Subramanya, Rakshith; Sierla, Seppo A.; Vyatkin, Valeriy

Exploiting Battery Storages With Reinforcement Learning: A Review for Energy Professionals

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2022.3176446

Published: 01/01/2022

Document Version
Publisher's PDF, also known as Version of record

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 2, 2022, accepted May 11, 2022, date of publication May 18, 2022, date of current version May 26, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3176446

Exploiting Battery Storages With Reinforcement Learning: A Review for Energy Professionals

RAKSHITH SUBRAMANYA[1], (Graduate Student Member, IEEE), SEPPO A. SIERLA[1], (Senior Member, IEEE), AND VALERIY VYATKIN[1,2], (Fellow, IEEE)

[1]Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland
[2]Department of Computer Science, Electrical and Space Engineering, Luleå Tekniska Universitet, 971 87 Luleå, Sweden

Corresponding author: Rakshith Subramanya (rakshith.subramanya@aalto.fi)

This work was supported in part by the Business Finland under Grant 7439/31/2018.

ABSTRACT The transition to renewable production and smart grids is driving a massive investment to battery storages, and reinforcement learning (RL) has recently emerged as a potentially disruptive technology for their control and optimization of battery storage systems. A surge of papers has appeared in the last two years applying reinforcement learning to the optimization of battery storages in buildings, energy communities, energy harvesting Internet of Things networks, renewable generation, microgrids, electric vehicles and plug-in hybrid electric vehicles. This article reviews these applications through 4 different perspectives. Firstly, the type of optimization problem is analyzed; the literature can be divided to approaches that optimize either financial targets or energy efficiency. Secondly, the approaches for handling user comfort are analyzed for applications that may impact a human user. Thirdly, this paper discusses the approach to model and reduce battery degradation. Fourthly, the articles are categorized by application context and applications likely to attract a high amount of research are identified. The paper concludes with a list of unresolved challenges.

INDEX TERMS Battery degradation, battery storage, electric vehicle, microgrid, reinforcement learning.

I. INTRODUCTION

The transition to renewable production and smart grids is driving a massive investment to battery storages, evidenced by numerous very recent reviews on the subject (e.g. [1]–[5]). Reinforcement learning (RL) has recently emerged as a potentially disruptive technology for the control and optimization of battery storage systems. For example, Lee & Choi [6] formulate the optimization of a domestic energy storage system as a mixed-integer linear programming (MILP) problem as well as a RL problem, reporting significant performance improvements with the RL approach. In the energy domain, Perera and Kamalaruban [7] note a dip in recent publications on model predictive control, mirrored by a rapid increase in RL publications. Yang *et al.* [8] note that RL is well suited for complex problems with nonlinearity and uncertainty, which is often the case in next generation electric systems.

The associate editor coordinating the review of this manuscript and approving it for publication was Vitor Monteiro.

A few reviews on RL applications to intelligent energy systems include battery storages among the energy resources that have been studied. Perera and Kamalaruban [7] review RL applications across six major sectors: building energy management system (BEMS), dispatch, vehicle energy systems, energy devices, grid and energy markets. Battery storages appear as subcategories of some of these. Yang *et al.* [8] have an even broader scope including RL applications to smart grid, microgrids, integrated energy systems and energy internet. Glavic [9] reviews RL applications for controlling power grids, so batteries are discussed only for the purpose of grid support. Vázquez-Canteli & Nagy [10] review RL applications to demand response and discuss battery storages to the extent that they are featured in such applications. Wang and Hong [11] review RL applications for building controls and note works involving batteries, without analyzing further the control and optimization problems involving batteries. Frikha *et al.* [12] identify significant interest in RL in Internet of Things (IoT) applications and identify battery energy consumption and lifetime management as one of the key challenges. In all of the above-mentioned reviews,

significant background knowledge of RL theory is required from the reader. Different RL techniques are discussed and their application to specific problems is analyzed critically. Further, none of these reviews has a section dedicated to battery storages. Rather, batteries are discussed if they appear in the context of energy systems.

A surge of papers on RL applications for battery storages has appeared in the last two years, so at the time of writing, a critical mass of literature exists, meriting a dedicated review. This review is targeted at researchers and practitioners applying battery storages in different areas of green electrification, who wish to understand the disruptive potential of RL to their field. RL technology consists of algorithms that are interfaced to real or simulated systems in such a way, that the algorithms learn to achieve specified optimization targets as the interact with the system. Since the great majority of our target audience are not RL experts, the objective of this paper is to review this research in a way that is understandable to this audience. The reviewed works apply batteries to a range of innovative applications in buildings, energy communities, energy harvesting IoT networks, renewable generation, microgrids, electric vehicles (EV), plug-in hybrid electric vehicles (PHEV) as well as hybrid electric vehicles (HEV). Thus, our presentation is aimed beyond battery experts to the broader energy community working on such applications. Key RL concepts are introduced within a general framework of a RL agent managing a battery storage system, without assuming prior knowledge of RL or machine learning from the reader. The reviewed works are analyzed with reference to this framework.

This paper is structured as follows. Section 2 introduces a general conceptual overview for RL agents managing a battery storage system, without assuming prior knowledge of RL or machine learning from the reader. Several examples of systems including batteries are presented in the context of this framework. Section 3 presents the methodology of the literature review and an overview of the papers that were included into the review. The objectives, scope and approach of each paper is studied through four different aspects in sections 4-7. Each paper is discussed in each of these sections to the extent that the relevant aspects were explicitly discussed in the paper. Section 4 assesses the literature with respect to the main optimization objective of the RL application, with three categories emerging: optimization of energy efficiency, minimization of operational costs and minimization of investment costs. Section 5 discusses how user comfort has been handled or ignored in applications that may impact a human user. Section 6 discusses the various levels of abstraction used to model the battery and how battery degradation has been included into the optimization. Section 7 summarizes the review by discussing the literature according to specific applications areas, so that readers interested in a specific area such as electric vehicle charging will understand the focus of the research and open challenges in their field. Section 8 discusses our proposal for handling the main problem that was encountered in the literature review: there is a rapidly growing body of research, but it is difficult to identify the works with breakthrough performance, due to the great diversity in problem formulations and experimental setups. Benchmark RL environments have successfully addressed this problem in other fields, and the closest such works to the topic of this paper are identified. Section 9 concludes the paper with recommendations for overcoming the main unresolved challenges.

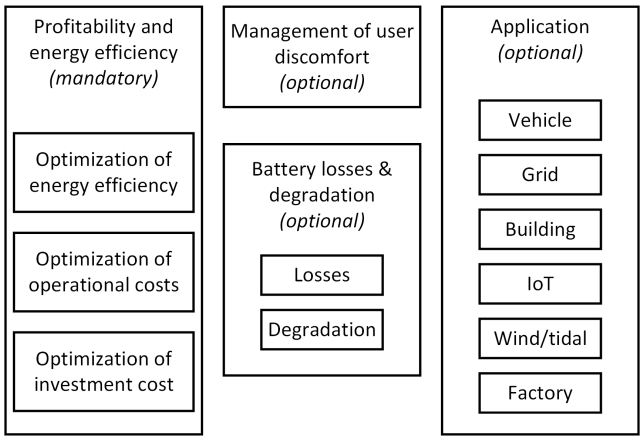

FIGURE 1. Taxonomy.

Figure 1 provides a graphical overview of the categorization in sections 4-7. Each of the four boxes corresponds to one of the four main categories analyzed in sections 4.7. Each of these categories has been indicated as being *mandatory* or *optional*. The optional category is not applicable to all of the papers selected for the review. The boxes within the main box are subcategories analyzed in their own subsection.

II. GENERAL CONCEPTUAL OVERVIEW FOR REINFORCEMENT LEARNING AGENTS MANAGING A BATTERY STORAGE

Major categories of machine learning methods include supervised, unsupervised and reinforcement learning. Supervised learning applications can be further categorized as regression and classification problems. Regression involves predicting a value based on several input datasets; for example, the price of an electricity market could be predicted based on weather and power system data. Classification involves choosing one out of several possible categories; for example, the categories could include a normal operating mode and several failure modes. In all cases, supervised learning methods require a training set, in which the correct output has been labelled for each input sample. If such labelled training data is not available, unsupervised learning can be applied to some problems. For example, if a time series dataset is available for a system running in a normal operating mode, an unsupervised learning algorithm can be trained to recognize a deviation from that normal operating mode, but it will not be able to classify the specific failure mode.

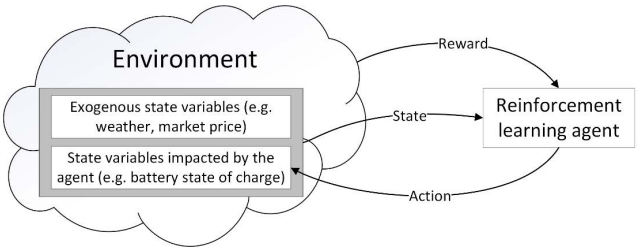

FIGURE 2. General framework of a RL agent managing a battery storage.

In the energy domain, supervised and unsupervised learning methods are used mainly for time series forecasting and condition monitoring, rather than decision making. RL differs from these methods in the sense that it learns to make better decisions by interacting with an environment and adjusting its actions according to feedback.

Figure 2 shows a general framework of a RL managing a battery storage. The figure introduces the key concepts that are used throughout this paper.

The environment consists of the battery storage and the system in which the storage is used. A few examples are presented in the following, to give the reader an idea of the great diversity of environments that researchers have developed to support their diverse RL problem formulations. If RL is used to minimize gasoline consumption of a PHEV, the gasoline tank and the engine should be modelled in the environment at a suitable level of abstraction [13]. The availability of V2G (vehicle-to-grid) needs to be considered when modelling the possibility to sell energy from the vehicle batteries to the grid [14], but details such as grid inverters may be abstracted away at the discretion of the authors [15]. For a wireless EV charging system, the EV characteristics and the traffic environment need to be considered [16]. For optimizing revenues of a wind farm with battery storage, the environment simulates the settlement scheme of the electricity market [17]. It is noted that the terms PHEV and HEV are used inconsistently in the literature. In this paper, all vehicles with an internal combustion engine and a battery are categorized as PHEVs. Further, EVs and HEVs are categorized so that the latter has another energy source such as a hydrogen fuel cell to complement the battery.

The RL agent takes actions, which impact the environment. The actions are specific to the application. Examples are bidding on various electricity markets [18], selecting between battery packs [19] or controlling the power of the engine in a PHEV [20].

The environment provides the RL agent with state information, which the agent considers when taking an action. The State of Charge (SoC) is a very commonly used state variable. Depending on the level of detail chosen by the authors, additional variables such as the temperature of batteries can be included [19]. Additional state variables depend on the specific application. For example, the energy management of PHEVs, EVs and HEVs is usually formulated in terms of a power demand state variable, which specifies the momentary power demand that must be jointly supplied by the on-board energy sources (e.g. [21]). As another example, relevant state information for the electricity management of a building's HVAC (Heating, Ventilation and Air Conditioning) includes indoor temperatures and occupancy [22].

The state may include additional exogenous variables that the RL agent cannot affect, but which are useful information for the RL agent as it determines the best action to take in the present state. For example, relevant market prices or weather data can be included if they are known at the time of taking the action, or if a forecast is available [23]. Otherwise, the price and weather can be treated as an unknown not included in the state information, but which can be taken into account in the reward [15]. Some studies use historical weather observation data instead of forecasts [24], so the system cannot be deployed as such to an online environment in which only uncertain weather forecasts are available.

The environment must implement a mapping to a next state given a current state and an action. The mapping can be constructed analytically with equations (e.g [15]). Another approach is to use an energy simulator and implement a wrapper around it to realize the state, action and reward interfaces [25]–[27]. In some cases, an energy simulator is not sufficient. For example, in self-driving vehicles that need to consider other vehicles and traffic lights, Wegener *et al.* [28] include a traffic simulator to the environment.

Finally, the RL agent requires feedback in the form of *rewards* to train its machine learning model, which determines the action based on the state information. The reward is generated by the environment and should penalize the agent for disadvantageous actions and reward it for advantageous actions. Depending on the objective of the paper, the reward is usually based on electricity costs [15], grid stability [29] or energy efficiency related criteria [23]. If the RL agent is allowed to impact the users comfort, for example by rescheduling appliances, adjusting indoor temperature or changing the charging behavior of EVs, a discomfort related penalty can be included to the reward [30]. A penalty for battery degradation can be included into the reward.

As training progresses, the RL agent learns to take actions that result in a high reward, but this may result in suboptimal solutions. To avoid this, the RL practitioner is able to force the training process to occasionally choose random actions, in a technique called *exploration*. Some authors may use their expertise of the specific battery energy management application to achieve more intelligent and computationally effective exploration. For example, in managing a HEV battery, Lian *et al.* [31] force the exploration to occur close to the Brake Specific Fuel Consumption curve, which is known to be the optimal region of operation for this kind of application. Zhou *et al.* [32] achieve a similar result with a heuristic algorithm developed to constrain the exploration.

The internals of the RL agent involve details understandable to machine learning practitioners. As the primary target audience of this article is energy practitioners, this review does not focus on these aspects. However, a brief overview

is provided as follows. The RL agent essentially implements a mapping from the state space to the action space. An early approach was Q-learning, in which the mapping was captured in a table, so the learning process involved updating the values in this table. More recently, due to increases in computational power and the resulting progress in deep neural networks, such networks are now commonly used to implement the mapping instead of the Q-table. The basic approach involves a single neural network that is used to make this mapping. However,, this approach does not always result in stable training performance. To overcome this, the concept of value was introduced: the value quantifies how good a particular state is, so this is a different concept than the reward. To exploit the value concept, the actor-critic network was introduced. The actor implements the mapping from the state space to the action space, and the critic computes the value, which is used in the training process of the actor. However, a weakness of actor-critic methods is that a small adjustment to the weights of the actor network may cause a jump to a region in which performance is poor, so the training may not converge, and thus fail to optimize the reward function. Several variants of the actor critic have been proposed to cope with issue. In particular, PPO (Proximal Policy Observation) limits the changes to the actor network parameters at each training step, improving the stability of the training process. Further innovations involving several neural networks have been developed, with Deep Deterministic Policy Gradient (DDPG) [73], [74] and Twin Delayed DDPG (TD3) being among the most commonly used. In general, these can be considered implementation details that are encapsulated in the ''Reinforcement learning agent'' box of Figure 2, so the choice of implementation method does not directly impact the formulation of state, action and reward. This is an encouraging observation, in the sense that battery domain experts could be more involved in the formulation in the future. However, there is a notable consideration in the choice of algorithm that will impact the formulation of state and action spaces. Some of the algorithms only support discrete state and action spaces, so if a state or action variable is of a continuous nature, the practitioner must define a limited number of discrete values for it.

Batteries are systems with complex chemical phenomena governing their charging, discharging and aging behavior. These phenomena are specific to the battery chemistry, and the development of such chemistries is an active area of research. However, as will be discussed in more detail in section VI, RL practitioners either explicitly or tacitly ignore these phenomena, or model them in a simplified way. For example, many authors assume that charging or discharging power can be expressed as the product of the battery capacity, SoC difference over a time period and a charging/discharging efficiency constant, so the battery chemistry is not considered or even mentioned (e.g. [116]). Such equations are implemented in the reinforcement learning environment, and they govern how the environment transitions from one state to the next upon receiving an action from the RL agent. Very few authors use more sophisticated models that are configured for a specific battery chemistry. For example, a lithium-ion battery model distinguishes between terminal and open circuit voltage and internal and trainset resistance [91]. In another example, the charge and discharge behavior of a lead acid battery is specified in the context of the system that the battery is used in, comprising a diesel generator and an inverter connected load [96]. The established way to capture aging in the reviewed articles was to add an aging penalty term to the reward function. The most common approaches are to penalize situations in which a minimum or maximum SoC threshold has been crossed (e.g. [130]), or to penalize deviations from a reference SoC (e.g. [70]). Unfortunately, the findings of such studies cannot be expressed in terms of equivalent full cycle, which is an established metric of battery lifetime.

