Technical Journal, University of Engineering and Technology (UET) Taxila, Pakistan Vol. 28 No. 4-2023
ISSN:1813-1786 (Print) 2313-7770 (Online)

Energy Management Optimization Through Conventional and AI Approaches for Efficient Electrical Energy Utilization

H. Hyder[1], K. H. Ali[2], A. Tahir[3]

[1,2]*Sukkur IBA University, Sindh, Pakistan*
[3]*Environment and Sustainability Institute (ESI), University of Exeter, Penryn Campus, Cornwall, TR10 9FE, United Kingdom*

[2]haiderali@iba-suk.edu.pk

Abstract- Reinforcement Learning (RL) is a promising technique for scheduling and planning storage systems in microgrids, which are small-scale power networks that can operate independently or in coordination with the main grid. RL can enhance the utilization of local renewable energy sources and reduce the operational costs of microgrids. In this comprehensive study and review the state-of-the-art applications of RL for Microgrid Energy Management (MEM), focusing on battery storage systems is discussed. This work also identify the main challenges, limitations, and future directions in this domain. Furthermore, this article present a novel benchmark algorithm that compares the performance of RL with mixed-integer linear programming (MILP), a widely used optimization technique. This complete work provides a valuable insight into the current status and future prospects of RL for MEM.

Keywords- Reinforcement Learning, Scheduling, Optimization, Energy Management, Q Learning, Microgrid.

I. INTRODUCTION

Energy companies worldwide are currently focused on reducing energy costs and managing load shedding through the following strategies:

- Maximum utilization of Renewable energy resources.
- Less dependency on the main grid.
- Use the storage system optimally.
- To sell power to the main grid when the utility tariff is high.
- Selection of generating units.
- Load-Management.

A microgrid is a comprehensive system that may include one or more renewable energy sources, a storage system, a charge controller, and an inverter [1]. The microgrid operates in two modes: Off-grid mode, functioning independently without connection to the main power grid, and On-grid mode, connected to the main power grid. In the On-grid mode, it can either draw power from or supply power to the grid as needed [2]. Examples of energy management in these modes include Autonomous Building energy management for Off-grid operation and scheduling of energy storage systems for a grid-tied microgrid in On-grid operation. For autonomous building energy management, control systems may require the integration of smart meters, Internet of Things (IoT) devices, and smart lighting. On the other hand, the management system in On-grid mode may necessitate the utilization of algorithms for selecting the optimal renewable energy source and mix, forecasting load and demand, determining the appropriate size for the storage system, scheduling storage devices, and strategic siting..

In the introduction section, we will delve into two pivotal dimensions of Microgrid Energy Management – Autonomous Building Energy Management and Grid-Tied Microgrid Energy Management. We will underscore the traditional approaches that have historically been applied in these domains. Additionally, an exploration of contemporary scheduling methods employed for Microgrid optimization will be provided. The main contribution of this research paper lies in a comprehensive review of Reinforcement Learning (RL) applications for Microgrid Energy Management, with a particular emphasis on battery storage systems. The paper not only identifies challenges and limitations but also introduces a novel benchmark algorithm, comparing RL performance against the widely used mixed-integer linear programming (MILP) technique. This study aims to offer valuable insights into the current landscape and future prospects of RL in Microgrid Energy Management.

1.1. Factors Influencing Building Energy Consumption and Optimization Strategies

It is well established in the literature that buildings consume around 40% of the world's total energy usage [3]. A notable strategy for minimizing overall energy consumption and cutting costs involves the adoption of Building Energy Management. Primary energy-consuming functions within buildings encompass water heating, ventilation and air conditioning (HVAC), and lighting. [3]. The utilization of energy in buildings is influenced by various factors such as building construction and design, insulation, building materials, and the direction of the sun shining on the windows during the day. By controlling these factors, the peak load can be reduced, leading to better energy management of the building. Moreover, optimizing the main sources of energy consumption can benefit both the providers and consumers of energy [4]. Customers pay less in terms of utility bills, and more importantly, the reduction of hazardous gas emissions benefits the environment [5].

1.1.1. Conventional Techniques for Autonomous Building Energy Management

HVAC needs a controller to turn it on and off to regulate the room temperature. Thermostat is an example of a control system used to switch HVAC systems [6]. Sensors of temperature, humidity and load, monitor the building constraints, which benefit the control system of HVAC. Internet of Things (IOT) facilitates communication between different devices of the control unit. There are lightning and energy intensive devices within the building such as television, washing machines, dryer, fridge, microwave oven, electric stoves that significantly affect the overall consumption of building. Studies showed that the use of smart household appliances reduces the intake of energy up to 15 % in the past few years [3]. The main challenge in this regard is variation of energy demand throughout the day. This leads to develop different tariff rates in a single day to change the behavior of customers. For example if the tariff in a specific hour of the day is high, consumers may switch its usage on that hour which has a low tariff rate. Smart meters for household users have gained popularity in recent years to optimize their usage by themselves. The contemporary electric vehicle sector within the automotive industry has a direct impact on residential electricity demand. This is due to the requirement for power from the utility provider to facilitate the charging of electric vehicle batteries, as explored [3]. .In Sweden additional 10 % daily electricity usage added as a load demand of the building. This significant increase is managed by using vehicle to grid (V2G) technology. It works by selling electricity to its respective grid when the car is not being used [3]. In addition, it may be managed as an electric vehicle charged at night time or when there is less demand for electricity.

