs to provide power to building loads by discharging the battery during high consumption peaks, it is possible to reduce energy consumption and greenhouse gas emissions, moving towards nearly-zero energy buildings. However, it should be noted that the arrival and departure times of EVs add uncertainty to the grid. As a result, it becomes clear that the development of suitable algorithms to control and optimize the charging/discharging process is crucial for the smooth integration of parked units into the building power grid.

In this study, we use a virtual environment to simulate the charging dynamics of grid-connected electric vehicles in a smart building. The model reproduces various conditions, such as disturbances, weather conditions, pricing models, and stochastic arrival and departure times of electric vehicles. By providing control over multiple charging points per vehicle, it provides insight into the underlying dynamics and their impact on the energy efficiency of the building. The main objective is to minimize the electricity costs that the smart building absorbs from the grid by employing parked EV batteries as flexible resources. A central energy manager is in charge of the power supply to the EVs and building loads, harnessing the EVs' stored energy to meet demands and avoid peak loads. Evaluation of the RL algorithms within the simulated environment showed savings of up to 4.5% in terms of energy and up to 13.2% in associated costs, indicating that these algorithms make more efficient charging station management decisions.

In the next section, we present the virtual environment that simulates the operation of a charging station considering random arrivals and departures of EVs. In section 3, we present the RL formulations aimed at achieving efficient charging control, optimizing energy consumption, reducing environmental impact and maximizing user satisfaction. In section 4, we present the experimental results obtained, and in section 5, we offer some concluding remarks.

2 Charging station model

This work implements a simulated model depicting the operations of a smart parking facility, interconnected to two core components: a photovoltaic generation system and a grid connection responsible for supplying unlimited energy to a series of electric vehicle charging stations within a building. The architecture of the simulated system is shown in Fig. 1.

The photovoltaic generation system is composed of 40 solar cells that provide the occupied parking spaces with a maximum energy of 60 kWh that depends solely on a fixed daily solar irradiance profile. When the system does not have sufficient capacity to self-supply the energy demand of the batteries in addition to the demand of the building, the smart manager has the option to purchase electricity from the connection to the grid. The grid distribution company sets a variable price profile throughout the day, independent of the consumption behavior of the smart parking system. In addition, due to the V2B functionality of EVs, the model supports the energy transfer between batteries connected to the parking spaces, in combination with the use of the electrical sources

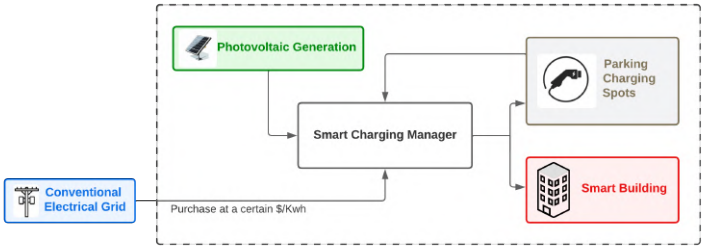

Fig. 1. Model energy flow diagram.

mentioned above. It is possible that the renewable energy generated is greater than the energy demand, in which case the remaining energy is considered wasted energy because the system is not able to use it.

The smart manager is in charge of interacting with the aforementioned sections and making decisions based on the energy flows, in order to meet its main objectives, minimizing the costs derived from energy purchases from the grid and correctly carrying out the charging processes to the parked EVs. This task is carried out by only defining which EV batteries to charge and which to discharge at each time step, thus implicitly interfering in the moments when purchases of electrical energy are made from the grid.