TABLE 1. Search results.

	Science Direct	MDPI	IEEEXplore
Hits	3128	30	359
Manually selected papers	66	22	56

III. LITERATURE REVIEW METHODOLOGY

Various terms for battery storages are used in the literature, such as energy storage system, battery storage, *battery energy storage system, battery* and *storage.* To capture these and other variants, the following search string was used: ''reinforcement learning'' AND (storage OR battery)

The search string was applied to all fields. The search results are shown in Table 1. The hits were studied manually to select the relevant papers to be included in the review. The ''storage'' term resulted in many irrelevant articles on data storage or industrial warehouse type of storage; however, including this term in the search was important to find several relevant articles not using the word ''battery'' but rather ''energy storage''. The Elsevier (Science direct) and IEEE (IEEEXplore) search engines returned a large number of hits. These were sorted by relevance and studied in batches of 25. The search was stopped upon encountering a batch with no relevant papers. The search was limited to papers published since 2016. The search in IEEEXplore was limited to journal articles, including early access.

It is notable that this approach of using a simple search string resulted in a larger number of articles, many of which were not considered relevant. Thus, the approach relies heavily on manual work and judgement on the part of the authors. An extensive list of criteria for article selection was developed for this purpose and it is discussed in the next paragraph. Since this is a new, incipient field, the number of hits was manageable, and the final number of papers selected for inclusion in the review was considered suitable for a review paper. The search was restricted to the publishers Science Direct, IEEE and MDPI. When the

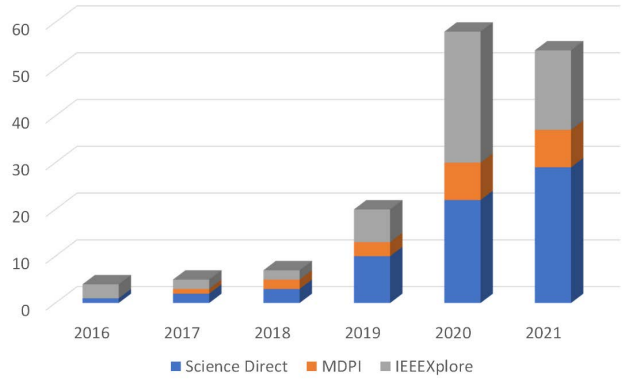

FIGURE 3. Papers that were selected manually for the literature review.

search was repeated in the Web of Science database and limited to journal articles, these publishers emerged as the top 3 publishers.

The following principles were used to guide the manual selection process:

- Several articles addressed non-battery energy storage systems such as fuel cells e.g. [33], ultracapacitors [34], natural gas storage tanks [35] and thermal storages such as hot water tanks [36], boilers [37], chilled water tanks [38], [39] and ice storage [40] or by exploiting the building structures themselves as a passive thermal energy storage [41]. Such works were not selected, unless these storages were used in addition to a battery.
- Papers planning to incorporate batteries in future work (e.g. [42]) were not selected.
- Approaches that were generally applicable to distributed energy resources, including batteries, were not selected if they did not explicitly consider battery energy management (e.g. [43], [44]).
- Papers only indirectly related to battery management were not selected. A few examples of such indirectly related works are as follows. Biemann *et al.* [45] optimize the temperature in a data center to ensure desirable operating temperature for batteries; Wang *et al.* [46] use a lightweight RL approach on IoT sensor nodes with limited battery capacity; Bing *et al.* [47] design an energy efficient gait for a battery powered mobile robot, but do not consider battery energy management.
- If the same authors published several highly similar papers, only one was selected.
- Although our search string covered all applications of RL to batteries, no recycling related application was encountered within the search results

Figure 3 plots the papers that were manually selected for the review according to the year of publication. An exponential growth in publications is observed, indicating that RL has good potential to become a disruptive technology in battery management. It is notable that the review was performed in the summer of 2021, so the numbers for 2021 in Figure 3 are expected to be significantly higher by the end of the 2021.

IV. PROFITABILITY AND ENERGY EFFICIENCY

In this section, each of the papers in Figure 3 is categorized either under "optimization of energy efficiency", "optimization of operational costs" or "optimization of investment cost".

A. OPTIMIZATION OF ENERGY-EFFICIENCY

A significant portion of the reviewed works aimed to optimize energy efficiency without considering financial objectives. Energy efficiency is understood in different ways depending on the application, and the applications attracting significant amounts of research as summarized as follows. In case of PHEVs, HEVs and EVs, many authors use RL to split the demand for driving power between different battery packs or between the battery and other power sources such as a fuel cell, ultracapacitor or internal combustion engine. The power split optimization is an instantaneous problem, but RL has also been used for energy optimization of EVs, including flying ones, over the course of a single trip. In built environments, batteries are optimized in conjunction with other energy resources such as PV, domestic loads and EV chargers, either in the scope of a single building or a microgrid. IoT sensor networks consist of potentially large numbers of battery-powered IoT nodes with no grid connection and limited or non-existent battery recharging possibilities. In such cases, electricity cost is not considered relevant, and most authors focus on minimizing energy consumption to maximize the battery lifetime. Table 2 summarizes all of the papers according to the 4 aspects around which this review is structured: profitability & energy efficiency, management of user discomfort, battery losses & degradation and context of use of the battery. Each of these aspects is discussed further in the text of the sections 4, 5, 6 and 7 respectively. The table summarizes these aspects for each paper, and the works are ordered according to the publication year.

B. OPTIMIZATION OF OPERATIONAL COSTS

The optimization of operational costs of systems with batteries is a popular objective for RL; however, since RL is capable of multi-objective optimization, the financial objectives are often complemented by other application specific objectives or battery degradation related objectives. Several authors are not explicit about what kind of electricity market is being considered, but an analysis of the papers that did specify this reveals the following list of markets against which optimizations have been performed: day-ahead, intra-day, real-time pricing, time-of-use pricing and frequency reserve markets. Using the battery storage to increase self-consumption of renewable generation and to reduce purchase of grid power is a common theme for RL applications to buildings, energy communities and microgrids. Such works have been included in Section IV.A if there was no financial element to the optimization Otherwise, they have been included in this section. The charging of EVs and fleets of

TABLE 2. Papers primarily aiming at optimizing energy-efficiency.

Paper	Profitability & energy efficiency	Management of user discomfort	Battery losses & degradation	Context of use of the battery
[48]	Minimize gasoline consumption	Not relevant	Min/max SoC limits	Tracked PHEV driving
[49]	Optimal control of heterogeneous battery types: vanadium redox and lead-acid	Not relevant	Hard constraints on SoC and charge/discharge power.	Microgrid with HVAC loads and heterogeneous batteries
[23]	Maximize PV self-consumption	Not relevant	Inverter efficiency	Building with PV
[50]	Minimize energy losses	Not relevant	Avoid high discharge current	HEV with battery and ultracapacitor
[51]	Maximize PV generation	Not relevant	Not considered	PV + load + battery
[52]	Optimize fog node's energy storage to reduce job loss probability	Not relevant	Charge/discharge efficiency	Fog-computing node
[53]	Optimize battery discharge time.	Maintain an acceptable Quality of Service	Not considered	Road side off-grid unit in vehicular network
[54]	Minimize gasoline consumption	Not relevant	Penalize deviations from reference SoC	PHEV bus driving
[13]	Minimize gasoline consumption	Not relevant	Limit discharge power	PHEV driving
[55]	Minimize energy losses	Not relevant	State-of-health is an optimization criterion	HEV with battery and ultracapacitor
[56]	Minimize gasoline consumption	Not relevant	Penalize deviations from ideal SoC	PHEV driving
[57]	Minimize gasoline consumption	Not relevant	Penalty when SoC under reference value	PHEV driving
[58]	Minimize gasoline consumption and tailpipe NOx emissions	Not relevant	Min/max SoC limits	PHEV driving
[59]	Minimize fuel consumption	Not relevant	Penalize deviation from reference SoC	PHEV bus driving
[60]	Minimize fuel consumption	Not relevant	Different reward functions for different SoC ranges	PHEV driving
[61]	Minimize the battery prediction loss for intelligent energy harvesting	Not considered	Not considered	Small cell IoT
[62]	Optimal scheduling of power transfer and data transmission	Not relevant	Not considered	Wireless energy harvesting for sensor network
[63]	Optimal policy for access and power control.	Not relevant	Not considered	Energy harvesting user equipment
[64]	Optimal power control policy	Not relevant	Not considered	Large energy harvesting networks
[65]	Real-time energy management of hybrid storage	Not relevant	Min/max limits for SoC	Wave energy conversion system with hybrid storage
[66]	Optimal policy allocation for ensuring quality of data transmission	Not relevant	Min/Max battery capacity	Energy harvesting in underwater relay network
[67]	Optimize online policy for wireless energy transfer	Not relevant	Not Considered	Energy harvesting RF-powered communication systems
[68]	Graceful degradation when grid connection is lost	Not relevant	Not considered	Islanded microgrid
[69]	Maximize PV self-consumption	Not affected	Charge/discharge efficiencies	House with PV, buffered heat pump & battery
[70]	Select between power sources to minimize consumption (hydrogen equivalent)	Not relevant	Penalize deviations of battery SoC from reference	PHEV with fuel cell, ultracapacitor & battery
[71]	Minimize fuel consumption	Not relevant	Charge/discharge efficiency & min/max SoC	PHEV driving
[72]	Minimize use of diesel generator	Not relevant	Detailed lead-acid battery model	Isolated microgrid with battery and fuel cell
[21]	Minimize gasoline consumption and SoC variations	Not relevant	Sophisticated battery model	PHEV driving
[73]	Minimize flight time	Not relevant	Not considered	Aerial crop scouting
[31]	Minimize gasoline consumption and SoC variations	Not relevant	Penalize deviations from reference SoC	HEV driving
[74]	Minimize gasoline consumption and battery losses	Not relevant	Charge/discharge losses converted to gasoline equivalent	PHEV driving
[75]	Optimize IoT node battery charging based on bandwidth need at the node and available wind power	Not relevant	Not considered	Drone fleet for charging IoT sensor nodes

TABLE 2. *(Continued.)* **Papers primarily aiming at optimizing energy-efficiency.**

[76]	Minimize acceleration/ deceleration cycles	Smoother traffic due to less stop-and-go traffic waves	Smoother acceleration should reduce degradation	Connected and automated electric vehicles
[77]	Energy efficient acceleration	Slower accelerations	Not considered	Self-driving EV
[78]	Minimize transmission power through selection of base station & subchannel	Not relevant	Not considered	Heterogeneous network of IoT devices
[79]	Minimize path length to cover the area	Not relevant	Not considered	Cleaning robot
[80]	Deactivate selected IoT nodes to minimize adverse effects of high ambient temperature and direct sunlight	Not relevant	Not considered	Outdoor IoT sensor network
[81]	Optimize Discharge Efficiency with algorithm toggling cells on/off	Improves the hardware safety by restricting the use of bypass circuits	Not considered	Reconfigurable battery consisting of several cells
[82]	Coordinated charge of grid-support batteries during PV peak	Not relevant	Charge/discharge efficiencies	Decreasing Overvoltage during high PV generation
[83]	Optimal power split between battery and supercapacitor	Not relevant	Equivalent circuit model in simulation package	Unmanned aerial vehicle
[84]	Optimal power split between battery and fuel cell	Not relevant	Charge and Discharge coefficients	All-electric ferry boat
[85]	Storage operation strategy to manage wind power forecast uncertainty	Not relevant	Charging and discharging power constraints.	Wind power forecast uncertainty
[86]	Optimal SoC management while participating on frequency reserves	Not relevant	Sophisticated battery model	Flywheel, battery and ultracapacitor for grid support
[87]	Decide which battery will charge/discharge when train breaks/drives	Not relevant	Min/max limits for SoC	Urban rail energy storage system with multiple batteries
[88]	Minimize grid energy consumption at the base station	Not relevant	Battery state is considered	Energy harvesting virtualized small cells with batteries
[89]	Minimize fuel consumption	Not relevant	Equivalent circuit model captures losses	PHEV driving
[90]	Energy saving & voltage stabilization for supercapacitor energy storage	Not relevant	Battery equivalent circuit model captures losses	Urban rail energy storage system with supercapacitor
[91]	Minimize energy consumption for a route	Not relevant	Detailed lithium ion battery model	EV route planning
[32]	Maintain SoC close to reference & minimize fuel consumption	Not relevant	Penalties when SoC out of ideal range	PHEV driving
[92]	Maximize reward of one tour of the mobile charger	Not relevant	Not considered	IoT sensor network with mobile wireless charger
[93]	Maintain quality-of-service while avoid battery depletion	Not relevant	Not considered	Energy harvesting device-to-device IoT
[94]	Minimize race time so that total energy from battery to motor is under 52kWh	Not relevant	Battery thermal model for fast moving vehicle	Formula-E
[20]	Minimize gasoline consumption and limit wear to battery and engine	Not relevant	Minimize charge/discharge cycles	PHEV driving
[29]	Reschedule EV charging to load valleys	Drivers are not required to give input	Linear charging losses	Scalable EV charging
[95]	Short and long-term battery optimization in microgrid	Supply to load is ensured	Charge/discharge efficiencies, power limits and degradation	Isolated microgrid
[96]	Optimize operating cost and pollution cost	Not relevant	Detailed charge/ discharge model of lead-acid battery	Microgrid with battery+PV+ diesel
[97]	Minimize wind power variance	Not relevant	Maximum depth of discharge	Wind farm
[19]	Energy-efficient operation of hybrid battery pack in EV	Not relevant	Focus on battery temperature	EV with high-energy and high-power battery packs
[98]	Battery lifetime for 5G IoT sensors	Not relevant	Not considered	IoT application for target tracking
[99]	Activate IoT nodes with best energy-efficiency for a tracking task	Not relevant	Not considered	Wireless energy harvesting for IoT nodes
[100]	Minimize energy loss and battery aging cost	Not relevant	Each battery pack has min/ max temperature constraints	EV with high-energy and high-power battery packs
[101]	Transfer computing tasks to IoT nodes closer to the access point	Not relevant	Not considered	Wireless energy harvesting for IoT nodes

TABLE 2. ***(Continued.)*** **Papers primarily aiming at optimizing energy-efficiency.**

[28]	Energy efficient pedal control	Safety logic can override the RL to prevent crashes or going against red lights	Not considered	Self-driving collaborative PHEV
[102]	Minimize fuel consumption	Safety logic can override the RL to prevent crashes or going against red lights	Not considered	Self-driving collaborative PHEV
[103]	Thermal and health-conscious energy management	Not considered	Overtemperature penalty and multistress-driven degradation	Hybrid electric bus
[104]	Minimize hydrogen fuel consumption	Not considered	Min/max limits for SoC	EV with fuel cell & battery
[105]	Reduce energy consumption of the UAV batteries	Not relevant	Limits for battery threshold	Unmanned aerial vehicle network in the presence of jammer
[106]	Minimize the overall data packet loss	Not relevant	Battery levels are considered.	Unmanned aerial vehicle flight control in wireless sensor networks
[107]	Minimize use of gasoline	Not relevant	Min/max limits for SoC	PHEV energy management
[108]	Reduce frequency disturbance caused by renewable sources	Not relevant	Min/max limits for SoC	Interconnected power grid with renewable energy

FIGURE 4. **Reviewed articles primarily aiming at optimizing energy-efficiency.**

FIGURE 5. **Reviewed articles primarily aiming at optimizing operational costs.**

EVs is another topic of interest; however, a wide variety of formulations for the optimization problem were encountered in the literature. Finally, a few authors propose a new market for prosumers, energy communities or microgrids, and perform their optimization against that market. The analyzed papers are summarized in Table 3, sorted by the year of publication.