1.2. Grid Tied Microgrid Energy Management

Microgrid structures are classified according to its connection with other grid-types of generating sources, voltage level of distribution system, peak load, energy production, generation capacity, number of customers served [7]. There are different techniques to optimize these different designs of microgrids in terms of its classification. For example, some methods apply to manage the production side and some are useful on the demand side. This section mainly focuses on the energy management of grid tied microgrids, which can divide into two subtypes according to its size and utility grid connection type. These types are large-grid connected to a microgrid and small grid connected to a microgrid. Both of these types of microgrids can operate and optimize independently or in conjunction with the main grid to get maximum benefit out of grid-connected microgrids. In the field of microgrid Energy Management, there are two main types of techniques: iterative and heuristic methods. Several scientists have proposed heuristic methods in their works [7-8]. This type finds a cost effective solution out of a large set of possible solutions. This requires less computational power than other optimization tools [7]. In addition to heuristic methods, other common approaches to solve the energy management problem in microgrid architecture include Linear Programming (LP), Integer Linear Programming (ILP), and Mixed Integer Linear Programming (MILP). However, it is a known fact that not all energy optimization methods, whether classical or novel, can always find the best solution due to the complexity of the problem. For instance, in the case of ILP and MILP, all or some unknown variables need to be in the integer form. If it is difficult or not possible to obtain integer variables, simple linear programming or other approaches may be used to manage energy. Linear programming and integer programming can be useful in many practical situations, such as in non-deterministic environments.

1.2.1. Approaches of Grid Tied Microgrid Energy Management

Several classical methods exist for achieving optimal management in On-grid microgrids [7]. One notable approach involves determining the power generation mix and selection, a process integral to the planning and implementation phases of the microgrid. In these stages, design engineers evaluate the available power sources, demand-side requirements, and appropriate electric power supplies within the designated installation area.. Critical decisions regarding initial investment, storage system sizing, and forecasting peak load demand are made to establish cost-effectiveness criteria. Strategic considerations for the overall

microgrid system are taken into account to achieve one or multiple objectives such as cost-effectiveness, environmental impact, and high reliability. In a summary, strategic issues for the overall system of microgrid installation are considered to achieve one or multiple objectives. Like, cost effectiveness, Environmental impact and high reliability. In literature, papers such as [8-9] play a crucial role in the strategic and planning level for power sources selection and sizing of microgrids. Another paper [10] contributes in this area and uses mixed integer nonlinear programming (MINLP) for selecting and sizing different power generating units to improve overall system efficiency of energy utilization to reduce the cost of implementation and operations.

The other traditional approach used in common to make grid tied microgrids cost efficient is "Sitting". Sitting problems deal with power source allocation and layout of power lines by maintaining quality constraints of the whole system. In these sitting methods, planning of the number of customers and potential increase of customers in that particular area are considered. In this regard, papers such as [11] provide cost effective solutions by decomposing the planning phase and annual reliability problem into two parts. Then by using software named Versatile Energy Resource (VERA) solve the overall optimization problem. In paper [11], the author also tries to forecast load demand by using weather conditions.

In the area of Sitting type Optimization, there is some work on multi-objective Optimization as well. In the paper [12], a two stage multi-objective efficient process for planning a microgrid is used to achieve the 1st objective. In the 1st stage microgrid area is proposed by using loss sensitivity factor. In the second stage Pareto algorithm is used to find the optimal area out of 1st stage proposed areas. The functions that applied in this paper to solve the problem were real power loss, load voltage deviation and investment cost per annum. At the end by using fuzzy logic, the tradeoff optimal solution was also evaluated in this paper as a second objective [12].

1.3. Scheduling Methods

These methodologies can also find application in managing energy for autonomous buildings and grid-tied microgrids. Recent research papers employ various scheduling algorithms to address the energy management challenges in both on-grid and off-grid scenarios. Contemporary scheduling methods predominantly rely on computational approaches, with mixed-integer linear programming and machine learning algorithms being the most prevalent choices. One of the novel methods of machine learning regarding optimization of microgrids is RL. Deep learning is also used in combination with RL and MIDP. The novelty of scheduling techniques are due to its application on both existing power grids and new electrical power projects, which are in the pipeline. Scheduling mainly focused to minimize operational costs by using optimization tools to plan and control the storage devices or generators attached to the microgrid. In the scheduling area, multi-objective optimization is getting more intention to achieve two or more objectives simultaneously. The example of multi-objective problems are optimization of cost, environment impact and quality assurance.

