Intelligent Dynamic Pricing and Integrated Demand Response for Multi-Energy Systems
Using Deep Reinforcement Learning

Gaddafi Abbas Almannouny

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF ENGINEERING

COLLEGE OF SCIENCE AND ENGINEERING

MARCH 2025

To my parents,

Abbas Almannouny and Salma Al Gaddafi

To my wife and my son
To my brothers and sisters

Abstract

The increasing penetration of renewable energy sources (RES) and distributed energy systems (DES) presents significant challenges for the power industry, particularly in ensuring grid stability and optimising energy market operations. This thesis investigates the integration of Dynamic Pricing Integrated Demand Response (IDR) into multi-energy systems using Deep Reinforcement Learning (DRL) algorithms to improve efficiency, grid stability, and stakeholder benefits in decentralised energy markets. The first study introduces a dynamic pricing mechanism for electricity and gas systems utilising the Deep Deterministic Policy Gradient (DDPG) algorithm. This mechanism optimises the supply-demand balance, enhances Distribution System Operators (DSOs) profitability, and reduces end-user costs. The second study expands this framework to manage multiple energy carriers—electricity, gas, and heat—through energy hubs (EHs). The DDPG-based IDR strategy promotes cost efficiency and operational flexibility while handling diverse energy demands sustainably. The third study integrates dynamic pricing IDR within a Peer-to-Peer (P2P) energy trading framework for microgrids, employing the Double Actors Regularized Critics (DARC) algorithm. This approach improves renewable energy utilisation, minimises energy deficits, and boosts profitability, outperforming traditional pricing models. The research includes case studies demonstrating the benefits of dynamic pricing and IDR, such as reduced peak loads, increased renewable integration, and enhanced consumer engagement. In conclusion, the thesis lays a foundation for intelligent energy management solutions and suggests future research avenues, including the potential of blockchain technology for P2P trading and advanced consumer behaviour modelling.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I extend my sincere gratitude to my supervisor, Dr. Shengrong Bu, for her invaluable guidance, unwavering support, and patience throughout my PhD journey. Her expertise, encouragement, and insightful feedback have been instrumental in shaping my research and helping me navigate the complexities of this field. I deeply appreciate her trust in my abilities and the countless hours he dedicated to mentoring me. I also express my appreciation to my second supervisor, Professor. Jin Yang, for his constructive advice and support, particularly during the challenging phases of my research. His insights and encouragement have been invaluable in keeping me motivated and focused. I gratefully acknowledge the financial support provided by the Libyan Government and the Ministry of Higher Education, which enabled me to pursue my PhD studies without financial burden. I would like to extend my heartfelt thanks to my parents, Abbas Almannouny and Salma Al Gaddafi, for their unconditional love, sacrifices, and unwavering belief in my potential. Their constant encouragement and support have been the driving force behind my academic pursuits. To my beloved wife, for her unwavering love, companionship, and understanding throughout this demanding journey. Her unwavering support, encouragement, and patience have been a source of strength and inspiration, allowing me to stay focused and motivated even during the most challenging times. I am grateful to my brothers and sisters for their constant encouragement and always being there for me, especially during my time abroad. Their support has been a source of comfort and motivation. I would like to thank my friends and colleagues at the University of Glasgow for their support and for fostering a positive research environment. I truly value the friendships that have enriched my time here. Lastly, I am grateful to Almighty Allah for His blessings, strength, and guidance in completing this thesis.

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution

Gaddafi Almannouny

Abbreviations

- `IDR` - Integrated Demand Response
- `ISO` - Independent System Operator
- `DSO` - Distribution System Operator
- `RES` - Renewable Energy Sources
- `EV` - Electric Vehicles
- `P2P` - Peer-to-Peer
- `EH` - Energy Hubs
- `IES` - Integrated Energy Systems
- `DRL` - Deep Reinforcement Learning
- `DDPG` - Deep Deterministic Policy Gradient
- `MADDPG` - Multi-Agent Deep Deterministic Policy Gradient
- `MATD3` - Multi-Agent Twin Delayed Deep Deterministic Policy Gradient
- `DARC` - Double-Actor Regularized Critic
- `TOU` - Time-of-Use
- `RTP` - Real-Time Pricing
- `CPP` - Critical Peak Pricing
- `DER` - Distributed Energy Resources
- `VPP` - Virtual Power Plants
- `ESS` - Energy Storage Systems
- `MARL` - Multi-Agent Reinforcement Learning
- `MESS` - Multi-Energy Systems

- `SP` - Service Provider
- `BES` - Building Energy Systems
- `TD3` - Twin Delayed Deep Deterministic Policy Gradients
- `CIES` - Community-Integrated Energy Systems
- `LSE` - Load-Serving Entity
- `CHP` - Combined Heat and Power
- `EHP` - Electric Heat Pumps
- `RES` - Residential
- `COM` - Commercial
- `IND` - Industrial
- `OEDI` - Open Energy Data Initiative
- `EES` - Electrical Energy Storage
- `TES` - Thermal Energy Storage
- `FC` - Fuel Cells
- `GB` - Gas Boilers
- `MEMG` - Multi-Energy Micro Energy
- `FC` - Fuel Cells

Chapter 1

Introduction

1.1 Background

The world is swiftly transitioning toward renewable energy sources (RES), such as solar and wind power, to combat climate change and reduce dependence on fossil fuels. Smart grids utilise digital technologies and advanced communication systems to modernise the electricity grid, facilitating RES's efficient and reliable integration.

The rising adoption of renewable energy sources, including solar and wind, alongside the increasing need for effective energy management, has led to a growing interest in demand response (DR) programs. This shift towards sustainable energy systems necessitates critically examining the synergies and challenges of RES and electric vehicles (EVs). These significant challenges to traditional energy management approaches are primarily due to the intermittency of RES and the increased complexity of grid operations [1].

DR, which involves modifying electricity consumption in response to price signals or grid operator instructions, has emerged as a crucial mechanism for enhancing grid flexibility and stability [2]. These programs incentivise consumers to adjust their energy consumption patterns in response to price signals or system conditions [3]. Price-based DR programs encourage consumers to shift their energy usage based on time-varying electricity prices. In contrast, incentive-based DR programs offer rewards for reducing consumption during peak demand periods [4]. DR is a strategy that enables the modification of electricity consumption in response to price signals or directives from grid operators. This approach has gained significance as a vital tool for enhancing the flexibility and stability of the power grid. By actively engaging consumers in energy management, DR programs aim to balance supply and demand effectively, especially during periods of high electricity usage.

