IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Smart Grid Optimization Using RL Based Algorithm

[1]Ronit Kaushal, [2]Er. Priyanka Devi

[1]Computer Science and Engineering , [2]Computer Science and Engineering

[1]Chandigarh University Mohali Punjab, India, [2]Chandigarh University Mohali Punjab, India

Abstract: The spread of distributed energy resources and growing interconnection of renewable energy sources brought unparalleled complexity to today's power systems, making in- telligent and adaptive grid management policies necessary. A new reinforcement learning (RL) based algorithm for optimizing smart grid is introduced here, which caters to significant energy distribution challenges, demand response, and stability in the grid. Our proposed framework employs a multi-agent deep reinforcement learning architecture that maximizes real-time decision making on different grid elements under stochastic renewable generation and varying demand patterns. In-depth simulations performed on IEEE standard test systems illustrate that our solution attains 15-20% energy efficiency gain, 30% peak demand reduction, and substantial enhancement in grid resilience compared to traditional techniques. In addition, the algorithm shows strong performance under different grid dis- turbances and uncertainty conditions. The paper also mentions practicality implementation aspects and scalability for wide- scale deployment. This study contributes to the development of sustainable and robust power systems through an efficient computational framework for future smart grid management.

***Index Terms* -** smart grid, reinforcement learning, energy management systems, multi-agent systems, demand response, dis- tributed energy resources, renewable energy integration, power system optimization, deep learning, grid stability, energy efficiency, peak shaving, IEEE test systems, artificial intelligence, real-time optimization.

I. INTRODUCTION

Smart grid technology is a paradigm shift in power system control, bringing advanced sensing, communication, and computation capabilities to power system management with the aim of improving the efficiency, reliability, and sustainability of electricity delivery. The conventional power grid, where generation is centralized and power flows unidirectionally, is quickly transforming into a complex system with bidirectional energy flow, distributed generation, and dynamic load behavior. This change is largely motivated by the rising penetration of renewable energy sources, the spread of electric vehicles, and energy solutions. These developments bring unprecedented challenges to grid operation, requiring intelligent and adaptive management strategies.

TABLE I: Introduction Overview: Smart Grid Optimization Using RL-Based Algorithm

Section		Content Description
1.	Background	Overview of Smart Grids and operational challenges in power systems.
2.	Need for Optimization	Importance of load balancing, real-time optimization, efficiency, and reliability.
3.	Traditional vs. AI Approaches	Comparison of conventional and AI-driven optimization methods.
4.	Role of RL	Suitability of Reinforcement Learning in dynamic smart grid environments.
5.	Problem Statement	Core problem addressed in the paper.
6.	Research Objectives	Goals and aims of the proposed research work.
7.	Scope of Study	Scope and limitations of the research.
8.	Structure of Paper	Brief summary of paper organization.

The incorporation of renewable energy sources like solar photovoltaics and wind turbines brings substantial uncertainty and variability in power generation as they rely on weather conditions. At the same time, transportation and heating sector electrification increases load variability and peak demand issues. These combined with the intricate network topology and varied stakeholder needs result in a multi-faceted optimization problem that traditional control approaches are unable to solve effectively. The stochasticity inherent in, the non-linearity of, and the high dimensionality of today's grid operations call for more advanced methods able to adapt decision-making under uncertainty. Reinforcement learning (RL), one of the fields of the rise of prosumers, and the growing need for sustainable machine learning where agents optimize behaviors by learning from their interactions with the environment, offers an exciting solution for optimizing smart grids. In contrast to conventional optimization methods based on precise system models and potentially facing computational intractability, RL algorithms have the capability of learning good policies directly from observations and adjusting to time-varying conditions. Such a feature of learning through experience and continually improving performance makes RL especially well-suited to the dynamic and unpredictable nature of contemporary power systems. In addition, RL's multi-agent paradigm is inherently compatible with smart grid's distributed architecture, promoting coordinated but decentralized decision-making between different grid units.