In this study, we will focus on the optimization of storage devices through scheduling techniques for both standalone and grid-connected microgrid systems, emphasizing on RL, deep neural networks, and forecasting techniques.

II. Model And Non-Model Based Scheduling / Planning

This planning and scheduling problem mainly consists of Model based and non-model based approaches. These are further divided into value based and policy based approaches, which are the types of RL under the umbrella of Artificial intelligence. The example of value based or value iteration approach in RL is Q learning, State Action Reward State Action (SARSA). Q learning is an off policy algorithm, which builds its value function or general policy independently off the policy [9]. It is used to gather the data [8]. SARSA is an on-policy approach in RL, which tries to evaluate or improve its policy that was already made to make decisions. On the other hand, the example of Policy based or actor-critic is Reinforce, Cross entropy method. It is not necessary that the Optimization problem can be handled by only one type of approach (model based, Non-model based) every time. It is the fact that in different kinds of Scenarios and architecture of microgrids, optimization may be done by different or combination of methods. This is because of the nature of the problem, sometimes it is due to stochastic behavior of the system or due to deterministic and non-deterministic characteristics of the environment. For example, apply a model free approach, which leads to give a model. After that, model based RL, value based RL or policy based RL or combination of all may be applied to achieve high performance. In a summary, these approaches also rely on each other as shown in Figure 1.

The model and non-model based approaches have individual strengths. Their respective strengths may be used in a combination to achieve high optimization targets.

It also depends on the approach in which the agent has access to the environment. It can either openly have access to a generative model of the environment in the form of a trainer with opportunity to gather data in a flexible way (infinitely), or it can identify the environment only through a (probably finite) number of paths within

that environment [13]. In the latter case, a model-based approach is possible but requires an extra step for model estimation. It is important to know that learning the model can share the hidden-state representations with a value-based approach [13].

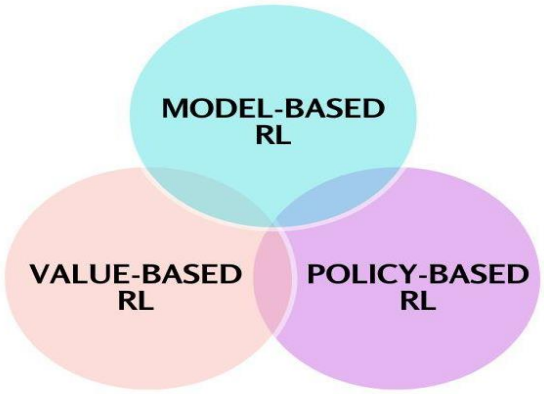

Figure 2 Venn diagram of different types of RL algorithms [2]

The model-based approach works best when combined with a forecasting algorithm, but this can be computationally intensive. When the agent has access to the model, a pure planning approach would be the most optimal, but it may not always be practical in real-time decision-making situations due to time limitations. In such cases, computing the decision-making policy in real-time becomes challenging [14].

In certain jobs, creating a policy or value function is easier to learn. But for other tasks, understanding the environment model might be more straightforward because of the task's specific structure (less complexity or more uniformity). The choice of the best approach also relies on the structure of the model policy or value function. Following are examples for better understanding:

In a maze where the agent can see everything, it can easily understand how its actions impact the next state. Even with just a few examples, the agent can figure out how the maze works. For instance, it learns that it gets stuck when trying to pass through a wall or moves forward when going in an open direction. Once the agent understands the maze rules, it can use a planning algorithm very effectively [13].

In another example, imagine an agent trying to cross a busy road with random events occurring everywhere. The best strategy might be to move forward unless there is an obstacle right in front of the agent. In this case, a model-free approach, which does not require building a detailed model of the environment, would work better. Trying to create a model and plan within it would be harder because of the unpredictable nature of the environment. The model can have various possible outcomes for the same set of actions due to its non-deterministic behavior [13].

2.1. Fundamentals of Q Learning

Q Learning is a method used in Reinforcement Learning to find the best actions to take in different situations. It works by using a function called Q, which represents the expected reward for each action in each state [3]. The main goal is to maximize this Q function. To do this, we create a table called the Q table, which helps us find the best action for each state. The Q table is continually updated as the agent learns from its experiences using a process called Q-Learning. This helps the agent make better decisions by selecting actions that lead to the highest expected rewards in each situation [14]. This equation permit to start solving these Morkov's decision processes (MDPs). The Bellman equations are crucial in Reinforcement Learning because they help us understand how RL algorithms work. These equations allow us to express the values of states and the values of their subsequent states. This means that if we know the value of the next state, we can easily calculate the value of the current state. This opens up possibilities for iterative approaches to calculate the value of each state, as we can use the values of future states to find the values of current states.

Having the Bellman equations is beneficial as it enables us to compute optimal policies and train RL agents. Initially, the agent explores the environment, updating its Q-Table as it learns from experiences. Once the Q-Table is ready, the agent switches to exploitation mode, making better decisions based on the knowledge gained [13-14].

