Chargym: An EV Charging Station Model for Controller Benchmarking

Georgios Karatzinis$^{1[0000-0003-4674-4163]}$, Christos Korkas$^{1,2[0000-0002-2491-2659]}$, Michalis Terzopoulos$^{1[]}$, Christos Tsaknakis$^{1,2[]}$, Aliki Stefanopoulou$^{1[0000-0002-0026-6779]}$, Iakovos Michailidis$^{1,2[0000-0001-7295-8806]}$, and Elias Kosmatopoulos$^{1,2[0000-0002-3735-4238]}$

[1] Democritus University of Thrace, Xanthi 67100, Greece
[2] Center for Research and Technology, Thessaloniki 57001, Greece

Abstract. This paper presents Chargym, a Python-based openai-gym compatible environment, that simulates the charging dynamics of a grid connected Electrical Vehicle (EV) charging station. Chargym transforms the classic EV charging problem into a Reinforcement Learning setup that can be used for benchmarking of various and off-the-shelf control and optimization algorithms enabling both single and multiple agent formulations. The incorporated charging station dynamics are presented with a brief explanation of the system parameters and function of the technical equipment. Moreover, we describe the structure of the used framework, highlighting the key features and data models that provide the necessary inputs for optimal control decisions. Finally, an experimental performance analysis is provided using two different state-of-the-art Reinforcement Learning (RL) algorithms validating the operation of the provided environment.

Keywords: Electric Vehicles · Charging Optimization · Deep Reinforcement Learning · Benchmarking

1 Introduction

Current and upcoming introduction of plug-in hybrid electric vehicles (PHEVs) and fully electric vehicles (EVs) in markets, will introduce large amounts of electrical loads and storage capacity into the electric grid requiring many efforts for the sufficient integration and management. Moreover, the EV charging and discharging will introduce different load profiles not only in terms of quantity, but also in terms of timely change. In US and North America in general, the EV deployment and introduction of EVs is more mature compared to other countries and Europe. The deployment of EV vehicles results to large electrical loads that reach the 18% of the total energy consumption [1]. Moreover, the introduction of EV's also adds uncertainty in the grid since, with the Vehicle-to-Grid (V2G) functionality [2], they can also provide energy to the power grid by discharging the battery. It is clear that developing appropriate algorithms to control and optimize the charging/discharging process is crucial in order to facilitate the smooth integration of EV units in the current electrical grid. Numerous works can be found in literature, focusing on optimal EV charging. Initial approaches for controlling the charging systems were formed by open loop strategies [6, 11], but the simple rule-based control approach that they follow, combined with possible computational complexity that they introduce, can lead to sub-optimal control behavior, since it is known that open-loop control strategies result to non-robust solutions [12], e.g. they require to re calibrate their decisions for different initial and exogenous conditions resulting to extensive and numerous simulations. Therefore, introducing close-loop efficient charging control provides the tools to reduce energy consumption, environmental impact and maximize the user satisfaction. Recent research approaches cover the fields of (robust) model predictive control (MPC) (see e.g., ([9, 10]), adaptive or learning-based approaches (see e.g., [3, 4]), and reinforcement learning (RL); see e.g., [5, 7]. A thorough review on electric vehicle technologies, charging methods, standards and optimization techniques can be found on [8].

1.1 Related Work

Alongside the recent developments in state-of-the-art RL and control algorithms for optimal charging and scheduling of EVs and other smart grid nodes, large efforts have been paid for the implementation of open-source tools and simulators offering robust and ease to use platforms supporting new research. Many of these platforms are focused on smart-grids, buildings and microgrid operation, however the research community has also developed simulators focused on EV operation and charging. In [16] the simulation aspects of energy consumption, available charging stations and charging duration are considered. The problem of the shortest path and travel planning is studied in [17], where the authors designed an approximation scheme to calculate the most energy-efficient path. In [18, 21] traffic simulations are presented utilizing EVs and investigating the optimal online charging based on highway available public chargers. Two of the most used simulators are V2GSim [14] and EVLibSim [13]. However both V2G-Sim and EVLibSim allow for precomputed charging schedules or simple control strategies. Another recently developed simulator, called ACN-Sim [15] is designed explicitly around evaluating online algorithms which adapt to changes in the system state over time. However, ACN-Sim is more oriented towards, energy system aspects of the problem rather than to the cost/penalty evaluation.