We assume that a single spot can perform the charging process of more than one EV throughout the day. It should be noted that the batteries of the EVs in the parking area always have the same manufacturing characteristics, therefore they have the same hyperparameters. Moreover, the maximum charging/discharging speed of the EVs is determined by the characteristics of the charging spot itself, i.e., all EVs entering the spots can support the charging/discharging speed of the parking area and from (1), the smart manager can predict the charging status of each vehicle in the resp. next hour.

$$SoC^i_{t+1} = SoC^i_t + \frac{E^i_{dem_t}}{C_{bat}} \tag{1}$$

where SoC is the state of charge, E_{dem} is the energy required or available from the battery and C_{bat} is the battery capacity. For each variable the superscript i indicates the corresponding vehicle and the subscripts t and $t+1$ the current and next hour. In this way the SoC is modified by the energy demanded over the battery capacity, this is a kind of propotion of the energy that the battery will be charged/discharged with respect to the maximum energy it can store.

3 Intelligent energy management

In this section, advanced RL methods for smart energy management will be explored. We will examine two specific approaches: Deep Deterministic Policy

Gradient (DDPG) algorithm, which combines deep learning techniques and deterministic policies, and Proximal Policy Optimization (PPO) algorithm, known for its ability to balance exploration and exploitation of policies, thereby enhancing efficiency and robustness in energy management.

3.1 Deep deterministic policy gradient (DDPG)

DDPG algorithm uses features of a method known as Deep Q-Network (DQN) and extends it to a multidimensional continuous action domain by combining it with the deterministic policy gradient (DPG) [11]. Thus, it obtains a deterministic action $\mu_\theta(s)$, where s is the current state, through an actor-critical approach without the need for prior knowledge of the policy gradient-based model to determine in which direction parameters should be adjusted in pursuit of improved performance outcomes.

The method is minimally conformed by two deep neural networks: the actor network $pi(s_t, \theta)$ that represents the agent's policy to determine the actions to be performed a_t to be executed in the state s_t, where θ is the network parameter, and the critical network that is used to evaluate the actions executed during the corresponding states, by calculating state-action functions $Q(s, a)$. In addition, DDPG adds two other networks called target networks in order to improve the learning stability in the actor-critical system.

The gradient policy is computed based on the critic's evaluation, and the actor parameters are adjusted to improve decision making. The optimization of the critic model is based on a loss function dependent on the difference between the actual reward and the critic's estimate.

3.2 Proximal Policy Optimization (PPO)

PPO is a policy gradient algorithm that uses in the objective function, the ratio of probabilities of the previous and current policy given a state-action pair, to improve training efficiency and reuse sampled data. In this way, quantifying the importance of the chosen action under the current policy compared to the previous policy limits the magnitude of policy updates during training, thus contributing to the stability of the algorithm [12, 13].

The general form of the optimization objective function is given by:

$$L^{CLIP}(\theta) = \mathbb{E}_t \left[\min \left(r_t(\theta)\hat{A}_t, \text{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t \right) \right] \tag{2}$$

where $r_t(\theta)$ is the probability ratio mentioned above, $\hat{A}_t$ is the advantage value representing the difference between the value of the action taken and the average value in the current state, and this is another term that improves the stability and convergence speed since it can reduce the variance of the gradient estimation.

Unlike the DDPG, the PPO algorithm focuses on stochastic policies, where the action is obtained randomly from a sample of a probability distribution. This implies a wider range of exploration since the agent can explore different actions according to the probability assigned by the policy.

4 Experiments

This section presents the validation process of the proposed system through an experimentation stage on the simulated environment, which maintains certain conditions and hyperparameters to obtain a consistent performance comparison.

The main hyperparameters of the environment are shown in Table 1. Concerning vehicle arrivals and departures, the movements are randomly defined at the beginning of the simulation, therefore, the times that the vehicles stay at the stations are variable. The times are limited both for entry (from 0 hs to 20 hs) and withdrawal (between 4 hs and 10 hs after arrival), as well as the SoC values at which vehicles arrive at the parking area. The maximum net energy that a parking spot can carry is 10 kWh, which means that a vehicle requires 3 hours to fully charge its battery (SoC from 0 to 100%, C_{bat} de 30kWh).