C. OPTIMIZATION OF INVESTMENT COST

Whereas most works focus on real-time operation or short-term planning of the operation of an existing system, a minority of works seek to minimize investment cost. The dimensioning or placement of the battery storages is a common optimization problem shared by these works. However, most authors do this optimization in the context of a larger problem that also considers investments to other kinds of energy storages. Another kind of investment cost optimization involves optimizing the lifespan of the battery by limiting the battery degradation. Such considerations have been incorporated as additional optimization criteria to several of the short-term optimization works in Section IV.A and Section IV.B. Works that have battery lifetime maximization

TABLE 3. Papers primarily aiming at optimizing operational costs.

Paper	Profitability & energy efficiency	Management of user discomfort	Battery losses & degradation	Context of use of the battery
[109]	Minimize the electricity cost over a billing period	Not affected	Charge/discharge efficiencies	Residential PV with battery
[15]	Minimize electricity cost	Not relevant	Explicitly not considered	Building with PV & V2G charger
[110]	Optimize charging to maximize revenue from carrying passengers	Not relevant	Maximize SoC after delivering passenger	EV taxi fleet
[111]	Balance the loads among generation units and batteries. Reduce electricity bills	Not relevant	Charge and Discharge coefficients	Microgrid with distributed generation and battery
[112]	Minimize peak load of the power plant	Not relevant	Min/Max battery capacity	Microgrids with PV, wind, battery & load
[113]	Minimize electricity cost and maximize manufacturing throughput	Not relevant	Operation and maintenance cost	Uninterrupted manufacturing in grid outages
[114]	Maximize self-sufficiency of microgrids & minimize cost of fuels	Not relevant	Penalize RL agent for deviations from reference SoC	Energy internet
[115]	Minimize diesel + electricity cost	Not relevant	Charge/discharge efficiency	PHEV bus driving
[6]	Minimize electricity cost	Included in optimization	Penalize RL agent for exceeding min/max SoC	Building with appliances & EV charger
[116]	New auction-based market	Not relevant	Charge/discharge efficiency	Community of microgrids
[117]	Minimize diesel + electricity cost	Not relevant	Charge/discharge efficiency	PHEV bus driving
[118]	Maximize PV self-consumption and minimize electricity costs	Ensure users electricity demand is satisfied.	Charge/discharge efficiencies	Smart energy community with P2P trading
[119]	Minimize energy costs & avoid transformer overloads	Follow consumer preferences	Reward for keeping SoC in 20-80% range	Fleet EV charging management
[120]	Minimize electricity cost of the charging station & degradation cost of the AEV	User dissatisfaction is considered	Penalty when outside min/max SoC	Autonomous Electric Vehicles charge scheduling
[121]	Minimize the electricity cost	Priority to Non-Shiftable Appliances	Not considered	Home energy management for appliances
[122]	Maximize profit of battery agent	Not relevant	Min/max SoC limits	Independent agents trading generation, battery and load resources
[123]	Optimal bidding strategy for EV owners	Minimize waiting time of EV owner	Not considered	EV charging station
[124]	Minimize vehicle travel and charging cost	Customer waiting cost and abandon penalty	Not considered	Automated EV taxi fleet
[17]	Maximize revenue under uncertain generation and price	Not relevant	Charge/discharge efficiencies & power limits	Wind farm
[24]	New market (similar to stock market)	Not relevant	Battery efficiency	Building with PV
[30]	Minimize electricity cost	Included in optimization	Penalize RL agent for exceeding min/max SoC	Building with appliances and EV charger
[125]	Minimize electricity cost	Not considered	Sophisticated degradation model	Factory production
[126]	Minimize charging cost under real-time electricity price	Not relevant	Not considered	General
[127]	Exploit Time-of-Use tariffs in microgrid	Not relevant	Consider operating cycles and SoC in each cycle	Microgrid with load, PV, EV charging, renewables
[128]	Energy arbitrage	Not relevant	Semi-empirical battery degradation model	Standalone battery
[129]	Minimize the operating cost of the battery swapping station	Not relevant	Ignored battery capacity degradation	Electric bus battery swapping station
[130]	Minimize consumption of power from grid	Considers penalties for the consumer thermal discomfort	Penalty for battery overcharging and undercharging.	A community of smart Homes
[131]	Minimize the operation costs of EV charging stations	Not relevant	Charge/discharge efficiencies	Vehicle charging stations with PV system
[132]	Minimize the charging cost under real-time pricing	Satisfy charging demand	Charge/discharge efficiencies	EV charging station
[133]	Trade surplus residential battery capacity on spot market	Supply to residential load is guaranteed	Charge/discharge efficiencies	Residential PV + battery

TABLE 3. ***(Continued.)*** **Papers primarily aiming at optimizing operational costs.**

[134]	Optimize fuel cells and battery pack to minimize voyage cost	Not relevant	Charge/discharge efficiencies and degradation	Ship powered with fuel cells and batteries
[135]	Minimize EV charging cost for the aggregators	Not relevant	Min/Max battery capacity	EV charging aggregators
[136]	Minimize the energy cost	Thermal comfort limits are defined	Charge/discharge efficiencies	Smart home energy management
[137]	Minimize electricity cost	Not relevant	Min/max limits for SoC	House energy management with PV and EV
[138]	Maximize the profits from trading aggregated battery capacity on several microgrid markets	Not relevant	Min/max limits for SoC	Interconnected microgrids energy trading platform
[139]	Exploit variable prices to reduce charging cost	Not relevant	Min/max limits for SoC	EV charge scheduling
[140]	Minimize operating cost of DSO & microgrids	Not relevant	Hard limits on min and max SoC and charge/ discharge power	Microgrids connected to a distribution network
[141]	Minimize cost of electricity on real-time markets for PV+battery+load	Not relevant	Nonlinear charging efficiency functions	Centrally managed microgrid
[18]	Participate on frequency reserve markets	Not relevant	Soft limits on min and max SoC	Grid support
[142]	Minimize electricity bill of microgrid	Price-responsive household loads	Charge/discharge efficiencies & power limits	Microgrid with wind, battery & residential
[22]	Minimize HVAC electricity bills under ToU pricing	Penalize indoor temperature limit violations	Penalize violation of min/max SoC	Community of buildings with shared battery
[143]	Maximize profits of a community battery on a local market	Not relevant	Financial benefit is afterwards compared to aging cost	Community of buildings with shared battery
[144]	Minimize cost of electricity from grid	Not relevant	Not considered	Household with PV & battery
[145]	Minimize electricity bills	Avoid any need for demand response	Penalize violation of min/max SoC	Community of end users & one battery
[146]	Minimize electricity cost	Not relevant	Degradation considered comprehensively	Building with PV
[147]	Passenger throughput	Penalty if the taxi collides	Not modelled	Aerial drone taxi
[14]	Minimize electricity bills	Not affected	Charge/discharge efficiencies for battery and V2G	Community of buildings in same low voltage grid
[148]	Minimize operating cost of microgrid	Not considered	Charge/discharge efficiencies	Multi-energy microgrid with power, heating, cooling & gas
[149]	Minimize operating cost of microgrid	Not affected	Battery deprecation cost is included to the multi-objective optimization	Multi-energy system with power, heating, renewables & gas
[150]	Minimize the cost of the power loss	Not relevant	Penalty when SoC out of bounds	Distribution network with battery store & wind power
[151]	Minimize storage maintenance cost	Not affected	Degradation considered	General
[152]	Maximize the net profit on spot and frequency reserve markets	Not affected	Degradation is considered as cost.	Photovoltaic-battery storage systems
[153]	Minimize charging expense	Satisfy user requirement of battery energy	Min/Max battery capacity	EV charging control
[154]	Minimize EV charging cost	Adaptively adjust to EV owners different requirements	Charge/discharge efficiencies	EV charging management
[155]	Reduce operating cost	Not relevant	Charge/discharge efficiencies	EV charging station with renewable generation
[156]	Scheduling to minimize cost & driver anxiety	Driver's anxiety is modeled	Charge/discharge efficiencies	EV charging control under variable price
[157]	Minimize the daily voltage regulation cost	Not relevant	Min/max limits for SoC	Distribution grid with PVs and batteries
[158]	Minimize grid power purchase with peer-to-peer trading	Meet the load demand of the user.	Min/max limits for SoC	Energy community with domestic PV and storage

TABLE 3. *(Continued.)* **Papers primarily aiming at optimizing operational costs.**

[159]	Minimize operating cost of microgrid with PV, wind, diesel and battery	Not relevant	Min/max limits for SoC	Microgrid with multiple distributed storages
[160]	Minimize maintenance cost	Not relevant	Gaussian process of degradation	Battery maintenance
[161]	Maximize revenue and minimize penalties from the market	Not relevant	Min/max limits for SoC	Primary frequency reserve market

TABLE 4. **Papers primarily aiming at optimizing investment cost.**

Paper	Profitability & energy efficiency	Management of user discomfort	Battery losses & degradation	Context of use of the battery
[162]	Maximize the battery lifespan	Not relevant	Thermal and SoH degradation model	Portable systems with battery & supercapacitor
[163]	Minimize the operating energy cost & identify the optimal battery sizing	Not relevant	Degradation is considered by usage & storage aging	Telecom base station with PV & battery
[6]	Minimize cost of wireless bus charging infrastructure	Not relevant	Not considered	Wireless charging of EV
[164]	Minimize cost of wireless tram charging infrastructure	Not relevant	Not considered	Wireless charging of tram
[165]	Optimize battery lifespan	Not relevant	Optimize the charge/discharge profile	Microgrid with PV, diesel, battery and ultracapacitor
[166]	Maximize day-ahead market revenue & minimize capital expenditure	Not relevant	Manufacturer's guaranteed energy capacity retention limit is used as worst case assumption for degradation	Timing and sizing of battery investment
[167]	Minimize battery capacity to cope with errors in wind generation forecasts	Not relevant	Hard limits on min and max SoC	Wind farm
[168]	Motivate investment to connecting microgrids	Not relevant	Not relevant	Off-grid microgrids
[169]	Minimize investment to battery, inverter, wind turbine, fuel cell, electrolyzer, diesel & PV	Not relevant	Note relevant	Isolated microgrid
[170]	Optimize battery lifespan	Not relevant	Impact of overheating on degradation	General
[171]	Reduce the battery investment cost and data transmission delay cost	Not relevant	Min/max limits for SoC	Energy Harvesting Machine-Type Communication
[172]	Minimize sensor data communication needs across subsystems	Not relevant	Not considered	Distributed energy system cybersecurity
[173]	Minimize cost for electricity, heat and fresh water	Not relevant	Limits to charge/discharge power	Multi-carrier water and energy system
[174]	Optimal battery investment strategy over the lifetime of a microgrid	Failure to supply loads is penalized	Not considered	Microgrid
[175]	Optimal battery sizing with lowest cost	Not relevant	Empirical battery degradation model	Fuel Cell – PHEV powertrain

as the main optimization criterion are covered in this section. The reviewed works are listed in Table 4, sorted by the year of publication.

V. MANAGEMENT OF USER DISCOMFORT

The reviewed papers can be categorized under three approaches for considering a human user. The first approach is to ignore the user, since the nature of the system and goals of the optimization problem are such that users are not impacted. The majority of the reviewed papers use this approach. The second approach is to impose constraints on the RL agent to ensure that users are not impacted. The third approach is to permit the RL agent to take actions that cause some inconvenience or discomfort to the users, and to penalize the agent for such impacts. In this analysis, all such impacts are referred to with the term user comfort.

For optimizing the battery usage of EVs and PHEVs as they are being driven, RL applications are concerned with energy efficiency. Most works are concerned with battery management decisions that do not impact passengers (e.g. [13], [19], [71]). However, a minority of works do consider the driving experience. The acceleration of the vehicle may be limited for reasons of energy efficiency [77] or safety [28], [102]. For traffic flow management with self-driving EVs, Qu *et al.* [76] aim to reduce stop-and-go traffic waves. In an autonomous flying taxi system, Yun *et al.* [147] penalize the RL system for any in-flight collisions.

FIGURE 6. Reviewed articles primarily aiming at optimizing investment cost.

EV charging systems have the potential to disrupt the lives of EV drivers. Tuchnitz *et al.* [29] define user comfort as avoiding the trouble to give input to a system that coordinates EV charging – this benefit is questionable since the EVs are not compensated. In an autonomous EV, Cao *et al.* [120] find a tradeoff between the total electricity cost and the waiting time to charge the vehicle. Zhang *et al.* [123] aim to minimize the time that the EV owner spends getting access to the charging point and waiting for the EV to charge. Yan *et al.* [156] propose a driver anxiety concept that captures the likelihood of the EV not having sufficient SoC for making an unexpected trip. For an EV taxi fleet, Tang *et al.* [124] define user comfort as customer waiting times. Reference [119], [153], [154], Li and Wan [132] define a user requirement for the SoC and minimize the charging cost within this constraint.

Residential sector applications require careful consideration of potential user comfort issues. When the battery is used to shift electricity purchases and sales, user comfort is not impacted [15], [24], [121]. This remains true if local PV generation is added to the mix [23], [146], [158]. Lee and Choi [6] and Lee and Choi [30] consider a smart home with a capability to reschedule appliances and EV battery charging, and penalize for dissatisfaction resulting from rescheduling. Nakabi and Toivanen [142] consider household loads that respond to a dynamic price signal, and assume that user comfort is incorporated to the load controllers as a price elasticity parameter. Lee *et al.* [22] optimize a HVAC system and minimize of the deviations of the indoor environment outside an ideal range. Lee and Choi [130] and Yu *et al.* [136] considers appliance agents for home energy management systems which are optimized against two criteria: reducing electricity bills while satisfying the consumer comfort level for heating and the consumer preferences for appliances.

VI. BATTERY LOSSES AND DEGRADATION

Some works assume that no loss occurs during the charging, discharging, and idling of the battery. The battery aging and degradation is also not considered by many works. In some case the authors state this explicitly (e.g. [15], [144]). In some kinds of applications, ignoring these effects can be justified, as they are not directly related to the optimization problem. For example, Sultan *et al.* [98] optimize the selection of active sensors in an IoT sensor network to achieve the required data communication task so that the total energy consumption at the IoT nodes is minimized. As another example, Sangoleye *et al.* [99] find the optimal base station for each IoT node to connect to. In both of these examples, authors assume that these optimizations achieve the ultimate goal of the research, which is to prolong the battery lifetime at the sensor nodes through reducing the energy consumption. Diverse approaches are used by the authors that do consider energy losses and degradation. With respect to the RL problem formulation, these approaches can be categorized as capturing losses in the environment, imposing constraints on the RL agent to avoid degradation, or including minimization of degradation as an optimization criterion in the reward function. These approaches are not mutually exclusive, and ideally authors capture charging and discharging inefficiencies in the environment and additionally consider degradation in the reward function (e.g. [18]).