1.2 Contributions

The main contribution of this work is to provide a framework that simulates the operation of a grid-connected EV charging station. Our main goal is to provide a generalised environment for charging/discharging EVs under various disturbances (weather conditions, pricing models, stochastic arrival-departure EV times and stochastic Battery State of Charge (BOC) at arrival). Thereby, by training multiple times in such generated environments, the controller will grasp/understand the underlining charging dynamics and leverage it to efficiently complete its goal, even in days/instances that it has never been trained. Moreover, Chargym offers control over multiple charging set-points (one per connected vehicle) enabling multi-agent formulations as well. Finally, both charging and discharging of the vehicles is offered, in order to achieve the optimal results. Within this work, a novel simulation environment for EV charging has been developed, based on the openai-gym format. All the core parameters and dynamics that describe and simulate a real EV charging setup have been included. Chargym (available at `https://github.com/georkara/Chargym-Charging-Station` (accessed on 9/3/2022), is a Python-based library, made for standardized comparison and evaluation of controller performances, based on predefined evaluation scenarios, and is inspired by the RL benchmarking library Gym. It should be emphasized that Chargym is one of the first simulators that provide highly realistic simulation of an EV charging station and it is also a framework upon where state of the art RL methods (and also other learning approaches) will be efficiently benchmarked in EV charging tasks aiming to optimize not only cost related rewards, but also penalties related to failing charge the EVs to the desired Battery State-of-Charge. Two state-of-the-art RL algorithms, namely DDPG and PPO, have been evaluated on the Chargym environment. To better comprehend these evaluation results, the average human-level performance (Rule-Based Controller) in the Chargym environment is also reported. However, the follow-up analysis utilizing the best-performing algorithm is not conducted with respect to the different levels of solar availability and pricing models and levels of performance, rather presents the easy integration of algorithms utilizing the station as a simulation environment.

The rest of this paper is organized as follow: Section 2 presents the physical description, characteristics and main assumptions of the charging station. Section 3 presents the details of the openai-gym/RL charging environment and formulation. Section 4 presents the main results of applying two state-of-the-art RL algorithms, evaluated against a Rule-Based-Controller, showcasing the integration and interoperability of Chargym with common RL libraries. Finally, Section 5 summarizes the main innovations of this work and the main future steps.

2 Charging Station Overview

Chargym simulates the operation of an electric vehicle charging station (EVCS) considering random EV arrivals and departures within a day. The main objective is to minimize the cost for the electricity absorbed by the power grid, while ensuring that all EVs reach their desired level of State of Charge (SoC) at departure. If an EV departs without reaching the desired SoC, a penalty cost is calculated and applied. The EVCS architecture is presented in Fig. 1. More specifically, the core components of the Chargym environment are given: i) a set of 10 charging spots; ii) one photovoltaic (PV) generation system; iii) power grid connection offering energy at a certain price and iv) the vehicle to grid operation (V2G),

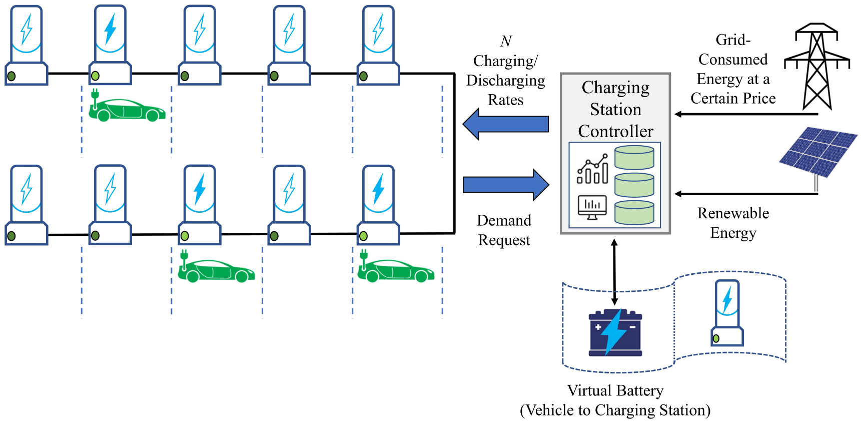


Fig. 1: Chargym interaction architecture

which adds a Vehicle to Charging Station functionality, allowing the usage of energy stored in EVs for charging other EVs, when necessary.

