

# Reinforcement Learning Techniques in Optimizing Energy Systems

Stefan Stavrev <sup>1,\*</sup>  and Dimitar Ginchev <sup>2</sup> 

<sup>1</sup> Department of Software Technologies, Faculty of Mathematics and Informatics, Plovdiv University “Paisii Hilendarski”, 4000 Plovdiv, Bulgaria

<sup>2</sup> Department of Air Transport, Faculty of Transport, Technical University of Sofia, 1000 Sofia, Bulgaria; dimginchev@tu-sofia.bg

\* Correspondence: stavrev@uni-plovdiv.bg; Tel.: +359-889-716-824

**Abstract:** Reinforcement learning (RL) techniques have emerged as powerful tools for optimizing energy systems, offering the potential to enhance efficiency, reliability, and sustainability. This review paper provides a comprehensive examination of the applications of RL in the field of energy system optimization, spanning various domains such as energy management, grid control, and renewable energy integration. Beginning with an overview of RL fundamentals, the paper explores recent advancements in RL algorithms and their adaptation to address the unique challenges of energy system optimization. Case studies and real-world applications demonstrate the efficacy of RL-based approaches in improving energy efficiency, reducing costs, and mitigating environmental impacts. Furthermore, the paper discusses future directions and challenges, including scalability, interpretability, and integration with domain knowledge. By synthesizing the latest research findings and identifying key areas for further investigation, this paper aims to inform and inspire future research endeavors in the intersection of reinforcement learning and energy system optimization.

**Keywords:** energy systems; reinforcement learning; optimization; deep learning



**Citation:** Stavrev, S.; Ginchev, D. Reinforcement Learning Techniques in Optimizing Energy Systems. *Electronics* **2024**, *13*, 1459. <https://doi.org/10.3390/electronics13081459>

Academic Editor: Cheng He

Received: 14 March 2024

Revised: 7 April 2024

Accepted: 10 April 2024

Published: 12 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The pursuit of energy efficiency embodies the strategic utilization of technology to minimize energy consumption while maintaining or enhancing the performance of systems across various domains, including industrial operations, power grids, civilian infrastructure, military applications, and Internet of Things (IoT) ecosystems. Achieving higher energy efficiency is paramount in the global endeavor towards sustainability, offering a multi-faceted spectrum of benefits such as significant cost reductions, diminished environmental footprints, and bolstered energy security. The automation of system operations and performance through cutting-edge artificial intelligence (AI) methodologies stands at the forefront of this quest, heralding a new era of efficiency and intelligence in energy management.

Reinforcement learning, a sophisticated branch of machine learning inspired by behavioral psychology, presents a paradigm where intelligent agents learn to make decisions autonomously to maximize cumulative rewards in novel and evolving environments. This methodology, particularly in its advanced form of deep reinforcement learning (DRL), which combines RL with the computational power of neural networks, has demonstrated remarkable proficiency in navigating complex challenges. Its applications span from mastering strategic games like chess and Go to advancing the fields of robotics and autonomous vehicular navigation, highlighting its versatility and potential.

In the realm of energy efficiency, the adaptive nature and online learning capabilities of DRL are drawing increasing interest for their ability to dynamically respond to the evolving demands of energy systems. Through varied neural network architectures, from the foundational multi-layer perceptrons to sophisticated recurrent networks like Long

Short-Term Memory (LSTM) or Gated Recurrent Units (GRUs), DRL offers a rich toolkit for modeling and optimizing energy systems under diverse and changing conditions.

As the global demand for energy surges amidst the urgency to mitigate environmental impacts and facilitate sustainable development, the complexity and unpredictability of modern energy systems have escalated. These systems, marked by variable demand, intermittent renewable sources, and the challenges of grid integration, call for innovative optimization strategies capable of navigating their inherent dynamics and uncertainties.

Reinforcement learning, with its ability to learn from direct interaction with the environment without relying on predefined models or assumptions, emerges as a potent solution to these challenges. Unlike conventional optimization techniques, RL's experiential learning approach equips it to adeptly manage the nonlinear dynamics and the unpredictability of contemporary energy systems.

This review critically assesses the efficiency of reinforcement learning in the optimization of energy systems, delving into RL's foundational principles, its relevance to energy optimization, and a comprehensive analysis of its application in this domain through the scholarly literature. This paper looks at different examples and evaluates different reinforcement learning algorithms in solving problems to highlight what RL does well and where it struggles when dealing with the complex issues of energy systems. Moreover, it ventures into the societal ramifications of deploying RL-driven optimization solutions and outlines prospects for future studies. In this review, we aim to add to the discussion about moving to cleaner energy sources and explore innovations in the optimization of energy systems.

## 2. Background and Motivation for Using RL Methods

### 2.1. Basic Formulation of Energy Efficiency

One of the base concepts in energy system optimization is the objective function. It seeks to either minimize costs or maximize efficiency, considering various operational and environmental factors. In the case of cost minimization, the formulation can be expressed as follows (1):

$$\min_x C(x) = \sum_{t=1}^T (C_t \cdot x_t + f(x_t)) \quad (1)$$

where:

- $C(x)$  is the total cost function;
- $C_t$  represents cost coefficients at time  $t$ ;
- $x_t$  are the decision variables related to energy production or consumption at time  $t$ ;
- $f(x_t)$  encapsulates other cost factors such as fuel costs and operational and maintenance costs.

Sometimes, there are capacity or operational constraints that need to be satisfied as follows (2):

$$a_t \leq x_t \leq b_t, \forall t \quad (2)$$

where  $a_t, b_t$  are lower and upper bounds for  $x_t$ . As for demand fulfillment, we use the following (3):

$$\sum_{i=1}^n x_{t,i} = D_t, \forall t \quad (3)$$

where  $D_t$  is the total demand at time  $t$ .

However, it is often the case that the optimization method needs to satisfy multiple constraints simultaneously. It is therefore difficult in practice to choose one of the conventional optimization techniques over the others.

### 2.2. Conventional Optimization Techniques

In energy system optimization, several conventional optimization methods are commonly used to address various operational and planning challenges. These methods are designed to improve efficiency, reliability, and cost-effectiveness in managing energy resources and demand.

Linear programming, for instance, is a widely used method in energy systems for optimizing a linear objective function, subject to linear equality and inequality constraints. It is particularly useful for tasks such as cost minimization and load dispatching where the relationships can be linearized. LP provides solutions that are globally optimal if the problem is convex. A similar technique is Integer Programming (IP). It extends linear programming by restricting some or all of the variables as integers. This is useful in energy system optimization for decisions that require discrete choices, like the number of generators to run or units of equipment to activate. A more advanced approach is Mixed-Integer Linear Programming (MILP)—it combines LP and IP to handle problems involving both continuous and discrete variables. MILP is extensively used in the planning and operation of power systems, including unit commitment and the scheduling of energy resources, where decisions about which power plants to run and their operating levels need to be made simultaneously. Quadratic Programming, on the other hand, is used when the objective function is quadratic, which is common in cost optimization problems involving power generation. It can optimize parabolic functions subject to linear constraints, applicable in the optimization of fuel consumption and emission levels. For problems that can be broken down into simpler subproblems and then solved recursively, dynamic programming is often used. For instance, it is used for solving multi-stage decision-making processes like hydrothermal scheduling, where the output from various power sources needs to be optimized over time. In nonlinear problems and constraints, there are techniques called Nonlinear Programming. Finally, when there are uncertainties in input data, such as future demand, fuel prices, or renewable output, stochastic optimization methods are employed. Techniques like Stochastic Programming can model these uncertainties as random variables to make more robust decisions under uncertainty.

