International Journal of Multiphysics
Volume 18, No. 4, 2024
ISSN: 1750-9548

A Reinforcement Learning Framework for Decentralized Decision-Making in Smart Energy Systems

Atef Gharbi[1*], Mohamed Ayari[2,3], Akil Elkamel[1], Mahmoud Salaheldin Elsayed[4], Zeineb Klai[4], Nuha Khedhiri[1]

Department of Information Systems, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia[1]
Department of Information Technology, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia[2]
SYSCOM Laboratory, National Engineering School of Tunis, University of Tunis El-Manar, Tunis 1068, Tunisia[3]
Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia[4]

*Corresponding author: E-mail address: atef.gharbi@nbu.edu.sa

Abstract

The complexity of modern smart grids decentralized energy systems and renewable energy sources has increased, requiring advanced energy management solutions. The paper presents a framework for reinforcement learning for decentralized energy management in smart grids. Based on production, consumption, and storage dynamics, the proposed model adapts unit costs to the individual prosumers' energy strategies. Meanwhile, external production operators (EPOs) have dynamically adjusted pricing in response to energy shortages and surpluses throughout the system. Through simulation, the framework demonstrates that actors with different energy profiles can independently design an optimized strategy, reducing the need for external energy supplies and stabilizing costs throughout the system. The research demonstrated the scalability and robustness of decentralized learning in energy management efficiency and adaptation and contributed to the development of smart grids.

Keywords: Smart Grid, Reinforcement Learning, Decentralized Energy Management, Dynamic Pricing, Prosumers, External Production Operator (EPO).

1. Introduction

Modern smart grids (SG) must take innovative approaches to energy management, as renewable energy is rapidly being used and energy systems become more complex. Smart grids facilitate decentralized energy systems by integrating different prosumers (producers and consumers) and optimizing resource allocation. However, these dynamic and decentralized systems face challenges balancing energy supply and demand and maintaining efficiency. Based on energy shortages and interactions with external production companies (EPOs), this paper examines how actors can adapt their cost-effective strategy over time. Using the proposed model, each actor optimizes its behavior according to production, consumption, and storage dynamics. In reinforcement learning, actors iteratively improve their unit cost estimates to reflect the actual cost of energy and reduce dependence on external providers. By dynamically adjusting its price to meet the entire system's energy deficits or surpluses, external production operators create an interactive price ecosystem. The decision-making process of decentralized actors and the energy supply mechanism of this dual-layer learning system interact and show the potential of adaptive pricing models for balance and efficiency.

Reinforcement learning (RL) has become a powerful tool for optimizing smart grid operations to address challenges such as decentralized decision-making, demand-supply balance, and dynamic pricing. Its applications include energy management, load scheduling, demand response, and market trading in a variety of smart grid-related areas. Below, we summarize the major advances made in the RL-based approach to smart grid systems.