A. CAPTURING BATTERY LOSSES IN THE REINFORCEMENT LEARNING ENVIRONMENT

The RL environment usually uses a set of equations that define how the SoC is impacted by the control action taken by the RL agent. The SoC is often a state variable and may in some cases be used in the reward function, for example in formulas that capture battery degradation. A common battery modelling approach in the environment of the RL agent is to capture energy losses resulting from battery operations with factors for charging and discharging efficiency, and to impose limits to charging and discharging power (e.g. [22], [24], [29],[116], [133], [142]). Whereas most authors capture losses as a simple coefficient for charging and discharging efficiency, a few authors use more detailed models. Chen *et al.* [71] use a non-linear battery model and Zhang *et al.* [96] model the charging and discharging dynamics in detail for a specific type of battery, the lead-acid battery. Kolodziejczyk *et al.* [141] model the maximum charging and discharging power as non-linear functions of SoC. Totaro *et al.* [95] model how charging and discharging efficiencies as well as the battery storage capacity degrade over time. For problem formulations that permit selling battery energy to the grid, the inverter efficiency as a function of discharging power is a significant factor taken into account only in the minority of the works [23]. Aljohani *et al.* [91] include temperature in their battery model to ensure an accurate tracking of SoC over the duration of a trip. Liu *et al.* [94] consider energy management in the

specific application of Formula-E races, and carefully model the impact of ambient temperature and vehicle speed on battery temperature.

In general, the battery is captured in the environment of the RL agent by a set of equations defined by the authors. An alternative approach would be to use a battery simulation [83]. A few authors use RL to improve such simulation models. Unagar *et al.* [176] use machine learning to infer the battery model's parameters. RL is used to avoid the need for labelled training data, as would be the case for supervised learning methods. Kim *et al.* [177] use RL to obtain a more accurate method for estimating the SoC of lithium-ion batteries than what has been possible with model-based methods.

B. IMPOSING CONSTRAINTS ON THE AGENT TO PREVENT BATTERY DEGRADATION

One approach for limiting battery degradation is to impose constraints, so that an external logic overrides the actions taken by the RL agent in case these constraints are violated. A simple approach is to define minimum and maximum SoC and charging and discharging power thresholds as hard constraints [49], [58], [140], [167]. Nyong-Bassey *et al.* [72] take this constraint as the starting point for power pinch analysis, which anticipates SoC threshold violations and takes actions ahead of time to ensure that the violations will not occur. For self-driving EVs, Tang *et al.* [124] implement a constraint that the vehicle must reach a charging station before its SoC drops below a minimum threshold.

C. INCORPORATING BATTERY DEGRADATION INTO THE REWARD FUNCTION

Reducing battery degradation is included to the multi-objective optimization problem by adding a penalty term to the reward function. A simple approach is to penalize situations in which the battery SoC exceeds a minimum or maximum threshold (e.g. [6], [18], [30], [130], [145]). Other authors penalize SoC deviations from a reference value [31], [54], [56], [59], [70], [89], [114]. Zhou *et al.* [32] do this only when the SoC is out of an ideal operating range of 60-85%, and Zhou *et al.* [57] do this only when the SoC is under the reference value. Qi *et al.* [60] add a penalty term to the reward function when the SoC is out of the 20-80% range. Cao *et al.* [150] is similar for the range 20-90% and Silva *et al.* [119] penalize when the SoC is less than 20%. Yang *et al.* [149] include a battery deprecation cost that is proportional to the charge/discharge power at each timestep. Chen *et al.* [13] include the minimization of the maximum battery discharge power as one of the optimization criteria. Cao *et al.* [128] determined that battery degradation is a linear function of charge/discharge cycles in the short term, and incorporate this penalty to the reward function. Shang *et al.* [127] consider the number of operating cycles and the SoC in individual cycles. Muriithi and Chowdhury [146] capture the degradation of a lithium-ion battery in terms of depth of discharge. Roesch *et al.* [125] use a sophisticated battery degradation model to capture the impacts of irregular charging and discharging cycles on battery degradation. Yang *et al.* [17] penalize the number of switches between charging, idle and discharging modes. Cao and Xiong [50] do not explicitly consider degradation, but formulate the RL problem to avoid energy losses by avoiding high discharge currents, an approach which will have side benefits related to mitigating degradation.

A minority of works considers the impact of temperature on aging and degradation. Sui and Song [170] consider a battery pack and propose an intelligent controller to select between batteries to avoid overheating caused by excessively frequent charging and discharging of any single battery. Li *et al.* [19] go further and consider diverse 'high energy' and 'high power' battery packs [100]. The above-mentioned approaches include temperature as an aspect of the optimization problem by incorporating the temperature effects into the reward function. Xie *et al.* [162] use a thermal model and SoH degradation for the aging of a lithium-ion battery.

The majority of works uses SoC in their reward formulations, but SoH (State of Health) is used in some papers. Xiong *et al.* [55] define SoH as the ratio between the present and rated battery capacity. Wu *et al.* [151] define SoH in terms of capacity fade. Mendil *et al.* [163] define the battery state jointly described by SoC and SoH.

VII. CATEGORIZATION BY APPLICATION

This section categorizes the reviewed works by application. Each paper is categorized under only one application, unless it strongly fits under several categories. Generic works that do not mention any application are not discussed in this section. The pie chart in Figure 7 categorizes the reviewed articles by application and gives an indication of which kinds of RL battery management applications are expected to receive a high number of publications in a future. However, in addition to the information in Figure 7, the following insights from the analysis of individual papers should be considered:

- The problem of managing the power split of one or more battery packs and other sources of power has become a well-established line of research, in which the RL problem formulation was quite similar across all the works in the 'EV & HEV driving' and 'PHEV driving' categories.
- Additional 'EV charging' applications are included in the 'Buildings' category.
- No works were found addressing stand-alone PV plants, so PV does not appear as a separate category. However, PVs are a central element in many of the works in the categories 'Buildings', 'Energy communities, 'Grid-connected microgrids', 'Isolated microgrids' and 'Multi-carrier systems'.
- A few works address wind farms. Wind power was also covered by works in some of the other categories, but to a much lesser extent than PV. This is unsurprising, since

rooftop PV is becoming increasingly common, whereas windmills are usually not welcomed in the vicinity of buildings.

- A huge body of research was encountered related to IoT, but only a small minority addressed battery management. As the IoT community begins to address the practical issues related to deploying and maintaining IoT systems, it is possible that there is a significant growth of research in this category.

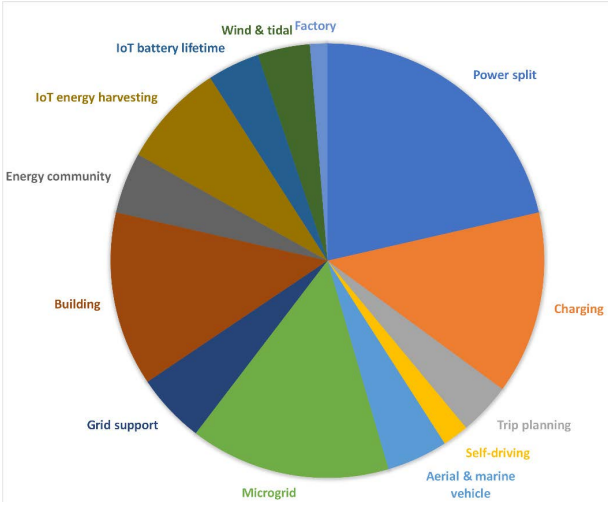

FIGURE 7. Reviewed articles per application.

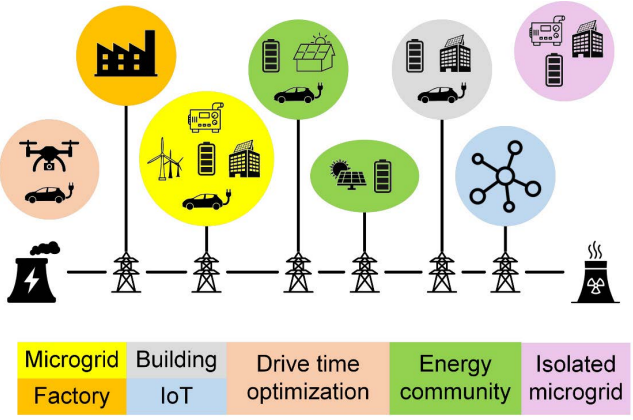

FIGURE 8. Applications overview.

A. VEHICLE

1) LAND

a: POWER SPLIT

i)PHEV

In contrast to EVs, the battery management of PHEVs has the additional consideration of switching between battery power and fuel. Most authors minimize fuel consumption [13], [32], [71], [107]. Other authors additionally penalize actions that wear down the battery [20], [31], [54], [56], [57], [60], [74] and the engine [21], [89]. RL formulations for optimizing the driving performance of PHEVs have power demand and SoC as the state variables. Some authors add velocity [71] and road slope [59]. The action involves controlling either the engine power (e.g. [20], [31], [56], [115]) or the battery power (e.g. [13], [71]). The problem is usually framed as a question of satisfying the power demand for moving the vehicle forward (heading demand) with the engine and the battery, but in case of a tracked vehicle, the power demand consists of the heading demand as well as the steering demand [48], [56]. In contrast to the majority of the research, Wu *et al.* [117] and Tan *et al.* [115] perform their optimization based on cost, taking into account the price of electricity and diesel. The above-mentioned authors consider emissions only indirectly through minimizing gasoline consumption. However, Hofstetter *et al.* [58] add tailpipe NOx emissions as a constraint to the optimization problem. For a Fuel Cell - PHEV hybrid powertrain, Li *et al.* [175] propose a framework for achieving optimal battery sizing parameters with minimal operation cost and component degradation.

ii) EV AND HEV

Most works on EVs and HEVs that do not have a gasoline engine involve selections between different types of battery packs [19], [100] or selections between the battery and other on-board power sources such as fuel cells [104] and ultracapacitors [70]. Cao & Xiong [50] aiming to reduce energy losses by avoiding high discharge currents and Xiong *et al.* [55] optimize the state of health of HEV batteries. Whereas most works assume a human driver, He *et al.* [77] and Wegener *et al.* [28] consider self-driving vehicles, with which it is possible to include energy efficient acceleration into the optimization.

b: CHARGING

i) CAR

EV charging optimization targets include the following: reducing peak load [29], reducing charging costs for the EV [119], [132], [139], [153], [154], reducing both charging cost and waiting time [123], reducing charging cost based on knowledge of user behavior [156], minimizing the cost for the charging station with a PV and battery storage [155], minimizing the cost of several such stations [131], and aggregating several stations within a local market operated by an aggregator [135].

ii) EV FLEET

For fleets of self-driving EVs, minimization targets include charging costs [124], charging times [120], time spent not carrying passengers [110] and battery exhaustion [147].

iii) RAIL

For rail applications, stationary batteries are a viable alternative to wireless charging. Regenerative breaking by the train can be used to charge the batteries, which are then

used to power the train when it drives. Zhu *et al.* [87] and Yang *et al.* [90] optimize such a system by minimizing power consumption from the grid and minimizing the losses from regenerative breaking.

Wireless charging is generally not investigated for rail transport, since established solutions for connecting to the grid are available. However, Ko [164] propose a wireless charging infrastructure for trams.

iv) BUS

Gao *et al.* [129] optimize the charging/discharging schedules of electric buses in battery swapping stations with V2G capability. The objective is to minimize the station's electricity bill. Wu *et al.* [103] optimize the energy management of hybrid electric buses by penalizing for overtemperature and degradation of the battery. Lee *et al.* [6] design a wireless charging system and minimize the battery size and charging times.

c: SELF-DRIVING VEHICLES

Self-driving vehicles could be coordinated to achieve smoother traffic flow than what is possible with human drivers. One goal formulation is to reduce stop-and-go traffic waves or other abrupt velocity changes, since this reduces acceleration/deceleration cycles and thus battery degradation [28], [76]. Guo *et al.* [102] minimize fuel consumption and travel time while having safety overrides to avoid hazardous actions.

d: TRIP PLANNING

RL has been applied for the trip planning of human driven EVs, HEVs or PHEVs [91], [94] and mobile robots [79]. This can include either route planning [79], [91] or optimizations made for a predetermined route [94].

2) AERIAL

Batteries in unmanned aerial vehicles (UAV) are used for flying and data transmission. Flying applications include the minimization of flight path length [73], maximizing flight time [83] and using only locally generated wind power for charging the fleet [75]. The following data transmission applications were encountered. Wang *et al.* [105] propose a framework for the UAVs to independently select their transmit power in the presence of a jammer. Li *et al.* [106] propose an RL-based flight resource allocation framework to minimize the overall data packet loss to avoid additional energy consumption from retransmission.

3) MARINE

Battery management for short distance electric ships involves optimization of decision making for battery usage and charging. The on-board energy storages include a battery and a fuel cell. When the ship is in port, on-shore power can be used to charge the battery, while when it is at sea only the fuel cell can be used to charge the battery. The authors minimize the total cost, which consists of hydrogen fuel cost, fuel cell degradation, battery degradation and on-shore electricity cost [84], [134].

B. GRID

1) MICROGRID

a: GRID CONNECTED

i) ELECTRIC

With respect to the RL applications reviewed in this article, grid-connected microgrids are very similar to the energy communities discussed in Section VII.C.2. The main difference is that a microgrid operates in a geographically constrained area, all energy resources must be physical connected to the microgrid, and power flow limits must be observed at the point of common coupling with the utility grid [127]. Nakabi and Toivanen [142] run a market for household loads within the microgrid, in which loads participate in microgrid-level demand response. Kolodziejczyk *et al.* [141] consider an aggregated load without specifying the type of load. Liu *et al.* [111] introduce a distributed framework to coordinate loads, distributed generation units and storage. Shuai *et al.* [159] perform a multi-objective optimization to minimize the operating cost of a microgrid with PV, wind and diesel, considering fuel prices, power exchange costs of the utility grid and curtailment costs of PV and wind. Nunna *et al.* [138] trade aggregated battery capacity on intra-microgrid markets as well as inter-microgrid markets. Lu *et al.* [112] minimize grid peak power consumption. Wang *et al.* [116] envision new auction-based markets in which microgrids can participate. Guo *et al.* [140] propose a new market to balance cost minimization objectives of the microgrids and the utility. Hua *et al.* [114] maximize self-sufficiency, minimize cost of non-renewable generation and minimize battery degradation. Qiu *et al.* [49] exploit the operational difference between batteries with different chemistries to achieve better efficiency. Duan *et al.* [165] optimize battery lifetime.

ii) MULTI-CARRIER

Multi-carrier systems involve the use of electricity along other forms of energy and, in some cases, freshwater production. Variants of a multi-energy microgrid involve electricity, heat and freshwater production [173], electricity and heat [122] and electricity, gas and heat [148], [149]. Nyong-Bassey *et al.* [72] designed an isolated microgrid with a battery, fuel cell and diesel generator, so that an electrolyzer can use excess PV to replenish the fuel cell, aiming to minimize the need for the diesel generator.

b: ISOLATED

In the case of isolated microgrids, purchases from an external electricity market are either not possible [95], [168] or a last resort to complement local fossil-fuel based emergency generation [96]. Phan and Lai [169] and Zhang *et al.* [96] note that the trend towards a decentralized electric power system should in some seashore regions be complemented

with a move to decentralized freshwater production, so a desalination plant is added to the microgrid. Nie *et al.* [68] curtail loads to keep the microgrid operational for as long as possible.

2) GRID SUPPORT

Applications for grid support can be categorized to market driven applications and to other applications in which the financial incentive has not been specified. Market driven applications include energy arbitrage [128] and frequency reserves participation [86], [152], [161]. Other applications include PV generation peak shaving [82], loss minimization in distribution networks [150], mitigating voltage deviations in low voltage distribution networks with high PV penetration [157] and frequency instability reduction not related to frequency reserve market participation [108].