In the beginning, we explore the environment to discover and gather information, which we then use to update the Q-Table [2].The reward function in Q learning helps in this regard. In the beginning, the sum of rewards for the whole time interval show random behaviour while Q-Table updates. But, after so many iterations/episodes, the sum of rewards for the complete time period starts converging. As shown in Figure 2. Once the initial exploration phase is completed and the Q-Table is updated, the agent will shift its focus towards exploiting the environment by selecting the best possible actions [7]. It converges the below mentioned Equation 1 as well. Equation 1 below represents the Q table.

$$Q(s,a) = Q(s,a) + \alpha(Reward + \gamma \times max(Q(s',a)) - Q(s,a)) \quad (1)$$

The proposed algorithm applied for the energy management of standalone or grid-tied microgrids to decrease the net energy cost. It can be done by considering the future prices, using the current system information during the training period. The training period required all information regarding the data such as PV production, Load demand. This data is forecasted by using different algorithms such as neural networks or LSTM. In the training session, the convergence of the Q table is done by tuning the hyper parameters of the RL such as Exploration vs Exploitation (ε), learning rate (∝),

and discount factor (γ). Once the algorithm learns pedagogy, policy has made followed by the data in real time. Nevertheless, the main problem is that the training that is being made on forecasted data sometimes fails or not up to the mark as change occurred at real time due to uncertain change in weather or load demand. Therefore, if training is done on forecasted data the forecasting should be appropriate or very near to real time data otherwise best optimization is not possible. This is the fact that in optimization problems the training performed off-line may not give proper results when tested On-line. So, forecasting real data is an actual challenge in this field. This challenge may be addressed by applying neural networks techniques or performing optimization tasks directly in real time. Both will discuss in the next section.

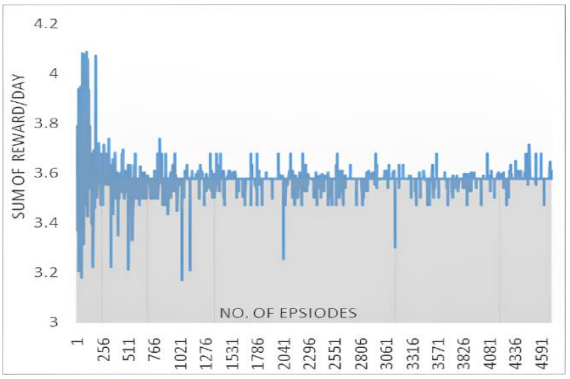

Figure 2 Sum of Rewards / day in each Episode

III. Literature Review Storage System Scheduling By Q Learning

Previous research on microgrid optimization, employing RL or Q learning, has predominantly centered around the scheduling of storage systems, particularly batteries, to reduce the overall cost of the microgrid. The scheduling of the battery involves the selection of optimal actions at each time interval throughout the day, such as charging, discharging, or remaining idle. Many researchers utilize RL to implement the following policies on the grid model, aiming to achieve cost reduction objectives..

1. The renewable energy for example PV has a priority to fulfill the demand of load first. If it is not enough then battery or utility grid or combination of all resources used to fulfil the demand.
2. Battery may charge from the PV directly. It can be charged from the utility grid as well.
3. It is also possible that at high tariffs batteries may discharge into the utility grid as a feed in tariff to earn money.

In the Paper [15] applied the two-step-ahead RL to optimize the battery of individual Customers who have a goal to utilize the storage system more during the high demand of electricity or at the time interval when tariff is high. This work [15] used local wind generators as a Renewable energy source to decrease the purchase of electricity from the external grid. Available wind power prediction; use by RL to train and get the optimal actions of the battery. The learning mechanism of RL may also be checked through different wind profiles achieved under different weather conditions. After training, a two-step-ahead decision was made to decide the optimal actions of the battery to save the utility cost at customers' end. This paper [15] applies RL for energy management successfully as it compares its methodology to other optimization approaches. However, it applied on a smaller scale leads to limiting the work. In addition, developed, two-step-ahead RL algorithms provide a way of energy management for the intelligent customers. Those who want to reduce their utility bills can achieve this by using Reinforcement Learning to understand the unpredictable behavior of their environment. They can then use this knowledge to schedule their battery usage two hours in advance from the current time.

In the paper [16], the author also used a stochastic, model free approach using Markov's Decision process (MDP) to optimize the building energy. This research proposes the optimal strategy of the battery actions by applying RL. In [16] uses extra renewable energy (PV) to charge the battery of the microgrid and discharge it when the utility tariff is high. As, the battery is only charged from the extra PV available, so due to uncertain behaviour of Renewable generation, the state of the charge of the battery becomes a non-deterministic variable which makes the problem more practical. In addition, the operator unlike paper [15] of the microgrid to give maximum benefit to all the customers who are getting electricity from this provider can use this suggested algorithm. The limitation of this paper is not charging the battery from the external grid. As there are certain situations, in which the main grid tariff is low and it is optimum to charge the battery and utilize this charging to fulfil the load demand when utility tariff is high. The other limitation in this work [15] is regarding the forecasting of PV a load profile. The training data (Load and PV) in RL, obtained from past years. However, due to climate changes and uncertain behavior of weather on current time can result in less optimization than the desired one. Another challenge in this paper [15] like [14] is the requirement of high computational power because of intensive training.