[1] In the Smart Grid application, we critically examined the safe reinforcement learning techniques and stressed the need for a robust framework to ensure stability while optimizing the performance of the system. The distributed RL approach proposed in [2] was applied to intelligent load scheduling, enabling households to optimize energy use while maintaining network stability. Similarly, [3] showed the use of multi-agent RL for industrial smart grids, focusing on coordination among various actors to maximize resource allocation. Dynamic pricing strategies are an important area of RL research. [4] The distributed real-time pricing mechanism has been developed using RL to optimize the prices of grid operators based on real-time demand and supply fluctuations. [5] The use of RL in the management of residential demand response demonstrates how price-based incentives can encourage consumers to adapt to their consumption patterns. In another study [6], an RL-based decision system proposed allows end-users to select the optimal electricity pricing plan, thereby further personalizing energy consumption. Energy storage and market participation are important applications for RL in smart grids. [7] combined deep learning and RL to develop profitable strategies for energy storage systems at the grid level, and [8] introduced the multiple agent RL algorithm (MARLA-SG) to optimize demand responses and improve grid flexibility and efficiency. [9] The RL's application is extended to the distributed energy market, enabling consumers to trade energy effectively in decentralized markets. [10] Using multiple agent RL for energy scheduling at vehicle charging stations, improved coordination and reduced costs are achieved. [11] Introduced a multi-objective RL framework based on preferences to optimize multi-microgrid systems, highlighting the potential of RL to manage competing objectives in smart grid environments. Similarly, [12] applied a multiagent deep RL for voltage control, achieving coordinated active and reactive power optimization. The hierarchical approach proposed by the RL for community energy trading [13] has demonstrated the power of the RL to promote local energy markets and to facilitate efficient energy exchange between households. Reinforcement learning technology is also applied to specific fields such as building energy management [14], selective power system application [15], and real-time autonomous control of multi-energy residential systems [16]. Despite these advances, RL remains challenged to extend to large-scale and complex smart grid systems, ensure safe and interpreting learning outcomes, and integrate RL frameworks into real-world constraints. The examined studies highlight the transformative potential of RL for smart grid operations and are paving the way for other innovations in decentralized energy management and system optimization.

Although these approaches have advanced in the field, they are often lacking in the scale, adaptability, robustness of dynamic and distributed environments. Moreover, much of the previous work has focused on theoretical developments without taking sufficiently into account the interactions between decentralized actors and centralized market mechanisms (e.g., EPOs). This limits our understanding of how multi-agent systems achieve optimal energy strategies in real scenarios. In this paper, we propose a decentralized RL-based framework integrating dynamic pricing mechanisms for smart grids and EPOs. In contrast to previous studies that focused on the individual aspects of smart grid management, this study integrated the adaptive behavior of multiple actors with heterogeneous energy profiles. We combine decentralized RL with a responsive EPO pricing model to ensure cost optimization and stability throughout the system. The study evaluated the scalability and robustness of an RL model by analyzing learning processes and convergence behaviors in different energy states, including deficits, balanced states, and surpluses. Due to the dynamic and distributed nature of the modern smart grid, innovative solutions to traditional centralization approaches must be sought. The research is applied to energy management by applying a decentralized RL framework that allows individual actors of smart grids to optimize their energy strategies adaptively over time. In particular, the study investigated how actors adjusted their unit costs based on energy deficits and interaction with EPO, how decentralized learning stabilized system costs and reduced dependence on external energy, and how the model was scalable when applied to different actors and dynamic energy environments. This work contributes to the development of multi-agent systems in smart grids by demonstrating the effectiveness of the RL to achieve decentralized optimization and providing valuable insight into energy adaptive management.

The rest of this paper is as follows: Part II introduces the Smart Grid Model, providing a comprehensive framework for the operation dynamics of the smart grid. Section III focuses on the application of reinforcement learning techniques within the smart grid. Finally, Section IV concludes the study and summarizes the main results and contributions.

2. Smart Grid Model

Our proposed model builds upon and extends the framework previously introduced in [17]. The model assumes that the day is composed of uniform continuous intervals for several hours. For each actor, two assumptions are made: (1) the consumption and production values remain unchanged within each interval, and (2) the consumption and production values of subsequent periods can be predicted accurately.

In this discrete-time framework, all energy parameters for each actor are treated as fixed within a single interval. During each period, the smart grid is represented as a multi-agent system. Let AG={ act_1,…, act_N} denote the set of N actors connected to the SG, where each actor act_i ($1 \le i \le N$) can potentially generate electricity, particularly from renewable sources, and store it. Each actor's production and storage capacity is limited and changes over time. The following parameters are defined for each actor act_i ($1 \le i \le N$) and the period t:

- $Prod_i^t$: The energy produced by the actor during t.
- $Cons_i^t$: The amount of electricity used by the actor during t.
- $Store_i^t$: the power stored by the actor at the beginning of the period t.
- $Store_i^{max}$ is the maximum storage capacity, while Rem_i^t is the residual storage capacity, which represents the maximum additional energy the actor can store during a period. $Rem_i^t = Store_i^{max} - Store_i^t$

In one period, actors cannot simultaneously consume and supply energy to storage.