Regarding the operational framework, the station is connected with the grid absorbing electricity at a fluctuating price, when the available amount of energy is inadequate. The station's available amount of energy (apart from the grid) is unfolded into two types:

- Stored energy in the cars that can be utilized under the V2G operation.
- Produced energy from the PV.

Note that the term *stored energy* refers to storage that is formed from the available energy storage of EVs in a Vehicle to Charging Station perspective. Therefore, the environment describes a case where the stored energy in EVs, can be utilized from the station (based on the control setpoints) to satisfy the demands of other EVs that have limited time until their departure time. In Table 1, the basic parameters related with EVCS and EVs are presented.

The following conditions describe the overall perception of the EVCS, providing clear insights concerning the implementation of the environment's operating framework.

Assumption 1. All EVs that arrive to the station are assumed to share the same characteristics related with their battery (type, capacity, charging/ discharging rate, charging/ discharging efficiency, battery efficiency).

Assumption 2: The desired State of Charge for every EV at departure time is 100%.

Assumption 3: If an EV departs with less than 100% State of Charge, a penalty score is calculated.

Table 1: Charging station parameters.

Station Parameters	Value
Timestep length (dk) (h)	1
EV Battery Capacity(B_{max}) (kWh)	30
Charging and Discharging Eff. (η_{ch}) (%)	91
Maximum Charging Output ($P_{ch,max}$) (kW)	11

Stochastic parameter	Minimum	Maximum
Arrival State of Charge (%)	10	80
Arrival Time (hour)	0	22
Departure Time (hour)	Arrival+2	Next Day

Assumption 4: There is no upper limit of supply from the power grid. This way, the grid can supply the Charging Station with any amount of requested energy.

Assumption 5: The maximum charging/discharging supply of each EV is dictated by charging/discharging rate of the station.

Assumption 6: Each charging spot, can be used more than once per day.

3 Framework Description

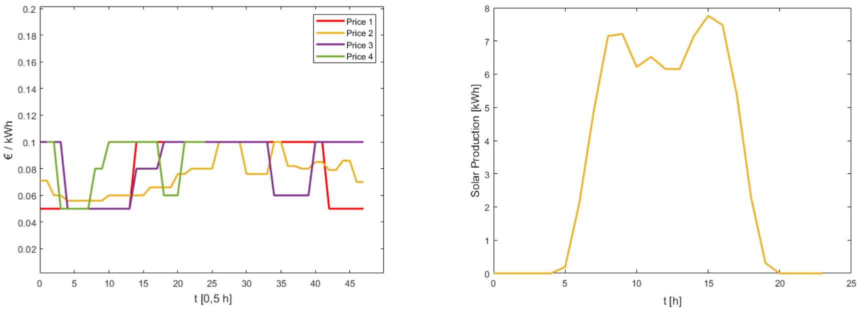
The real-time EVCS scheduling problem can be formulated, in a Reinforcement Learning context, as a Markov Decision Process (MDP) with a 4-tuple $(\mathbf{S}, \mathbf{A}, \mathbf{P}, \mathbf{R})$, where $\mathbf{S}$ is the set of states, $\forall s \in S$; $\mathbf{A}$ is a finite set of actions, $\forall a \in A$; $\mathbf{P}$ is the state transition probability with $P : S \times A \times S \to [0, 1]$ being the transition function with the probability of the transition from state s by choosing action a to state s' at time $t+1$, such that $p_a(s, s') = p(s_{t+1} = s' | s_t = s, a_t = a)$; $R : S \times A \times S \to \mathbb{R}$ is the reward function, where $R_a(s, s')$ is the reward received by the agent after transition from state s to state s' occurs.