These programs incentivise residential and commercial consumers to adjust their energy consumption patterns based on real-time price signals or prevailing system conditions. This flexibility not only helps manage energy costs but also supports the grid's reliability.

There are two primary types of Demand Response programs. Price-based DR programs encourage consumers to shift their energy usage in accordance with time-varying electricity prices. For instance, during times when electricity prices are low, consumers may be motivated to use more electricity, while they might reduce usage during peak pricing periods to mitigate costs.

On the other hand, incentive-based DR programs offer financial rewards or other incentives to consumers who voluntarily reduce their electricity consumption during identified peak demand periods. These periods often coincide with extreme weather or heightened energy usage. By participating in such programs, consumers help alleviate stress on the grid, ensuring a more stable and reliable electricity supply for all users. Overall, Demand Response plays a crucial role in promoting energy efficiency and sustainability within the electricity market.

1.1. Background

Integrated Energy Systems (IES) have gained significant attention in recent years due to their ability to seamlessly coordinate energy management across multiple energy sectors, specifically electricity, heat, and gas networks. By integrating these diverse energy carriers, IES enhance overall system flexibility and operational efficiency, capitalising on each energy form's unique characteristics and strengths [5].

Multiple studies have highlighted the advantages of implementing IES in various contexts, such as microgrids, which provide localised energy solutions, and energy hubs, which serve as central points for energy distribution and management. These systems optimise resource utilisation and contribute to increased reliability, reduced operational costs, and greater resilience in the face of fluctuating energy demands and supply constraints [6]. Integrating energy systems offers a promising pathway towards a more sustainable and efficient energy future.

Integrated Demand Response (IDR) represents a sophisticated evolution of traditional DR methods, providing a comprehensive and optimised strategy for energy management that addresses the complexities of modern energy demand [7].

Unlike conventional DR, which often operates in isolated segments, IDR integrates multiple demand response activities across various energy systems. This includes residential and commercial buildings and industrial facilities [8].

By coordinating these diverse energy sectors, IDR fosters a more resilient, efficient, and flexible energy framework. This holistic approach enhances energy efficiency and improves grid reliability, ultimately leading to a more sustainable energy future as it effectively accommodates energy demand and supply fluctuations. IDR can respond dynamically to real-time energy needs using advanced technologies and data analytics, creating a more responsive and adaptable energy ecosystem [9].

1.1. Background

In traditional power systems, consumers typically respond passively to electricity prices, mainly adjusting their consumption based on price signals rather than actively engaging in DR programs. While price fluctuations and incentives influence consumption patterns, the level of involvement in traditional DR initiatives is often limited. However, the advent of energy hubs (EHs) has transformed how consumers interact with the energy system. EHs function as central nodes integrating various energy forms, including electricity, heat, natural gas, and renewable sources [10]. This integration enables all types of energy consumers, even those with historically inelastic loads, to actively participate in DR programs.

Peer-to-peer (P2P) energy trading represents a transformative shift in energy systems, allowing individuals, including both consumers and prosumers—those who both consume and produce energy—to engage in direct energy transactions. This innovative approach is centred around eliminating traditional intermediaries, such as utilities and energy companies, enabling participants to trade energy directly with one another [11].

The foundational principle of P2P energy trading revolves around creating a more decentralised energy market that empowers users to sell excess energy generated from renewable sources, such as solar panels or wind turbines, directly to their neighbours or the broader community. This system enhances energy access and promotes the use of renewable resources by facilitating local energy circulation.

Several key models are utilised within P2P energy trading frameworks. One prominent model is based on dynamic supply-demand ratios, where energy prices fluctuate according to real-time consumption and availability. Another approach involves auction-based frameworks, which allow participants to bid on energy prices, fostering a competitive marketplace. Additionally, game-theoretic strategies are employed to understand and predict trade interactions, enhancing efficiency and optimising transaction outcomes [12].

1.1. Background

Overall, P2P energy trading has the potential to create a more resilient, sustainable, and consumer-driven energy ecosystem, aligning with broader goals of energy independence and reduced carbon footprints.

Dynamic pricing aims to align energy costs with real-time fluctuations in demand and supply. While it has been widely adopted in single-energy systems, its implementation within multi-energy frameworks remains underexplored. The interaction among electricity, gas, and heat pricing represents a critical area for further development [13].

Dynamic pricing, a key component of IDR, involves varying electricity prices in real-time or near-real-time to reflect the actual cost of generation and grid conditions [14]. It provides economic incentives for consumers to shift their consumption to periods of lower demand, thereby reducing peak loads, improving system efficiency, and facilitating RES integration. However, effectively implementing dynamic pricing and managing IDR requires sophisticated optimisation techniques that can handle the complexities of the dynamic grid environment, diverse consumer behaviours, and the stochastic nature of RES generation [15].

Deep Reinforcement Learning (DRL), a powerful class of machine learning algorithms that combines deep learning with reinforcement learning, has emerged as a promising solution for complex decision-making problems in dynamic environments [16]. DRL agents learn optimal policies through trial-and-error interactions with an environment, making them well-suited for optimising energy management strategies in the context of dynamic pricing and IDR.

1.1. Background

DRL has become a significant and practical approach for addressing complex challenges in energy optimisation. Various techniques within this domain, including Deep Deterministic Policy Gradient (DDPG), Multi-Agent Deep Deterministic Policy Gradient (MADDPG), and Multi-Agent Twin Delayed Deep Deterministic Policy Gradient (MATD3), have demonstrated considerable potential in the context of energy trading [17]. These methods leverage advanced algorithms to facilitate decision-making in dynamic environments, enabling more efficient trading strategies.

1.2 Problem Statement

The transition to renewable energy sources and the growing adoption of smart grid technologies underscore the necessity for efficient energy management strategies to navigate the complexities of modern energy systems. The intermittent nature of RES, combined with the increasing integration of electric vehicles (EVs), presents challenges in maintaining grid reliability and operational efficiency. While traditional DR programs have proven effective in balancing supply and demand, they often lack the scalability and adaptability needed to address the dynamic interactions among electricity, heat, and gas networks.