The smart grid also interfaces with an EPO, capable of supplying electricity to the grid as needed or purchasing surplus electricity from the grid. Importantly, the actors themselves do not directly engage with the EPO.

The smart grid functions as a centralized energy container connecting all actors act_i ($1 \le i \le N$). During each period t, the actor act_i contributes $Prod\_ToSG_i^t$ units of electricity to the grid and consumes $Cons\_FromSG_i^t$ units of electricity from the grid, ensuring that $Prod\_ToSG_i^t$ and $Cons\_FromSG_i^t$ cannot simultaneously be greater than zero. The aggregate electricity injected into the grid is denoted by $I_{SG}^t = \sum_{i=1}^{N} \text{Prod\_ToSG}_i^t$, while the total electricity consumed from the grid is $Out_{SG}^t = \sum_{i=1}^{N} \text{Cons\_FromSG}_i^t$.

If $\text{In}_{SG}^t < Out_{SG}^t$, the EPO supplies a shortfall of energy $q = In_{SG}^t - Out_{SG}^t$, incurring costs determined by the linear increasing cost function $\phi_{EPO}^{-}(q)$. Conversely, if $Out_{SG}^t \le In_{SG}^t$, the smart grid sells surplus energy $q = Out_{SG}^t - In_{SG}^t$ to the EPO, generating revenue based on the linear increasing benefit function $\phi_{EPO}^{+}(q)$.

At each t period, each actor act_i ($1 \le i \le N$) selects one of the four possible operating modes, known as $mode_i^t \in \{CONS^+, CONS^-, DIS, PROD\}$, based on energy requirements and production capacity.

- $CONS^-$ and $CONS^+$: These modes show that actors need energy from SGs to meet consumption requirements ($Cons\_FromSG_i^t > 0$), since their production is insufficient.

  The distinction lies in whether the actor utilizes its storage:

  - $CONS^+$: The actor consumes stored energy in addition to energy from the SG.
  - $CONS^-$: The actor refrains from using stored energy, relying solely on the SG.
- DIS: This mode signifies the actor's decision to operate independently of the SG, such that $Prod\_ToSG_i^t = Cons\_FromSG_i^t = 0$.
- PROD: This mode indicates that the actor is contributing energy to the SG ($Prod\_ToSG_i^t > 0$), with its production being sufficient to meet its consumption.

The choice of actors' modes depends on three different states: production ($Prod_i^t$), storage ($Store_i^t$), and consumption ($Cons_i^t$):

- **State Deficit:** In this state, the actor's production and storage are insufficient to meet its consumption needs. $Prod_i^t + Store_i^t \le Cons_i^t$

As a result:

- $Prod\_ToSG_i^t = 0$, and the actor must rely on the SG for energy.
- The actor can choose:
  - $CONS^+$: Consumes from both the SG and storage. In this case, $Cons\_FromSG_i^t = Cons_i^t - (Prod_i^t + Store_i^t)$, and the storage is depleted ($Store_i^{t+1} = 0$).
  - $CONS^-$: Consumes solely from the SG without using storage. Here, $Cons\_FromSG_i^t = Cons_i^t - Prod_i^t$, and storage remains unchanged ($Store_i^{t+1} = Store_i^t$).

➢ **State Self:** In this state, the actor has sufficient resources to meet its consumption needs but does not have surplus production.
$Prod_i^t + Store_i^t > Cons_i^t$ and $Prod_i^t \leq Cons_i^t$
The actor can choose:

- $CONS^-$: Consumes only from production ($Cons\_FromSG_i^t = Cons_i^t - Prod_i^t$), leaving storage unchanged ($Store_i^{t+1} = Store_i^t$).
- DIS: Operates independently of the SG. In this case, $Cons\_FromSG_i^t = 0$, and storage is partially depleted to cover the deficit ($Store_i^{t+1} = Store_i^t - (Cons_i^t - Prod_i^t)$).