These conventional methods have provided the backbone for decision-making in energy systems for decades, offering robust frameworks for optimizing the complex operations and planning tasks required in the energy sector.

### *2.3. Disadvantages of Conventional Optimization Methods*

Conventional optimization methods typically rely on predefined models and assumptions that may not accurately capture the complexities and dynamics of real-world energy systems. These methods often struggle to adapt to changes in the environment, such as fluctuating demand, renewable energy integration, and unforeseen operational disruptions. In addition, many conventional methods are model-dependent, requiring a precise and comprehensive understanding of all system variables and interactions. In the context of energy systems, where variables and conditions can change unpredictably (e.g., weather impacts on renewable sources), maintaining up-to-date models can be both challenging and resource-intensive. Furthermore, scaling conventional optimization methods to large, complex systems such as national power grids can be computationally expensive and inefficient. These conventional methods often face difficulties in handling the high dimensionality and the multi-objective nature of modern energy systems without significant simplifications. Finally, energy systems are increasingly influenced by stochastic elements like renewable energy sources, which introduce variability and uncertainty. Conventional methods often require complex and computationally expensive stochastic optimization techniques to address these elements, which can still fall short in real-time or highly unpredictable contexts.

### *2.4. Motivation for RL-Based Methods*

Reinforcement-learning-based methods offer several compelling advantages for optimizing energy systems, particularly due to their inherent flexibility and adaptability. Unlike conventional methods that require complete models of the environment, RL can derive optimal strategies directly through system interactions. This feature enables RL to continuously adapt to new data and changing conditions, which is essential in dynamic and complex energy systems where variables frequently change. Additionally, RL is uniquely suited to handle environments characterized by high uncertainty and variability. This capability is

crucial for effectively integrating intermittent renewable energy sources such as wind and solar, which experience significant output fluctuations due to changing weather conditions.

Moreover, RL methods excel in making real-time decisions based on current state observations, providing significant operational benefits for tasks such as demand response and real-time grid balancing. This is a notable improvement over traditional optimization methods, which often require model reruns or recalculations that are not feasible on a minute-to-minute basis. Furthermore, RL can operate effectively with minimal information about the system's dynamics, an advantage in scenarios where complete data may not be available or practical to collect.

Lastly, RL facilitates the simultaneous optimization of multiple objectives, enabling the balancing of cost, reliability, and sustainability in energy management. This capacity for multi-objective optimization aligns well with the complex trade-offs required in modern energy systems, making RL an increasingly preferred approach in the field of energy system optimization. A visual comparison is structured in Table 1.

**Table 1.** A comparison between conventional and RL-based optimization methods.

Aspect	Traditional Optimization Methods	Reinforcement Learning (RL)
Flexibility	Constrained by the need for accurate models; struggle with adapting to changes in the environment such as fluctuating demand and renewable integration.	Highly adaptable to dynamic and complex environments. Capable of continuous learning and adjustment as system variables change.
Handling Uncertainty	Often require complex and resource-intensive stochastic optimization to address uncertainties like fluctuating fuel prices or renewable outputs.	Inherently designed to handle high uncertainty and variability, making it ideal for integrating intermittent renewable energy sources like wind and solar.
Decision-Making	Decision-making is based on reruns of models or recalculations, which may not be feasible in real-time scenarios.	Excels in making real-time decisions based on current-state observations, which provide operational benefits for tasks like demand response and grid balancing.
Information Requirements	Requires detailed, comprehensive models of all system variables and interactions, which can be challenging and resource-intensive in maintaining up-to-dateness.	Can operate effectively even with minimal information about the system's dynamics, which is an advantage when complete data may not be available.
Optimization Objectives	Typically focus on a single objective like cost minimization or load dispatching; handling multiple objectives can require significant simplifications.	Supports simultaneous optimization of multiple objectives, such as minimizing cost while maximizing reliability and sustainability, aligning well with the complex trade-offs required in modern energy systems.
Computational Efficiency	Scaling to large, complex systems such as national power grids can be computationally expensive and inefficient.	Although computationally intensive, RL's ability to learn and adapt can offset the computational demand through more targeted and efficient processing.
Suitability	Well suited for stable environments with well-understood dynamics where changes are gradual and predictable.	Best suited for environments where rapid adaptation is needed, as well as for managing systems with high levels of unpredictability and operational dynamics.

### 3. Reinforcement Learning

RL encompasses foundational concepts crucial to its operation. An agent, representing the decision-maker, is described as "an abstract entity (usually a program) that can make observations, takes actions, and receives rewards for the actions taken, and transitioning to new states based on actions taken. The overarching objective for the agent lies in learning a policy that dictates optimal actions in various states, with the aim of maximizing cumulative rewards over time [1]. Given a history of such interactions, the agent must make the next

choice of action to maximize the long-term sum of rewards. To do this well, an agent may take suboptimal actions which allow it to gather the information necessary to later take optimal or near-optimal actions with respect to maximizing the long term sum of rewards" [2]. Software agents have the capability to act either by following hand-coded rules or by learning how to act through the utilization of machine learning algorithms. Reinforcement learning constitutes one such subarea of machine learning.

Formally, most reinforcement learning problems can be described through a Markov Decision Process (MDP). An MDP is delineated as the tuple  $\{S, A, T, R\}$ , where  $S$  denotes a set of states,  $A$  stands for a set of actions,  $T$  represents the transitional probability, and  $R$  denotes the reward function. Within MDP environments, a learning agent selects and executes action  $a_t \in A$  at a current state  $s \in S$  at time  $t$ . Upon transitioning to the state  $s_{t+1} \in S$  at time  $t + 1$ , the agent receives a reward  $r_t$ . The primary objective of the agent is to maximize the discounted sum of rewards from time  $t$  to infinity, referred to as the return  $R_t$ , which is defined as follows:

$$R_t = r_{t+1} + \gamma^1 r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (4)$$

where  $\gamma$  is a discount factor, specifying the degree of importance of future rewards.

The agent chooses its actions according to a policy  $\pi$ , which is a mapping from states to actions. Each policy is associated with a state-value function  $V^\pi(s)$ , which predicts the expected return for state  $s$ , when following policy  $\pi$ :

$$V^\pi(s) = E[R_t | s_t = s] \quad (5)$$

where  $E[\cdot]$  indicates the expected value. The optimal value of state  $s$ ,  $V^*(s)$ , is defined as the maximum value over all possible policies:

$$V^*(s) = \max_{\pi} E \left[ \sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s, a_t = \pi(s_t) \right] \quad (6)$$

Related to the state-value function is the action-value function  $Q^\pi(s)$ , which gives the expected return when taking action  $a$  in state  $s$  and following policy  $\pi$  thereafter:

$$Q^\pi(s, a) = E[R_t | s_t = s, a_t = a] \quad (7)$$

The optimal  $Q$ -value of state–action pair  $(s, a)$ , is the maximum  $Q$ -value over all possible policies:

$$Q^*(s, a) = \max_{\pi} E \left[ \sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s, a_t = a, a_{t>0} = \pi(s_t) \right] \quad (8)$$

An optimal policy  $\pi^*(s)$  is a policy whose state-value function is equal to (6).