3.1 State

The EVCS state at each time step t is defined as:

$$\begin{aligned} s_t = \big(&G_t, pr_t, G_{t+1}, G_{t+2}, G_{t+3}, pr_{t+1}, pr_{t+2}, pr_{t+3}, \quad \cdots \\ \cdots \quad &SoC_t^1, SoC_t^2, ..., SoC_t^{10}, Tleave_t^1, Tleave_t^2 ..., Tleave_t^{10}\big) \end{aligned} \tag{1}$$

This vector contains four types of information: (i) G_t is current value of solar radiation and $(G_{t+1}, G_{t+2}, G_{t+3})$ implies solar radiation ahead predictions for the next three hours; (ii) pr_t is current value of electricity price that the utility company charges the station for a requested amount of energy and $(pr_{t+1}, pr_{t+2}, pr_{t+3})$ are the three hour price predictions ahead. Although price changes dynamically throughout the day simulating tariff, it is not a function of supply and demand. Thus, price is dynamic but independent of the requested



(a) (b)

Fig. 2: Price and solar production simulation profiles. (a) Different profiles of dynamic pricing, used in test cases; (b) Solar production on a typical day.

amount; (iii) SoC_t^i denotes the SOC of the EV at i_{th} charging spot at timestep t and (iv) $Tleave_t^i$ indicates the number of hours until departure for the EV at i_{th} charging spot. The latter states, SoC_t^i and $Tleave_t^i$, can be considered as the physical states of the EVCS. In Figure 2(a), the four pricing profiles that are available for simulation are presented, whereas in Figure 2(b) a typical day of electricity production is shown.

Regarding the states two main points should be highlighted:

- If charging spot i is empty at t, then SOC_t^i and $Tleave_t^i$ are 0.

- All the states presented in Equation 1, are normalized between 0 and 1.

3.2 Action

As described above, the charging station is composed by 10 charging spots, all able to charge or discharge (V2G capability) the connected EVs. Therefore, there are 10 actions defining the charging or discharging rate of each vehicle spot. These 10 action set-points ($action^i$) are defined as continuous variables, which are constrained in the $[-1, 1]$ space. The charging/discharging power for each vehicle i at timestep t is defined as :

$$MaxEnergy_t^i = \begin{cases} (1-\mathrm{SOC}_t^i)*B_{max} & action_t^i >= 0 \\ (\mathrm{SOC}_t^i)*B_{max} & action_t^i < 0 \end{cases} \tag{2}$$

$$MaxEnergy_t^i <= P_{ch,max} * \eta_{ch} \tag{3}$$

$$P_t^{dem,i} = action_t^i * MaxEnergy_t^i \tag{4}$$

where SoC_t^i is the state of charge of each EV (i) at time t and B_{max}, $P_{ch,max}$, η_{ch} are given in Table 1.

Thus, the three equations above, describe the calculation of the demand for each charging spot in timestep t based on the actions that are taken by the controller. If action is a positive number, the $P_t^{dem,i}$ is positive (charging mode), whereas if action is negative $P_t^{dem,i}$ is negative (discharging mode). The value of $action^i$ affects directly the demand as shown in Equation 4. Equations 3 and 2, describe the constraints on the maximum charging/discharging energy that can be allocated in one timestep.

3.3 Reward

The main objective of the EVCS's controller / agent is to adopt a scheduling policy towards minimizing the cost for the electricity absorbed by the power grid. The reward function observed at each timestep t is the electricity bill being payed by EVCS to the utility company. However, an additional term is incorporated in order to present a more realistic and complete description ensuring that the controller will exploit effectively the available resources as well as fulfil the defined requirements. The second term considers penalizing situations involving EVs that are not completely charged. The equation describing this specific formulation is the following:

$$r_t(S_t, A_t) = \sum_{i \in \Omega_t} (pr_t \cdot P_t^{dem}) + \sum_{i \in \Psi_t} [2 \cdot (1 - SoC_t^i)]^2 \tag{5}$$

where $P_t^{dem} \in \Omega_t$ stands for the total charging demand that the EVCS requests to receive from the utility company as mentioned above. The electricity price, pr_t, follows a varying bill profile that the utility company provides / charges the EVCS at each timestep in €/KWh and presented in Figure 2(a). The second term is related with the state of charge of those EVs that are expected to departure the next hour. The goal is to fully charge the EVs that depart. However, in future realizations, the EV owner could choose the desired SOC at departure (to reduce the charging cost), making the formulation even more realistic.