In both modes, the actor does not produce energy for the SG ($Prod\_ToSG_i^t = 0$).

➢ **State Surplus:** In this state, the actor's production exceeds its consumption requirements.
$Prod_i^t > Cons_i^t$
The actor can choose:

- PROD: Supplies excess energy to the SG, where $Prod\_ToSG_i^t = Prod_i^t - Cons_i^t$, while storage remains unchanged ($Store_i^{t+1} = Store_i^t$).
- DIS: Prioritizes storing excess energy. In this case:
  - Storage is updated to $Store_i^{t+1} = \min(Simax, Store_i^t + (Prod_i^t - Cons_i^t))$.
  - Any energy that cannot be stored ($|Prod_i^t - Cons_i^t - Rem_i^t|$) is provided to the SG ($Prod\_ToSG_i^t$). In both cases, the actor does not consume energy from the SG ($Cons\_FromSG_i^t = 0$).

In the State Deficit, each actor act_i has the option to choose its mode from {CONS+,CONS-} at each period t. In the State Self, the actor can select its mode from {CONS−,DIS }. In the State Surplus, the choice is between {PROD,DIS}. Thus, in any state, each actor always has exactly two strategic options available.

These modes are designed to prioritize the actor's production for its consumption. Additionally, the stored energy of act_i is exclusively replenished using its production and never imported from the SG. Essentially, each actor ai utilizes its total production (and possibly its current storage) before importing electricity from the SG. Therefore, the decision-making for actors focuses solely on the policy governing the use or replenishment of their storage.

3. Reinforcement learning

On the one hand, we examine the price evolution of individual actors or prosumers, driven by reinforcement learning processes. On the other hand, we focus on the price evolution of the Smart Grid (SG) and the External Production Operator (EPO).

3.1. Adaptive Price Evolution of Prosumers Through Reinforcement Learning

Based on the reinforcement learning framework, each actor can adjust its unit costs over multiple learning periods. Each actor evaluates performance by analyzing energy deficits and energy costs purchased from EPO. Various parameters influence the actors' learning dynamics. First, their production capacity ($Prod_i^t$) represents the amount of energy they generate over each period. Second, their Consumption Requirement **(**$Cons_i^t$**)** reflects the energy demand in the same period. Thirdly, their storage dynamics are defined by the energy available at the beginning

of the period (current storage ($Store_i^t$)) and the maximum storage ($Store_i^{max}$) which sets the maximum amount of energy storage. In addition, each actor has a learning rate (α_i) that determines how quickly it adapts to changes in energy costs, and an initial unit cost ($U_{Initial}$) that represents its baseline energy consumption cost. These characteristics allow actors to dynamically adapt and optimize energy strategies within the grid.

In the smart grid system, each actor determines its actions based on the relationship between its production ($Prod_i^t$), consumption ($Cons_i^t$), and storage ($Store_i^t$). In the State Deficit ($Prod_i^t + Store_i^t < Cons_i^t$), the actor's production and storage are insufficient to meet its consumption requirements. Consequently, the actor must import energy from the smart grid, which may depend on the EPO to meet the demand. During this state, the actor employs a reinforcement learning process to evaluate the cost incurred from importing energy. It updates its unit cost estimate using the formula: $U_i^{t+1} = U_i^t + \alpha_i \left(\frac{\text{Cost Incurred}}{\text{Energy Deficit}} - U_i^t \right)$,

where α_i is the actor's learning rate, and the adjustment reflects the real cost of energy. In addition, the storage of the actor is fully exhausted in this state to give priority to immediate consumption needs. In a state balance or surplus ($Prod_i^t + Store_i^t \geq Cons_i^t$), the production and storage of the actor are sufficient to meet or exceed its consumption. However, this simulation focuses on deficit management and does not adjust unit costs in these states in the learning process. This approach enables actors to dynamically optimize their energy strategy, particularly in deficit scenarios where cost management is important.