Another appreciated work in the field of Energy management of a microgrid is a paper published in the year 2017-18 [17]. It tries to solve the demand of high computational need for RL algorithms by using Regressor. The neural network or Regressor approximates the Q function, which helps the computation. In [17], the author used batch

reinforcement learning to solve the optimization of microgrid problems which also increase the computational efficiency especially when the dimension of the space variables are large. Batch Reinforcement Learning aims to discover or learn the most optimal control policy from a set of training data provided in batch format. Once the policy is learned, it can be applied to the current real-time environment to make decisions and take actions accordingly. In the paper [18], also proposed an RL based algorithm to schedule the operational modes of the battery attached to a grid tied microgrid system in an efficient way. The idea behind batch RL is regarding the observable state information, which contains input data such as PV, Load demand, Tariff rate, Time interval. This data information sent to the batch of RL, the agent learns to extract features of the system. These features are necessary for the learning process because they contain the state information of the system/environment. In this study, the Fitted Q Iteration (FQI) algorithm is employed to derive a closed-loop policy that depends on the current state of the system. The policy is learned from a batch of four tuples, consisting of state, action, next state, and the corresponding cost. The primary objective of the control policy is to minimize electricity costs by maximizing the utilization of the battery and locally produced renewable energy, which is photovoltaic (PV) energy in this particular research [17]. In Q or FQI, learning there is a major challenge regarding dimensionality of the system state vs Actions. As, Q table consists of state vs actions which may consist of large or continuous spaces (State/Actions).It may lead to the requirement of high computational power and time. This paper [17] used one of the regressor technique named as extremely randomized trees to solve this problem. Then this method, compared by using the commercial available optimization solver software CPLEX (OPL) by formalizing the problem as mixed integer linear programing (MILP). It [17] suggested that the FQI algorithm is 19% less efficient than MILP. Similarly [15], [16], this work [17] only charges the battery from the PV. However, microgrids are connected to the main grid and used only when the load demand is not fulfilled by the battery or Renewable source. It may limit the cost saving and flexibility to apply on different architectures of the current microgrid. The scope of this work [17] is also limited as it uses a backup controller, which is unknown to learning agent (outside the FQI) due to respecting the battery and microgrid constraints such as battery maximum and minimum charging, discharging capacity. For example if the agent receive optimal actions of the battery after learning and want to dispatch them in real time, but this back up controller may influence to change the best possible action. It may result in decreasing efficiency in terms of overall operating cost of the microgrid. The papers [15-17] discussed above needs forecasted data such as renewable energy production, load demand and utility tariff at each time step. So the agent is trained by using this data either by Q learning or FQI. The optimal actions achieved from this training may be dispatched in real time. The main challenge for above-mentioned approaches suggested in papers [15-17] needs appropriate forecasted data profiles otherwise the actions obtained from this training showed inefficient or less cost savings on real time.

Algorithm 1 Fitted Q-iteration with function approximation (Regressor) [17]
Initialize: Discount factor γ, control period T
Generate samples $F(s_l , a_l , s_{l',Cl}) \vert\rightarrow l = F = \{0, \ldots , F .. 1\}$ *Where,* F the number of batches of tuples. **"c"** is SOC of the battery which is dependent on control actions of the battery.
$s_{l'} \leftarrow (\text{SOC}, s_{x'})$ observed exogenous component of the state which are Non-Controllable $s_{x'} = \{S_x^{load}, S_x^{Pv}\}$
Initialize Q_T to zero for all state-action pairs, $Q_T \leftarrow 0$
For $K = T-1,....0$ do
For $l = F-1,....0$ do
$Q_{k,l} \leftarrow c_l + \min_{a \in A} Q_{k+l}(s_{l'}, a)$ *Where,* Actions=A=a1, a2, a3, …
end for
use a regression algorithm to build Q_k from $T_S = \{((s_l, a_l), Q_{k,l}), l = \{0....., F-1\}\}$
end for
Output $\rightarrow Q^* = Q_0$

The large difference between forecasted and real profile may divert the optimality. To address this problem one of the research [17] in the field of Q learning suggested a solution. This strategy of Q learning, applied directly in real time. So, it does not require Predicted day- ahead information. The paper [17] suggested, giving average optimal cost for a whole year rather than a single day. The proposed algorithm [17] suggested a simple Q table initialization procedure, in which each value of Q table is set to an instantaneous reward obtained with gamma=0 at time step 0.