IDR and dynamic pricing present promising avenues for enhancing grid flexibility and sustainability. However, implementing these approaches within a multi-energy framework remains underexplored, especially in systems incorporating P2P energy trading and diverse consumer participation via EHs. Current DR and dynamic pricing methods frequently do not account for the interdependencies among various energy carriers or their ability to adapt to real-time energy demand and supply fluctuations.

1.2. Problem Statement

This research seeks to address these challenges by exploring the integration of IDR with dynamic pricing mechanisms within a decentralised multi-energy framework. By utilising advanced optimisation techniques, including DRL, the study aims to develop adaptive energy management strategies that promote efficient resource allocation, enhance grid stability, and increase consumer involvement in sustainable energy practices.

1.3 Aim of the Study

The overarching aim of this research is to develop and evaluate intelligent energy management strategies by integrating dynamic pricing with IDR for complex, decentralised multi-energy systems. This study leverages advanced DRL algorithms to create adaptive frameworks that can navigate the challenges posed by high renewable energy penetration and interconnected energy markets. The goal is to demonstrate that these DRL-driven solutions can significantly enhance system efficiency, improve grid stability, and deliver quantifiable economic benefits for all stakeholders, from service providers to end-users, within electricity, gas, and heat networks.

1.4 Research Objectives

The objectives of this thesis are as follows:

1. Develop a dynamic pricing strategy integrated with IDR mechanisms for a multi-energy system.
2. Utilise a modified DRL framework to optimise energy trading strategies.
3. compares the proposed approach with traditional methods in terms of efficiency, cost, and sustainability.

1.5 Research Questions

To guide this research and address the identified gaps, the following research questions are formulated:

- **RQ1:** How effective is a DRL-based dynamic pricing strategy for IDR in simultaneously optimising DSO profitability and reducing end-user costs within an integrated electricity and gas system?
- **RQ2:** To what extent can a DRL-based IDR framework be extended to manage the complexities of multi-carrier energy systems (electricity, gas, and heat) that incorporate EHs, and what is its impact on energy source utilisation and peak demand reduction?
- **RQ3:** How does an advanced DRL algorithm, specifically the modified Double-Actor Regularised Critic (DARC), perform in optimising a P2P multi-energy trading framework compared to other DRL models and traditional pricing schemes, particularly concerning overall system welfare, stakeholder profitability, and energy balance?

1.6 Research Contributions

This thesis is based on three key contributions, each represented by a distinct research paper:

1. Deep Reinforcement Learning for Integrated Demand Response Dynamic Pricing of Electricity and Gas Systems:

- This contribution presents a dynamic pricing mechanism based on DRL specifically designed for the integrated demand response in electricity and gas systems. The proposed approach utilises the DDPG algorithm to effectively address supply-demand mismatches, enhance system reliability, and optimise benefits for both consumers and service providers.

2. Dynamic Pricing Integrated Demand Response for Multiple Energy Carriers with Deep Reinforcement Learning:
   - This work extends the dynamic pricing framework to manage interactions among multiple energy carriers, including electricity, gas, and heating systems. It introduces integrated energy management strategies that improve system efficiency and accommodate diverse energy demands across interconnected systems.
3. Dynamic Pricing IDR in P2P Multi-Energy Trading Systems:
   - The third contribution integrates dynamic pricing IDR strategies within a P2P energy trading framework. Using the DARC algorithm, this work optimises energy trading across residential, commercial, and industrial microgrids, enhancing renewable energy utilisation, reducing deficits, and improving stakeholder profitability.

The collective contributions presented in Table 1.1 aim to enhance the comprehension and practical application of dynamic pricing IDR mechanisms within multi-energy systems. By leveraging DRL-based solutions, these contributions introduce innovative frameworks specifically designed to tackle the complexities and challenges faced by decentralised energy markets. Through a detailed exploration of various strategies and methodologies, the findings provide valuable insights into optimising energy distribution and pricing while accommodating consumers' and producers' diverse and fluctuating demands. This work not only paves the way for more efficient energy management but also contributes to the overall sustainability and resilience of energy systems in a rapidly evolving market landscape.

1.6. Research Contributions

Table 1.1: Summary of contributions from PhD research papers

Contribution	Paper Title	Key Features
Dynamic Pricing IDR in Electricity and Gas Systems	*Deep Reinforcement Learning for Integrated Demand Response Dynamic Pricing of Electricity and Gas Systems*	• Utilises DDPG algorithm for dynamic pricing. • Addresses supply and demand mismatches. • Enhances reliability for multi-energy systems.
Dynamic Pricing IDR for Multiple Energy Carriers	*Dynamic Pricing Integrated Demand Response for Multiple Energy Carriers with Deep Reinforcement Learning*	• Extends dynamic pricing IDR to electricity, gas, and heat. • Promotes energy carrier interaction by using EHs. • Improves overall system efficiency.
Dynamic Pricing IDR in P2P Multi-Energy Trading Systems	*Dynamic Pricing IDR in P2P Multi-Energy Trading Systems*	• Employs DARC algorithm for P2P trading. • Optimises energy distribution across residential, commercial, and industrial microgrids. • Maximises renewable utilisation and stakeholder profitability.

1.7 Thesis Structure

The rest of the thesis is organised as follows:

- **Chapter 2** presents a comprehensive review of relevant literature on demand response DR, integrated demand response IDR, P2P energy trading and Deep Reinforcement Learning applications in Energy Systems.
- **Chapters 3, 4, and 5** are dedicated to the three papers, which collectively form the core of this research.
- **Chapter 6** provides a synthesis of findings and highlights future research directions.

Chapter 2

Literature Review

2.1 demand response DR

The literature on DR emphasises its crucial role in contemporary energy systems, particularly in improving efficiency, facilitating the integration of renewable energy, and ensuring grid reliability. As an extension of Demand-Side Management (DSM), DR enables consumers to actively engage in electricity markets by adjusting their consumption patterns in response to price signals or incentives. Research highlights the diverse approaches to DR, exploring its theoretical frameworks, technological implementations, and practical effects on power systems [18] [19].

One area of the literature examines the categorisation of DR into price-based and incentive-based programs. Price-based DR encompasses Time-of-Use (TOU) rates, Real-Time Pricing (RTP), and Critical Peak Pricing (CPP), which enable consumers to modify their electricity usage in response to varying prices [18]. On the other hand, incentive-based DR includes programs such as direct load control and emergency demand response, where participants receive financial compensation for adjusting their consumption during critical periods [18] [19].