Figure 1 shows the convergence of unit costs between five different intelligent actors and OEAs over 50 learning periods. Initially, the unit costs show significant differences due to the random initialization of the unit costs of the actors and different energy profiles such as production, consumption, and storage capacity. In response to the actual energy cost during the period of deficit, all actors adjust their unit costs as unit costs stabilize over time. The level of learning and energy deficits affects the convergence rate, indicating that actors have reached the best estimates of their energy costs. At the end of the learning process, all actors match the EPO costs reflecting the system balance. As a result, reinforcement learning is effective in reducing external energy dependence and optimizing cost strategies.

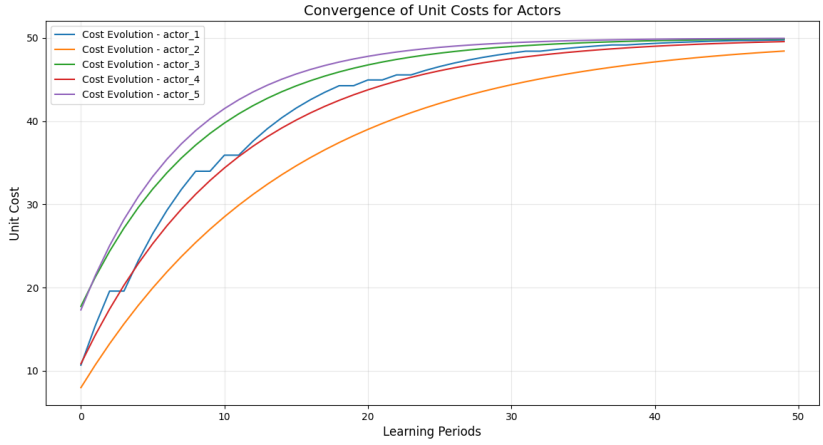

Figure 1. Convergence of Unit Costs for Smart Grid Actors Over Learning Periods

In Figure 1, some key lessons can be drawn from the proposed learning framework and its impact on smart grid energy management. As a first step, the actors demonstrate how their unit costs adapt over time and achieve convergence despite different starting conditions. Furthermore, each actor learns independently and shows how decentralized learning can help to effectively manage energy in smart grids. In addition, unit cost convergence reflects the actors' ability to estimate and adapt to external energy prices, which leads to more efficient energy use and lower costs. Finally, it illustrates the scalability of learning models in which several actors with different energy profiles independently converge on optimal strategies without central control. Because all actors align

costs with the conditions of the external market, the model demonstrates its ability to achieve equilibrium in dynamic energy environments and demonstrates the robustness of learning mechanisms.

3.2 Price Evolution of the Smart Grid (SG) and External Production Operator (EPO)

The learning dynamics of SGs can be classified as decentralized RL, and SGs adjust their internal prices dynamically based on the behavior of individual actors (prosumers) and the aggregated state of the grid, such as energy shortages.

The SG does not dictate actor decisions but instead reacts to their behaviors, creating a decentralized optimization process. The SG's pricing mechanism evolves through feedback: it observes the total energy deficit (state), adjusts its unit cost (action), and responds to changes in demand and actor behavior over time (feedback). Through improved efficiency or locally generated energy, actors can reduce deficits and stabilize SG prices. Through demand-driven optimization, SG indirectly learns from the network to manage costs dynamically and ensure network efficiency.