C. BUILDING

1) SINGLE BUILDING

Buildings are a common context for RL agents managing battery storages in coordination with other energy resources. The main difference is the types of other energy resources available and their flexibility in terms of possibilities for rescheduling or curtailment. Only PV is considered in [24], [23], [69], [109], [133], [136], [144], [146],. Liu *et al.* [137] and Kim and Lim [15] consider EV chargers along with PV. Lee and Choi [6] and Lee and Choi [30] include reschedulable appliances and an EV charger; Alfaverh *et al.* [121] only consider appliances. The optimization objectives for PV related works can be categorized either as maximizing the PV generation through Maximum Power Point Tracking [51], maximizing PV self-consumption [23], [69] or minimizing electricity bills. The latter requires assuming a specific type of electricity contract, such as real-time pricing [144], [146], day-ahead markets [133] or Time-of-Use pricing [15].

2) ENERGY COMMUNITY

Communities of buildings offer further optimization opportunities with shared batteries. An aggregator can trade the capacity of the batteries and other flexible energy resources on utility markets [22], [130], [145]. Alternatively, a local market can be established to avoiding buying and selling from the grid [14], [118], [143], [158]. Recent regulation such as the EU Directive on Common Rules for the Internal Energy Market (EU Directive 2019/944 [178]), and its Article 16 in particular, indicate a change in the regulatory environment that is favorable to such local markets.

D. IoT

1) BATTERY LIFETIME MANAGEMENT

IoT sensor networks are a field of research with diverse applications. Maximizing battery lifetime is generally important in such applications. RL approaches indirectly achieve this by innovative and often application specific solutions for minimizing power consumption. Sultan *et al.* [98] track several targets and activate a minimal number of nearby sensors that can perform the tracking energy-efficiently. Ding *et al.* [78] reduce transmission power by optimizing the selection of the base station and subchannel. Huang *et al.* [93] take battery management as one criterion in a multi-objective optimization that aims to reduce the latencies and dropped packets for the IoT computation tasks. Banerjee *et al.* [80] selectively activate IoT nodes in an outdoor network to minimize the increased energy requirement for data transmission when the node is exposed to high outdoor temperature and direct sunlight. Teng *et al.* [171] reduce both the battery investment cost and data transmission delay with an intelligent power transmission policy. Conti *et al.* [52] allow IoT nodes to offload computation to a fog-computing node.

2) ENERGY HARVESTING

If minimizing energy consumption is not sufficient or practical for prolonging the battery lifetime, energy harvesting approaches are used to recharge the battery. Sangoleye *et al.* [99] identify the best base station to connect to for energy harvesting, whereas Chen *et al* [101] migrate computation tasks to nodes that are best positioned for harvesting. Elmagid *et al.* [67] schedule packet transmissions in a way that is optimal for energy harvesting. Chu *et al.* [61] use battery forecasts to optimize access of IoT nodes to energy harvesting. Chu *et al.* [63] optimize the access and power control policies. Li *et al.* [62] perform simultaneous energy harvesting and data transfer by finding a transmission scheduling strategy to minimize data loss. For maximizing the throughput of large multiple-access channel energy harvesting networks, Sharma *et al.* [64] propose an optimal power control policy. Temesgene *et al.* [88] perform an optimization at virtual small cells that jointly minimizes harvested energy and the volume of dropped traffic. Cao *et al.* [120] minimize the distance travelled, and thus the energy consumption, of a battery powered mobile wireless sensor charger. Faraci *et al.* [75] go further, using a fleet of drones as the mobile wireless charger, and using only locally generated wind power for charging the drones. In a V2I (vehicle-to-infrastructure) roadside unit, a battery is periodically recharged and RL can be used to optimize the quality of service of the communication link without draining the battery before the next recharging period [53]. For energy harvesting in an underwater relay network, Wang *et al.* [66] propose an optimal online power allocation policy to ensure the quality of data transmission.

E. WIND FARM AND TIDAL

Power production from wind and tidal needs to be traded ahead of time, based on forecasts. RL applications to batteries in this context include management of uncertain generation forecasts [85], [167], management of uncertain generation and market forecasts (Yang *et al.*, [17]) smoothing fluctuations in generation [65], [97] and optimizing the revenue of a wind farm with other generation resources on site [140].

F. FACTORY

Batteries are emerging as an element of factory energy systems, either for rescheduling production tasks to lower electricity price periods [125] or to ensure the continuity of production during outages [113].

VIII. DISCUSSION

Comparisons between original research works and attempts to synthesize them are hindered by the fact that each author has a unique formulation of the RL problem, resulting in unique environments, state and action spaces and reward formulations. This field could greatly benefit from the availability of benchmark environments for the different applications of batteries identified in section 7. The OpenAI Gym is an open-source project for creating such environments that implement a standard interface for the RL agent to connect to [179]. A range of benchmark environments implementing the OpenAI interface are available for video games [180]. Similar benchmarks are not available for the energy domain, although a few works in the energy domain implement the OpenAI interface for the following applications: maximum power point tracking of PV installations [181], building energy management [25]–[27], microgrid energy management [142], demand response for building cooling [182]. Building on such works, the emergence of a range of open-source benchmark environments for diverse battery applications could greatly speed up the research on RL applications for battery management and improve the possibilities to comparatively assess similar works and identify the superior RL designs. The closest work to this direction that was found is by Henry & Ernst [183], who published precisely such an environment for electricity distribution systems, but it does not involve batteries.

Specific areas of research that are expected to see significant numbers of publications in the future have been discussed in conjunction with Figure 7. The following unsolved challenges have been identified for further research:

- A number of solutions exist for the problem of managing a battery in conjunction with diverse local energy resources in a building or microgrid. Approaches are split into two bodies of research: optimizing energy efficiency goals and minimizing electricity bills. As any deployments will require financial investments, proponents of the former approach should consider adjusting their research targets to obtain benefits that can serve as the basis for a return-on-investment calculation. *Further research challenge: a cost-benefit perspective should be included in RL problem formulations motivated by energy efficiency.*
- The battery is modelled as part of the environment used in the RL agent's training process. Various levels of abstraction have been used in the modelling, and only a minority of works try to capture the characteristics of a specific type of battery, such as a lithium ion or lead-acid battery. The chosen level of abstraction can cause a significant difference between the performance of the RL agent that has been reported in a scientific publication, and the performance of the same agent when it is deployed to manage a physical battery. *Further research challenge: the trained agents should be deployed to physical batteries and the performance should be compared to the performance achieved against the battery model.*
- Long term battery degradation is captured in a minority of works, which use diverse ways to define the degradation and to incorporate it to a multi-objective optimization problem. This issue, in combination with the varying levels of abstraction in modelling the battery, prevents direct comparisons between the performance reported in different works. Thus, researchers will have difficulties in identifying the most promising lines of research. Developers and implementers cannot be expected to assess how these issues will impact the performance of a RL agent, should it be deployed. *Further research challenge: a benchmark battery model is needed to assess the efficacy of RL solutions aiming to mitigate battery degradation.*
- Innovative battery management solutions can cause inconvenience or discomfort to human users of the system that contains the battery. The identification and resolution of these issues remains largely an unsolved issue. Some authors ignore these issues, some define constraints on user comfort, and some include comfort as one aspect of a multi-objective optimization problem. As these issues receive more attention from researchers, it is possible that original and unique formulations of user comfort will further complicate the comparisons between the performance of different research works. *Further research challenge: the end user of the system that contains the battery needs to be identified and standard approaches for quantifying user comfort are needed; for example, if the battery is used in conjunction with smart building loads, established standard metrics for indoor air quality and thermal comfort should be identified and adapted to the RL problem formulation.*

IX. CONCLUSION

The objective of this manuscript has been to provide an application-oriented review of RL applications to battery systems. In particular, this review aims to introduce energy domain experts to RL and to describe the diverse applications that have been recently published involving batteries. A four-fold approach has been undertaken for this purpose. Firstly, the motivations of the RL research have been analyzed either from an energy-efficiency or financial perspective. Secondly, any efforts to identify and mitigate impacts on end users were analyze. Thirdly, approaches for modelling charging and discharging losses as well as battery degradation were analyzed. Fourthly, the reviewed literature was categorized according to the application.

One key finding is that the batteries are modelled at a high level of abstraction. The great majority of works do not specify the battery chemistry. The RL solutions are trained and validated against these simplified battery models, and there is a lack of further validation against high fidelity models or physical batteries. Further multi-disciplinary research involving battery experts is needed. This article intends to provide such experts with necessary background knowledge and an understanding of the state-of-the-art.

Our literature search was general and thus covered all lifecycle phases of the battery. The great majority of articles addressed real-time control or short-term optimizations. Thus, the focus of the research is on the operation phase of the battery lifecycle. A few works addressed the planning phase, in order to optimize the battery investment cost. None of the reviewed works addressed second-life battery applications, decommissioning or recycling.

REFERENCES

[1] L. Zhang, X. Hu, Z. Wang, J. Ruan, C. Ma, Z. Song, D. G. Dorrell, and M. G. Pecht, "Hybrid electrochemical energy storage systems: An overview for smart grid and electrified vehicle applications," *Renew. Sustain. Energy Rev.*, vol. 139, Apr. 2021, Art. no. 110581, doi: 10.1016/j.rser.2020.110581.

[2] W. Zhang, J. Lu, and Z. Guo, "Challenges and future perspectives on sodium and potassium ion batteries for grid-scale energy storage," *Mater. Today*, vol. 50, pp. 400–417, Nov. 2021, doi: 10.1016/j.mattod.2021.03.015.

[3] K. M. Tan, T. S. Babu, V. K. Ramachandaramurthy, P. Kasinathan, S. G. Solanki, and S. K. Raveendran, "Empowering smart grid: A comprehensive review of energy storage technology and application with renewable energy integration," *J. Energy Storage*, vol. 39, Jul. 2021, Art. no. 102591, doi: 10.1016/j.est.2021.102591.

[4] N. McIlwaine, A. M. Foley, D. J. Morrow, D. Al Kez, C. Zhang, X. Lu, and R. J. Best, "A state-of-the-art techno-economic review of distributed and embedded energy storage for energy systems," *Energy*, vol. 229, Aug. 2021, Art. no. 120461, doi: 10.1016/j.energy.2021.120461.

[5] L. M. S. de Siqueira and W. Peng, "Control strategy to smooth wind power output using battery energy storage system: A review," *J. Energy Storage*, vol. 35, Mar. 2021, Art. no. 102252, doi: 10.1016/j.est.2021.102252.

[6] S. Lee and D.-H. Choi, "Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system, energy storage system, and home appliances," *Sensors*, vol. 19, no. 18, p. 3937, Sep. 2019, doi: 10.3390/s19183937.

[7] A. T. D. Perera and P. Kamalaruban, "Applications of reinforcement learning in energy systems," *Renew. Sustain. Energy Rev.*, vol. 137, Mar. 2021, Art. no. 110618, doi: 10.1016/j.rser.2020.110618.

[8] T. Yang, L. Zhao, W. Li, and A. Y. Zomaya, "Reinforcement learning in sustainable energy and electric systems: A survey," *Annu. Rev. Control*, vol. 49, pp. 145–163, 2020, doi: 10.1016/j.arcontrol.2020.03.001.

[9] M. Glavic, "(Deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annu. Rev. Control*, vol. 48, pp. 22–35, 2019, doi: 10.1016/j.arcontrol.2019.09.008.

[10] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Appl. Energy*, vol. 235, pp. 1072–1089, Feb. 2019, doi: 10.1016/j.apenergy.2018.11.002.

[11] Z. Wang and T. Hong, "Reinforcement learning for building controls: The opportunities and challenges," *Appl. Energy*, vol. 269, Jul. 2020, Art. no. 115036, doi: 10.1016/j.apenergy.2020.115036.

[12] M. S. Frikha, S. M. Gammar, A. Lahmadi, and L. Andrey, "Reinforcement and deep reinforcement learning for wireless Internet of Things: A survey," *Comput. Commun.*, vol. 178, pp. 98–113, Oct. 2021, doi: 10.1016/j.comcom.2021.07.014.

[13] Z. Chen, H. Hu, Y. Wu, R. Xiao, J. Shen, and Y. Liu, "Energy management for a power-split plug-in hybrid electric vehicle based on reinforcement learning," *Appl. Sci. Basel*, vol. 8, p. 2494, Dec. 2018, doi: 10.3390/app8122494.

[14] D. Qiu, Y. Ye, D. Papadaskalopoulos, and G. Strbac, "Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach," *Appl. Energy*, vol. 292, Jun. 2021, Art. no. 116940, doi: 10.1016/j.apenergy.2021.116940.

[15] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, no. 8, p. 2010, Aug. 2018, doi: 10.3390/en11082010.

[16] H. Lee, D. Ji, and D.-H. Cho, "Optimal design of wireless charging electric bus system based on reinforcement learning," *Energies*, vol. 12, no. 7, p. 1229, Mar. 2019, doi: 10.3390/en12071229.

[17] J. J. Yang, M. Yang, M. X. Wang, P. J. Du, and Y. X. Yu, "A deep reinforcement learning method for managing wind farm uncertainties through energy storage system control and external reserve purchasing," *Int. J. Electr. Power Energy Syst.*, vol. 119, Jul. 2020, Art. no. 105928, doi: 10.1016/j.ijepes.2020.105928.

[18] Y. Dong, Z. Dong, T. Zhao, and Z. Ding, "A strategic day-ahead bidding strategy and operation for battery energy storage system by reinforcement learning," *Electr. Power Syst. Res.*, vol. 196, Jul. 2021, Art. no. 107229, doi: 10.1016/j.epsr.2021.107229.

[19] W. Li, H. Cui, T. Nemeth, J. Jansen, C. Ünlübayir, Z. Wei, L. Zhang, Z. Wang, J. Ruan, H. Dai, X. Wei, and D. U. Sauer, "Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles," *J. Energy Storage*, vol. 36, Apr. 2021, Art. no. 102355, doi: 10.1016/j.est.2021.102355.

[20] N. Yang, L. Han, C. Xiang, H. Liu, and X. Li, "An indirect reinforcement learning based real-time energy management strategy via high-order Markov chain model for a hybrid electric vehicle," *Energy*, vol. 236, Dec. 2021, Art. no. 121337, doi: 10.1016/j.energy.2021.121337.

[21] G. Du, Y. Zou, X. Zhang, T. Liu, J. Wu, and D. He, "Deep reinforcement learning based energy management for a hybrid electric vehicle," *Energy*, vol. 201, Jun. 2020, Art. no. 117591, doi: 10.1016/j.energy.2020.117591.

[22] S. Lee, L. Xie, and D.-H. Choi, "Privacy-preserving energy management of a shared energy storage system for smart buildings: A federated deep reinforcement learning approach," *Sensors*, vol. 21, no. 14, p. 4898, Jul. 2021, doi: 10.3390/s21144898.

[23] B. Mbuwir, F. Ruelens, F. Spiessens, and G. Deconinck, "Battery energy management in a microgrid using batch reinforcement learning," *Energies*, vol. 10, no. 11, p. 1846, Nov. 2017, doi: 10.3390/en10111846.

[24] J.-G. Kim and B. Lee, "Automatic P2P energy trading model based on reinforcement learning using long short-term delayed reward," *Energies*, vol. 13, no. 20, p. 5359, Oct. 2020, doi: 10.3390/en13205359.

[25] Z. Zhang, A. Chong, Y. Pan, C. Zhang, and K. P. Lam, "Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning," *Energy Buildings*, vol. 199, pp. 472–490, Sep. 2019, doi: 10.1016/j.enbuild.2019.07.029.