2.1. demand response DR

Recent advancements highlight the integration of DR within microgrid systems and smart buildings. Microgrid energy management systems utilise DR to align generation from Distributed Energy Resources (DERs), such as solar and wind, with variable demand. For example, optimising DER operations through DR has significantly reduced operating costs and emissions [20] [21]. Additionally, the adoption of artificial intelligence (AI)-driven frameworks, including Multi-Agent Reinforcement Learning (MARL), enhances adaptive and efficient load management, effectively addressing the uncertainties associated with renewable energy generation [22].

The role of Virtual Power Plants (VPPs) in enhancing the effectiveness of DR is a key focus area. VPPs aggregate distributed energy resources (DERs), energy storage systems (ESSs), and flexible loads, facilitating coordinated participation in electricity markets and offering ancillary services such as demand-side frequency control. This aggregation enables real-time balancing of demand and supply, thereby contributing to both grid stability and efficiency. Nevertheless, challenges such as the uncertainty of renewable generation and market prices highlight the need for sophisticated forecasting and optimisation techniques [23].

Peer-to-peer (P2P) energy trading exemplifies a cutting-edge application of DR, particularly within community microgrids. This decentralised model allows prosumers to engage in direct energy trading, leveraging pricing mechanisms that encourage participation and reduce reliance on the grid. Research indicates that P2P trading not only lowers costs but also bolsters the resilience of local energy systems [24].

Decentralised frameworks are essential for implementing DR systems. Algorithms designed to optimise energy consumption and generation using local data safeguard privacy and help reduce costs and discomfort for participants. These frameworks effectively balance supply and demand, even in the face of challenges like grid congestion and the intermittency of renewable energy sources [25].

2.1. demand response DR

Economic analyses throughout the literature consistently highlight the cost-effectiveness of DR programs. Notably, these programs have achieved peak load reductions of up to 5.13%, resulting in substantial cost savings and improved system reliability [20]. Nevertheless, the effectiveness of DR initiatives hinges on consumer participation as well as the design of incentives and pricing mechanisms that align with market dynamics. Furthermore, integrating DR with renewable energy systems, supported by advanced technologies such as the Internet of Things (IoT), has paved the way for smarter, more responsive grids [23].

Future research directions highlighted include enhancing the scalability and adaptability of DR technologies, especially in light of the growing share of renewable energy and the electrification of sectors such as transportation. Additionally, integrating advanced communication technologies and machine learning algorithms could significantly improve DR systems' predictive capabilities and operational efficiency [22] [25].

To summarise, DR continues to be a fundamental element in sustainable energy approaches, with uses ranging from conventional grid oversight to creative decentralised market structures. Research highlights its ability to connect supply and demand, promoting a more robust and efficient energy system.

2.2 Integrated Demand Response (IDR)

IDR has emerged as a transformative strategy in demand-side management (DSM), enhancing the flexibility and efficiency of multi-energy systems (MESs). Unlike traditional DR programs primarily concentrate on electricity, IDR incorporates multiple energy carriers such as electricity, natural gas, and thermal energy. This integration facilitates dynamic interactions across various energy systems, enabling a more holistic response to energy demands. This literature review offers a comprehensive analysis of IDR, emphasising its applications, modelling approaches, and implications for energy hubs and MESs.

The concept of IDR expands upon the traditional DR model by harnessing the synergies among various energy carriers. IDR enables coordinated demand-side management within energy hubs, where electricity, heating, and natural gas systems function in harmony. These hubs, defined by their interconnected generation, conversion, and storage systems, play a vital role in contemporary energy infrastructure, according to Kamwa et al. IDR programs within these energy hubs promote energy switching and demand shifting, thereby optimising energy use, reducing costs, and enhancing grid reliability along with the integration of renewable energy sources [26].

Dynamic pricing mechanisms are essential for enabling IDR. By reflecting real-time market conditions and supply and demand dynamics, these mechanisms encourage users to modify their energy consumption patterns. Nguyen et al. investigated optimal pricing strategies for IDR, illustrating how dynamic tariffs can help balance load profiles and improve economic efficiency in Microgrid Energy Systems (MESs) [15]. Likewise, Chen et al. highlighted the importance of price elasticity in developing IDR strategies that accommodate a variety of consumer preferences while also reducing financial risks for energy providers [27].

2.2. Integrated Demand Response (IDR)

Incorporating IDR into MESs necessitates advanced modelling and optimisation techniques to navigate the inherent complexities of multi-energy interactions. Mansouri et al. introduced a multi-stage joint planning and operation model for energy hubs, utilising stochastic programming to manage uncertainties related to renewable energy output and demand variability [28]. Their findings underscored the potential of IDR to enhance operational efficiency and reduce costs by promoting demand-side flexibility [29]. Additionally, Zheng et al. proposed an incentive-based IDR model that considers the behavioural coupling effects among consumers. This approach improves the model's accuracy and applicability, ensuring that IDR strategies are aligned with consumer behaviour and the constraints of the energy system [30].

Furthermore, IDR plays a crucial role in promoting energy resilience and sustainability. By incorporating RESs into MESs, IDR improves system flexibility and decreases reliance on fossil fuels. For example, Bahrami and Sheikhi [31] illustrated how IDR programs in smart energy hubs could facilitate energy switching between electricity and natural gas, optimising resource utilisation during peak demand periods and lowering greenhouse gas emissions. Similarly, Shao et al. [32] emphasised the advantages of IDR in integrated electricity and natural gas systems, demonstrating how demand-side flexibility can alleviate the operational constraints of interconnected energy networks.

Recent studies have broadened the understanding of IDR by applying game-theoretic approaches to optimise energy pricing and consumption. Gao et al. [33] employed an evolutionary game model to analyse the participation behaviours of residential users in IDR programs. Their results indicated that adaptive pricing strategies significantly influence user participation and enhance overall energy efficiency. Yang et al. [34] introduced a Stackelberg game-based pricing strategy for multiple energy providers, highlighting the integration of IDR with home energy management systems (HEMS) to optimise multi-energy loads and improve economic outcomes for both providers and consumers.

2.2. Integrated Demand Response (IDR)

Advanced control and optimisation frameworks are critical for the implementation of IDR in community-integrated energy systems (CIESs). Li et al. [35] introduced a scheduling model that incorporates IDR alongside renewable energy uncertainties, leveraging power-to-gas (P2G) and micro-turbine technologies to improve system flexibility and enhance user satisfaction. Their research highlights the significance of coordinated IDR strategies in achieving a balance between economic efficiency and system reliability.