The SG updates its price (π_0^-) dynamically based on the total energy deficit in the grid. The formula for the update can be expressed as: $\pi_0^- = \text{Base Cost} + \alpha \cdot \frac{\text{Total Deficit}}{\text{Number of Actors}}$

Where: π_0^- is the unit price set by the SG for the current period, dynamically adjusted to reflect real-time grid conditions. The Base Cost represents the baseline price, accounting for the SG's fixed operating costs, such as infrastructure and maintenance. α is the adjustment factor, a scaling parameter that determines how sensitive the SG price is to changes in the total energy deficit, ensuring the pricing mechanism reacts proportionally to system demands. Due to the unsatisfactory energy demand of the grid, a total deficit means a cumulative energy deficit experienced by all actors in deficits. To ensure a fair distribution of price adjustments, SG can adjust pricing dynamically based on the real-time state of the network. When the total deficit increases, prices rise, encouraging actors to optimize energy use and increase local production. In addition, when the deficit is low, prices stabilize close to the basic cost and ensure the equilibrium of the system.

The EPO employs Decentralized Reinforcement Learning through a dynamic pricing mechanism defined by the cost function: $\phi_{EPO}^-(q) = \alpha \cdot \ln(1 + \beta \cdot q)$

This logarithmic function ensures that the cost grows linearly with the energy deficit (q), which is a realistic approach for bulk energy pricing. The parameters α and β control the scale and sensitivity of the cost, allowing the EPO to adapt pricing dynamically to changing energy demands.

This structure prevents runaway costs while maintaining responsiveness to energy shortfalls. In contrast, the SG also employs reinforcement learning, dynamically adjusting its internal pricing to reflect the collective behavior of prosumers and the total energy deficit. Unlike the EPO, the SG indirectly learns from actor responses, making its approach more adaptive and decentralized as it bridges the gap between local prosumers and the external energy market. Table 1 highlights the distinctions between the pricing mechanisms of SG and EPO.

Table 1. Comparison of Pricing Mechanisms Between SG and EPO

Aspect	**Smart Grid (SG)**	**External Production Operator (EPO)**
State	Aggregate system deficit or surplus	Energy flow (deficit or surplus)
Action	Adjust internal pricing	Adjust external pricing
Feedback	System-wide energy dynamics	Immediate energy flow (demand/supply)
Objective	Minimize costs and stabilize prices	Respond dynamically to supply and demand
Adaptability	High (influenced by actors)	Medium (driven by predefined cost/benefit functions)

The SG learns adaptively, optimizing its pricing strategy through reinforcement learning to balance internal efficiency while responding to external conditions, such as the pricing set by the EPO. SG's energy demand and supply are dynamically adjusted by the EPO based on predefined cost-benefit functions. Although the EPO does not "learn" as SG does, its reactive price influences SG's reinforcement learning and creates a feedback loop between internal and external energy systems.

Through dynamic multilayer systems, SG and EPO are interdependent to maintain network efficiency and stability. For EPOs and SG, this graph shows the evolution of unit costs over time. As a result of the fluctuating demand for smart grids, the blue line represents the unit cost of energy purchased from EPO. The peak in EPO costs is reflected in periods of high energy shortages where the SG relies heavily on external energy supply. On the other hand, a solid green line shows the energy unit cost of SG, which is always lower than EPO costs. The SG is devoted to local production and storage energy and minimizes the costs of participants. SG costs are relatively stable and less volatile, reducing dependence on EPO and providing local energy sources to meet demand. Figure 2 shows the ability of SG to stabilize energy costs while reducing dependence on expensive external sources.