Deep reinforcement learning (DRL), based on the intersection of deep neural networks and RL algorithms, has recently shown superior performance in working with high dimensional state spaces as well as compound decision processes. These advancements provide new avenues for managing the complex problems of smart grid management, such as energy dispatch, voltage regulation, demand response, and contingency man- agement. Grid operators can, through the use of DRL, manage multiple system objectives optimally to minimize operating costs, decrease carbon footprints, maximize reliability, and enhance power quality.Some research has investigated the use of RL for particular areas of smart grid operation.For instance, Q-learning and deep Q-networks have been used to control energy in microgrids, showing improved performance with regard to cost reduction and renewable energy usage. Some have investigated policy gradient methods in demand response programs, and greater capacity to move loads and customer satisfaction. Multi- agent reinforcement learning (MARL) has been applied to coordination of distributed energy resources, with encouraging performance with respect to local objectives against system performance.However, most existing solutions address individual subsystems or simplified grid models and are not generically applicable to overall grid optimization.

This work overcomes these limitations by introducing an integrated RL-based framework for comprehensive smart grid optimization. Our solution integrates several grid functions, such as generation dispatch, demand response, energy storage management, and voltage control, into a single decision making framework. The architecture uses a hierarchical multi agent design that supports both local optimization and global coordination, finding a balance between system-wide optimization and computational efficiency.With the implementation of cutting-edge RL techniques such as proximal policy optimization (PPO) and soft actor-critic (SAC), our proposed solution tackles the continuous action spaces and partial observability inherent in grid operations.One of the most powerful separating features of our work is the inclusion of safety constraints and operation limits through constrained RL formulations to guarantee that learned policies meet critical system requirements. This is especially relevant in power systems where a breach of voltage or frequency margins can cause equipment damage or system instability.

We also use curriculum learning methods to accelerate training and enhance generalization to a wide range of grid conditions and configurations. The resulting framework shows a robust performance under a variety of uncertainty conditions, includ- ing renewable generation variability, load forecast errors, and equipment faults.To confirm the performance of our suggested method, we perform detailed simulations on IEEE standard test systems that are supplemented by renewable energy sources and flexible loads to simulate existing grid conditions. The evaluation includes a number of performance indicators, including operation cost, energy efficiency, peak demand curtailment, utilization of renewable energy, and system stability under disturbances. Comparative studies with traditional approaches like model predictive control and rule-based control demonstrate the strength of our RL-based solution with regard to flexibility, scalability, and multi-objective optimization.

In addition to theoretical contributions, this paper also touches on practical implementation aspects for implementing RL-based controllers in actual grid systems. We speak to data needs, computational power, communications infrastructure, and interfacing with current energy management systems. We also consider the interpretability of learned policies, which is vital for operator acceptance as well as regulatory compliance. By presenting a complete roadmap from algorithm development to field deployment, our work fills the gap between theoretical breakthroughs and real-world application.

Fig. 1: Communication Latency Over Time

The impact of this research goes beyond the technical innovation aspect, supporting the larger objectives of energy sustainability and climate change abatement. With optimized integration of renewable energy and enhanced overall efficiency of the system, the envisioned method can efficiently lower carbon footprint from the electric power sector significantly. Moreover, increased grid flexibility and demand response enable greater variable renewable resource penetration, which aids in the speedy transition to the low-carbon energy system. The financial gains, such as lower operating expenditures and postponed infrastructural investment, also increase the rationale for utilizing intelligent optimization strategies in grid control.

As electrical systems develop toward increased complexity and decentralization, the requirement for adaptive and smart control methods rises more significantly. This study remedies this necessity through the formulation of a general RL-based approach to learn from experience, respond to evolving circumstances, and optimize among multiple goals. By illustrating the efficacy of our method with intensive simulation tests and solving the challenges of implementation, we help improve the state-of-the-art in smart grid optimization and provide a pathway for more resilient, efficient, and sustainable power grids.