In refining the RL framework and its application, researchers have emphasized the importance of accurately modeling the environment's dynamics through MDPs and continually refining the estimation of value functions and policies based on the agent's experiences. This process enables RL agents to adapt to complex, dynamic environments effectively, paving the way for innovative solutions in various domains, including energy system optimization [1].

### 3.1. Model-Based Learning

Model-based learning, a pivotal subset of reinforcement learning strategies, emphasizes the construction and utilization of an environmental model to inform decision-making processes. This approach hinges on the agent's ability to estimate or learn a model that encapsulates the dynamics of the environment, specifically the transition probabilities between states and the outcomes associated with various actions. By using this model, the agent can engage in sophisticated planning and decision-making to optimize long-term rewards. This section delves into the intricacies of model-based learning, exploring its

mechanisms, advantages, and the challenges it faces, particularly in complex domains like energy system optimization.

### 3.1.1. Estimation of Environmental Dynamics

At the heart of model-based learning lies the construction of a model that accurately represents the environment's dynamics. This model typically includes the following:

- Transition probabilities  $T(s' | s, a)$  which predict the likelihood of transitioning from a current state  $s$  to a new state  $s'$  due to the given action  $a$ .
- Reward functions  $R(s, a, s')$ , which estimate the immediate reward received after transitioning from state  $s$  to state  $s'$  due to action  $a$ .

These components are derived from observed interactions within the environment, allowing the agent to forecast future states and rewards based on its actions. The fidelity of these estimates is crucial, as it directly affects the agent's ability to make informed decisions.

### 3.1.2. Planning and Decision-Making

With a model of the environment in place, the agent employs planning algorithms to navigate the decision space efficiently. Techniques such as dynamic programming, Dyna, and Prioritized Sweeping offer structured methods for iteratively improving policy decisions based on the model's predictions [1]. More sophisticated approaches like Monte Carlo Tree Search (MCTS) and rollout algorithms expand the agent's capability to explore and evaluate complex action sequences, leading to the formulation of optimal or near-optimal policies. These methods balance the exploration of uncharted actions with the exploitation of known strategies to maximize cumulative rewards.

### 3.1.3. Advantages of Model-Based Learning

The strategic advantage of model-based learning resides in its predictive capacity, which enables an agent to anticipate the consequences of actions without needing to physically execute them in the environment. This foresight allows for a more efficient use of available data, reducing the sample size required to achieve effective policies. Furthermore, by capturing the environment's dynamics, model-based methods can adapt to changes or uncertainties in the environment with greater agility, enhancing the agent's performance and robustness in dynamic settings.

### 3.1.4. Challenges and Considerations

Despite these advantages, model-based learning faces significant challenges, particularly in environments characterized by complexity and uncertainty, such as energy systems. The accuracy of the environmental model is paramount; however, capturing the intricate dynamics of complex systems can be exceedingly difficult. Errors in the model can lead to suboptimal decision-making, undermining the efficacy of the approach. Moreover, the memory and computational requirements for maintaining and updating the model can be substantial, particularly for large-state and action spaces, where the complexity of the model may grow quadratically with the size of the state space and linearly with the action space, noted as  $O(|S|^2|A|)$ .

Recent advancements in model-based RL, including the integration of deep learning techniques, have shown promise in addressing some of these challenges. Algorithms like Deep Dyna-Q [3] and Model-Based Policy Optimization (MBPO) [4] incorporate neural networks to enhance the accuracy of environmental models and the efficiency of policy optimization, even in complex and high-dimensional spaces. These innovations have extended the applicability of model-based learning, opening new avenues for optimizing energy systems through more accurate and scalable modeling techniques.

## 3.2. Model-Free Learning

Model-free learning strategies represent a pivotal branch of reinforcement learning, being particularly effective when an explicit model of the environment is not available or is

too complex to formulate. Unlike model-based methods that require an understanding of the environment's dynamics for planning and decision-making, model-free approaches learn optimal policies through direct interaction with the environment. In this learning paradigm, an agent updates state or state–action values based on observed samples. In that way, the need for a pre-learned model of the environment is not required. The agent relies instead on the accumulation of samples to guide its decision-making processes towards optimizing the cumulative rewards over time [1].

Central to model-free learning is the notion of learning through trial and error, where the agent iteratively refines its policy based on the feedback received from the environment in the form of rewards. This method enables the agent to adaptively navigate the decision space and converge towards an optimal or near-optimal policy, even in the face of uncertainty and complexity inherent in the environment. The ability to learn without a model is particularly advantageous in dynamic systems, such as sustainable energy and electric systems, where the state space can be vast or continuous, and the environment's dynamics are influenced by stochastic elements like renewable energy sources and variable consumer demand [5].

### Prominent Algorithms in Model-Free Learning

A classic example of model-free learning algorithm is Q-learning. It learns by updating the value of state–action pairs using observed rewards and the estimated future value, without requiring a model of the environment's dynamics. The evolution of Q-learning into Deep Q-Networks (DQNs) employs deep neural networks to manage high-dimensional state spaces, thereby extending the applicability of Q-learning to more complex scenarios. Double Q-learning, introduced by van Hasselt [6], and further refined in the Double DQN (DDQN) framework [7], addresses the overestimation bias observed in traditional Q-learning by maintaining two separate estimators (Q-tables) and updating them alternately, enhancing the accuracy of value approximation. Another example is Policy Gradient Methods. They offer an alternative approach within model-free learning, optimizing policy parameters directly to maximize expected cumulative rewards. This category includes algorithms like Proximal Policy Optimization (PPO) and Trust Region Policy Optimization (TRPO), which have been instrumental in advancing the application of RL in continuous action spaces. These methods are particularly suited for optimizing the control and operation of energy systems, where actions can be continuous, and the objective is to enhance system efficiency and reliability [8]. Yet, another example is True Online Temporal-Difference Learning, as detailed by [9]. This method exemplifies the continuous refinement of model-free methods, offering a more sophisticated approach to updating value functions that combines the benefits of both TD ( $\lambda$ ) and Q-learning. This method represents a significant advancement in the efficiency and effectiveness of model-free learning algorithms, further broadening their potential application areas.

The application of model-free learning to energy systems is motivated by the need for adaptive solutions capable of managing the uncertainty and variability introduced by renewable energy sources and fluctuating demand. By employing algorithms like DQNs and PPO, RL can optimize energy consumption, improve demand response, and enhance the overall efficiency of energy systems without relying on precise models of system dynamics, thus overcoming the limitations of traditional optimization techniques [5]. Despite their effectiveness, model-free methods come with their computational and memory requirements, typically necessitating space proportional to the product of the number of states and actions, denoted as  $O(|S| |A|)$ . However, advancements in computational resources and algorithmic efficiency have made these methods increasingly feasible for real-world applications, including those in sustainable energy and electric systems.