4 Performance Evaluation

This section presents an experimental evaluation of the Chargym environment. The analysis begins with all the implementation details that are important for realizing the Chargym experimental setup. We employ two state-of-the-art Deep Reinforcement Learning algorithms, namely Deep Deterministic Policy Gradient (DDPG) [19] and Proximal Policy Optimization (PPO) [20] , in order to evaluate their performances using the Chargym environment. Moreover, a simple rule based controller (RBC) is presented to perform as a baseline model providing a reasonable operational strategy, simulating human-operated decisions. However, as stated in introduction, the goal of this section is to evaluate the operation and the ease-of-use of Chargym environment, and not to perform a deep analysis on the performance of state-of-the-art algorithms.

4.1 Implementation details and key experimental attributes

Stable Baselines 3 framework [22] was utilized to perform all the experiments. The fact that Stable Baselines 3 is a well-documented, highly-robust library also eases the build-on developments (e.g., apply a different RL pipeline), as it follows a common framework. Furthermore, such an experimental setup may also leverage the interoperability with other powerful frameworks and showcases the ease-to-use nature of the Chargym environment.

On the other hand Chargym was designed to bridge the gap between realistic charging/discharging EV operations and powerful state-of-art control and RL solutions. The experiments/simulations that were conducted on Chargym, as well as, the whole formulation of the EVCS design were based on the following key attributes:

- **Schedule Diversity**: For each episode, the general dynamics and EV schedules are determined by a specific automated and random process. These levels correspond to the randomness in the number, arrival and departure rates, and schedules of the EVs, the initial SoC of each EV upon arrival and of course the different solar and pricing conditions. This approach forces the control algorithms to be trained and tested in multiple/diverse layouts, producing robust solutions which are of paramount importance in real-life applications where unknown schedules appear.
- **Partial Observability**: At each timestep, the EVCS is only aware of the attributes of the connected EVs, and the forecasts for the next three hours for the solar and pricing tariffs. The station can not utilize information of the EVs that are going to arrive in the future, therefore, any long-term plan should be agile enough to be adjusted on the fly, based on future information about the newly arrived vehicles.
- **Real Life Applicability**: One of the fundamental advantages of Chargym is that any learned policy can be straightforwardly applied to an appropriate EVCS case, since the whole problem formulation is using common and relevant EV knowledge and assumptions. Our goal is to create policies that calculate the optimal charging/discharging schedule based on the generic perception of the environment. Thus, assuming that a smooth integration with the sensor's readings, can be used to represent the environment as in 1, and no elaborate simulation model of the dynamics is required to adjust the RL algorithm into the specifics of the station.

4.2 Rule-Based Controller

The main comparison for the used RL algorithms, will be conducted with respect to a Rule-Based Controller (RBC), that makes human-based decisions regarding the charging-discharging of the EVs. The RBC offers simple and fast decisions, however far from optimal, decisions. The RBC consists of two simple rules that are presented in Equation 6:

$$action_t^i = \begin{cases} 1 & Tleave_t^i <= 3 \\ \frac{(G_t + G_{t+1})}{2} & Tleave_t^i > 3 \end{cases} \tag{6}$$