Regarding the proposed RL algorithms, the main hyperparameters used during the training stage, optimized by means of heuristic processes, are detailed in the tables 2-3.

Table 1. Main hyperparameters of the environment.

Symbol	Variable	Range
SoC_0	Initial Charge State	[20%, 50%]
h_0	Arrival time range	[0, 20]
h_f	Departure time range	[h_0+4, h_0+10]
C_{bat}	Individual capacity of EVs	30KWh
$E^i_{u_t}$	Maximum net energy	10 KWh
$P_{ch_{max}}$	Maximum load power	11KW
η_{ch}	Charging/discharging efficiency	0.91
N_{pv}	Number of solar cells	20
P_{pv}	Nominal photovoltaic power	10KW

4.1 Environment formulation

Since the proposed system was formulated under an MDP process, the following sections specify the elements that define their corresponding tuple.

State space: according to (3), the state at each time step is defined by the current value of the radiation and the prediction of what values it will have for the next 3 hours (G_t), similarly happens with the purchase price of energy from the grid and building consumption, incorporating current values and their predictions (Pr_t and Bc_t). All these values do not vary across episodes the state

Table 2. Main hyperparameters of the DDPG algorithm.

Symbol	Designation	Range
σ	Actor network layers	[356, 128]
ρ	Critical network layers	[356, 128, 32]
e_{DDPG}	DDPG epochs	800
m_{DDPG}	Steps by epochs in training	1200
B_{DDPG}	Buffer memory size	100000000
N_{DDPG}	Batch size	2048
γ	Discount factor	0.99
η_{ch}	Actor and critic learning rate	0.0003
ϵ_{DDPG}	Action noise range	0.1
	Optimizer	Adam

Table 3. Main hyperparameters of the PPO algorithm.

Symbol	Designation	Range
H	Number of hidden layers	128
s_{PPO}	PPO steps	480000
	Policy distance	Gaussian
N_{PPO}	Batch size	2048
η_{ch}	Actor and critic learning rate	0.0003
γ	Discount factor	0.99

of charge of each car SoC_t^i, $Tleave_t^i$, unlike SoC_t^i and $Tleave_t^i$ which represent the state of charge of the i-th position at the t-th step and the number of hours remaining until each car's departure, in case any spot is empty, both variables acquire a zero value. All these values are normalized between 0 and 1 before starting the agent training, in order to facilitate a fast and efficient learning optimization, reducing the computational cost of the algorithms.

$$s_t = (G_t, pr_t, Bc_t, G_{t+1}, G_{t+2}, G_{t+3}, Pr_{t+1}, Pr_{t+2}, Pr_{t+3}, \\ Bc_{t+1}, Bc_{t+2}, Bc_{t+3}, SoC_t^i, Tleave_t^i) \quad (3)$$

Action space: the action set A is defined by the charge/discharge rates of each post, with a total of 10 continuous variables constrained in the space [-1, 1], the action expresses the degree of energy transfer at a given post in a given time, where each value is positive during the charging mode of the corresponding EV battery, and negative during the discharging mode. If the agent interprets that a maximum charging mode is required then it will assign an action of 1, on the other hand, if it decides to discharge the battery as fast as possible, the action will be -1. Based on this action, the environment calculates, via (4), the net energy of the battery, which represents the energy required (how much remains to be charged when the action $a > 0$) or available (how charged it is when the action $a < 0$).

$$E^i_{n_t} = \begin{cases} (1 - SoC^i_t)C_{bat} \; if \; a^i_t \geq 0 \\ \\ (SoC^i_t)C_{bat} \qquad if \; a^i_t < 0 \end{cases} \tag{4}$$

At the same time, the net energy that can be transferred in the i-th vehicle in a time t is bounded by the characteristics of the charging spot and is represented as:

$$E^i_{n_t} \leq P_{ch_{max}} \eta_{ch} t \tag{5}$$

here $P_{ch_{max}}$ is the maximum charging power, η_{ch} is the charge/discharge efficiency and t is the time.