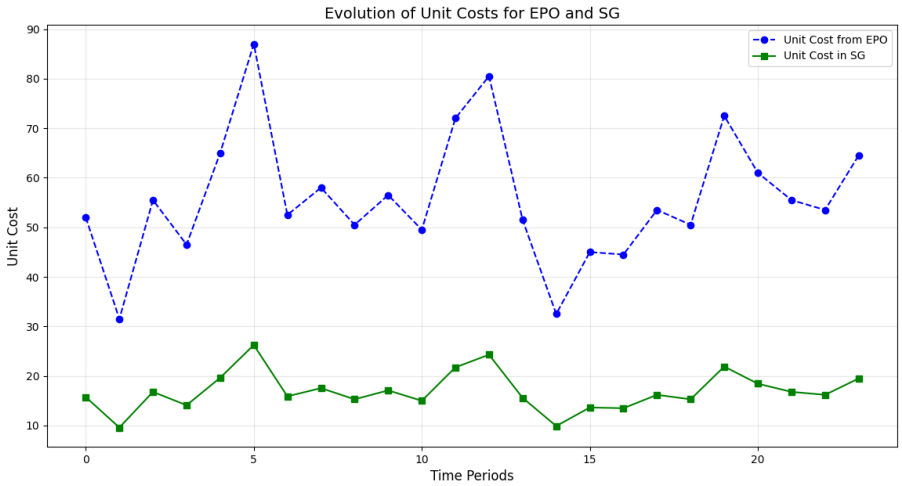

Figure 2. Comparison of Unit Costs Between External Production Operator (EPO) and Smart Grid (SG) Over Time

There are several reasons why the price of the SG is consistently lower than the EPO. First, the SG is a buffer system that optimizes resource allocation between prosumers and uses energy generated and stored internally, resulting in lower operational costs than EPO purchases. As a second factor, the pricing logic of the SG is weighed, combining a basic cost with an EPO part when external energy is needed. SG prices reflect operational efficiency and are still lower unless the network becomes too dependent on the EPO. Thirdly, the smart grid reduces the overall cost of prosumers by minimizing dependence on external suppliers and deliberately keeping internal costs lower than those of the EPO. Finally, the location of energy through the contributions of consumers reduces the average internal costs of SG by reducing the need for external energy purchases. As a result, SG efficiently manages energy and reduces member costs.

The convergence of the unit costs of actors demonstrates the adaptive learning and optimization of energy strategy. Initially, because actors are relying on external energy from the EPO to offset deficits, unit costs rise. Each actor follows a unique trajectory, influenced by its learning rate and specific energy conditions, such as production, consumption, and storage capacity. This variation in convergence rates reflects the diversity in actors' behaviors and initial states. Over time, the unit costs stabilize, indicating that actors have successfully learned to balance their energy usage and minimize reliance on external resources. The convergence highlights the effectiveness of the learning process, as actors in deficit states gradually refine their strategies to achieve greater efficiency in energy management.

4. Conclusion

The study aims to develop a reinforcement learning framework for the decentralized energy management of smart grids, allowing actors to optimize energy strategies adaptively while interacting with external production operators (EPOs). This work provides for the development of multi-agent reinforcement learning models with scalable and robust features to support decentralized decision-making, the integration of dynamic price mechanisms for smart grids and EPOs, and the complete analysis of the convergent behaviors of actors with different energy profiles. Based on the results, decentralized learning is effective in stabilizing unit costs, minimizing external energy dependence, and achieving cost reductions throughout the system. The practical applications of the proposed framework include improving energy allocation, encouraging participants, and balancing local energy production with market requirements in smart grid systems in the real world. Through adaptive pricing and specific learning, this model supports dynamic energy environments, making it particularly relevant to grids that integrate renewable energy sources and local storage.

To strengthen the proposed framework, future research can explore several directions. Among these, the extension of the model incorporates elastic demand responses, the integration of advanced actor interactions, such as cooperative energy sharing, and the assessment of performance in large-scale systems with real-world constraints. Additional research is needed to address the need for interpretation and safety in the reinforcement of learning in critical energy systems. Future research can use this research to improve the efficiency, scale, and sustainability of smart grids.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number “NBU-FFR-2024-2441-04“.

References

[1] Bui, V. H., Das, S., Hussain, A., Hollweg, G. V., & Su, W. (2024). A Critical Review of Safe Reinforcement Learning Techniques in Smart Grid Applications. arXiv preprint arXiv:2409.16256.