[26] D. Azuatalam, W.-L. Lee, F. de Nijs, and A. Liebman, "Reinforcement learning for whole-building HVAC control and demand response," *Energy AI*, vol. 2, Nov. 2020, Art. no. 100020, doi: 10.1016/j.egyai.2020.100020.

[27] S. Brandi, M. S. Piscitelli, M. Martellacci, and A. Capozzoli, "Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings," *Energy Buildings*, vol. 224, Oct. 2020, Art. no. 110225, doi: 10.1016/j.enbuild.2020.110225.

[28] M. Wegener, L. Koch, M. Eisenbarth, and J. Andert, "Automated eco-driving in urban scenarios using deep reinforcement learning," *Transp. Res. C, Emerg. Technol.*, vol. 126, May 2021, Art. no. 102967, doi: 10.1016/j.trc.2021.102967.

[29] F. Tuchnitz, N. Ebell, J. Schlund, and M. Pruckner, "Development and evaluation of a smart charging strategy for an electric vehicle fleet based on reinforcement learning," *Appl. Energy*, vol. 285, Mar. 2021, Art. no. 116382, doi: 10.1016/j.apenergy.2020.116382.

[30] S. Lee and D.-H. Choi, "Energy management of smart home with home appliances, energy storage system and electric vehicle: A hierarchical deep reinforcement learning approach," *Sensors*, vol. 20, no. 7, p. 2157, Apr. 2020, doi: 10.3390/s20072157.

[31] R. Lian, J. Peng, Y. Wu, H. Tan, and H. Zhang, "Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle," *Energy*, vol. 197, Apr. 2020, Art. no. 117297, doi: 10.1016/j.energy.2020.117297.

[32] J. Zhou, S. Xue, Y. Xue, Y. Liao, J. Liu, and W. Zhao, "A novel energy management strategy of hybrid electric vehicle via an improved TD3 deep reinforcement learning," *Energy*, vol. 224, Jun. 2021, Art. no. 120118, doi: 10.1016/j.energy.2021.120118.

[33] J. Li and T. Yu, "A new adaptive controller based on distributed deep reinforcement learning for PEMFC air supply system," *Energy Rep.*, vol. 7, pp. 1267–1279, Nov. 2021, doi: 10.1016/j.egyr.2021.02.043.

[34] W. Zhang, J. Wang, Y. Liu, G. Gao, S. Liang, and H. Ma, "Reinforcement learning-based intelligent energy management architecture for hybrid construction machinery," *Appl. Energy*, vol. 275, Oct. 2020, Art. no. 115401, doi: 10.1016/j.apenergy.2020.115401.

[35] B. Zhang, W. Hu, J. Li, D. Cao, R. Huang, Q. Huang, Z. Chen, and F. Blaabjerg, "Dynamic energy conversion and management strategy for an integrated electricity and natural gas system with renewable energy: Deep reinforcement learning approach," *Energy Convers. Manage.*, vol. 220, Sep. 2020, Art. no. 113063, doi: 10.1016/j.enconman.2020.113063.

[36] P. Lissa, C. Deane, M. Schukat, F. Seri, M. Keane, and E. Barrett, "Deep reinforcement learning for home energy management system control," *Energy AI*, vol. 3, Mar. 2021, Art. no. 100043, doi: 10.1016/j.egyai.2020.100043.

[37] S. Zhong, X. Wang, J. Zhao, W. Li, H. Li, Y. Wang, S. Deng, and J. Zhu, "Deep reinforcement learning framework for dynamic pricing demand response of regenerative electric heating," *Appl. Energy*, vol. 288, Apr. 2021, Art. no. 116623, doi: 10.1016/j.apenergy.2021.116623.

[38] A. Kathirgamanathan, E. Mangina, and D. P. Finn, "Development of a soft actor critic deep reinforcement learning approach for harnessing energy flexibility in a large office building," *Energy AI*, vol. 5, Sep. 2021, Art. no. 100101, doi: 10.1016/j.egyai.2021.100101.

[39] M. Weigold, H. Ranzau, S. Schaumann, T. Kohne, N. Panten, and E. Abele, "Method for the application of deep reinforcement learning for optimised control of industrial energy supply systems by the example of a central cooling system," *CIRP Ann.*, vol. 70, no. 1, pp. 17–20, 2021, doi: 10.1016/j.cirp.2021.03.021.

[40] T. Schreiber, C. Netsch, M. Baranski, and D. Müller, "Monitoring data-driven reinforcement learning controller training: A comparative study of different training strategies for a real-world energy system," *Energy Buildings*, vol. 239, May 2021, Art. no. 110856, doi: 10.1016/j.enbuild.2021.110856.

[41] Z. Jiang, M. J. Risbeck, V. Ramamurti, S. Murugesan, J. Amores, C. Zhang, Y. M. Lee, and K. H. Drees, "Building HVAC control with reinforcement learning for reduction of energy cost and demand charge," *Energy Buildings*, vol. 239, May 2021, Art. no. 110833, doi: 10.1016/j.enbuild.2021.110833.

[42] X. Zhang, R. Lu, J. Jiang, S. H. Hong, and W. S. Song, "Testbed implementation of reinforcement learning-based demand response energy management system," *Appl. Energy*, vol. 297, Sep. 2021, Art. no. 117131, doi: 10.1016/j.apenergy.2021.117131.

[43] J. Sun, Z. Zhu, H. Li, Y. Chai, G. Qi, H. Wang, and Y. H. Hu, "An integrated critic-actor neural network for reinforcement learning with application of DERs control in grid frequency regulation," *Int. J. Electr. Power Energy Syst.*, vol. 111, pp. 286–299, Oct. 2019, doi: 10.1016/j.ijepes.2019.04.011.

[44] X. Wang, Y. Liu, J. Zhao, C. Liu, J. Liu, and J. Yan, "Surrogate model enabled deep reinforcement learning for hybrid energy community operation," *Appl. Energy*, vol. 289, May 2021, Art. no. 116722, doi: 10.1016/j.apenergy.2021.116722.

[45] M. Biemann, F. Scheller, X. Liu, and L. Huang, "Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control," *Appl. Energy*, vol. 298, Sep. 2021, Art. no. 117164, doi: 10.1016/j.apenergy.2021.117164.

[46] D. Wang, J. Liu, and D. Yao, "An energy-efficient distributed adaptive cooperative routing based on reinforcement learning in wireless multimedia sensor networks," *Comput. Netw.*, vol. 178, Sep. 2020, Art. no. 107313, doi: 10.1016/j.comnet.2020.107313.

[47] Z. Bing, C. Lemke, L. Cheng, K. Huang, and A. Knoll, "Energy-efficient and damage-recovery slithering gait design for a snake-like robot based on reinforcement learning and inverse reinforcement learning," *Neural Netw.*, vol. 129, pp. 323–333, Sep. 2020, doi: 10.1016/j.neunet.2020.05.029.

[48] Y. Zou, T. Liu, D. X. Liu, and F. C. Sun, "Reinforcement learning-based real-time energy management for a hybrid tracked vehicle," *Appl. Energy*, vol. 171, pp. 372–382, Jun. 2016, doi: 10.1016/j.apenergy.2016.03.082.

[49] X. Qiu, T. A. Nguyen, and M. L. Crow, "Heterogeneous energy storage optimization for microgrids," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1453–1461, May 2016, doi: 10.1109/TSG.2015.2461134.

[50] J. Cao and R. Xiong, "Reinforcement learning-based real-time energy management for plug-in hybrid electric vehicle with hybrid energy storage system," *Energy Proc.*, vol. 142, pp. 1896–1901, Dec. 2017, doi: 10.1016/j.egypro.2017.12.386.

[51] P. Kofinas, S. Doltsinis, A. I. Dounis, and G. A. Vouros, "A reinforcement learning approach for MPPT control method of photovoltaic sources," *Renew. Energy*, vol. 108, pp. 461–473, Aug. 2017, doi: 10.1016/j.renene.2017.03.008.

[52] S. Conti, G. Faraci, R. Nicolosi, S. A. Rizzo, and G. Schembra, "Battery management in a green fog-computing node: A reinforcement-learning approach," *IEEE Access*, vol. 5, pp. 21126–21138, 2017, doi: 10.1109/ACCESS.2017.2755588.

[53] R. F. Atallah, C. M. Assi, and J. Y. Yu, "A reinforcement learning technique for optimizing downlink scheduling in an energy-limited vehicular network," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4592–4601, Jun. 2017, doi: 10.1109/TVT.2016.2622180.

[54] J. Wu, H. He, J. Peng, Y. Li, and Z. Li, "Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus," *Appl. Energy*, vol. 222, pp. 799–811, Jul. 2018, doi: 10.1016/j.apenergy.2018.03.104.

[55] R. Xiong, J. Cao, and Q. Yu, "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle," *Appl Energy*, vol. 211, pp. 538–548, Feb. 2018, doi: 10.1016/j.apenergy.2017.11.072.

[56] X. Han, H. He, J. Wu, J. Peng, and Y. Li, "Energy management based on reinforcement learning with double deep Q-learning for a hybrid electric tracked vehicle," *Appl. Energy*, vol. 254, Nov. 2019, Art. no. 113708, doi: 10.1016/j.apenergy.2019.113708.

[57] Q. Zhou, J. Li, B. Shuai, H. Williams, Y. He, Z. Li, H. Xu, and F. Yan, "Multi-step reinforcement learning for model-free predictive energy management of an electrified off-highway vehicle," *Appl. Energy*, vol. 255, Dec. 2019, Art. no. 113755, doi: 10.1016/j.apenergy.2019.113755.

[58] J. Hofstetter, H. Bauer, W. Li, and G. Wachtmeister, "Energy and emission management of hybrid electric vehicles using reinforcement learning," *IFAC-PapersOnLine*, vol. 52, no. 29, pp. 19–24, 2019, doi: 10.1016/j.ifacol.2019.12.615.

[59] Y. Li, H. He, A. Khajepour, H. Wang, and J. Peng, "Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information," *Appl. Energy*, vol. 255, Dec. 2019, Art. no. 113762, doi: 10.1016/j.apenergy.2019.113762.

[60] X. Qi, Y. Luo, G. Wu, K. Boriboonsomsin, and M. Barth, "Deep reinforcement learning enabled self-learning control for energy efficient driving," *Transp. Res. C, Emerg. Technol.*, vol. 99, pp. 67–81, Feb. 2019, doi: 10.1016/j.trc.2018.12.018.

[61] M. Chu, H. Li, X. Liao, and S. Cui, "Reinforcement learning-based multiaccess control and battery prediction with energy harvesting in IoT systems," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2009–2020, Apr. 2019, doi: 10.1109/JIOT.2018.2872440.

[62] K. Li, W. Ni, M. Abolhasan, and E. Tovar, "Reinforcement learning for scheduling wireless powered sensor communications," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 2, pp. 264–274, Jun. 2019, doi: 10.1109/TGCN.2018.2879023.

[63] M. Chu, X. Liao, H. Li, and S. Cui, "Power control in energy harvesting multiple access system with reinforcement learning," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 9175–9186, Oct. 2019, doi: 10.1109/JIOT.2019.2928837.

[64] M. K. Sharma, A. Zappone, M. Assaad, M. Debbah, and S. Vassilaras, "Distributed power control for large energy harvesting networks: A multi-agent deep reinforcement learning approach," *IEEE Trans. Cognit. Commun. Netw.*, vol. 5, no. 4, pp. 1140–1154, Dec. 2019, doi: 10.1109/TCCN.2019.2949589.

[65] J. N. Forestieri and M. Farasat, "Integrative sizing/real-time energy management of a hybrid supercapacitor/undersea energy storage system for grid integration of wave energy conversion systems," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 8, no. 4, pp. 3798–3810, Dec. 2020, doi: 10.1109/JESTPE.2019.2926061.

[66] R. Wang, A. Yadav, E. A. Makled, O. A. Dobre, R. Zhao, and P. K. Varshney, "Optimal power allocation for full-duplex underwater relay networks with energy harvesting: A reinforcement learning approach," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 223–227, Feb. 2020, doi: 10.1109/LWC.2019.2948992.

[67] M. A. Abd-Elmagid, H. S. Dhillon, and N. Pappas, "A reinforcement learning framework for optimizing age of information in RF-powered communication systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4747–4760, Aug. 2020, doi: 10.1109/TCOMM.2020.2991992.

[68] H. Nie, Y. Chen, Y. Xia, S. Huang, and B. Liu, "Optimizing the post-disaster control of islanded microgrid: A multi-agent deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 153455–153469, 2020, doi: 10.1109/ACCESS.2020.3018142.

[69] A. Soares, D. Geysen, F. Spiessens, D. Ectors, O. De Somer, and K. Vanthournout, "Using reinforcement learning for maximizing residential self-consumption—Results from a field test," *Energy Buildings*, vol. 207, Jan. 2020, Art. no. 109608, doi: 10.1016/j.enbuild.2019.109608.

[70] H. Sun, Z. Fu, F. Tao, L. Zhu, and P. Si, "Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles," *J. Power Sources*, vol. 455, Apr. 2020, Art. no. 227964, doi: 10.1016/j.jpowsour.2020.227964.

[71] Z. Chen, H. Hu, Y. Wu, Y. Zhang, G. Li, and Y. Liu, "Stochastic model predictive control for energy management of power-split plug-in hybrid electric vehicles based on reinforcement learning," *Energy*, vol. 211, Nov. 2020, Art. no. 118931, doi: 10.1016/j.energy.2020.118931.

[72] B. E. Nyong-Bassey, D. Giaouris, C. Patsios, S. Papadopoulou, A. I. Papadopoulos, S. Walker, S. Voutetakis, P. Seferlis, and S. Gadoue, "Reinforcement learning based adaptive power pinch analysis for energy management of stand-alone hybrid energy storage systems considering uncertainty," *Energy*, vol. 193, Feb. 2020, Art. no. 116622, doi: 10.1016/j.energy.2019.116622.

[73] Z. Zhang, J. Boubin, C. Stewart, and S. Khanal, "Whole-field reinforcement learning: A fully autonomous aerial scouting method for precision agriculture," *Sensors*, vol. 20, no. 22, p. 6585, Nov. 2020, doi: 10.3390/s20226585.

[74] B. Xu, D. Rathod, D. Zhang, A. Yebi, X. Zhang, X. Li, and Z. Filipi, "Parametric study on reinforcement learning optimized energy management strategy for a hybrid electric vehicle," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114200, doi: 10.1016/j.apenergy.2019.114200.

[75] G. Faraci, A. Raciti, S. A. Rizzo, and G. Schembra, "Green wireless power transfer system for a drone fleet managed by reinforcement learning in smart industry," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114204, doi: 10.1016/j.apenergy.2019.114204.

[76] X. Qu, Y. Yu, M. Zhou, C.-T. Lin, and X. Wang, "Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 114030, doi: 10.1016/j.apenergy.2019.114030.

[77] H. He, J. Cao, and X. Cui, "Energy optimization of electric vehicle's acceleration process based on reinforcement learning," *J. Cleaner Prod.*, vol. 248, Mar. 2020, Art. no. 119302, doi: 10.1016/j.jclepro.2019.119302.

[78] H. Ding, F. Zhao, J. Tian, D. Li, and H. Zhang, "A deep reinforcement learning for user association and power control in heterogeneous networks," *Ad Hoc Netw.*, vol. 102, May 2020, Art. no. 102069, doi: 10.1016/j.adhoc.2019.102069.

[79] A. K. Lakshmanan, R. E. Mohan, B. Ramalingam, A. V. Le, P. Veerajagadeshwar, K. Tiwari, and M. Ilyas, "Complete coverage path planning using reinforcement learning for tetromino based cleaning and maintenance robot," *Autom. Construct.*, vol. 112, Apr. 2020, Art. no. 103078, doi: 10.1016/j.autcon.2020.103078.