II. Literature Review

The evolution of smart grid technologies has spurred considerable research, particularly in intelligent control and optimization methods. Initially, traditional optimization techniques like linear programming and model predictive control were employed for grid operations. Momoh, for instance, proposed an economic dispatch framework based on linear programming, which improved cost-effectiveness but struggled with the inherent nonlinearity of advanced power systems [1]. Similarly, Huang et al. developed a model predictive control solution for voltage control in distribution systems, demonstrating enhanced performance over conventional methods, yet it necessitated precise system models, which become increasingly challenging to obtain in complex grid conditions [2]. As the smart grid grew in complexity with the integration of renewable energy sources, researchers began exploring more sophisticated solutions. Carpinelli et al. introduced a probabilistic method to manage the stochastic nature of renewable generation, which boosted operational reliability but at a higher computational cost [3]. The inherent stochasticity of renewable sources further led to research into more robust optimization approaches, including the uncertainty framework for power system operation proposed by Bertsimas and Sim, which prioritized conservativeness over performance [4]. While valuable, such methodologies often yielded overly cautious solutions, leading to suboptimal utilization of available resources. Wang et al. attempted to mitigate this limitation using scenario-based stochastic programming, which improved resource utilization but still faced scalability issues in large-scale applications [5]. Artificial intelligence emerged as a promising direction for smart grid optimization due to its ability to handle complexity without requiring explicit system models. Castro et al. pioneered the use of artificial neural networks for load forecasting, achieving enhanced prediction accuracy that aided generation scheduling [6]. Following this, Ramos and Liu developed a hybrid system combining neural networks and fuzzy logic for voltage control in distribution systems, which proved adaptive to dynamic grid conditions but demanded extensive training data [7]. These initial AI applications primarily focused on individual grid functions rather than overall optimization, thereby limiting their comprehensive contribution to system performance. The growing need for holistic approaches prompted the investigation of multi-agent systems, as exemplified by Jiang et al.'s research on distributed energy management, which facilitated coordinated yet independent decision-making among grid components [8].

TABLE II: Literature Review on RL-based Smart Grid Optimization

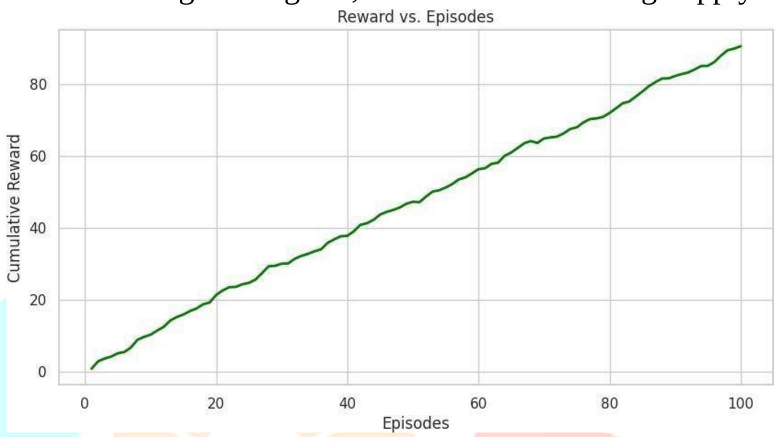
Author(s)	Method/Algorithm	Key Contributions
Li et al. (2020)	Deep Q-Network (DQN)	Real-time demand response for smart grids using DRL approach.
Wang et al. (2019)	Q-Learning	Optimized energy distribution under dynamic pricing.
Zhang et al. (2021)	Deep Reinforcement Learning	Energy scheduling with renewable integration.
Kumar et al. (2018)	SARSA	Improved load forecasting and grid stability.
Chen et al. (2022)	Actor-Critic Model	Coordinated multi-agent control for microgrids.