### 3.3. RL Relevance to Energy System Optimization

Reinforcement learning techniques have garnered considerable attention for their potential applicability to energy system optimization tasks. One area of interest lies in

demand-side management, where RL algorithms can dynamically adjust energy consumption patterns in response to changing conditions, thereby enhancing overall system efficiency and reliability [10]. For instance, RL-based control strategies have been explored for load balancing in smart grids, optimizing the operation of distributed energy resources (DERs), and scheduling energy-intensive tasks in industrial facilities [11,12]. Moreover, RL has shown promise in addressing complex optimization problems in power generation and distribution. By learning optimal control policies from historical data and real-time feedback, RL algorithms can improve the dispatch of renewable energy sources, reduce transmission losses, and enhance grid stability [13]. Recent studies have investigated the integration of RL with advanced control techniques to optimize the operation of wind turbines, microgrids, and energy storage systems [13,14]. Furthermore, RL-based approaches have been applied to energy market optimization, where they can facilitate strategic decision-making and risk management for market participants. By learning from historical market data and simulating future scenarios, RL models can support energy traders in optimizing bidding strategies, portfolio management, and hedging against price volatility [15,16]. Research in this area has emphasized the need for RL algorithms capable of handling large-scale, uncertain environments and adapting to evolving market dynamics [15].

In addition to operational optimization, RL holds promise for addressing long-term planning and policy-making challenges in the energy sector. By modeling the interactions between different stakeholders, infrastructure investments, and regulatory frameworks, RL-based simulations can inform decision-makers about the potential impacts of alternative strategies on energy affordability, environmental sustainability, and social equity [10]. Future research in this domain is expected to focus on integrating RL with system dynamics models, multi-agent simulations, and optimization algorithms to support holistic energy planning and policy analysis [10,17].

Overall, the versatility and adaptability of RL make it a promising tool for addressing diverse optimization challenges in energy systems. By applying advances in algorithmic techniques, computational resources, and domain-specific knowledge, researchers can harness the full potential of RL to drive innovation and transformation in the energy sector.

#### 4. Challenges in Energy System Optimization

Optimizing energy systems is a multifaceted endeavor critical for ensuring sustainable energy supply, reducing environmental impacts, and meeting growing global energy demands. However, this pursuit is fraught with challenges stemming from the inherent complexity, dynamics, and uncertainties characterizing energy systems. The complexity arises from the interplay of various factors, including fluctuating energy demand patterns, the integration of intermittent renewable energy sources, and the complexities of grid management [18]. The integration of renewable energy sources, such as solar and wind power, presents a particularly daunting challenge. These sources exhibit inherent intermittency and variability, making their integration into the grid a nontrivial task [19]. The unpredictable nature of renewable energy generation necessitates advanced forecasting techniques and robust management strategies to ensure grid stability and reliability. Furthermore, the growing decentralization of energy generation, coupled with the proliferation of distributed energy resources like rooftop solar panels and electric vehicles, adds layers of complexity to grid integration efforts [20]. In addition, grid integration issues extend beyond managing renewable energy variability to encompass the broader challenge of balancing energy supply and demand in real time while maintaining grid stability. The increasing complexity of the grid, coupled with the need to accommodate diverse energy resources and fluctuating demand patterns, underscores the importance of grid flexibility and resilience [20]. However, traditional optimization approaches, while useful in many contexts, often fall short in addressing the complexities and uncertainties inherent in modern energy systems. These approaches typically rely on simplified models and assumptions that may not capture the nuances of real-world energy dynamics [21]. As

a result, they may struggle to adapt to dynamic changes in energy supply and demand, leading to suboptimal solutions.

In the face of these challenges, innovative approaches are needed to enhance energy system optimization. Reinforcement learning emerges as a promising solution for addressing the complexities inherent in sustainable energy and electric systems. As a versatile class of optimal control methods, experiences, simulations, or searches may use reinforcement learning to derive value estimates in highly dynamic, stochastic environments. Its interactive nature fosters robust learning capabilities and the ability to adapt without the need for explicit models of system dynamics. This property makes reinforcement learning particularly well suited for addressing the complex nonlinearities and uncertainties present in sustainable energy and electric systems.

## 5. Applications of Reinforcement Learning in Energy Systems

In this section, we outline the prominent applications of reinforcement learning techniques in energy systems, providing examples and insight from recent papers.

### 5.1. Demand Response Optimization

Demand response represents a pivotal strategy within modern energy systems and aims to adjust consumer power consumption to match supply availability, enhance grid stability, and optimize energy costs. The integration of reinforcement learning into demand response optimization processes has emerged as a transformative approach, using the capability of RL algorithms to dynamically adapt to the fluctuating nature of energy markets and grid conditions [22].

RL's core strength in demand response optimization lies in its ability to learn and adapt strategies based on real-time data and feedback loops. By continuously interacting with the energy system, RL agents develop strategies that encourage or discourage power usage during specific periods, effectively shifting demand to off-peak times or when renewable energy availability is high. This adaptive learning process is crucial for maintaining grid balance and optimizing energy costs in the face of renewable energy integration and varying demand patterns [23]. Several examples underscore the potential of RL in demand response optimization. For instance, a study by Zhang et al. [24] demonstrated how an RL-based system could effectively reduce peak demand and energy costs in a residential community by dynamically adjusting the operation of HVAC systems and electric vehicles in response to price signals. Similarly, research by Mocanu et al. [25] employed deep Q-networks to optimize the charging schedules of electric batteries, highlighting significant improvements in energy savings and peak shaving.

The application of RL in demand response optimization offers a promising avenue for enhancing the efficiency and sustainability of energy systems. Through its capacity for adaptive learning and dynamic decision-making, RL facilitates the development of sophisticated DR strategies that can respond effectively to the challenges posed by renewable integration, variable demand, and the evolving landscape of energy markets. As RL algorithms continue to advance, their role in demand response and broader energy system optimization is poised to expand, heralding a new era of intelligent energy management.

One of the primary challenges is the accurate modeling of consumer behaviors, which are influenced by a wide range of factors including personal preferences, habits, and responsiveness to DR signals. Traditional RL models may struggle to capture this complexity, leading to suboptimal DR strategies. A promising solution lies in the adoption of multi-agent reinforcement learning (MARL) frameworks, where each agent can represent an individual consumer or a specific group of devices within the energy system [19]. This granular approach allows for the modeling of diverse consumer behaviors and interactions within the system, enhancing the overall effectiveness of DR strategies. In addition, studies by [26] provide a robust framework for managing power distribution across networked microgrids efficiently and safely, ensuring that both local and global constraints are met. The later research presents a Supervised Multi-Agent Safe Policy Learning (SMAS-PL)

method aimed at optimizing power management in networked microgrids with a focus on maintaining safe operational practices. It addresses the common challenges in reinforcement learning where black-box models may not adhere to operational constraints, potentially leading to unsafe grid conditions. Unlike conventional RL that might overlook crucial operational constraints, this method integrates constraints directly into the policy-learning process. It utilizes gradient information from these constraints to ensure that the policy decisions are both optimal and feasible under grid operational limits. A distributed consensus-based optimization algorithm is introduced for training policy functions across multiple agents. The approach significantly reduces the need for re-solving complex optimization problems and offers a scalable solution to real-time decision-making in power distribution.