In the beginning of the 1st day, the Q table is initialized at time step 0 by the technique mentioned above before the actual Q learning process starts. After the initialization of Q table at time zero, the Q table will be updated on day 1 using regular Q learning mechanism and parameters (alpha, epsilon, gamma) by the help of

real time data profiles (generated Renewable, Load demand).Like other training algorithms of Q learning this approach is not repetitive over same training data. Rather, it suggest and dispatch the actions of the battery after one iteration. As, in real time delay on dispatch of battery actions are undesirable.

The learning of the agent, which may be very less, passes to the second day by updating the Q table. The Q table updated again by the same process as of day one. This process continues from day one to the last day of the year. In the beginning, days of the year the agents learning abilities are low as it is exploring the environment in one go (One iteration). However, as days progress, Q table start exploiting the actions vs each state and converged. Maybe the convergence time, achieved after 90 days (3 months) but after that it gives the best optimal actions for the battery and in real time. Before convergence of Q table the battery action (e.g. 1st, 2nd -------90th day) may not be the best in terms of cost saving but can save at least some cost. However, the claimed aim of this algorithm [18] to save average annual cost without forecasting data profile may be achieved. The drawback of this technique of RL is annual based optimization rather than a single day contra to paper [14-16]discussed above.

Algorithm 2 Q-learning [15][16][18]
Initialize $Q(s,a) \rightarrow 0$ in case of off line Q learning [10][11] *OR:* Online Q learning [13] Initialize $Q(s,a)$ by total discounted rewards with $\gamma = 0$
Initialize learning parameters with ∝, γ, ϵ,
for each time step t do
Determine possible action set A_{st}
Obtain greedy action a_t
Select action at from A_{st} by policy π
Take action at and observe $r(s_t, a_t), s_t + t$
Update $Q(s_t, a_t)$
$t+1 \rightarrow t_{NEXT}$
$s_{t+1} \rightarrow s'$
end for

Another work [14] to deal with real time energy management of the battery actions by using RL in combination of neural networks contributed in this area published in June 2019. This study focused on the real-time scheduling of a microgrid, taking into account the uncertainties related to load demand, renewable energy generation, and electricity prices. The main goal was to minimize the daily operating cost using a Markov Decision Process (MDP) framework. To solve this MDP, a deep reinforcement learning (DRL) approach was developed.

In the DRL approach, a deep feed-forward neural network was designed to approximate the optimal action-value function. The deep Q-network (DQN) algorithm was then applied to train this neural network [18]. By taking the state of the microgrid as inputs, the suggested approach generated real-time outputs, enabling efficient and cost-effective decision-making.

The researcher also compared its approach by developing the problem in YALMIP toolbox, using mixed integer programming and then solved via a built-in solver named "BMIBNB" to get the best generation schedules [18]. He claimed his approach (DRL) is 2.2% less efficient than the other one.

Algorithm 3 DQN Algorithm [19]
Initialize : Q-network $Q(s,a;\vec{\theta}_0)$ with random parameters $\vec{\theta}_0$ *Where:* $Q(s,a;\vec{\theta}_0)$ Denote the approximate of the optimal action-value function of Q. And: Where, $\vec{\theta}_0$, represents the set of all connection weights of the neural network
for episode = 1, do
Initialize the state (s_0)
for t = 1,do
Select an action at using the *e*-greedy policy $\pi(s)$
Execute action a_t and observe reward r then go to Next state $(st+1)$
Store transition $(st, at, rt, st+1)$ in D
Sample random mini batch of transitions $(sj, aj, rj, sj+1)$ from D
Set $r_j + \gamma \max_a Q(s_j+1, a'; \theta_{i-1} \mid s_j, a_j)$ for terminal s_j+1 And: Set $yj = rj$ for non-terminal $sj+1$
Perform a gradient descent step on $(yi - Q(s,a;\vec{\theta}))^2$
end for end for

IV. Comparison of Mixed Integer Linear Programming (MILP) With RL (Q Learning)

In this section, we compared MILP with RL. Fig. 3 shows Load, PV & Tariff profiles per hour of the day to establish a benchmark, the data either for training and assumed real data are the same. The

data (Load, PV and Tariff) as shown in Fig. 3 is taken from [20]. MILP and RL both show optimization with respect to daily cost. While the daily cost achieved through MILP and RL methods compared with the cost achieved when there is no optimizer used in the microgrid system. The non-optimized cost is attained at min instant reward by taking random actions of the battery (without training). The graph in Figure 4 and 5 show that the hourly cost and imported power from the main grid is much lower in case of MILP and RL method than the technique at min instant reward (Non optimized). While there is approximately no difference in terms of daily average cost between MILP and RL techniques. The hourly cost achieved by MILP and RL may be different as shown in Figure 4 but at the end of the day total average cost per day is the same (both MILP & RL). In the literature, it has claimed that MILP is more efficient than RL. However, the data (PV and load demand) set for training and assumed real time are not the same. Here, in this work we assumed both forecasted and real time data are the same to compare MILP and RL.