[80] P. S. Banerjee, S. N. Mandal, D. De, and B. Maiti, "RL-sleep: Temperature adaptive sleep scheduling using reinforcement learning for sustainable connectivity in wireless sensor networks," *Sustain. Comput., Informat. Syst.*, vol. 26, Jun. 2020, Art. no. 100380, doi: 10.1016/j.suscom.2020.100380.

[81] S. Jeon, J. Kim, J. Ahn, and H. Cha, "Optimizing discharge efficiency of reconfigurable battery with deep reinforcement learning," *IEEE Trans. Comput. Design Integr. Circuits Syst.*, vol. 39, no. 11, pp. 3893–3905, Nov. 2020, doi: 10.1109/TCAD.2020.3012230.

[82] A.-S. Mohammed and M. Petr, "Reinforcement learning-based distributed BESS management for mitigating overvoltage issues in systems with high PV penetration," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 2980–2994, Jul. 2020, doi: 10.1109/TSG.2020.2972208.

[83] J. Kim, S. Baek, Y. Choi, J. Ahn, and H. Cha, "Hydrone: Reconfigurable energy storage for UAV applications," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 11, pp. 3686–3697, Nov. 2020, doi: 10.1109/TCAD.2020.3013052.

[84] S. Hasanvand, M. Rafiei, M. Gheisarnejad, and M.-H. Khooban, "Reliable power scheduling of an emission-free ship: Multiobjective deep reinforcement learning," *IEEE Trans. Transport. Electrific.*, vol. 6, no. 2, pp. 832–843, Jun. 2020, doi: 10.1109/TTE.2020.2983247.

[85] E. Oh and H. Wang, "Reinforcement-learning-based energy storage system operation strategies to manage wind power forecast uncertainty," *IEEE Access*, vol. 8, pp. 20965–20976, 2020, doi: 10.1109/access.2020.2968841.

[86] F. S. Gorostiza and F. M. Gonzalez-Longatt, "Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5039–5050, Nov. 2020, doi: 10.1109/TSG.2020.2996274.

[87] F. Zhu, Z. Yang, F. Lin, and Y. Xin, "Decentralized cooperative control of multiple energy storage systems in urban railway based on multiagent deep reinforcement learning," *IEEE Trans. Power Electron.*, vol. 35, no. 9, pp. 9368–9379, Sep. 2020, doi: 10.1109/TPEL.2020.2971637.

[88] D. A. Temesgene, M. Miozzo, D. Gunduz, and P. Dini, "Distributed deep reinforcement learning for functional split control in energy harvesting virtualized small cells," *IEEE Trans. Sustain. Comput.*, vol. 6, no. 4, pp. 626–640, Oct. 2021, doi: 10.1109/TSUSC.2020.3025139.

[89] H. Lee and S. W. Cha, "Reinforcement learning based on equivalent consumption minimization strategy for optimal control of hybrid electric vehicles," *IEEE Access*, vol. 9, pp. 860–871, 2021, doi: 10.1109/ACCESS.2020.3047497.

[90] Z. Yang, F. Zhu, and F. Lin, "Deep-reinforcement-learning-based energy management strategy for supercapacitor energy storage systems in urban rail transit," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1150–1160, Feb. 2021, doi: 10.1109/TITS.2019.2963785.

[91] T. M. Aljohani, A. Ebrahim, and O. Mohammed, "Real-time metadata-driven routing optimization for electric vehicle energy consumption minimization using deep reinforcement learning and Markov chain model," *Electric Power Syst. Res.*, vol. 192, Mar. 2021, Art. no. 106962, doi: 10.1016/j.epsr.2020.106962.

[92] X. Cao, W. Xu, X. Liu, J. Peng, and T. Liu, "A deep reinforcement learning-based on-demand charging algorithm for wireless rechargeable sensor networks," *Ad Hoc Netw.*, vol. 110, Jan. 2021, Art. no. 102278, doi: 10.1016/j.adhoc.2020.102278.

[93] B. Huang, X. Liu, S. Wang, L. Pan, and V. Chang, "Multi-agent reinforcement learning for cost-aware collaborative task execution in energy-harvesting D2D networks," *Comput. Netw.*, vol. 195, Aug. 2021, Art. no. 108176, doi: 10.1016/j.comnet.2021.108176.

[94] X. Liu, A. Fotouhi, and D. J. Auger, "Formula-E race strategy development using distributed policy gradient reinforcement learning," *Knowl.-Based Syst.*, vol. 216, Mar. 2021, Art. no. 106781, doi: 10.1016/j.knosys.2021.106781.

[95] S. Totaro, I. Boukas, A. Jonsson, and B. Cornélusse, "Lifelong control of off-grid microgrid with model-based reinforcement learning," *Energy*, vol. 232, Oct. 2021, Art. no. 121035, doi: 10.1016/j.energy.2021.121035.

[96] G. Zhang, W. Hu, D. Cao, W. Liu, R. Huang, Q. Huang, Z. Chen, and F. Blaabjerg, "Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach," *Energy Convers. Manage.*, vol. 227, Jan. 2021, Art. no. 113608, doi: 10.1016/j.enconman.2020.113608.

[97] Z. Yang, X. Ma, L. Xia, Q. Zhao, and X. Guan, "Reinforcement learning for fluctuation reduction of wind power with energy storage," *Results Control Optim.*, vol. 4, Sep. 2021, Art. no. 100030, doi: 10.1016/j.rico.2021.100030.

[98] S. M. Sultan, M. Waleed, J.-Y. Pyun, and T.-W. Um, "Energy conservation for Internet of Things tracking applications using deep reinforcement learning," *Sensors*, vol. 21, no. 9, p. 3261, May 2021, doi: 10.3390/s21093261.

[99] F. Sangoleye, N. Irtija, and E. E. Tsiropoulou, "Smart energy harvesting for Internet of Things networks," *Sensors*, vol. 21, no. 8, p. 2755, Apr. 2021, doi: 10.3390/s21082755.

[100] W. Li, H. Cui, T. Nemeth, J. Jansen, C. Ünlübayir, Z. Wei, X. Feng, X. Han, M. Ouyang, H. Dai, X. Wei, and D. U. Sauer, "Cloud-based health-conscious energy management of hybrid battery systems in electric vehicles with deep reinforcement learning," *Appl. Energy*, vol. 293, Jul. 2021, Art. no. 116977, doi: 10.1016/j.apenergy.2021.116977.

[101] M. Chen, W. Liu, T. Wang, A. Liu, and Z. Zeng, "Edge intelligence computing for mobile augmented reality with deep reinforcement learning approach," *Comput. Netw.*, vol. 195, Aug. 2021, Art. no. 108186, doi: 10.1016/j.comnet.2021.108186.

[102] Q. Guo, O. Angah, Z. Liu, and X. Ban, "Hybrid deep reinforcement learning based eco-driving for low-level connected and automated vehicles along signalized corridors," *Transp. Res. C, Emerg. Technol.*, vol. 124, Mar. 2021, Art. no. 102980, doi: 10.1016/j.trc.2021.102980.

[103] J. Wu, Z. Wei, W. Li, Y. Wang, Y. Li, and D. U. Sauer, "Battery thermal- and health-constrained energy management for hybrid electric bus based on soft actor-critic DRL algorithm," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 3751–3761, Jun. 2021, doi: 10.1109/TII.2020.3014599.

[104] H. Lee and S. W. Cha, "Energy management strategy of fuel cell electric vehicles using model-based reinforcement learning with data-driven model update," *IEEE Access*, vol. 9, pp. 59244–59254, 2021, doi: 10.1109/ACCESS.2021.3072903.

[105] W. Wang, Z. Lv, X. Lu, Y. Zhang, and L. Xiao, "Distributed reinforcement learning based framework for energy-efficient UAV relay against jamming," *Intell. Converged Netw.*, vol. 2, no. 2, pp. 150–162, Jun. 2021, doi: 10.23919/ICN.2021.0010.

[106] K. Li, W. Ni, and F. Dressler, "LSTM-characterized deep reinforcement learning for continuous flight control and resource allocation in UAV-assisted sensor network," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4179–4189, Mar. 2022, doi: 10.1109/JIOT.2021.3102831.

[107] G. Du, Y. Zou, X. Zhang, L. Guo, and N. Guo, "Heuristic energy management strategy of hybrid electric vehicle based on deep reinforcement learning with accelerated gradient optimization," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 4, pp. 2194–2208, Dec. 2021, doi: 10.1109/TTE.2021.3088853.

[108] L. Xi, L. Zhou, Y. Xu, and X. Chen, "A multi-step unified reinforcement learning method for automatic generation control in multi-area interconnected power grid," *IEEE Trans. Sustain. Energy*, vol. 12, no. 2, pp. 1406–1415, Apr. 2021, doi: 10.1109/TSTE.2020.3047137.

[109] Y. Wang, X. Lin, and M. Pedram, "A near-optimal model-based control algorithm for households equipped with residential photovoltaic power generation and energy storage systems," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 77–86, Jan. 2016, doi: 10.1109/TSTE.2015.2467190.

[110] C. X. Jiang, Z. X. Jing, X. R. Cui, T. Y. Ji, and Q. H. Wu, "Multiple agents and reinforcement learning for modelling charging loads of electric taxis," *Appl. Energy*, vol. 222, pp. 158–168, Jul. 2018, doi: 10.1016/j.apenergy.2018.03.164.

[111] W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2192–2203, Jun. 2018, doi: 10.1109/TNNLS.2018.2801880.

[112] X. Lu, X. Xiao, L. Xiao, C. Dai, M. Peng, and H. V. Poor, "Reinforcement learning-based microgrid energy trading with a reduced power plant schedule," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10728–10737, Dec. 2019, doi: 10.1109/JIOT.2019.2941498.

[113] W. Hu, Z. Sun, Y. Zhang, and Y. Li, "Joint manufacturing and onsite microgrid system control using Markov decision process and neural network integrated reinforcement learning," *Proc. Manuf.*, vol. 39, pp. 1242–1249, Jan. 2019, doi: 10.1016/j.promfg.2020.01.345.

[114] H. Hua, Y. Qin, C. Hao, and J. Cao, "Optimal energy management strategies for energy internet via deep reinforcement learning approach," *Appl. Energy*, vol. 239, pp. 598–609, Apr. 2019, doi: 10.1016/j.apenergy.2019.01.145.

[115] H. Tan, H. Zhang, J. Peng, Z. Jiang, and Y. Wu, "Energy management of hybrid electric bus based on deep reinforcement learning in continuous state and action space," *Energy Convers. Manage.*, vol. 195, pp. 548–560, Sep. 2019, doi: 10.1016/j.enconman.2019.05.038.

[116] N. Wang, W. Xu, W. Shao, and Z. Xu, "A Q-cube framework of reinforcement learning algorithm for continuous double auction among microgrids," *Energies*, vol. 12, no. 15, p. 2891, Jul. 2019, doi: 10.3390/en12152891.

[117] Y. Wu, H. Tan, J. Peng, H. Zhang, and H. He, "Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus," *Appl. Energy*, vol. 247, pp. 454–466, Aug. 2019, doi: 10.1016/j.apenergy.2019.04.021.

[118] S. Zhou, Z. Hu, W. Gu, M. Jiang, and X.-P. Zhang, "Artificial intelligence based smart energy community management: A reinforcement learning approach," *CSEE J. Power Energy Syst.*, vol. 5, no. 1, pp. 1–10, 2019, doi: 10.17775/CSEEJPES.2018.00840.

[119] F. L. D. Silva, C. E. H. Nishida, D. M. Roijers, and A. H. R. Costa, "Coordination of electric vehicle charging through multiagent reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2347–2356, May 2020, doi: 10.1109/TSG.2019.2952331.

[120] Y. Cao, D. Li, Y. Zhang, and X. Chen, "Joint optimization of delay-tolerant autonomous electric vehicles charge scheduling and station battery degradation," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8590–8599, Sep. 2020, doi: 10.1109/JIOT.2020.2992133.

[121] F. Alfaverh, M. Denai, and Y. Sun, "Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management," *IEEE Access*, vol. 8, pp. 39310–39321, 2020, doi: 10.1109/ACCESS.2020.2974286.

[122] E. Samadi, A. Badri, and R. Ebrahimpour, "Decentralized multi-agent based energy management of microgrid using reinforcement learning," *Int. J. Electr. Power Energy Syst.*, vol. 122, Nov. 2020, Art. no. 106211, doi: 10.1016/j.ijepes.2020.106211.

[123] Y. Zhang, Z. Zhang, Q. Yang, D. An, D. Li, and C. Li, "EV charging bidding by multi-DQN reinforcement learning in electricity auction market," *Neurocomputing*, vol. 397, pp. 404–414, Jul. 2020, doi: 10.1016/j.neucom.2019.08.106.

[124] X. Tang, M. Li, X. Lin, and F. He, "Online operations of automated electric taxi fleets: An advisor-student reinforcement learning framework," *Transp. Res. C, Emerg. Technol.*, vol. 121, Dec. 2020, Art. no. 102844, doi: 10.1016/j.trc.2020.102844.

[125] M. Roesch, C. Linder, R. Zimmermann, A. Rudolf, A. Hohmann, and G. Reinhart, "Smart grid for industry using multi-agent reinforcement learning," *Appl. Sci.*, vol. 10, no. 19, p. 6900, Oct. 2020, doi: 10.3390/app10196900.

[126] F. Chang, T. Chen, W. Su, and Q. Alsafasfeh, "Control of battery charging based on reinforcement learning and long short-term memory networks," *Comput. Electr. Eng.*, vol. 85, Jul. 2020, Art. no. 106670, doi: 10.1016/j.compeleceng.2020.106670.

[127] Y. Shang, W. Wu, J. Guo, Z. Ma, W. Sheng, Z. Lv, and C. Fu, "Stochastic dispatch of energy storage in microgrids: An augmented reinforcement learning approach," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114423, doi: 10.1016/j.apenergy.2019.114423.

[128] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, and K. Li, "Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4513–4521, Sep. 2020, doi: 10.1109/TSG.2020.2986333.

[129] Y. Gao, J. Yang, M. Yang, and Z. Li, "Deep reinforcement learning based optimal schedule for a battery swapping station considering uncertainties," *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 5775–5784, Sep. 2020, doi: 10.1109/TIA.2020.2986412.

[130] S. Lee and D.-H. Choi, "Federated reinforcement learning for energy management of multiple smart Homes with distributed energy resources," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 488–497, Jan. 2022, doi: 10.1109/TII.2020.3035451.

[131] M. Shin, D.-H. Choi, and J. Kim, "Cooperative management for PV/ESS-enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3493–3503, May 2020, doi: 10.1109/TII.2019.2944183.

[132] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2427–2439, May 2020, doi: 10.1109/TSG.2019.2955437.

[133] G. C. Chasparis and C. Lettner, "Reinforcement-learning-based optimization for day-ahead flexibility extraction in battery pools," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 13351–13358, 2020, doi: 10.1016/j.ifacol.2020.12.170.

[134] P. Wu, J. Partridge, and R. Bucknall, "Cost-effective reinforcement learning energy management for plug-in hybrid fuel cell and battery ships," *Appl. Energy*, vol. 275, Oct. 2020, Art. no. 115258, doi: 10.1016/j.apenergy.2020.115258.

[135] D. Qiu, Y. Ye, D. Papadaskalopoulos, and G. Strbac, "A deep reinforcement learning method for pricing electric vehicles with discrete charging levels," *IEEE Trans. Ind. Appl.*, vol. 56, no. 5, pp. 5901–5912, Sep. 2020, doi: 10.1109/TIA.2020.2984614.