The fluctuating nature of energy prices, driven by market dynamics and the variability of renewable energy production, poses another significant challenge. Accurate price prediction is crucial for effective DR, as it informs the decision-making process regarding when to encourage or discourage energy consumption. Advanced RL models that incorporate deep learning techniques, such as Deep Q-Networks (DQN), have shown promise in improving the accuracy of price predictions. These models can process high-dimensional data and learn complex patterns, enabling more precise forecasting of energy prices and better informed DR strategies [27]. Furthermore, another model-free methodology [28] explores the utilization of reinforcement learning framework for managing power in networked microgrids under incomplete information scenarios, making traditional model-based optimization challenging. This approach employs a distinctive bi-level hierarchical structure, unlike the typical single-agent, flat-structure setup in standard RL, and innovatively handles incomplete information through aggregated data and predictive modeling. Additionally, it incorporates advanced adaptive learning techniques with a forgetting factor, allowing it to adjust to changing system conditions in real time, a significant enhancement over simpler RL methods that cannot dynamically adapt without retraining. This adaptive capability makes the system robust to changes in system parameters and operational conditions. In addition, the RL approach respects the privacy of microgrid data by not requiring detailed user or operational data, aligning with concerns about data privacy and confidentiality in smart grids. Compared to traditional optimization methods, the RL approach shows better adaptability and faster computational times because it learns from past experiences and predicts optimal power distribution without detailed system models.

Maintaining user comfort while optimizing energy consumption is a critical concern in DR. Aggressive DR strategies may lead to discomfort or inconvenience, undermining user participation and satisfaction. To address this issue, RL models must be designed with mechanisms to balance energy savings against comfort criteria. This can be achieved by incorporating user feedback into the learning process, allowing the RL model to adjust its strategies based on user preferences and comfort levels. Furthermore, reward functions in RL algorithms can be carefully designed to penalize actions that significantly compromise user comfort, ensuring that the optimization process remains aligned with user satisfaction goals.

The integration of renewable energy sources introduces additional variability and uncertainty into the energy system. RL algorithms must be capable of adapting to rapid changes in energy availability, especially from sources like solar and wind power, which are highly dependent on environmental conditions. Techniques such as robust reinforcement learning and stochastic optimization models have been developed to enhance the resilience of RL-based DR strategies to the uncertainties associated with renewable energy [23].

Overall, despite the considerable challenges, ongoing advancements in algorithmic techniques and model architectures offer promising solutions in this energy domain. Through the adoption of multi-agent systems, deep learning enhancements, user-centric models, and robust optimization techniques, RL can effectively address the complexities of demand response, paving the way for more efficient, reliable, and user-friendly energy management systems.

### 5.2. Renewable Energy Integration

The integration of renewable energy sources into existing power grids represents a crucial step towards achieving sustainability and reducing dependency on fossil fuels. However, the intermittent and unpredictable nature of renewable energy sources like wind and solar power poses significant challenges to energy grid stability and efficiency. Reinforcement learning offers a suite of methodologies for addressing these challenges, enabling more effective integration of renewable energies through adaptive and intelligent control systems.

Accurate forecasting of renewable energy production is important for effective grid integration. Deep learning techniques have shown significant promise in predicting energy output from renewable sources. By continuously learning from historical data and real-time environmental conditions, these algorithms adaptively improve their predictions, accounting for the variability inherent in renewable energy production, such as in the work of [24]. Furthermore, energy storage systems play a critical role in mitigating the intermittency of renewable energy sources. There are examples of RL algorithms used to optimize the charging and discharging cycles of these storage systems, ensuring that stored energy is available during periods of high demand or low renewable production. As we discussed, model-free RL methods are able to learn optimal storage management policies, thus helping to smooth out fluctuations into the grid [29].

The integration of renewables requires adaptive grid control mechanisms capable of managing the variability and uncertainty associated with these energy sources. RL algorithms offer a dynamic solution by learning to adjust grid operations in response to changing energy production and demand patterns. This includes optimizing the dispatch of renewable energy, managing load balancing, and ensuring grid stability. Multi-agent systems, where different agents represent various components of the energy system, enable a coordinated approach to grid control [27].

Another application of RL in renewable energy integration is demand-side management (DSM). By influencing energy consumption patterns on the demand side, RL algorithms can help match demand with the availability of renewable energy. This might involve shifting energy-intensive processes to times of high renewable production or incentivizing reduced consumption during shortages. RL-driven DSM strategies not only support the integration of renewables but also promote energy efficiency and conservation among consumers [19].

While RL offers powerful tools for renewable energy integration, challenges remain in terms of scalability, data quality, and model interpretability. Future research directions include the development of more sophisticated RL models that can handle complex, multi-dimensional energy systems, the integration of RL with other artificial intelligence techniques for enhanced prediction and optimization, and the exploration of ways to make RL models more transparent and interpretable to facilitate their adoption in energy system management.

### 5.3. Smart Grid Applications

Smart grid technologies aim to modernize traditional power systems by integrating advanced communication, control, and monitoring capabilities. The advent of smart grids heralds a significant leap toward enhancing the efficiency, reliability, and sustainability of power systems. Smart grid optimization encompasses a wide array of functionalities including but not limited to real-time demand response, distributed energy resource management, and advanced metering infrastructure. Reinforcement learning is a pivotal technology in this arena, offering dynamic and adaptive solutions to the multifaceted challenges faced by smart grids.

Smart grids facilitate a more interactive approach to demand response and load balancing, which is crucial for maintaining grid stability and efficiency. There are several examples in that area that excel in balancing energy supply and demand in real time. Through continuous interaction with the grid and analysis of consumption patterns, RL

models can predict peak-load periods and adjust demand accordingly, either by directly controlling smart appliances or through pricing incentives to consumers [30]. The best DRL model, as identified by Gallego et al. [31], achieves a complete listing of optimal actions for the forthcoming hour 90% of the time. This level of precision underscores the potential of DRL in enhancing the flexibility of smart grids, providing a robust mechanism for adjusting grid operations in response to real-time conditions and forecasts. This predictive prowess is pivotal for maintaining operational efficiency and optimizing the dispatch of energy resources within the grid.

Furthermore, RL enables the dynamic allocation of energy resources to where they are most needed, ensuring optimal load distribution across the grid. The management of distributed energy resources, such as rooftop solar panels, wind turbines, and battery storage systems, is another critical aspect of smart grid optimization. RL algorithms are adept at optimizing the operation of DERs, enhancing grid resilience and facilitating the integration of renewable energy sources. By dynamically adjusting the dispatch of DERs based on current grid conditions and forecasted demand, RL contributes to a more flexible and responsive grid system [32]. Gallego et al. [31] illustrate the application of deep reinforcement learning techniques, specifically Deep Q-Networks (DQNs), to select optimal actions for managing grid components.