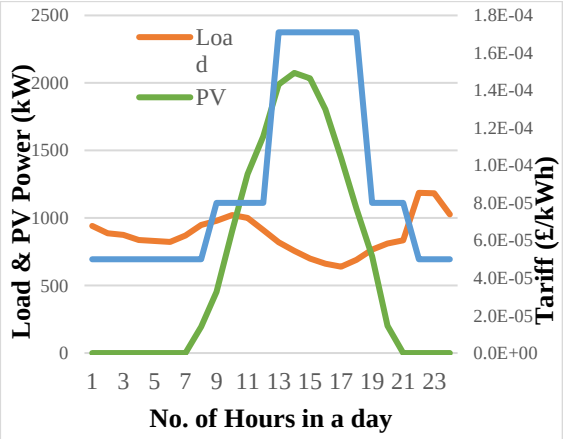

Figure 3 Load, PV & Tariff profiles per

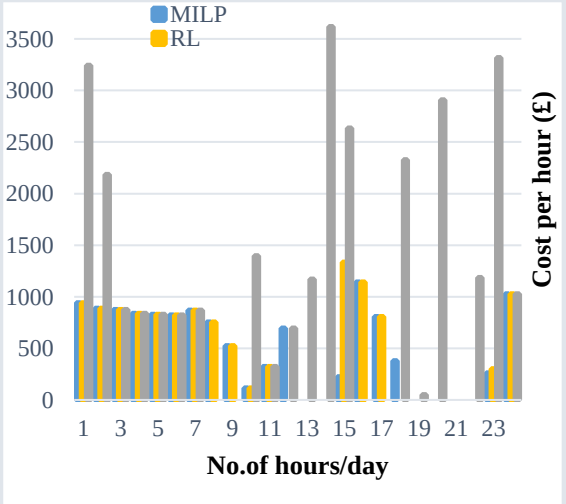

Figure 4 Comparison of Cost per hour of the day between MILP, RL & Min Instant Reward

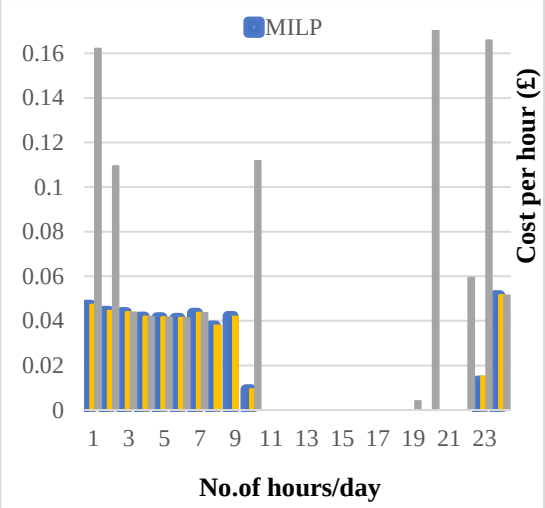

Figure 5 Comparison of Imported Power from utility grid between MILP, RL & Min Instant Reward hour of the day.

V. Limitations & Future Directions

1. Most of the work done in this area tried to optimize the battery to save overall cost of the microgrid. RL may also apply to other parts of the microgrid to manage it efficiently.
2. It is essential to investigate additional applications of Reinforcement Learning in the realm of Microgrid Energy Management, including endeavors to enhance the longevity of batteries or storage systems.
3. The existing gap in achieving multiple objectives through RL in energy management, specifically in the realm of Multi-Objective RL. This area requires focused attention, as addressing it has the potential to yield valuable and fruitful results. In the near future, Microgrid Energy Management should consider incorporating alternative algorithms such as Multi-Agent RL, Meta-RL, or Multi-Task approaches. When comparing RL with other optimization techniques like Mixed Integer Linear Programming, results showed that RL is less efficient than MILP. There is a need to explore more to see in which system, architecture or model RL show better results than MILP.
4. The RL in combination with other optimization methods like MILP can be used to solve different energy management problems for example cost saving. Combination of both algorithms may provide better optimal solutions or results.
5. In future, RL approach should be tested on different data sets (Predicted and assumed real) to investigate its performance in comparison with other well-known optimization approaches. The different data sets may be

generated randomly by developing some function in Matlab.

VI. CONCLUSION

The realm of Microgrid Energy Management is experiencing rapid growth, with Reinforcement Learning (RL) emerging as a promising tool for optimizing Microgrids and reducing expenses. The efficient scheduling and planning of storage systems within Microgrids, facilitated by RL, can lead to significant cost savings and increased reliance on renewable energy sources. The unique contribution of this literature review lies in its comparison of RL with the mixed-integer linear programming (MILP) benchmark algorithm, offering novel insights to the field.

However, the implementation of RL for Microgrid optimization presents challenges. Notably, the need for extensive datasets to effectively train RL models poses a primary obstacle. Moreover, the intricate decision-making processes within Microgrids make identifying optimal policies using RL challenging.

Moving forward, the development of efficient RL algorithms for Microgrid optimization requires continued research and exploration. Addressing these challenges is crucial for fully realizing the benefits of RL in enhancing Microgrid operations and achieving cost-effective, sustainable energy management.