On the other hand, the energy demanded from an EV in a time t is given by the maximum energy transfer it can perform and the charging or discharging rates of the EV, in other words it depends on the energy of the battery and what the manager deems necessary to transfer.

$$E^i_{dem_t} = a^i_t E^i_{n_t} \tag{6}$$

As can be seen in (6), the action directly affects the calculation of the energy demand, and therefore, in an implicit way allows to control the moments to make purchases of electric energy from the grid.

Reward function: the reward function under state S_t and performing action A_t has a first term associated to the energy purchase, which penalizes the agent each time it gets energy from the grid, the same is formed by the amount of energy purchased from the grid multiplied by the price at each time step t; and a second term related to the charging process, where the Evs must reach the desired state of charge (SoC = 100 %) before departing to avoid being penalized. The summation shown in (7) over the price of purchased energy takes into account the i-th spot belonging to Ω, being this the set of all charging spots and the i-th EV belonging to Φ, which is the set of EVs that must depart at the current time step, i.e. only the charging state of those cars is taken into account.

$$r_t(S_t, A_t) = -((\sum_{i \in \Omega_t} (pr_t E^i_{dem_t}) + Bc_t - Gt_{ren_t}) + \sum_{i \in \Phi_t} [2.(1 - SoC^i_t)]^2) \tag{7}$$

where Bc_t and Gt_{ren_t} are the building consumption and renewable energy generation at this time step.

The fact that both terms are larger when the situation is less desirable implies that the agent must be trained to learn an optimal policy that minimizes the function. A very large negative value may imply that the demand for energy coming from the grid is very large and/or that cars left the parking area with a very low SoC value; and as a contraposition, a minimum reward value means that at a time step t there was no energy purchase and the departing EVs are fully charged (SoC = 100 %).

4.2 Reference method

To compare the effectiveness of the proposed methods, a reference algorithm known as "Rule-Based Controller" (RBC) [12], which is characterized by low computational requirements, good real-time performance and stability, based on the definition of a set of deterministic rules, was included in the experimental stage.

In this case, the method is composed of only two rules:

$$a_t^i = \begin{cases} 1, & \text{if } Tleave_t^i \leq 3 \\ 1/2(G_t + G_{t+1}), & \text{otherwise} \end{cases} \tag{8}$$

At each spot the departure time of each vehicle is reviewed. If the remaining time to departure time is less than 3 hours then the charging rate is maximized so that the EV can charge to maximum capacity and thus be able to reach the maximum SoC before the EV proceeds to departure. On the other hand, if the departure time is more than 3 hours away, the action is dependent solely on the current irradiance values and their prediction in the next hour. The hourly values defined for the RBC were selected from empirical tests.

4.3 Experimental results

This section shows the results obtained during the experimentation of the algorithms presented in Section 3 in comparison with the reference method proposed in Sect. 4.2.

To compare the effectiveness between the models, we first took into account the reward obtained by the RL agents during training performed under random conditions, since they are methods that require a learning process. Whereas, the comparison method (rule-based control) is exempt from this stage.

Fig. 2 shows the evolution of the average episodic reward of the RL methods throughout a training process, where although it can be noticed that clearly the PPO has a considerably more stable curve, with less uncertainty in the first half of the learning and stabilizes much earlier than the DDPG, the same one reaches a slightly suboptimal policy.

Once the agents converged to a given policy during learning, a comparison was made with the RBC method during the evaluation of a full day simulated with the same conditions for the three algorithms, calculating the average of the rewards obtained during this process. The values achieved are shown in Table 4, where an improvement of both algorithms belonging to deep RL with respect to RBC can be seen, representing a reduction of up to 17.49% in the average episodic penalty.