Maintaining optimal voltage and frequency levels is essential for grid stability and the efficient operation of electrical devices. RL algorithms, through their ability to learn and adapt from environmental feedback, can be effectively employed for real-time voltage and frequency regulation. By continuously monitoring grid conditions and adjusting control mechanisms, such as capacitor banks and voltage regulators, RL ensures that voltage and frequency remain within desired ranges, even under fluctuating demand and generation conditions [22].

Despite the promising applications of RL in smart grid optimization, several challenges remain, including the scalability of RL solutions, integration with existing grid infrastructure, and the protection of consumer privacy in data-driven applications. Addressing these challenges requires ongoing research and development, as well as collaboration between industry, academia, and regulatory bodies.

Future directions in smart grid optimization with RL include the exploration of more advanced RL techniques, such as deep reinforcement learning and multi-agent systems, to manage the increasing complexity of smart grids. In conclusion, RL offers a versatile and powerful toolset for optimizing smart grid operations across a range of dimensions. Through its adaptive and predictive capabilities, RL not only enhances grid efficiency and reliability but also plays a crucial role in the transition toward more sustainable and resilient energy systems.

#### *5.4. Grid Management and Control*

Grid management and control are essential for ensuring the stability, reliability, and efficiency of modern power systems. Reinforcement learning techniques offer valuable tools for optimizing grid operation, addressing challenges such as voltage control, reactive power management, and distribution system optimization.

One notable application of RL in grid management is voltage control, which involves regulating voltage levels within acceptable limits to ensure the safe and efficient operation of the grid. RL algorithms can optimize voltage control strategies by adjusting the set points of voltage regulators and reactive power devices in real time. For example, a study by [33] applied RL techniques to optimize voltage control in distribution networks, achieving improved voltage stability and reduced energy losses.

Another important aspect of grid management is reactive power management, which involves the control of reactive power flow to maintain system stability and voltage regulation. RL-based controllers can optimize reactive power dispatch strategies, minimizing system losses and improving voltage stability. For instance, a study by Wang et al. [34] developed an RL-based reactive power dispatch algorithm for power systems, achieving

an improved voltage profile and reduced system losses. Furthermore, RL techniques can be applied to optimize distribution system operation, particularly in the context of integrating distributed energy resources such as solar panels, wind turbines, and energy storage systems. RL-based controllers can optimize DER dispatch strategies, maximizing renewable energy utilization and grid reliability. For example, a study by Kumar et al. [35] used RL techniques to optimize the operation of a microgrid with renewable energy sources, achieving improved grid stability and reduced energy costs. Moreover, RL algorithms can be applied to address challenges related to grid congestion and load balancing. By optimizing the scheduling of energy flows and grid assets, RL-based controllers can alleviate congestion and improve grid efficiency. For example, a study by [36] applied RL techniques to optimize energy scheduling in distribution networks, achieving reduced congestion and improved grid reliability.

### 5.5. Summary

The summarized RL techniques and their applications are presented in Table 2.

**Table 2.** Reinforcement learning applications in energy systems.

RL Method	Applicable Energy System Scenarios	Benefits	Challenges
Deep Learning Techniques	All energy systems, including microgrids, transmission, and distribution systems.	Enhance predictive accuracy for demand and supply variations; optimize operational efficiency.	Require extensive training data; computationally intensive; may overfit without proper tuning.
Deep Q-Networks (DQNs)	Smart grids, particularly in demand response and battery management.	Offer robust decision-making under uncertainty; effective in policy optimization for load balancing.	Prone to overestimation of Q-values; require large and diverse datasets to train effectively.
Multi-Agent RL (MARL)	Distributed systems including microgrids and decentralized smart grids.	Facilitates cooperative control and decentralized decision-making; improves resilience and flexibility.	Coordination complexity increases with number of agents; risk of conflicting objectives.
Policy Gradient Methods	Voltage and frequency regulation in smart grids.	Directly optimize policy; capable of handling continuous action spaces.	Suffer from high variance in gradient estimates; slow convergence in environments with high-dimensional action spaces.

The table provides a comprehensive comparison of various RL methods and their applications across different energy system scenarios, outlining both benefits and challenges associated with each method. Deep learning techniques are applied universally across energy systems, enhancing predictive accuracy and operational efficiency, but require extensive data and significant computational resources. Deep Q-Networks (DQNs) are particularly effective in smart grids for demand response and battery management, offering robust decision-making capabilities, though they are susceptible to the overestimation of Q-values. Multi-Agent RL (MARL) is ideal for distributed systems like microgrids, promoting cooperative control and decentralized decision-making, yet faces challenges with coordination complexity and potential objective conflicts. Lastly, Policy Gradient Methods are utilized in smart grids for voltage and frequency regulation due to their direct policy optimization, but their application is hindered by slow convergence and high variance in gradient estimates. The presented comparison underscores the suitability of each RL approach depending on the specific needs and constraints of the energy system scenario, highlighting the critical balance between their benefits and the inherent challenges.

## 6. Discussion

In recent years, the application of reinforcement learning techniques in optimizing energy systems has witnessed several notable trends and phenomena, driven by advancements in technology, shifts in energy policies, and emerging challenges in the energy sector.

One trend is the increasing integration of RL algorithms with advanced data analytics techniques, such as machine learning and deep learning [3,7,24,37]. This integration enables RL-based controllers to use large volumes of data to learn complex patterns and relationships in energy systems, leading to improved decision-making and optimization performance. Furthermore, the rise of edge computing and Internet of Things (IoT) devices has enabled RL algorithms to be deployed directly at the device level, allowing for real-time control and optimization of distributed energy resources and smart grid components.

Another trend is the growing emphasis on decentralized and distributed energy systems, driven by the proliferation of renewable energy sources, advancements in energy storage technologies, and evolving consumer preferences. RL techniques play a crucial role in optimizing the operation of distributed energy resources, microgrids, and virtual power plants, enabling greater grid flexibility, resilience, and sustainability. Furthermore, the emergence of peer-to-peer energy trading platforms and community-based energy initiatives presents new opportunities for RL-based optimization and coordination among energy prosumers.

Effective RL applications depend significantly on the quality, granularity, and timeliness of the data collected. Diverse data sources, from real-time sensor outputs in smart grids to historical energy consumption records, provide the necessary inputs for training and refining RL models. The integration of IoT devices [38] and smart meters has revolutionized data collection, enabling more precise and continuous streams of information. These technologies not only facilitate the accurate modeling of energy demand and supply dynamics [39] but also support the training of RL algorithms that can predict and adapt to complex energy patterns efficiently. Additionally, advanced data preprocessing techniques such as normalization, anomaly detection, and feature engineering are essential to prepare raw data for effective learning and performance optimization [40]. This comprehensive data infrastructure supports the adaptive and predictive capabilities of RL models, ultimately driving their success in optimizing energy systems.