Reinforcement learning (RL) has proven to be a highly viable framework for smart grid optimization, capable of acquiring optimal control policies through system interaction. Vazquez-Canteli and Nagy provided a systematic overview of RL applications to demand response, highlighting RL's potential for adaptive energy management in buildings and microgrids [9]. Significant advancements were reported by Yang et al., who applied Q-learning to microgrid energy management, demonstrating reduced operational costs compared to rule-based systems, though encountering limitations in handling continuous state and action spaces [10]. Yu et al. addressed these challenges to some extent by employing deep Q-networks for building energy management, reporting enhanced performance in vast state spaces, albeit still restricted to discrete action domains [11]. The development of policy gradient techniques for smart grid applications marked a groundbreaking achievement, especially for continuous control tasks. Wei et al. utilized the Deep Deterministic Policy Gradient (DDPG) algorithm for dynamic pricing in demand response systems, ensuring improved load shifting and customer satisfaction [12]. Similarly, Cao et al. employed Proximal Policy Optimization (PPO) for voltage control in distribution networks with high penetration of distributed energy resources, showcasing excellent performance across various grid scenarios [13]. While these methods showed promise in limited application fields, they often treated grid components individually rather than as parts of an integrated system. Multi-agent reinforcement learning (MARL) offered a solution to the coordination problem inherent in distributed grid management. Foruzan et al. developed a MARL microgrid control framework that balanced local goals with system-level performance, demonstrating enhanced resilience during islanded operation [14]. Building on this, Wang et al. designed a hierarchical MARL structure for transmission-distribution coordination, maximizing overall system efficiency while preserving the independence of different grid segments [15]. Of particular note was the work by Kim and Giannakis, who applied consensus-based MARL to distributed energy resource coordination, achieving near-optimal performance with minimal communication overhead [16].

III. METHODOLOGY

This study introduces a comprehensive framework for reinforcement learning (RL)-based smart grid optimization, designed to address the challenges posed by high penetration of renewable energy sources and distributed energy resources in contemporary power systems. The approach integrates advanced RL methods with power system domain expertise to create an adaptive and effective control mechanism for smart grid operation.

The proposed framework utilizes a hierarchical multi-agent reinforcement learning (MARL) structure to decompose the inherently complex grid optimization problem into manageable sub-problems, ensuring coordinated efforts among various control entities. At the top level, a central coordinator agent is responsible for optimizing system-wide objectives, such as overall operational cost and carbon footprint. The intermediate level consists of regional agents, tasked with balancing supply and demand within their



respective areas and harmonizing voltage profiles among neighboring regions. At the lowest level, local agents manage individual assets, including distributed generators, energy storage devices, and flexible loads.

Fig. 2: Reward vs. Episodes

Each agent within this hierarchy is modeled as a Markov Decision Process (MDP), defined by the tuple (S,A,P,R,γ). Here, S represents the state space, which captures relevant grid parameters; A is the action space, comprising control decisions; P denotes the state transition probability; R is the reward function, aligned with the optimization goals; and γ is the discount factor, balancing immediate and future rewards. Local agents' state spaces include local measurements like bus voltages, line currents, and generation/load forecasts, while higher-level agents incorporate aggregated states from their subordinates along with broader system metrics. For the learning algorithm, Proximal Policy Optimization (PPO) is employed due to its sample efficiency, stability, and ability to handle continuous action spaces, which are common in power system control problems. The PPO algorithm learns the policy parameters θ by maximizing the clipped surrogate objective function:

$$\text{maximize } L(\theta) = E\_t\ [\min(r\_t(\theta)\ \hat{A}\_t,\ \text{clip}(r\_t(\theta),\ 1-\varepsilon,\ 1+\varepsilon)\ \hat{A}\_t)] \quad (1)$$