Despite its many advantages, IDR encounters challenges regarding implementation and scalability. The intricate nature of modelling interactions among various energy carriers, combined with the uncertainties of renewable energy outputs, necessitates the development of robust computational tools and methodologies. Furthermore, consumer engagement is a vital aspect, as the success of IDR programs depends on active participation and behavioural adaptation. Future research should prioritise the creation of user-centric IDR frameworks that leverage advanced communication technologies and machine learning algorithms for real-time decision-making.

IDR represents a notable advancement in demand-side management. It provides a comprehensive approach to optimising energy consumption across various carriers. By harnessing the capabilities of energy hubs and Market Energy Systems (MESs), IDR improves system efficiency, facilitates the integration of renewable energy, and supports sustainable energy development. Ongoing research and innovation in this area are crucial to addressing current challenges and realising the full potential of IDR in contemporary energy systems.

2.3 Peer-to-peer (P2P) energy trading

P2P energy trading has emerged as a transformative model within contemporary energy markets, presenting a decentralised and consumer-centric approach to energy exchange. In contrast to traditional hierarchical systems dominated by centralised utilities, P2P energy trading allows consumers and prosumers to directly participate in energy transactions. This review examines the existing literature on P2P energy trading, emphasising its technical, economic, and social aspects while also addressing the challenges and opportunities that arise from this innovative approach.

The foundation of P2P energy trading lies in the widespread adoption of Distributed Energy Resources (DERs), including solar panels, wind turbines, and energy storage systems. As highlighted by Zhang et al. [11], the integration of DERs has transformed energy consumers into prosumers, who both generate and consume energy. P2P trading empowers these prosumers to sell excess energy to their neighbours, thereby cultivating localised energy markets that enhance grid resilience and optimise resource utilisation. The emergence of P2P platforms, such as Elecbay, illustrates how information and communication technologies (ICTs) can facilitate direct energy trading, reducing transmission losses and reliance on centralised utilities.

Blockchain technology has been instrumental in overcoming the operational challenges associated with P2P energy trading. By enabling secure, transparent, and tamper-proof transactions, blockchain establishes trust and accountability among participants. Esmat et al. [36] designed a decentralised P2P energy trading platform that is anchored in blockchain, which incorporates market-clearing algorithms and smart contracts to automate energy transactions. This system not only minimises transaction costs but also safeguards user privacy, showcasing blockchain's potential to transform energy markets. Similarly, Alskaif et al. [37] demonstrated how blockchain-enabled smart contracts can optimise bilateral trading preferences, ensuring alignment between energy supply and consumer demand while maintaining grid stability.

From an economic standpoint, P2P energy trading has been demonstrated to enhance market efficiency by fostering competitive local energy markets. Soto et al. [38] emphasise that P2P trading incentivises the adoption of renewable energy by providing prosumers with the opportunity to monetise their surplus generation. This approach not only lowers energy costs for participants but also facilitates the global transition to low-carbon energy systems. Furthermore, Guerrero et al. [33] investigated decentralised P2P trading within the context of network constraints, highlighting its potential to reduce grid congestion and optimise the use of local resources.

Game theory and optimisation models have been widely utilised to develop effective P2P energy trading mechanisms. Morstyn and McCulloch [39] proposed a multi-class energy management framework that considers prosumer preferences, enabling differentiated pricing based on factors such as energy source and location. This strategy enhances consumer satisfaction and encourages market participation by customising energy trading to individual requirements. Similarly, Paudel et al. [40] introduced a decentralised market-clearing mechanism that takes into account power losses and network fees, thereby ensuring fair and efficient energy allocation among participants.

Although P2P energy trading holds significant promise, it encounters various technical and regulatory challenges. A primary concern is ensuring network stability while managing a high volume of decentralised transactions. Wu et al. [41] highlighted the importance of microgrids and blockchain technology in addressing these challenges, offering a coordinated framework for energy management and market integration. Their research emphasises that regulatory support and technological innovation are crucial for the expansion of P2P energy trading. Furthermore, [39] pointed out the issues of privacy and the necessity for robust data security measures to build consumer trust and encourage participation.

P2P energy trading carries significant social implications, fostering energy democracy and enhancing community resilience. By enabling individuals to take an active role in energy markets, P2P trading cultivates a sense of ownership and engagement in the energy transition. Case studies, such as the Brooklyn Microgrid project, demonstrate how community-driven energy initiatives can strengthen social cohesion while addressing local energy needs [37]. Moreover, the integration of P2P trading with demand-side management and renewable energy sources amplifies its environmental advantages, contributing to the achievement of sustainable development goals [42].

Future research on P2P energy trading should prioritise addressing scalability and interoperability challenges, particularly in the integration of diverse energy systems and market structures. Developing standardised protocols and frameworks can facilitate cross-border energy trading, thereby expanding the reach and impact of P2P markets. Furthermore, advancements in artificial intelligence and machine learning can potentially enhance the efficiency of market operations, enabling real-time optimisation and informed decision-making [38].

P2P energy trading signifies a transformative shift within the energy sector. It promotes a decentralised, consumer-centric model that aligns economic, environmental, and social goals. Although challenges remain, continual technological advancements and supportive policy measures will unlock its full potential, paving the way for a more resilient and equitable energy future.

2.4 DRL Applications in Energy Systems

DRL has emerged as a groundbreaking technology for optimising energy systems. It provides innovative solutions to the complex and dynamic challenges faced by modern energy grids. Unlike traditional model-based optimisation techniques, DRL capitalises on the interactions between agents and their environments, enabling the creation of adaptive and scalable strategies. This literature review examines the application of DRL in energy systems, with a particular focus on its use in energy management, demand response, and peer-to-peer energy trading.

One of the most significant contributions of DRL to energy systems is in energy management for integrated energy systems (IES). Traditional energy management methods often struggle with the stochastic nature of renewable energy sources and the multi-energy coupling characteristics of IES. Han et al. demonstrated the efficacy of a DRL-based framework in optimising the interaction between supply and demand in IES. The proposed model effectively captured the uncertainties in renewable energy generation and dynamic user behaviours by integrating a proximal policy optimisation (PPO) algorithm with long short-term memory (LSTM) networks. The simulation results highlighted a substantial reduction in operational costs and improved utilisation of renewable energy sources [43].