Since the comparison of algorithms via reward only gives us a general overview of which methods perform better than others based on the defined reward function, the electrical behaviors of the system under each smart manager were additionally evaluated. Particularly the energy flow (renewable, purchased and stored energy) and the vehicle charging process were examined in more detail.

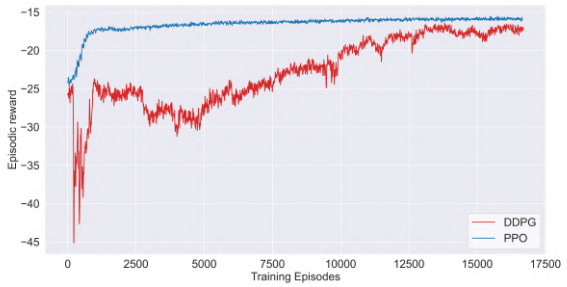

Fig. 2. Average episodic reward through the learning process.

Table 4. Average reward of the methods during evaluation.

Method	Average reward
DDPG	-16.08
PPO	-16.44
RBC	-19.49

Starting with the RBC algorithm, Fig. 3(a) shows the useful energy (energy from EVs, grid energy and renewable energy plotted in cyan, blue and green bars respectively), and the management behaviour of the system (orange curve is the energy consumption by smart building (SB), cyan curve is the stored energy in EVs and gray curve is the total consumption). In addition, the energy purchase price is shown as a red curve. As can be seen, the energy acquired by EVs is directly proportional to the renewable energy available due to the dependency on the irradiance, given the rules defined in (8). In other words, during sun exposure, the EVs are charged even if they are not close to retiring.

Figure 4 shows graphs corresponding to the 10 charging stations, where each one has a blue curve representing the SoC of the EVs parked in the respective station, and a green curve describing the presence of the vehicle. It is important to hightlight that there is a drastic increase in the slope of the SoC when the vehicle is close to leaving ($Tleave < 3$), corresponding with the first term of (8). In most cases, the EV is fully charged long before departure. However, if the EV exits before the panels start generating energy or if the time in which the EV remains charging is very short, the vehicle may not be fully charged.

As a consequence of RBC behaviour, the energy demand is too high, as can be seen in Table 5, where the agent required to purchase 529 kWh of energy, with a cost of $38.21, which represents an average purchase of 0.072 $/kWh. It can be deduced that the policy prioritizes vehicle charging over energy purchase management.

According to Fig. 3(b), unlike the RBC, the PPO agent decides to buy more energy when the price is lower. At the same time, the agent has no choice but to

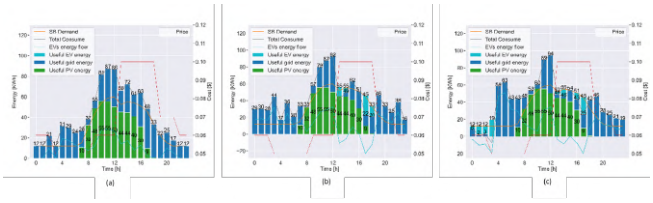

Fig. 3. (a) Percentage of energy consumption by RBC; (b) Percentage energy consumption by PPO; (c) Percentage energy consumption by DDPG.

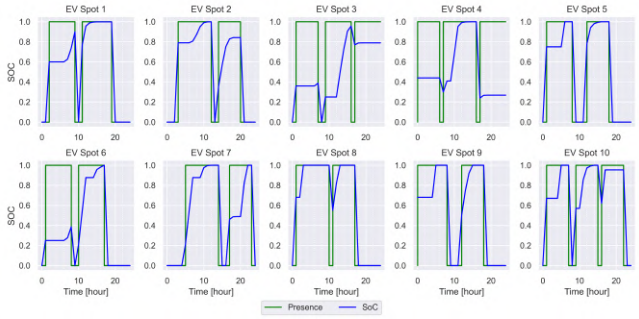

Fig. 4. EV charging process using RBC. Presence=1 means the existence of EV in the parking area and Presence=0 means the absence of EV in the parking area.