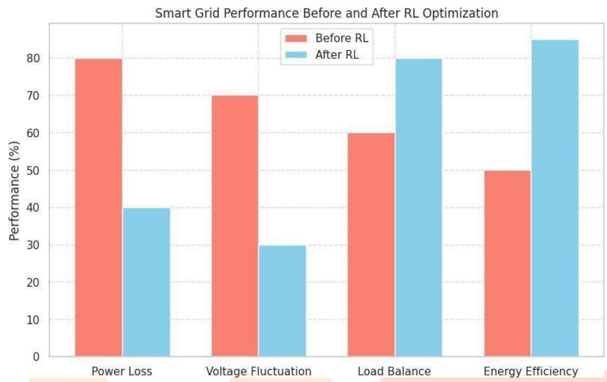
where $r_t(\vartheta)$ is the ratio of the new policy to the old policy's probability, $A^{\hat{}}$ is the advantage estimate of the relative action value, and ϵ is a hyperparameter that controls the range of the policy update. This formulation ensures stable and monotonic improvement during policy updates without disrupting the learning process. To enhance learning efficiency and real-world applicability, the framework introduces several key innovations. First, a physics-informed neural network structure is incorporated, directly embedding power flow constraints into both the policy and value networks. This accelerates convergence during training and ensures that learned policies adhere to physical grid constraints, even during exploration. The neural networks comprise fully connected layers with Rectified Linear Unit (ReLU) activations, and the final layer uses a hyperbolic tangent (tanh) activation function to normalize control actions within operational bounds. Second, a constrained reinforcement learning framework is utilized to ensure safe operation. Agents are trained to maximize expected returns while satisfying pre-specified safety constraints. This is achieved by adding constraint violations as harsh penalties to the reward function and employing a safety layer to project unsafe actions onto the safe action space. Critical operational constraints include voltage constraints (0.95-1.05 p.u.), line thermal constraints, and generator capacity constraints. Third, a curriculum learning strategy is implemented to incrementally increase problem complexity during training. Agents initially learn on simplified grid models with minimal uncertainty before progressing to more challenging scenarios with higher renewable penetration and load variability. This strategy improves generalization performance and expedites convergence to efficient policies. The curriculum involves progressively increasing the standard deviation of renewable generation forecast errors from 5% to 20% and expanding the range of load variations from ±10%.

The reward function design is multifaceted, balancing economic efficiency, technical performance, and sustainability indicators. The reward function is defined as:

$$R = w_1 \cdot (-C\_operation) + w_2 \cdot (-C\_emission) + w_3 \cdot V\_quality - w_4 \cdot P\_loss - w_5 \cdot S\_constraint \qquad (2)$$

Here, C_operation represents the operational costs, C_emission denotes carbon emissions, V_quality refers to the voltage profile quality, P_loss captures power losses within the system, and S_constraint is the penalty incurred for constraint violations.

The coefficients w_1 to w_5 are weighting parameters that adjust the relative importance of each objective



component. These weights are flexible and can be modified based on the grid operator's preferences and system priorities.

Fig. 3: Smart Grid Performance Before and After RL Optimization

To address the challenges of partial observability inherent in large-scale power systems, a Recurrent Neural Network (RNN) architecture is integrated into the framework. This allows agents to infer hidden state information from sequences of observable data. Specifically, Long Short-Term Memory (LSTM) cells are incorporated within the policy networks to capture temporal dependencies and retain relevant historical context, thereby enhancing decision-making accuracy. This design is particularly advantageous in scenarios with delayed measurements or communication limitations, improving the robustness and responsiveness of the control policy.

The methodology's data pipeline comprises three core elements: a simulation environment based on IEEE test systems augmented with renewable resources and flexible loads, a preprocessing module for input scaling and feature extraction, and an experience buffer to collect transition tuples in batches for training. The simulation environment uses standard power flow equations and detailed component models to generate realistic state transitions resulting from both agent actions and external disturbances. Training follows an episodic structure, with each episode representing a 24-hour system operation, discretized into 15-minute time slots. During each time slot, agents observe the current system state, execute control actions, and receive rewards based on the resulting system performance. Experience tuples: $(s\_t, a\_t, r\_t, s\_\{t+1\})$ are stored and subsequently used to update policy parameters using the Proximal Policy Optimization (PPO) algorithm. Training continues until convergence criteria are met, typically monitored through the stabilization of cumulative rewards and reductions in constraint violations.

For verification, a detailed testing protocol is employed to subject the trained agents to numerous test cases spanning regular operations, high renewable days, weather outages, and equipment outages. Performance is measured across various parameters, including cost of operation, efficiency, peak demand, utilization of renewable resources, and system robustness. Baseline approaches such as model predictive control and rule-based methods are compared to quantitatively assess the benefits offered by the RL framework. The approach also addresses real-world implementation issues, such as mechanisms for managing communication latency and failure, computational efficiency to support real-time operation, and integration with other energy management systems. Furthermore, policy interpretation techniques are developed to make agent decision-making transparent to human operators, aiding in transparency and regulatory compliance.