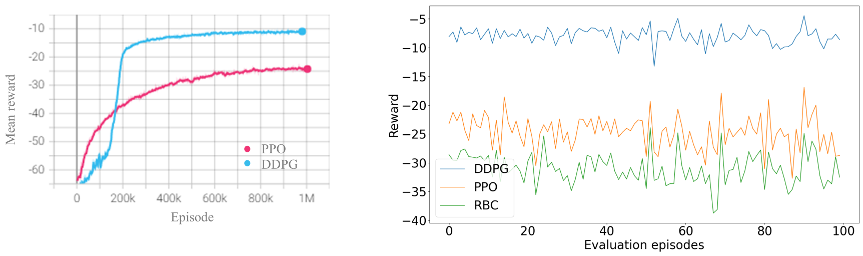
The controller checks each charging spot and collects the Departure timeplan of each connected EV. If an EV is going to depart during the next three hours, then the station is charging in full capacity this specific EV. On the other hand, if an EV does not depart during the next three hours, the station checks the current availability of the solar energy and charges the EV, based on that availability. The three hour time-limit, is selected based on the EVCS attributes, since the EVs utilize 30kWh batteries, and the maximum charging ability of the station is 10kW. Thus, an EV needs three hours to charge from 0 to 100 % SoC.

4.3 State-of-the-Art RL Algorithms Comparison

Regarding the training process, each episode concerns a different simulated day in terms of solar energy production, pricing profile and EV schedules and demands. Also the RL implementations, DDPG and PPO, are trained utilizing diverse training sets. The hyperparameter configurations of both RL algorithms are shown in Table 2. A larger hidden layer structure has been chosen for the case of DDPG. Figure 3(a) presents the transition learning performance from day to day between the adopted algorithms. As it can be noted, DDPG converges a little bit earlier achieving higher levels of reward in respect with PPO. In order to test the performance of the RL algorithms, we recall the trained models, indicatively after 940K episodes, and evaluate them in a new set of 100 simulated days with diverse configurations. The evaluation comparison between the RL algorithms and the RBC involves the same day configuration with identical operating conditions among algorithms for each episode, but different configurations between episodes (days). Both RL algorithms outperform the RBC, as depicted in Figure 3(b) and Table 3, while also the DDPG implementation consistently attains higher reward versus PPO.

Table 2: DDPG and PPO hyperparameter configuration

DDPG		*PPO*	
Parameters	Value	Parameters	Value
Memory buffer size	1e5	Learning rate	3e-4
Batch size	100	Batch size	64
Optimizers	Adam	Optimizers	Adam
Layer structure	[400,300]	Layer structure	[64,64]
Actor, Critic learning rate	0.001	GAE lambda, Clipping ratio	0.95, 0.2
Discount factor (γ)	0.99	Discount factor (γ)	0.99
Target update factor (τ)	0.005	Entropy and value function coefficients	0, 0.5
Ornstein Uhlenbeck (μ, σ, θ)	$0_{1\times10}, 0.5, 0.15$	Max grad norm	0.5



(a) (b)

Fig. 3: Training and evaluation reward comparison. (a) Rolling mean of the average episodic reward of PPO and DDPG; (b) Evaluation reward over 100 days between RL algorithms and RBC.

Table 3: Mean evaluation reward

Approach	Mean Reward
RBC	-30.99
PPO	-24.64
DDPG	-7.93

5 Conclusions and Future Work

This paper presents a new simulation enviroment, called Chargym, based on openai-gym format that bridges the gap between reinforcement learning and the real-life charging/discharging strategies in the EV domain. The environment simulates the problem of EV charging/discharging under multiple stochastic parameters (weather, electricity pricing, EV arrival, departure, SOC) into a reinforcement learning setup that can be tackled by a wide range model-free and model based RL and optimal control algorithms. An experimental evaluation was also conducted and presented, with 2 state-of-the-art RL algorithms, namely DDPG and PPO evaluated in Chargym, and their training results were also compared with a Rule Based Controller's performance for the task at hand. Future work, will aim on extensive comparison tests between state of the art RL and control algorithms (such as MPC) for the optimal scheduling of charging/discharging set-points, and special effort will be given in multi-agent frameworks enabling distributed and coordinated control of the charging setpoints/slots.

Acknowledgements We acknowledge support of this work by the project "Study, design, development and implementation of a holistic system for upgrading the quality of life and activity of the elderly (ASPiDA)" (MIS 5047294) which is implemented under the Action "Support for Regional Excellence", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund) and by the European Commission H2020-EU.2.1.5.2., Turning traditional reactive buildings into proactive ones, under contract 958284 (PRECEPT).