Moreover, the increasing complexity and interconnectedness of modern energy systems pose significant challenges for traditional optimization methods, which often struggle to handle nonlinear dynamics, uncertainty, and stochastic environments. RL algorithms offer a promising alternative by providing adaptive, model-free optimization approaches that can learn and adapt to changing system conditions over time. Additionally, the application of RL techniques in multi-agent systems and game-theoretic frameworks opens up new avenues for addressing strategic interactions and market dynamics in energy systems.

One interesting phenomenon is the convergence of RL with other emerging technologies, such as blockchain and quantum computing, to address key challenges in energy system optimization. Blockchain technology offers decentralized and transparent mechanisms for peer-to-peer energy trading, while quantum computing promises exponential gains in computational power for solving complex optimization problems. By integrating RL with these technologies, researchers and practitioners can explore novel approaches for optimizing energy systems at scale, leveraging the strengths of each technology to address specific challenges and constraints.

Furthermore, the increasing focus on energy efficiency and sustainability has led to the development of novel RL-based approaches for optimizing energy consumption and reducing environmental impact. RL algorithms can learn adaptive control strategies for energy-intensive processes, such as industrial manufacturing and HVAC systems, to minimize energy waste and improve overall efficiency. Additionally, RL-based controllers can optimize energy consumption in smart buildings and homes, by using real-time data and user preferences to achieve significant energy savings without compromising comfort or functionality.

In addition to traditional RL algorithms, there is a growing interest in meta-learning and transfer learning techniques for energy system optimization. Meta-learning enables RL agents to learn how to learn, adapting quickly to new environments and tasks with limited data. Transfer learning allows RL models trained on one task or domain to be transferred and fine-tuned for related tasks or domains, accelerating the learning process and improving generalization performance. These approaches hold promise for addressing data scarcity and domain-specific challenges in energy system optimization, particularly in scenarios where labeled data or expert knowledge is limited.

Moreover, the democratization of RL tools and platforms has made them more accessible to researchers, engineers, and practitioners in the energy sector. Open source RL libraries, such as TensorFlow 2.16.1, PyTorch 2.2, and OpenAI Gym 0.26.2, provide user-friendly interfaces and pre-trained models that enable rapid prototyping and experimentation. Additionally, cloud-based RL platforms offer scalable computing resources and collaborative environments for developing and deploying RL-based solutions for energy system optimization. This democratization of RL technology is driving innovation and empowering stakeholders across the energy value chain to explore new opportunities for efficiency improvement and sustainability.

However, despite the significant progress and advancements in RL-based energy system optimization, several challenges and limitations remain. One major challenge is the scalability and computational complexity of RL algorithms, particularly in large-scale energy systems with millions of decision variables and uncertain dynamics. Addressing this challenge requires the development of scalable RL algorithms, distributed optimization techniques, and efficient approximation methods that can handle the complexity and heterogeneity of real-world energy systems. Another challenge is the interpretability and transparency of RL models, which are often viewed as black boxes due to their complex decision-making processes and nonlinear dynamics. Ensuring the accountability and trustworthiness of RL-based controllers is crucial for gaining acceptance and adoption in safety-critical applications, such as energy grid management and control. Developing explainable RL techniques and model validation frameworks is essential for providing insights into the decision-making process and fostering trust among stakeholders.

In summary, while RL holds great promise for optimizing energy systems and advancing sustainability goals, addressing the remaining challenges and limitations will require interdisciplinary collaboration, innovative research, and real-world experimentation. By harnessing the power of RL algorithms, energy stakeholders can unlock new opportunities for efficiency improvement, grid reliability, and environmental stewardship, paving the way for a more sustainable and resilient energy future.

## 7. Conclusions

The application of reinforcement learning in optimizing energy systems offers transformative potential, addressing key challenges within the sector such as efficiency, reliability, and sustainability. This research has demonstrated that RL's capability to adapt to dynamic environments and handle complex, multi-objective optimization tasks can significantly enhance how energy systems operate. Particularly, RL's proficiency in real-time decision-making and its robustness against uncertainties prove advantageous over traditional optimization methods. These attributes facilitate the integration of renewable energy sources, optimize demand response, and improve grid management, thereby supporting the transition to more sustainable energy systems. However, challenges such as scalability, computational demand, and the need for interpretable models remain and must be addressed through continued interdisciplinary research. Looking forward, the implementation of RL could reshape energy management practices, making them more adaptive, efficient, and aligned with global sustainability goals.

**Author Contributions:** Conceptualization, S.S.; literature search, S.S.; review and analysis, S.S. and D.G.; citations and references, D.G.; discussion, D.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Research and Development Sector at the Technical University of Sofia.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** This research is supported by the Aerospace Equipment and Technologies Laboratory at the Technical University of Sofia and the Human–Computer Interactions and Simulation Laboratory (HSL) at the University of Plovdiv “P. Hilendarski”.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
2. Langford, J. Efficient exploration in reinforcement learning. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2011; pp. 1–5. [\[CrossRef\]](#)
3. Liu, Y.; Swaminathan, A.; Liu, Z. Deep Dyna-Q: Integrating Planning for Task-Completion Dialogue Policy Learning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019. [\[CrossRef\]](#)
4. Janner, M.; Fu, J.; Zhang, M.; Levine, S. When to Trust Your Model: Model-Based Policy Optimization. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 12519–12530. [\[CrossRef\]](#)
5. Yang, T.; Zhao, L.; Li, W.; Zomaya, A.Y. Reinforcement learning in sustainable energy and electric systems: A survey. *Annu. Rev. Control* **2020**, *49*, 145–163. [\[CrossRef\]](#)
6. Van Hasselt, H. Double Q-learning. Advances in Neural Information Processing Systems 23. In Proceedings of the 24th Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–9 December 2010; Volume 2010, pp. 2613–2621.
7. Van Hasselt, H.; Guez, A.; Silver, D. Deep reinforcement learning with double Q-learning. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; p. 30. [\[CrossRef\]](#)
8. Schulman, J.; Levine, S.; Abbeel, P.; Jordan, M.; Moritz, P. Trust region policy optimization. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; Volume 37, pp. 1889–1897.
9. Van Seijen, H.; Mahmood, A.R.; Pilarski, P.M.; Machado, M.C.; Sutton, R.S. True online temporal-difference learning. *J. Mach. Learn. Res.* **2016**, *17*, 1–40.
10. Iqbal, S.; Sarfraz, M.; Ayyub, M.; Tariq, M.; Chakraborty, R.K. A comprehensive review on residential demand side management strategies in smart grid environment. *Sustainability* **2021**, *13*, 7170. [\[CrossRef\]](#)
11. Ali, K.H.; Sigalo, M.; Das, S.; Anderlini, E.; Tahir, A.A.; Abusara, M. Reinforcement Learning for Energy-Storage Systems in Grid-Connected Microgrids: An Investigation of Online vs. Offline Implementation. *Energies* **2021**, *14*, 5688. [\[CrossRef\]](#)
12. Paudel, A.; Hussain, S.A.; Sadiq, R.; Zareipour, H.; Hewage, K. Decentralized cooperative approach for electric vehicle charging. *J. Clean. Prod.* **2022**, *364*, 132590. [\[CrossRef\]](#)
13. Puech, A.; Read, J. An improved yaw control algorithm for wind turbines via reinforcement learning. In *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2022*; Amini, M.R., Canu, S., Fischer, A., Guns, T., Novak, P.K., Tsoumakas, G., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2023; Volume 13717. [\[CrossRef\]](#)
14. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M.; Slootweg, J.G. On-line building energy optimization using deep reinforcement learning. *IEEE Trans. Smart Grid* **2018**, *9*, 3254–3264. [\[CrossRef\]](#)
15. Zhang, Z.; Zhang, D.; Qiu, R.C. Deep reinforcement learning for power system applications: An overview. *Front. Inf. Technol. Electron. Eng.* **2019**, *20*, 1358–1372. [\[CrossRef\]](#)
16. Glavic, M. (Deep) Reinforcement learning for electric power system control and related problems: A short review and perspectives. *Annu. Rev. Control* **2019**, *48*, 22–35. [\[CrossRef\]](#)
17. Alabi, T.M.; Aghimien, E.I.; Agbajor, F.D.; Yang, Z.; Lu, L.; Adeoye, A.R.; Gopaluni, B. A review on the integrated optimization techniques and machine learning approaches for modeling, prediction, and decision making on integrated energy systems. *Renew. Energy* **2022**, *194*, 822–849. [\[CrossRef\]](#)
18. DeCarolis, J.; Daly, H.; Dodds, P.; Keppo, I.; Li, F.; McDowall, W.; Pye, S.; Strachan, N.; Trutnevyte, E.; Usher, W.; et al. Formalizing best practice for energy system optimization modelling. *Appl. Energy* **2017**, *194*, 184–198. [\[CrossRef\]](#)
19. Palensky, P.; Dietrich, D. Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads. *IEEE Trans. Ind. Inform.* **2011**, *7*, 381–388. [\[CrossRef\]](#)
20. Cicilio, P.; Glennon, D.; Mate, A.; Barnes, A.; Chalishazar, V.; Cotilla-Sanchez, E.; Vaagensmith, B.; Gentle, J.; Rieger, C.; Wies, R.; et al. Resilience in an evolving electrical grid. *Energies* **2021**, *14*, 694. [\[CrossRef\]](#)
21. Rehman, A.U.; Wadud, Z.; Elavarasan, R.M.; Hafeez, G.; Khan, I.; Shafiq, Z.; Alhelou, H.H. An optimal power usage scheduling in smart grid integrated with renewable energy sources for energy management. *IEEE Access* **2021**, *9*, 9448087. [\[CrossRef\]](#)
22. Vázquez-Canteli, J.R.; Nagy, Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Appl. Energy* **2019**, *235*, 1072–1089. [\[CrossRef\]](#)
23. Ruelens, F.; Claessens, B.J.; Vandael, S.; De Schutter, B.; Babuska, R.; Belmans, R. Residential demand response of thermostatically controlled loads using batch Reinforcement Learning. *IEEE Trans. Smart Grid* **2017**, *8*, 2149–2159. [\[CrossRef\]](#)