Future research endeavors should prioritize the refinement of advanced RL techniques and the resolution of existing challenges. In summary, this study underscores the pivotal role of RL in Microgrid Energy Management and its potential to contribute to a sustainable energy future.

REFERENCES

[1] Suwi, O., & Justo, J. J. (2024). Comprehensive discussions on energy storage devices: modeling, control, stability analysis with renewable energy resources in microgrid and virtual power plants. In Modelling and Control Dynamics in Microgrid Systems with Renewable Energy Resources (pp. 139-177). Academic Press.

[2] Rani, P., Parkash, V., & Sharma, N. K. (2024). Technological aspects, utilization and impact on power system for distributed generation: A comprehensive survey. Renewable and Sustainable Energy Reviews, 192, 114257.

[3] Nagy, Z., Henze, G., Dey, S., Arroyo, J., Helsen, L., Zhang, X., & Bernstein, A. (2023). Ten questions concerning reinforcement learning for building energy management. Building and Environment, 110435.

[4] Youssef, H., Kamel, S., Hassan, M. H., Yu, J., & Safaraliev, M. (2024). A smart home energy management approach incorporating an enhanced northern goshawk optimizer to enhance user comfort, minimize costs, and promote efficient energy consumption. International Journal of Hydrogen Energy, 49, 644-658.

[5] Salam, I. U., Yousif, M., Numan, M., & Billah, M. (2024). Addressing the Challenge of Climate Change: The Role of Microgrids in Fostering a Sustainable Future-A Comprehensive Review. Renewable Energy Focus, 100538.
Meimand, M., & Jazizadeh, F. (2024). A personal touch to demand response: An occupant-centric control strategy for HVAC systems using personalized comfort models. Energy and Buildings, 303, 113769.

[6] Shahgholian, G. (2021). A brief review on microgrids: Operation, applications, modeling, and control. International Transactions on Electrical Energy Systems, 31(6), e12885.

[7] Chreim, B., Esseghir, M., & Merghem-Boulahia, L. (2024). Recent sizing, placement, and management techniques for individual and shared battery energy storage systems in residential areas: A review. Energy Reports, 11, 250-260.

[8] Sandelic, M., Peyghami, S., Sangwongwanich, A., & Blaabjerg, F. (2022). Reliability aspects in microgrid design and planning: Status and power electronics-induced challenges. Renewable and Sustainable Energy Reviews, 159, 112127.

[9] Ouramdane, O., Elbouchikhi, E., Amirat, Y., & Sedgh Gooya, E. (2021). Optimal sizing and energy management of microgrids with vehicle-to-grid technology: A critical review and future trends. Energies, 14(14), 4166.

[10] Cabral, C., Andiappan, V., Aviso, K., & Tan, R. (2021). Equipment size selection for optimizing polygeneration systems with reliability aspects. Energy, 234, 121302.

[11] Wood, M., Ogliari, E., Nespoli, A., Simpkins, T., & Leva, S. (2023). Day Ahead Electric Load Forecast: A Comprehensive LSTM-EMD Methodology and Several Diverse Case Studies. Forecasting, 5(1), 297-314.

[12] Zhang, M., & Li, Y. (2020). Multi-objective optimal reactive power dispatch of power systems by combining classification-based multi-objective evolutionary algorithm and integrated decision making. IEEE Access, 8, 38198-38209.

[13] V. François-Lavet, "Contributions to deep reinforcement learning and its applications in smartgrids," Univ. Liège, p. 177, 2017, [Online]. Available: http://hdl.handle.net/2268/214216

[14] M. Castronovo, "Offline Policy-search in

Bayesian Reinforcement Learning," no. March, p. 115, 2017, [Online]. Available: http://hdl.handle.net/2268/208421

[15] Perera, A. T. D., & Kamalaruban, P. (2021). Applications of reinforcement learning in energy systems. Renewable and Sustainable Energy Reviews, 137, 110618.

[16] C. Essayeh, M. Raiss El-Fenni, and H. Dahmouni, "Cost-Effective Energy Usage in a Microgrid Using a Learning Algorithm," Wirel. Commun. Mob. Comput., vol. 2018, 2018, doi: 10.1155/2018/9106430.

[17] B. V. Mbuwir, F. Ruelens, F. Spiessens, and G. Deconinck, "Battery energy management in a microgrid using batch reinforcement learning," Energies, vol. 10, no. 11, pp. 1–19, 2017, doi: 10.3390/en10111846.

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

[19] Y. Ji, J. Wang, J. Xu, X. Fang, and H. Zhang, "Real-time energy management of a microgrid using deep reinforcement learning," Energies, vol. 12, no. 12, 2019, doi:0.3390/en1.

[20] A. Khawaja Haider, Moh. Abusara, A. Ali Tahir, and S. Das. 2023. "Dual-layer Q-Learning strategy for energy management of battery storage in grid-connected microgrids" *Energies* 16, no. 3, 2023: 1334. https://doi.org/10.3390/en16031334