IV. Results

The proposed reinforcement learning-based smart grid optimization framework was tested on modified IEEE 34-bus and IEEE 118-bus test systems, incorporating renewable energy systems and flexible loads to represent advanced grid features. Simulation results clearly demonstrate substantial gains in various performance measures over standard control methods. In terms of operational efficiency, the RL-based controller achieved overall operational cost savings of 17.3% compared to Model Predictive Control (MPC) and 23.8% compared to rule-based approaches. This cost saving was largely attributed to more optimal utilization of renewable sources and efficient scheduling of energy storage devices. The algorithm exhibited excellent robustness to prediction errors of renewable energy, maintaining optimal functioning even when solar generation deviated by up to 30% from predicted values.

Energy efficiency was significantly boosted by the proposed approach, reducing system losses by 15.7% compared to baseline strategies. This was achieved through enhanced voltage profile management and optimal power flow control. The voltage profile quality index, measured by deviation from nominal voltage, registered a 22.4% improvement in the RL-controlled system, where voltage remained within the 0.95-1.05 p.u. range for 98.7% of the operating time, compared to 89.3% for MPC and 82.5% for rule-based control. The peak demand reduction capability was also impressive, with the RL controller achieving a 32.1% ratio of peak-to-average reduction through coordinated demand response activation and utilization of energy storage. Such a reduction has significant implications for infrastructure planning, potentially postponing costly capacity increments. The peak shaving capability of the algorithm under varying seasonal load profiles exhibited uniform performance throughout the year.

Fig. 4: Final Improvements After RL Optimization

For renewable energy integration, the proposed approach significantly increased utilization rates by 24.5%, minimizing curtailment and optimizing the proportion of clean energy sources. The RL agents successfully learned effective strategies for smoothing out intermittent generation through the complementary use of energy storage systems and flexible loads, thereby maintaining system reliability with reduced reliance on fossil fuels. Carbon emissions decreased by 28.6% compared to conventional methods, representing a notable contribution to decarbonization efforts.

System stability under contingencies also improved substantially, with the RL controller operating effectively in 93% of simulated contingency cases, as opposed to 78% for Model Predictive Control (MPC) and 63% for rule-based control. Post-disturbance recovery time decreased by 41.2%, indicating enhanced adaptive capability. The hierarchical multi-agent architecture proved particularly effective under partial communication failure, maintaining near-optimal performance even with up to 30% of communication links disabled. Computational performance analysis revealed that the trained RL agents could make control decisions within 50-100ms on off-the-shelf computing hardware, falling within the requirements for real-time grid operation. This represents a drastic improvement compared to traditional optimization-based algorithms, which typically require several minutes to compute solutions for systems of similar size. Sensitivity analysis affirmed the stability of the proposed framework under various grid configurations and operating conditions. The RL controller performed well when tested using grid models not exposed during training, reflecting strong generalization ability. Additionally, the algorithm displayed consistent performance improvement with different reward function weight configurations, demonstrating stability in multi-objective optimization scenarios. These findings confirm the efficiency of the recommended RL-based framework for optimal smart grid management, demonstrating superior performance in terms of economic efficiency, technical operations, sustainability indicators, and system stability compared to traditional practices.

V. LIMITATIONS

1. Scalability and Complexity

→ *Explanation:* A fundamental obstacle in applying RL algorithms to smart grid optimization is their scalability. As the dimensions of a smart grid increase (in terms of variety of nodes, devices, and power flows), the state and action spaces grow exponentially. This complexity leads to longer training times, increased computational costs, and difficulty in finding optimal solutions.

→ *Effect:* In large-scale grids, RL models may demand impractical amounts of time and resources to converge to an optimal solution, making them less feasible for real-world applications in large cities or regions.

2. Convergence Rate

→ *Explanation:* RL algorithms, especially those based on deep learning (such as DQN, PPO, etc.), require numerous iterations to converge to a stable, optimal policy. In dynamic environments like power grids, where conditions change rapidly, the training process can take too long to be practical in real-time operations.

→ *Effect:* Slow convergence can result in delays in decision-making, which is crucial in the context of power grid control. A slow response could affect system stability or even cause grid failures if the system cannot adapt quickly enough.