The application of DRL in building energy systems (BES) represents a promising advancement. Shen et al. developed a multi-agent DRL framework utilising duelling double deep Q-networks (D3QN) to optimise energy consumption in buildings equipped with renewable energy sources. This framework incorporated prioritised experience replay and value decomposition, enhancing learning efficiency while ensuring system stability. When compared to traditional control methods, the DRL method notably decreased energy expenses and unused renewable energy, emphasising its capability to enhance sustainable building operations [44].

2.4. DRL Applications in Energy Systems

In the field of electric vehicle (EV) charging, DRL has demonstrated significant effectiveness in tackling challenges associated with dynamic pricing and load balancing. Lee and Choi proposed a federated DRL approach for smart EV charging stations, employing soft actor-critic algorithms to optimise charging schedules and dynamic pricing. This privacy-preserving framework not only ensured data security but also facilitated collaborative learning among distributed agents. As a result, the approach led to increased profitability for charging station operators while alleviating grid stress during peak demand periods [45].

P2P energy trading exemplifies the potential of DRL in decentralising energy systems. In their research, Chen et al. introduced a multi-agent DRL model based on twin delayed deep deterministic policy gradients (TD3) to manage P2P energy trading among interconnected multi-energy microgrids. The study illustrated how the model effectively coordinated trading strategies among residential, commercial, and industrial microgrids, reducing operational costs and improving energy efficiency. This application highlights DRL's capacity to navigate high-dimensional decision-making in dynamic environments [46].

The integration of DRL with blockchain technology offers additional opportunities for securing and automating energy trading processes. For instance, DRL-enhanced smart contracts can enable transparent and tamper-proof transactions in decentralised energy markets, as demonstrated in research on blockchain-based peer-to-peer trading frameworks. This integration improves trust among participants while ensuring adherence to system constraints [47].

Although there are benefits, implementing DRL in energy systems comes with its challenges. According to Cao et al., several primary concerns include the high computational demands for training DRL models, the necessity for extensive datasets to achieve reliable learning, and the potential for settling on suboptimal policies. To overcome these challenges, improvements in algorithm development are needed, such as blending DRL with heuristic optimisation methods [48]. Additionally, enhancing the scalability of DRL frameworks is a vital area for investigation, especially in effectively managing the integration of various energy systems and market stakeholders.

The sustainability implications of DRL are significant. Forootan et al. highlighted the importance of machine learning and DRL in advancing the adoption of renewable energy and mitigating greenhouse gas emissions. By optimising energy consumption patterns and enabling demand-side management, DRL plays a vital role in achieving broader objectives of energy efficiency and carbon neutrality. The adaptability of DRL algorithms to evolving environmental and market conditions positions them as crucial tools in the transition toward sustainable energy systems [49].

Deep reinforcement learning signifies a transformative change in how energy systems are managed and operated. Its utilisation in energy management, demand response, and peer-to-peer trading highlights its ability to improve efficiency, sustainability, and resilience. Future studies should concentrate on overcoming scalability and computational issues while investigating innovative integrations with emerging technologies such as blockchain and Internet of Things (IoT) platforms. By harnessing these advancements, DRL can significantly contribute to the development of intelligent and sustainable energy systems in the future.

An essential analysis of doctoral research thesis in comparison to comprehensive literature reviews is presented in Table 2.1.

Table 2.1: Comparison of PhD Research Papers with Literature Review

Aspect	**Literature Review Insights**	**PhD Research Papers Contributions**

Dynamic Pricing	Widely explored in single-energy systems. Limited focus on multi-energy systems.	Extends dynamic pricing to multi-energy systems including electricity, gas, and heat, optimising supply-demand dynamics.
IDR	Focused on electricity-only systems. Limited integration across energy carriers.	Incorporates IDR into multi-energy systems, enabling energy carrier interaction and holistic energy management.
P2P Energy Trading	Emphasises decentralised markets with auction-based and game-theoretic models. Limited real-world implementation.	Develops a DARC-based P2P trading platform tailored for interconnected microgrids, enhancing renewable utilisation.
DRL	Demonstrated potential in energy trading and demand response but with high computational costs and scalability issues.	Implements DRL (DDPG and DARC) for efficient dynamic pricing IDR in multi-energy systems, addressing computational challenges.
Energy Resilience and Sustainability	Explores renewable integration but lacks emphasis on operational reliability in complex systems.	Enhances system resilience by integrating renewable sources into multi-energy frameworks, reducing reliance on fossil fuels.
Multi-Energy Systems (MES)	discusses the theoretical benefits of MES but lacks practical optimisation models for real-time operations.	Provides a practical framework for MES optimisation, leveraging DRL to balance real-time operations and stakeholder benefits.
Consumer Engagement	Limited focus on incentive mechanisms and behavioural modelling.	Introduces dynamic pricing IDR mechanisms that actively engage consumers through tailored incentives and demand flexibility.

2.5 Research Gap

A comprehensive review of the literature on DR, IDR, P2P trading, and DRL applications in energy systems reveals significant advancements. However, several critical gaps persist, which this thesis aims to address directly.

1. **Limited Scope of Dynamic Pricing in Truly Multi-Energy Systems:** As discussed in the literature, dynamic pricing has been widely explored, but predominantly within single-energy (electricity-only) systems. While the concept of multi-energy systems is acknowledged, there is a distinct lack of research that develops and evaluates dynamic pricing mechanisms across tightly coupled electricity, gas, and heat networks simultaneously. Existing models often fail to fully capture the operational and economic synergies that can be unlocked through cross-carrier substitution and the integrated management capabilities of EHs.
2. **Insufficient Integration and Optimisation in IDR Frameworks:** The evolution from DR to IDR is a key theme in recent literature. However, many proposed IDR frameworks remain theoretical or electricity-centric in their practical application. There is a significant gap in developing holistic IDR strategies that not only integrate multiple energy carriers but are also optimised using advanced, adaptive control methods. The potential for DRL to manage the high-dimensional, continuous decision spaces inherent in multi-energy IDR remains largely underexplored.
3. **Nascent Development of P2P Trading in Complex Multi-Energy Environments:** The literature on P2P energy trading highlights its potential for decentralised markets, often focusing on auction-based or game-theoretic models for electricity trading. A clear gap exists in the design and evaluation of P2P platforms that operate within interconnected multi-energy microgrids, where participants can trade electricity, gas, and heat. Furthermore, integrating dynamic pricing and IDR strategies into such complex P2P frameworks to optimise overall system welfare, rather than just individual transactions, has not been thoroughly investigated.