[136] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2751–2762, Apr. 2020, doi: 10.1109/JIOT.2019.2957289.

[137] Y. Liu, D. Zhang, and H. B. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE J. Power Energy Syst.*, vol. 6, no. 3, pp. 572–582, 2020, doi: 10.17775/CSEE-JPES.2019.02890.

[138] H. S. V. S. K. Nunna, A. Sesetti, A. K. Rathore, and S. Doolla, "Multiagent-based energy trading platform for energy storage systems in distribution systems with interconnected microgrids," *IEEE Trans. Ind. Appl.*, vol. 56, no. 3, pp. 3207–3217, May 2020, doi: 10.1109/TIA.2020.2979782.

[139] F. Wang, J. Gao, M. Li, and L. Zhao, "Autonomous PEV charging scheduling using dyna-Q reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12609–12620, Nov. 2020, doi: 10.1109/TVT.2020.3026004.

[140] C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107048, doi: 10.1016/j.ijepes.2021.107048.

[141] W. Kolodziejczyk, I. Zoltowska, and P. Cichosz, "Real-time energy purchase optimization for a storage-integrated photovoltaic system by deep reinforcement learning," *Control Eng. Pract.*, vol. 106, Jan. 2021, Art. no. 104598, doi: 10.1016/j.conengprac.2020.104598.

[142] T. A. Nakabi and P. Toivanen, "Deep reinforcement learning for energy management in a microgrid with flexible demand," *Sustain. Energy, Grids Netw.*, vol. 25, Mar. 2021, Art. no. 100413, doi: 10.1016/j.segan.2020.100413.

[143] H. Zang and J. Kim, "Reinforcement learning based peer-to-peer energy trade management using community energy storage in local energy market," *Energies*, vol. 14, no. 14, p. 4131, Jul. 2021, doi: 10.3390/en14144131.

[144] S. Abedi, S. W. Yoon, and S. Kwon, "Battery energy storage control using a reinforcement learning approach with cyclic time-dependent Markov process," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107368, doi: 10.1016/j.ijepes.2021.107368.

[145] G. Han, S. Lee, J. Lee, K. Lee, and J. Bae, "Deep-learning- and reinforcement-learning-based profitable strategy of a grid-level energy storage system for the smart grid," *J. Energy Storage*, vol. 41, Sep. 2021, Art. no. 102868, doi: 10.1016/j.est.2021.102868.

[146] G. Muriithi and S. Chowdhury, "Optimal energy management of a grid-tied solar PV-battery microgrid: A reinforcement learning approach," *Energies*, vol. 14, no. 9, p. 2700, May 2021, doi: 10.3390/en14092700.

[147] W. J. Yun, S. Jung, J. Kim, and J.-H. Kim, "Distributed deep reinforcement learning for autonomous aerial eVTOL mobility in drone taxi applications," *ICT Exp.*, vol. 7, no. 1, pp. 1–4, Mar. 2021, doi: 10.1016/j.icte.2021.01.005.

[148] L. Yin and S. Li, "Hybrid metaheuristic multi-layer reinforcement learning approach for two-level energy management strategy framework of multi-microgrid systems," *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104326, doi: 10.1016/j.engappai.2021.104326.

[149] T. Yang, L. Zhao, W. Li, and A. Y. Zomaya, "Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning," *Energy*, vol. 235, Nov. 2021, Art. no. 121377, doi: 10.1016/j.energy.2021.121377.

[150] D. Cao, W. Hu, X. Xu, Q. Wu, Q. Huang, Z. Chen, and F. Blaabjerg, "Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices," *J. Mod. Power Syst. Clean Energy*, vol. 9, no. 5, pp. 1101–1110, 2021, doi: 10.35833/MPCE.2020.000557.

[151] Q. Wu, Q. Feng, Y. Ren, Q. Xia, Z. Wang, and B. Cai, "An intelligent preventive maintenance method based on reinforcement learning for battery energy storage systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 8254–8264, Dec. 2021, doi: 10.1109/TII.2021.3066257.

[152] B. Huang and J. Wang, "Deep-reinforcement-learning-based capacity scheduling for PV-battery storage system," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2272–2283, May 2021, doi: 10.1109/TSG.2020.3047890.

[153] F. Zhang, Q. Yang, and D. An, "CDDPG: A deep-reinforcement-learning-based approach for electric vehicle charging control," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3075–3087, Mar. 2021, doi: 10.1109/jiot.2020.3015204.

[154] S. Li, W. Hu, D. Cao, T. Dragičević, Q. Huang, Z. Chen, and F. Blaabjerg, "Electric vehicle charging management based on deep reinforcement learning," *J. Mod. Power Syst. Clean Energy*, vol. 10, pp. 1–12, Jun. 2021, doi: 10.35833/MPCE.2020.000460.

[155] J. Jin, L. Hao, Y. Xu, J. Wu, and Q.-S. Jia, "Joint scheduling of deferrable demand and storage with random supply and processing rate limits," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5506–5513, Nov. 2021, doi: 10.1109/TAC.2020.3046555.

[156] L. Yan, X. Chen, J. Zhou, Y. Chen, and J. Wen, "Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5124–5134, Nov. 2021, doi: 10.1109/TSG.2021.3098298.

[157] S. Wang, L. Du, X. Fan, and Q. Huang, "Deep reinforcement scheduling of energy storage systems for real-time voltage regulation in unbalanced LV networks with high PV penetration," *IEEE Trans. Sustain. Energy*, vol. 12, no. 4, pp. 2342–2352, Oct. 2021, doi: 10.1109/TSTE.2021.3092961.

[158] D. Wang, B. Liu, H. Jia, Z. Zhang, J. Chen, and D. Huang, "Peer-to-peer electricity transaction decision of user-side smart energy system based on SARSA reinforcement learning method," *CSEE J. Power Energy Syst.*, pp. 1–11, 2020, doi: 10.17775/CSEEJPES.2020.03290.

[159] H. Shuai, F. Li, H. Pulgar-Painemal, and Y. Xue, "Branching dueling Q-network-based online scheduling of a microgrid with distributed energy storage systems," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5479–5482, Nov. 2021, doi: 10.1109/TSG.2021.3103405.

[160] S. Peng and Q. Feng, "Reinforcement learning with Gaussian processes for condition-based maintenance," *Comput. Ind. Eng.*, vol. 158, Aug. 2021, Art. no. 107321, doi: 10.1016/j.cie.2021.107321.

[161] H. Aaltonen, S. Sierla, R. Subramanya, and V. Vyatkin, "A simulation environment for training a reinforcement learning agent trading a battery storage," *Energies*, vol. 14, no. 17, p. 5587, Sep. 2021, doi: 10.3390/en14175587.

[162] Q. Xie, D. Shin, N. Chang, and M. Pedram, "Joint charge and thermal management for batteries in portable systems with hybrid power sources," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 4, pp. 611–622, Apr. 2016, doi: 10.1109/TCAD.2015.2474410.

[163] M. Mendil, A. De Domenico, V. Heiries, R. Caire, and N. Hadjsaid, "Battery-aware optimization of green small cells: Sizing and energy management," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 3, pp. 635–651, Sep. 2018, doi: 10.1109/TGCN.2018.2829344.

[164] Y. D. Ko, "An efficient integration of the genetic algorithm and the reinforcement learning for optimal deployment of the wireless charging electric tram system," *Comput. Ind. Eng.*, vol. 128, pp. 851–860, Feb. 2019, doi: 10.1016/j.cie.2018.10.045.

[165] J. Duan, Z. Yi, D. Shi, C. Lin, X. Lu, and Z. Wang, "Reinforcement-learning-based optimal control of hybrid energy storage systems in hybrid AC–DC microgrids," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5355–5364, Sep. 2019, doi: 10.1109/TII.2019.2896618.

[166] J. J. Kelly and P. G. Leahy, "Optimal investment timing and sizing for battery energy storage systems," *J. Energy Storage*, vol. 28, Apr. 2020, Art. no. 101272, doi: 10.1016/j.est.2020.101272.

[167] E. Oh, "Reinforcement-learning-based virtual energy storage system operation strategy for wind power forecast uncertainty management," *Appl. Sci.*, vol. 10, no. 18, p. 6420, Sep. 2020, doi: 10.3390/app10186420.

[168] D. Sidorov, D. Panasetsky, N. Tomin, D. Karamov, A. Zhukov, I. Muftahov, A. Dreglea, F. Liu, and Y. Li, "Toward zero-emission hybrid AC/DC power systems with renewable energy sources and storages: A case study from lake Baikal region," *Energies*, vol. 13, no. 5, p. 1226, Mar. 2020, doi: 10.3390/en13051226.

[169] B. C. Phan and Y. C. Lai, "Control strategy of a hybrid renewable energy system based on reinforcement learning approach for an isolated microgrid," *Appl. Sci.*, vol. 9, no. 19, p. 4001, Sep. 2019, doi: 10.3390/app9194001.

[170] Y. Sui and S. Song, "A multi-agent reinforcement learning framework for lithium-ion battery scheduling problems," *Energies*, vol. 13, no. 8, p. 1982, Apr. 2020, doi: 10.3390/en13081982.

[171] Y. Teng, M. Yan, D. Liu, Z. Han, and M. Song, "Distributed learning solution for uplink traffic control in energy harvesting massive machine-type communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 485–489, Apr. 2020, doi: 10.1109/LWC.2019.2959583.
[172] Z. Zhu, F. Zeng, G. Qi, Y. Li, H. Jie, and N. Mazur, "Power system structure optimization based on reinforcement learning and sparse constraints under DoS attacks in cloud environments," *Simul. Model. Pract. Theory*, vol. 110, Jul. 2021, Art. no. 102272, doi: 10.1016/j.simpat.2021.102272.
[173] H. Zou, J. Tao, S. K. Elsayed, E. E. Elattar, A. Almalaq, and M. A. Mohamed, "Stochastic multi-carrier energy management in the smart islands using reinforcement learning and unscented transform," *Int. J. Electr. Power Energy Syst.*, vol. 130, Sep. 2021, Art. no. 106988, doi: 10.1016/j.ijepes.2021.106988.
[174] S. Tsianikas, N. Yousefi, J. Zhou, M. D. Rodgers, and D. Coit, "A storage expansion planning framework using reinforcement learning and simulation-based optimization," *Appl. Energy*, vol. 290, May 2021, Art. no. 116778, doi: 10.1016/j.apenergy.2021.116778.
[175] J. Li, H. Wang, H. He, Z. Wei, Q. Yang, and P. Igic, "Battery optimal sizing under a synergistic framework with DQN-based power managements for the fuel cell hybrid powertrain," *IEEE Trans. Transport. Electrific.*, vol. 8, no. 1, pp. 36–47, Mar. 2022, doi: 10.1109/TTE.2021.3074792.
[176] A. Unagar, Y. Tian, M. A. Chao, and O. Fink, "Learning to calibrate battery models in real-time with deep reinforcement learning," *Energies*, vol. 14, no. 5, p. 1361, Mar. 2021, doi: 10.3390/en14051361.
[177] M. Kim, K. Kim, J. Kim, J. Yu, and S. Han, "State of charge estimation for lithium ion battery based on reinforcement learning," *IFAC-PapersOnLine*, vol. 51, no. 28, pp. 404–408, 2018, doi: 10.1016/j.ifacol.2018.11.736.
[178] *EU Directive 2019/944 of the European Parliament and of the Council of 5 June 2019 on Common Rules for the Internal Market for Electricity*. Accessed: Sep. 22, 2021. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2019/944/oj
[179] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI gym," Jun. 2016, *arXiv:1606.01540*.
[180] C. Bamford, "Griddly: A platform for AI research in games," *Softw. Impacts*, vol. 8, May 2021, Art. no. 100066, doi: 10.1016/j.simpa.2021.100066.
[181] L. Avila, M. De Paula, M. Trimboli, and I. Carlucho, "Deep reinforcement learning approach for MPPT control of partially shaded PV systems in smart grids," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106711, doi: 10.1016/j.asoc.2020.106711.
[182] T. Schreiber, S. Eschweiler, M. Baranski, and D. Müller, "Application of two promising reinforcement learning algorithms for load shifting in a cooling supply system," *Energy Buildings*, vol. 229, Dec. 2020, Art. no. 110490, doi: 10.1016/j.enbuild.2020.110490.
[183] R. Henry and D. Ernst, "Gym-ANM: Reinforcement learning environments for active network management tasks in electricity distribution systems," *Energy AI*, vol. 5, Sep. 2021, Art. no. 100092, doi: 10.1016/j.egyai.2021.100092.

RAKSHITH SUBRAMANYA (Graduate Student Member, IEEE) was born in 1988. He received the B.E. degree in electronics and instrumentation engineering from the Malnad College of Engineering, Hassan, India, in 2009, and the M.S. degree in software design and engineering from the Manipal Institute of Technology, Manipal, India, in 2014. He is currently pursuing the Ph.D. degree in electrical and automation engineering with Aalto University, Finland.

He has more than ten years of experience in the energy and healthcare industry and has spent the last five years as a Research and Development Engineer. His research interests include the IoT, artificial intelligence, and web technologies.

SEPPO A. SIERLA (Senior Member, IEEE) was born in 1977. He received the M.Sc. degree in technology with the major in embedded systems and the D.Sc. degree in technology from the Helsinki University of Technology, in 2003 and 2007, respectively, and the title of docent in software design for industrial automation from the Aalto University School of Electrical Engineering, in 2013.

From 2003 to 2011, he was a Research Scientist with Aalto University, Finland; and the Helsinki University of Technology. Since 2011, he has been a University Lecturer with Aalto University. His research interests include applications of simulation, machine learning, and ICT technologies to the energy sector.

Dr. Sierla is/will be the Vice-Chair of the IEEE Finland Section, from 2022 to 2023. Before that, he was the Section Treasurer, from 2020 to 2021. He is the Finance Chair of several IEEE conferences, such as INDIN2019, MLSP2020, ISGT2021 Europe, ISIT2022, and ISIE2023. Since 2020, he has been the Co-Chair of the "Industrial Digitalization, Digital Twins in Industrial Applications" Track in the IEEE INDIN Conference Series.

VALERIY VYATKIN (Fellow, IEEE) received the Dr.Sc. and Ph.D. degrees in applied computer science from the Taganrog Radio Engineering Institute, Taganrog, Russia, in 1992 and 1999, respectively, the Dr.Eng. degree from the Nagoya Institute of Technology, Nagoya, Japan, in 1999, and the Habilitation degree from the Ministry of Science and Technology of Sachsen-Anhalt, in 2002.

He is on joint appointment as the Chair of dependable computations and communications with the Luleå University of Technology, Luleå, Sweden; and a Professor of information technology in automation with Aalto University, Finland. He is also the Co-Director of the International Research Laboratory Computer Technologies, ITMO University, Saint Petersburg, Russia. Previously, he was a Visiting Scholar with Cambridge University, Cambridge, U.K.; and had permanent appointments with the University of Auckland, New Zealand; Martin Luther University, Germany; as well as in Japan and Russia. His research interests include dependable distributed automation and industrial informatics; software engineering for industrial automation systems; artificial intelligence; distributed architectures; and multiagent systems in various industries: smart grid, material handling, building management systems, data centers, and reconfigurable manufacturing.

Dr. Vyatkin was awarded the Andrew P. Sage Award for the Best IEEE Transactions paper, in 2012. He has been the Chair of the IEEE IES Technical Committee on Industrial Informatics for two terms, from 2016 to 2019.