[2] Chung, H. M., Maharjan, S., Zhang, Y., & Eliassen, F. (2020). Distributed deep reinforcement learning for intelligent load scheduling in residential smart grids. IEEE Transactions on Industrial Informatics, 17(4), 2752-2763.

[3] Roesch, M., Linder, C., Zimmermann, R., Rudolf, A., Hohmann, A., & Reinhart, G. (2020). Smart grid for industry using multi-agent reinforcement learning. Applied Sciences, 10(19), 6900.

[4] Zhang, L., Gao, Y., Zhu, H., & Tao, L. (2022). A distributed real-time pricing strategy based on reinforcement learning approach for smart grid. Expert systems with applications, 191, 116285.

[5] Wan, Y., Qin, J., Yu, X., Yang, T., & Kang, Y. (2021). Price-based residential demand response management in smart grids: A reinforcement learning-based approach. IEEE/CAA Journal of Automatica Sinica, 9(1), 123-134.

[6] Lu, T., Chen, X., McElroy, M. B., Nielsen, C. P., Wu, Q., & Ai, Q. (2020). A reinforcement learning-based decision system for electricity pricing plan selection by smart grid end users. IEEE Transactions on Smart Grid, 12(3), 2176-2187.

[7] Han, G., Lee, S., Lee, J., Lee, K., & Bae, J. (2021). Deep-learning-and reinforcement-learning-based profitable strategy of a grid-level energy storage system for the smart grid. Journal of Energy Storage, 41, 102868.

[8] Aladdin, S., El-Tantawy, S., Fouda, M. M., & Eldien, A. S. T. (2020). MARLA-SG: Multi-agent reinforcement learning algorithm for efficient demand response in smart grid. IEEE access, 8, 210626-210639.

[9] Ghasemi, A., Shojaeighadikolaei, A., Jones, K., Hashemi, M., Bardas, A. G., & Ahmadi, R. (2020, November). A multi-agent deep reinforcement learning approach for a distributed energy marketplace in smart grids. In 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) (pp. 1-6). IEEE.

[10] Zhang, Y., Yang, Q., An, D., Li, D., & Wu, Z. (2022). Multistep multiagent reinforcement learning for optimal energy schedule strategy of charging stations in smart grid. IEEE Transactions on Cybernetics, 53(7), 4292-4305.

[11] Xu, J., Li, K., & Abusara, M. (2022). Preference based multi-objective reinforcement learning for multi-microgrid system optimization problem in smart grid. Memetic Computing, 14(2), 225-235.

[12] Hu, D., Ye, Z., Gao, Y., Ye, Z., Peng, Y., & Yu, N. (2022). Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization. IEEE Transactions on Smart Grid, 13(6), 4873-4886.

[13] Yan, L., Chen, X., Chen, Y., & Wen, J. (2022). A hierarchical deep reinforcement learning-based community energy trading scheme for a neighborhood of smart households. IEEE Transactions on Smart Grid, 13(6), 4747-4758.

[14] Yu, L., Qin, S., Zhang, M., Shen, C., Jiang, T., & Guan, X. (2021). A review of deep reinforcement learning for smart building energy management. IEEE Internet of Things Journal, 8(15), 12046-12063.

[15] Chen, X., Qu, G., Tang, Y., Low, S., & Li, N. (2022). Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. IEEE Transactions on Smart Grid, 13(4), 2935-2958.

[16] Ye, Y., Qiu, D., Wu, X., Strbac, G., & Ward, J. (2020). Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. IEEE Transactions on Smart Grid, 11(4), 3068-3082.

[17] Barth, D., Cohen-Boulakia, B., & Ehounou, W. (2022). Distributed reinforcement learning for the management of a smart grid interconnecting independent prosumers. Energies, 15(4), 1440.