Table 5. Purchased energy quantity and its cost during evaluation.

Method	Purchased energy [kWh]	Total Cost [$]	Average cost [$]
RBC	529.86	38.21	0.072
PPO	530.63	33.97	0.064
DDPG	505.94	33.15	0.065

buy energy given the overlap of the decrease of available renewable energy with the massive departure of many vehicles. Despite this, through V2G functionality, the agent produces a marked difference with RBC regarding the charging stability of the EVs, even generating that at times the 10 charging spots are delivering more energy than they consume, that energy added to the renewable energy still available allows to supply almost completely the consumption of the building. This action is clearly noticeable in the hours where the energy flow curve of the EVs has negative values, at those times the amount of energy delivered by the vehicles is the same amount absorbed by the building.

The agent takes greater preponderance in managing energy purchases from the grid, hence some vehicles may retire without reaching the maximum SoC. This behavior is observed in Fig. 5 and Table 5, where the amount of energy purchased from the grid is the highest obtained in the tests, but with a cost of

$33.97, which is a decrease of 11% with respect to the RBC. Consequently, the average purchase price is 0.064 $/kWh.

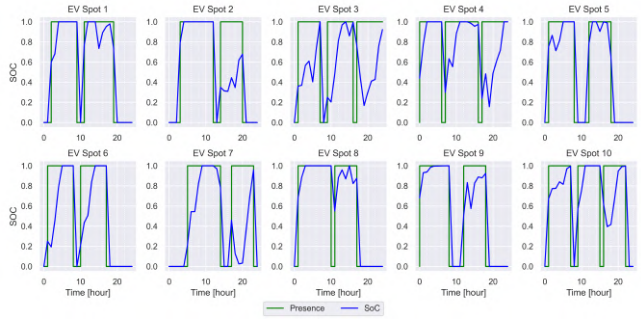

Fig. 5. EV charging process using PPO. Presence=1 means the existence of EV in the parking area and Presence=0 means the absence of EV in the parking area.

Finally, the DDPG algorithm also takes advantage of hours with lower prices to purchase energy and, during higher cost hours, grid energy use is reduced, reaching zero on several occasions (see Fig. 3(c)). However, higher demand is observed during hours with low renewable energy and a large number of vehicles about to leave the parking area. It can be seen that the energy stored in the EVs is very low during high price hours, not exceeding 20 kWh. Thus, the most part of the energy purchased by the agent during these hours is intended to supply the building consumption, based on the prediction of the agent that the renewable energy is not able to supply.

Meanwhile, similar stability to that shown in PPO can be observed in Fig. 6, which despite not reaching a SoC of 100 % in all vehicles, the behavior shows that the agent has learned to regulate the charging speed of each stall depending on its departure time without neglecting the efficiency of energy purchase. Therefore, like the PPO algorithm, the DDPG has a tendency to buy cheaper energy above the EV load. In the table 5, it is observed that the DDPG agent purchased 505 kWh of energy, 4.5% less than RBC, with a total cost of $33.15, a decrease of 13%. This is reflected in the average cost of energy purchased of $0.065/kWh.

A detail to take into account is that both the DDPG and the RBC have EV energy storage peaks in hours 4 and 5, while the PPO algorithm does not seem to suffer this increase in demand, added to the fact that it has the graph with the lowest variation in EV energy consumption, it could mean that it is able to manage more efficiently the energy flow throughout the episode.