24. Zhang, C.; Wang, X.; Li, F.; He, Q.; Huang, M. Deep learning-based network application classification for SDN. *Trans. Emerg. Telecommun. Technol.* **2018**, *29*, e3302. [[CrossRef](#)]
25. Mocanu, D.C.; Mocanu, E.; Stone, P.; Nguyen, P.H.; Gibescu, M.; Liotta, A. Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. *Nat. Commun.* **2018**, *9*, 2383. [[CrossRef](#)]
26. Zhang, Q.; Dehghanpour, K.; Wang, Z.; Qiu, F.; Zhao, D. Multi-Agent Safe Policy Learning for Power Management of Networked Microgrids. *IEEE Trans. Smart Grid* **2021**, *12*, 1048–1062. [[CrossRef](#)]
27. Francois-Lavet, V.; Henderson, P.; Islam, R.; Bellemare, M.G.; Pineau, J. An introduction to deep reinforcement learning. *Found. Trends Mach. Learn.* **2018**, *11*, 219–354. [[CrossRef](#)]
28. Zhang, Q.; Dehghanpour, K.; Wang, Z. A Learning-Based Power Management Method for Networked Microgrids under Incomplete Information. *IEEE Trans. Smart Grid* **2020**, *11*, 1193–1204. [[CrossRef](#)]
29. Deng, R.; Yang, Z.; Chow, M.-Y.; Chen, J. A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches. *IEEE Trans. Ind. Inform.* **2019**, *11*, 570–582. [[CrossRef](#)]
30. Siano, P. Demand response and smart grids—A survey. *Renew. Sustain. Energy Rev.* **2014**, *30*, 461–478. [[CrossRef](#)]
31. Gallego, F.; Martín, C.; Díaz, M.; Garrido, D. Maintaining flexibility in smart grid consumption through deep learning and deep reinforcement learning. *Energy AI* **2023**, *13*, 100241. [[CrossRef](#)]
32. Dall’Anese, E.; Simonetto, A. Optimal Power Flow Pursuit. *IEEE Trans. Smart Grid* **2018**, *9*, 942–959. [[CrossRef](#)]
33. Meng, X.; Zhang, P.; Xu, Y.; Xie, H. Construction of decision tree based on C4.5 algorithm for online voltage stability assessment. *Int. J. Electr. Power Energy Syst.* **2020**, *117*, 105668. [[CrossRef](#)]
34. Wang, N.; Li, J.; Hu, W.; Zhang, B.; Huang, Q.; Chen, Z. Optimal reactive power dispatch of a full-scale converter based wind farm considering loss minimization. *Renew. Energy* **2019**, *136*, 317–328. [[CrossRef](#)]
35. Kumar, S.; Krishnasamy, V.; Kaur, R.; Kandasamy, N.K. Virtual energy storage-based energy management algorithm for optimally sized DC nanogrid. *IEEE Syst. J.* **2022**, *16*, 231–239. [[CrossRef](#)]
36. Jiang, X.; Wu, L. Residential power scheduling based on cost efficiency for demand response in smart grid. *IEEE Access* **2020**, *8*, 197379–197388. [[CrossRef](#)]
37. Christoff, N.; Bardarov, N.; Nikolova, D. Automatic Classification of Wood Species Using Deep Learning. In Proceedings of the 2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Biarritz, France, 8–12 August 2022; pp. 1–5. [[CrossRef](#)]
38. Al-Fuqaha, A.; Guizani, M.; Mohammadi, M.; Aledhari, M.; Ayyash, M. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2347–2376. [[CrossRef](#)]
39. Pourbehzadi, M.; Niknam, T.; Kavousi-Fard, A.; Yilmaz, Y. IoT in Smart Grid: Energy Management Opportunities and Security Challenges. In *Internet of Things. A Confluence of Many Disciplines, IFIP IoT 2019*; Casaca, A., Katkooi, S., Ray, S., Strous, L., Eds.; Springer: Cham, Switzerland, 2020; Volume 574, pp. 236–249. [[CrossRef](#)]
40. Alasali, F.; Haben, S.; Foudeh, H.; Holderbaum, W. A Comparative Study of Optimal Energy Management Strategies for Energy Storage with Stochastic Loads. *Energies* **2020**, *13*, 2596. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.