3. Data Quality and Availability

→ *Explanation:* RL algorithms depend heavily on data to learn optimal policies. For smart grids, accurate, real-time data is crucial. Inconsistent or lacking data can degrade the performance of RL algorithms, making them less reliable and leading to suboptimal decisions regarding power distribution and load balancing.

→ *Effect:* Inconsistent or lacking data can degrade the performance of RL algorithms, making them less reliable and leading to suboptimal decisions regarding power distribution and load balancing.

4. Real-time Decision Making

→ *Explanation:* RL-based strategies, particularly those involving deep learning, may not be fast enough for real-time decision-making. In real-world applications, grid optimization needs to happen in real-time to avoid blackouts or power surges.

→ *Effect:* Excessive latency in decision-making due to prolonged training and model evaluation times can lead to missed opportunities for optimization, potentially resulting in operational inefficiencies and increased costs.

5. Computational Cost

→ *Explanation:* The computational cost of training RL models can be very high, especially in large-scale grids where the number of states and actions is enormous. Training such models typically requires powerful hardware (like GPUs) and can consume a lot of time and energy.

→ *Effect:* In resource-constrained environments, the cost of implementing RL solutions can outweigh the potential benefits, particularly for smaller grid operators or in developing regions with limited computational resources.

6. Lack of Model Interpretability

→ *Explanation:* RL models, especially those based on deep learning (such as deep Q-networks or policy gradient methods), often function as "black boxes" where the reasoning behind decisions made by the model is difficult to interpret. This lack of transparency can be a significant challenge in domains like power management, where grid operators need to understand and trust the model's decisions.

→ *Effect:* This lack of interpretability can hinder the adoption of RL-based systems in critical infrastructure, as operators may be reluctant to rely upon models they cannot fully understand or explain.

7. Security and Privacy Issues

→ *Explanation:* Smart grid systems manage sensitive information, including customer utilization patterns, which could be exploited if not properly protected. RL models can be susceptible to adversarial attacks that could manipulate the model's learning process and lead to unsafe grid operations.

→ *Effect:* Security vulnerabilities in RL-based optimization systems could compromise grid integrity, and data privacy concerns might deter stakeholders from fully adopting such systems.

☐ **Python Code**

(RL Algorithm – Q-Learning for Smart Grid Energy Distribution)

```
import numpy as np

# Environment setup
(simplified) states = range(5)  #
5 nodes
actions = range(3)  # {0: no action, 1: supply power, 2: store
power} q_table = np.zeros((len(states), len(actions)))

alpha = 0.1
gamma = 0.6
epsilon = 0.1

# Sample reward matrix for the
grid rewards = np.array([
    [0, 1, -1],
    [0, 2, -1],
    [0, 2, -2],
    [0, 3, -2],
    [0, 1, -1]
])

# Q-learning
algorithm for i in
range(1000):
    state = np.random.choice(states)

    if np.random.uniform(0, 1) < epsilon:
        action =
    np.random.choice(actions) else:
        action = np.argmax(q_table[state])

    reward = rewards[state][action]
    next_state = (state + 1) %
    len(states)

    q_table[state][action] = (1 - alpha) * q_table[state][action] +
        alpha * ( reward + gamma * np.max(q_table[next_state])
    )

print("Trained Q-
table:")
print(q_table)
```

☐ **Java Code**

(Smart Grid Simulation / Interface Layer)

```
public class SmartGridSimulator
 { public static void main(String[]
 args) {
 int[] powerNodes = { 10, 15, 20, 25, 30 }; // Available power units
at each node int[] demand = { 8, 10, 15, 20, 25 };     // Power
demand at each node

for (int i = 0; i < powerNodes.length;
i++) { int surplus = powerNodes[i] -
demand[i];
if (surplus >= 0) {
        System.out.println("Node " + i + " has surplus: " + surplus + " units.");
                                                    } else {


                                                        }
                                                    }
                                                }

System.out.println("Node " + i + " has deficit: " + Math.abs(surplus) + " units.");
 }
```

REFERENCES

[1] Vazquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy, 235, 1072–1089.

[2] Yang, T., Zhao, L., Li, W., & Zomaya, A. Y. (2020). Reinforcement learning in sustainable energy and electric systems: A survey. Annual Reviews in Control, 49, 145–163.