References

1. Ma, Zhongjing, Duncan S. Callaway, and Ian A. Hiskens. "Decentralized charging control of large populations of plug-in electric vehicles." IEEE Transactions on control systems technology 21.1 (2011): 67-78
2. Han, Sekyung, Soo Hee Han, and Kaoru Sezaki. "Design of an optimal aggregator for vehicle-to-grid regulation service." 2010 Innovative Smart Grid Technologies (ISGT). IEEE, 2010.
3. Korkas, C. D., Baldi, S., Michailidis, P. and Kosmatopoulos, E. B. (2017, July). A cognitive stochastic approximation approach to optimal charging schedule in electric vehicle stations. In 2017 25th Mediterranean Conference on Control and Automation (MED) (pp. 484-489). IEEE.
4. Korkas, C. D., Baldi, S., Yuan, S. and Kosmatopoulos, E. B. (2017). An adaptive learning-based approach for nearly optimal dynamic charging of electric vehicle fleets. IEEE Transactions on Intelligent Transportation Systems, 19(7), 2066-2075.
5. Qian, T., Shao, C., Wang, X. and Shahidehpour, M. (2019). Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system. IEEE transactions on smart grid, 11(2), 1714-1723.
6. Bhatti, A. R., Salam, Z., Sultana, B., Rasheed, N., Awan, A. B., Sultana, U. and Younas, M. (2019). Optimized sizing of photovoltaic grid-connected electric vehicle charging system using particle swarm optimization. International Journal of Energy Research, 43(1), 500-522.
7. Wan, Z., Li, H., He, H. and Prokhorov, D. (2018). Model-free real-time EV charging scheduling based on deep reinforcement learning. IEEE Transactions on Smart Grid, 10(5), 5246-5257.
8. Arif, S. M., Lie, T. T., Seet, B. C., Ayyadi, S. and Jensen, K. (2021). Review of electric vehicle technologies, charging methods, standards and optimization techniques. Electronics, 10(16), 1910.
9. Zheng, Y., Song, Y., Hill, D. J. and Meng, K. (2018). Online distributed MPC-based optimal scheduling for EV charging stations in distribution systems. IEEE Transactions on Industrial Informatics, 15(2), 638-649.
10. Tang, W. and Zhang, Y. J. (2016). A model predictive control approach for low-complexity electric vehicle charging scheduling: Optimality and scalability. IEEE transactions on power systems, 32(2), 1050-1063.
11. Zhang, M. and Chen, J. (2014). The energy management and optimized operation of electric vehicles based on microgrid. IEEE Transactions on Power Delivery, 29(3), 1427-1435.
12. Bardi, M. and Dolcetta, I. C. (1997). Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations (Vol. 12). Boston: Birkhäuser.

13. Rigas, E. S., Karapostolakis, S., Bassiliades, N., and Ramchurn, S. D. (2018). EVLibSim: A tool for the simulation of electric vehicles' charging stations using the EVLib library. Simulation Modelling Practice and Theory, 87, 99-119.
14. Saxena, S. (2013). Vehicle-to-grid Simulator (No. V2G-Sim; 005701MLTPL00). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).
15. Lee, Zachary J., Daniel Johansson, and Steven H. Low. "ACN-sim: An open-source simulator for data-driven electric vehicle charging research." 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). IEEE, 2019.
16. Díaz de Arcaya, A., et al. "Simulation Platform for Coordinated Charging of Electric Vehicles." (2015).
17. Strehler, Martin, Sören Merting, and Christian Schwan. "Energy-efficient shortest routes for electric and hybrid vehicles." Transportation Research Part B: Methodological 103 (2017): 111-135.
18. Mou, Yuting, et al. "Decentralized optimal demand-side management for PHEV charging in a smart grid." IEEE Transactions on Smart Grid 6.2 (2014): 726-736.
19. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., and Wierstra, D. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.
20. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.
21. Bae, S., and Kwasinski, A. (2011). Spatial and temporal model of electric vehicle charging demand. IEEE Transactions on Smart Grid, 3(1), 394-403.
22. Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., and Dormann, N. (2021). Stable-Baselines3: Reliable Reinforcement Learning Implementations. Journal of Machine Learning Research.