4. **Need for More Advanced and Adapted DRL Algorithms:** While DRL is recognised as a powerful tool for energy system optimisation, many studies apply standard algorithms without tailoring them to the specific challenges of the problem. As energy systems become more complex, such as in P2P multi-energy markets, the limitations of algorithms like DDPG become more apparent. There is a need to explore and adapt more sophisticated DRL algorithms, such as the DARC approach, which are specifically designed to handle challenges like value overestimation and inefficient exploration in complex, multi-faceted environments.

This thesis addresses these interconnected gaps by proposing a progressive series of DRL-based frameworks. It begins by establishing a dynamic pricing IDR model for a dual-carrier system, extends it to a multi-carrier system with Energy Hubs, and culminates in an advanced DRL-driven solution for a P2P multi-energy trading market, thereby providing a comprehensive and novel contribution to the field.

Chapter 3

DRL for IDR: A Dynamic Pricing Strategy for Electricity and Gas Systems

The traditional concept of demand response has evolved into integrated demand response, benefiting from the advanced capabilities offered by energy integration technologies. The interdependence between critical electricity and gas systems sectors has intensified due to the increasing reliance on natural gas for electricity generation. This chapter presents an innovative approach to IDR within the energy sector, employing DRL to price energy dynamically across an integrated electricity and gas system. This method optimises energy consumption patterns while considering both the profitability of Distribution System Operators (DSOs) and the costs faced by end users. The potential advantages for the energy sector include enhanced profitability for DSOs, reduced energy costs for consumers, a balanced supply and demand within the integrated energy market, and improved reliability of energy systems. Simulation results demonstrate that the proposed approach significantly boosts the profitability of DSOs, decreases energy costs for end users, and effectively balances supply and demand within the integrated market. Additionally, the algorithm enhances the reliability of the energy system and increases consumer satisfaction by dynamically adapting to changing market conditions. These findings highlight the potential of DRL-driven IDR mechanisms as a mutually beneficial strategy for both energy providers and consumers.

3.1 Introduction and Background

In light of recent developments in energy resources, it is imperative to acknowledge the critical necessity for a more complex, robust, efficient, and sustainable energy system. Considering the environmental concerns along with the economic aspects [50], it is pertinent to note that the smart grid is being actively developed to facilitate the increased integration of renewable generation [51], [52]. One significant advantage of the smart grid is its ability to enhance the integration of variable and uncertain RES through diverse energy storage solutions, in contrast to conventional energy systems where operational sectors function independently [53]. Nonetheless, despite the advancements, the limited capacity of smart grids continues to pose challenges, particularly with respect to renewable energy curtailment in existing energy systems.

A recent advancement in energy systems presents a simpler and more effective approach to enhancing grid reliability and reducing energy costs through DR. This approach leverages modern advanced information and communication technologies within smart grid systems to improve the capability to swiftly address supply-demand mismatches by adjusting flexible loads on the demand side [54].

3.1. Introduction and Background

DR can be defined as a structured program or tariff designed to provide incentive payments that encourage reduced electricity consumption during peak demand periods, when market prices increase, or when there is a threat to grid reliability. Furthermore, it seeks to incentivise fluctuations in electricity pricing over time [55].

The various DR programs fall into two primary categories: price-based and incentive-based [56]. Participants in price-based DR are encouraged to modify their energy consumption patterns in response to fluctuating electricity prices. In contrast, incentive-based DR offers either fixed or time-variable incentives to participants who decrease their energy usage during periods of stress within the power system [57].

IES have emerged as a pivotal solution to address the escalating demand for coordinated energy management across electricity, heat, and gas networks [58]. Mancarella [59] provides a detailed overview of multi-energy systems, highlighting their potential to enhance flexibility and efficiency by leveraging the complementary properties of various energy carriers. The integration of electricity and heat systems has been extensively studied, as evidenced by the work of Arteconi et al. [60], which demonstrates the economic and operational benefits of coupling electric heating systems with thermal energy storage. Similarly, Liu et al. [61] investigated distributed energy management for combined heat and power (CHP) based microgrids, underscoring the importance of integrated electricity and heat systems in meeting dynamic energy demands.

The interaction between electricity and gas systems has received considerable attention in recent years. Shahidehpour et al. [62] investigated the influence of natural gas infrastructure on power systems. In contrast, He et al. [63] developed robust operational models for integrated electricity-natural gas systems, aiming to improve reliability in uncertain conditions. Additionally, Wang et al. [64] introduced a bilateral market framework to analyse the interdependencies between electricity and gas markets, with a particular focus on marginal price-based trading mechanisms.

3.1. Introduction and Background

DR has historically been limited to single energy carriers, which constrains its effectiveness in systems characterised by diverse energy requirements. Siano [57] identified DR as an essential tool for managing uncertainties in energy supply and demand, while Nolan and O'Malley [65] addressed the challenges faced in implementing DR, particularly regarding inflexible loads. IDR builds on traditional DR by incorporating multiple energy carriers, allowing users to shift between different energy types and adjust consumption patterns in response to price signals [66]. Zhao et al. [67] proposed a multi-energy DR model that utilises time-of-use pricing to optimise energy consumption across electricity and gas systems.

Reinforcement Learning (RL) is gaining traction in energy systems for its ability to optimise decision-making in complex situations. Researchers have explored its use in various applications, such as demand response, battery management, and energy trading between microgrids [68] [57] [69] [70]. These studies demonstrate RL's potential to improve efficiency, reduce costs, and enhance user satisfaction in the energy sector.

The increasing integration of renewable energy sources and the deregulation of energy markets present new challenges [71] [72]. RL offers a promising approach to address these challenges by enabling dynamic pricing and demand response programs that balance the needs of consumers and service providers [73] [74] [75]. For example, RL can optimise residential load scheduling considering factors like consumer preferences, renewable energy availability, and costs, or determine dynamic pricing based on real-time electricity demand and wholesale prices.