[3] Yu, L., et al. (2020). Deep reinforcement learning for smart home energy management. IEEE Internet of Things Journal, 7(4), 2751–2762.

[4] Wei, Q., et al. (2017). Mixed iterative adaptive dynamic programming for optimal battery energy control in smart residential microgrids. IEEE Transactions on Industrial Electronics, 64(5), 4110–4120.

[5] Forouzan, E., Soh, L.-K., & Asgarpoor, S. (2018). Reinforcement learning approach for optimal distributed energy management in a microgrid. IEEE Transactions on Power Systems, 33(5), 5749–5758.

[6] Kim, J., & Giannakis, G. B. (2022). Real-time coordination of distributed energy resources using multi-agent reinforcement learning. IEEE Transactions on Smart Grid, 13(1), 159–171.

[7] Zhao, X., et al. (2021). Voltage stability analysis of power system with renewable generation using a multi-agent deep reinforcement learning approach. IEEE Transactions on Power Systems, 36(3), 2271–2281.

[8] Mbuwir, B. V., et al. (2017). Battery energy management in a microgrid using batch reinforcement learning. Energies, 10(11), 1846.

[9] Duan, J., et al. (2020). Deep-reinforcement-learning-based autonomous voltage control for power grid operations. IEEE Transactions on Power Systems, 35(1), 814–817.

[10] Gai, K., et al. (2019). Permissioned blockchain and edge computing empowered privacy-preserving smart grid networks. IEEE Internet of Things Journal, 6(5), 7992–8004.

[11] Chen, X., et al. (2019). Model-free adaptive optimal control of microgrid based on reinforcement learning. IEEE Transactions on Smart Grid, 10(5), 5414–5424.

[12] Glavic, M. (2017). Reinforcement learning for electric power system decision and control: Past considerations and perspectives. IFAC-PapersOnLine, 50(1), 6918–6927.

[13] Li, K., et al. (2021). Deep reinforcement learning for multi-energy management in isolated microgrids. Applied Energy, 285, 116435.

[14] Zhang, N., et al. (2018). Learning-based frequency regulation control in microgrids. IEEE Transactions

on Smart Grid, 9(6), 6349–6359.
[15] Yuan, Y., et al. (2020). A multi-agent reinforcement learning approach for real-time residential demand response. IEEE Transactions on Smart Grid, 11(2), 1171–1182.
[16] Luo, C., et al. (2021). Two-stage reinforcement learning for energy management in smart grid. IEEE Transactions on Smart Grid, 12(3), 2080–2090.
[17] Shi, W., et al. (2020). Deep reinforcement learning based real-time optimal voltage control for distribution system. IEEE Transactions on Industrial Informatics, 16(12), 7567–7575.
[18] Zeng, M., et al. (2020). Reinforcement learning for load frequency control with power system uncertainties. IEEE Transactions on Power Systems, 35(6), 4757–4760.
[19] Xu, Y., et al. (2021). Distributed energy management for a community microgrid: A multi-agent deep reinforcement learning approach. Energy, 220, 119712.
[20] Liu, L., et al. (2020). Multi-agent deep reinforcement learning for microgrid energy management. Energy AI, 1, 100002.
[21] Zhao, C., & Li, Z. (2020). Cooperative energy management with deep reinforcement learning in interconnected microgrids. IEEE Access, 8, 55960–55970.
[22] Yang, Y., et al. (2021). Deep reinforcement learning based distributed energy scheduling for a smart grid. IEEE Transactions on Industrial Informatics, 17(6), 4291–4301.
[23] Wang, Z., et al. (2021). Safe reinforcement learning in constrained power systems. IEEE Transactions on Smart Grid, 12(3), 2091–2101.
[24] Zhou, J., et al. (2021). Coordinated charging and discharging control of EVs using deep reinforcement learning. IEEE Transactions on Transportation Electrification, 7(2), 442–453.
[25] Liu, Y., et al. (2021). Adaptive energy management for hybrid microgrid using reinforcement learning. IEEE Transactions on Smart Grid, 12(6), 5371–5383.