5 Conclusion

The integration of EV charging stations within V2B frameworks signifies not just technological evolution but also a crucial advancement toward the creation

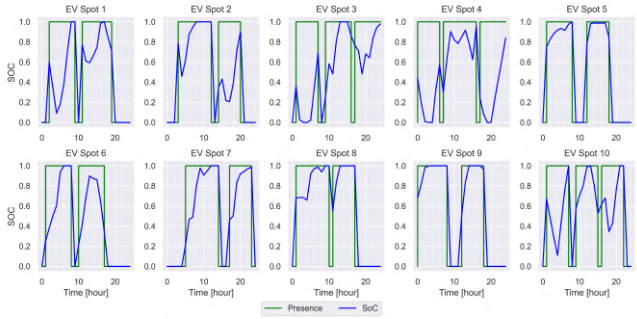

Fig. 6. EV charging process using DDPG. Presence=1 means the existence of EV in the parking area and Presence=0 means the absence of EV in the parking area.

of smarter and more eco-friendly cities. This effort actively contributes to the optimization of system resources and, meanwhile, facilitates the integration of renewable energy sources. In this study, we implemented RL based control strategies for the management of smart parking lots. The main goal was to minimize building energy purchases from the grid while ensuring the efficient charging of EVs.

The RL algorithms demonstrated substantial improvement compared to the reference controller, achieving improvements in evaluation reward from 17 % to 19 %. Savings in energy metrics were also achieved, with energy cost reductions of over 9 % and up to 11 %. In essence, these algorithms showed more efficient decision making compared to other energy allocation methods, ensuring efficient EV charging.

Future work involves incorporating dynamic prices mechanisms that enable the design of strategies aiming to balance supply and demand in the building. This includes the rapid integration of renewable energy resources, mitigating demand peaks, and preventing power blackouts. Given that communication is crucial for optimizing information flows, fundamental concepts of established and emerging descentralized protocols based on blockchain for communication in smart grids will be evaluated. Transactive energy management may enable faster bidirectional energy transfer and optimize demand-side resources through the use of decentralized intelligent equipment.

References

1. M. Zuccalà, E. S. Verga, Enabling energy smart cities through urban sharing ecosystems, Energy Procedia 111 (2017) 826–835.
2. B. Vaidya, H. T. Mouftah, Smart electric vehicle charging management for smart cities, IET Smart Cities 2 (1) (2020) 4–13.
3. M. Inci, M. M. Savrun, Ö. Çelik, Integrating electric vehicles as virtual power plants: A comprehensive review on vehicle-to-grid (v2g) concepts, interface topologies, marketing and future prospects, Journal of Energy Storage 55 (2022) 105579.

4. M. Trimboli, L. Avila, Optimal battery charge with safe exploration, Expert Systems with Applications 237 (2024) 121697.
5. Z. Moghaddam, I. Ahmad, D. Habibi, Q. V. Phung, Smart charging strategy for electric vehicle charging stations, IEEE Transactions on transportation electrification 4 (1) (2017) 76–88.
6. M. Bose, B. R. Dutta, N. Shrivastava, S. R. Sarangi, Pc-ilp: A fast and intuitive method to place electric vehicle charging stations in smart cities, Smart Cities 6 (6) (2023) 3060–3092.
7. A. Ouammi, Peak load reduction with a solar pv-based smart microgrid and vehicle-to-building (v2b) concept, Sustainable Energy Technologies and Assessments 44 (2021) 101027.
8. Z. He, J. Khazaei, J. D. Freihaut, Optimal integration of vehicle to building (v2b) and building to vehicle (b2v) technologies for commercial buildings, Sustainable Energy, Grids and Networks 32 (2022) 100921.
9. S. M. Ahsan, H. A. Khan, S. Sohaib, A. M. Hashmi, Optimized power dispatch for smart building and electric vehicles with v2v, v2b and v2g operations, Energies 16 (13) (2023) 4884.
10. M. Ghafoori, M. Abdallah, S. Kim, Electricity peak shaving for commercial buildings using machine learning and vehicle to building (v2b) system, Applied Energy 340 (2023) 121052.
11. T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, arXiv preprint arXiv:1509.02971 (2015).
12. D. B. Leake, Case-Based Reasoning: Experiences, Lessons and Future Directions, 1st Edition, MIT Press, Cambridge, MA, USA, 1996.