3.1. Introduction and Background

Despite substantial advancements, current research frequently neglects the integrated operation of electricity, heat, and gas systems within a unified market framework. Furthermore, the potential of RL to dynamically optimise IDR strategies remains largely unexamined. This study seeks to bridge these gaps by proposing a tri-layer multi-energy market model that incorporates RL-based optimisation, aiming to enhance system reliability, economic efficiency, and consumer satisfaction. The primary contributions of this chapter are as follows:

1. The tri-layer market structure comprises Independent System Operators (ISOs), DSOs, and end-users. This framework aims to enhance coordination among energy sources, facilitate efficient energy transactions, and optimise market operations.
2. Building on IDR concepts, this research introduces mechanisms for energy users to shift consumption patterns and switch between energy carriers. This approach maximises demand-side flexibility while ensuring consumer satisfaction, even for inflexible, must-run loads.
3. Development of an IDR algorithm that employs DRL to dynamically price energy within an integrated electricity and gas system. This algorithm takes into account both the profitability of DSOs and the costs incurred by end-users. The DDPG algorithm is applied to devise optimal bidding and pricing strategies in a continuous state and action space, addressing a complex challenge in energy markets.

3.2 Integrated Market Framework and System Assumptions

3.2.1 Market Settings and Assumptions

In this framework, the hierarchical energy market consists of ISOs, DSOs, and end-users. The integrated electricity and natural gas systems are managed by a single operator, with increasing reliance on natural gas for electricity generation, enhancing the interdependence between these sectors. WANG et al. [64] present a model for interdependent gas and electricity markets that includes a marginal price-based bilateral trading system. ISOs facilitate these coupled markets, enabling bidirectional transactions and implementing bidding strategies in the Gas-Electricity wholesale markets. Figure 3.1 depicts the tri-layer market architecture—ISOs, DSOs and end-users—showing how the DRL agent interacts with each layer to set dynamic prices.

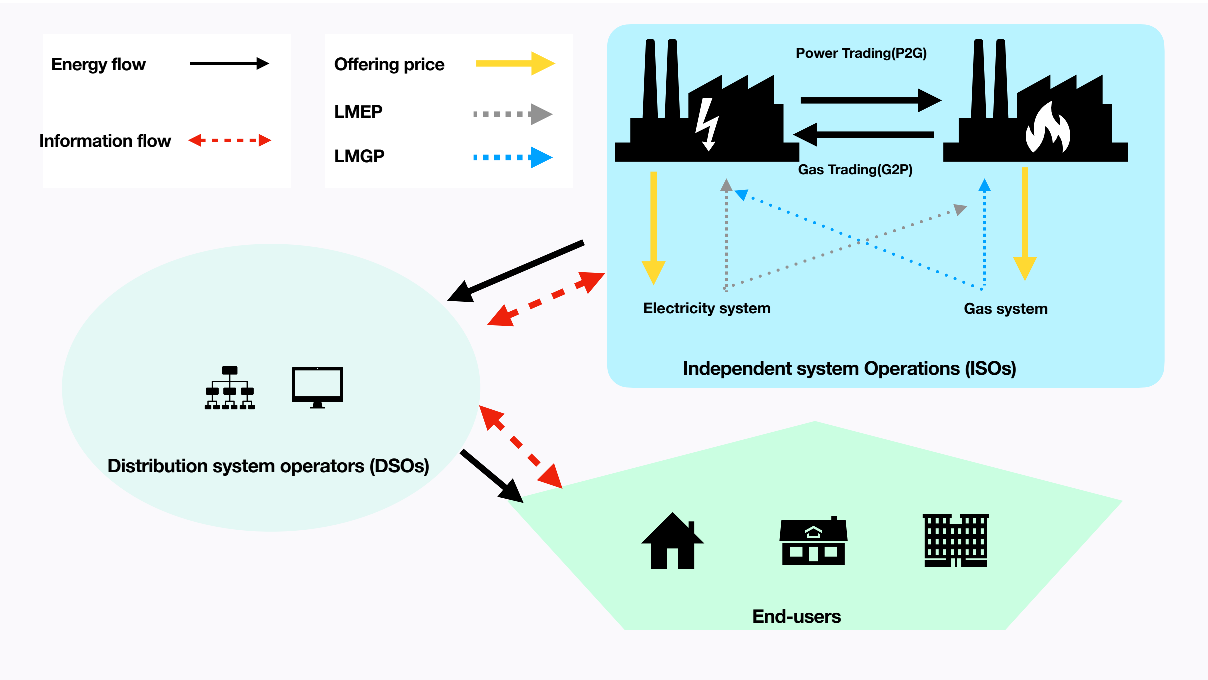

Figure 3.1: Hierarchical energy market model with ISOs, DSOs, and end-users.

3.2. Integrated Market Framework and System Assumptions

DSOs operate with a profit-driven approach, acquiring gas and electricity from ISOs and selling it to end-users. They establish dynamic retail pricing policies that aim to encourage efficient energy consumption while maximising profits. End-users are equipped with IEMS that enable them to analyse and manage their electricity, heating, and cooling demands. Additionally, these consumers have the opportunity to participate in dynamic pricing programs, which help them balance their energy use and lower costs.

The DSOs are responsible for implementing IDR. They develop dispatch plans that align with the load curtailment index, ensuring effective coordination between the end-users and the DSO. The comprehensive interaction process of the proposed model is outlined as follows:

- The DSO conducts a day-ahead load forecast and reports it to the ISO or submits bids into the energy market.
- The DSO receives the load curtailment index from the ISO.
- The DSO develops the dispatch plan utilising a dynamic pricing IDR algorithm and communicates the results to the end-users.
- If any end-users are unable to respond as directed, the DSO will modify the dispatch plan; otherwise, the response will proceed as intended.

3.2.2 End-Users Model

This study assumes that end-users possess the flexibility to choose between gas and electricity energy carriers within specified ranges. Based on their preferences and load characteristics, end-users' energy demands are categorised into critical loads and curtailable loads.

3.2. Integrated Market Framework and System Assumptions

Critical loads refer to the energy requirements that Distribution System Operators (DSOs) must meet consistently and reliably to ensure uninterrupted service. These loads are represented mathematically as:

$$e_{t,n}^{\text{critic}} = E_{t,n}^{\text{critic}} \tag{3.1}$$

Where $t \in \{1,2,3\ldots T\}$ denotes time slot t. t is the final time slot of a day, considering that the price is updated every hour, then $t = 24$. $n \in \{1,2,3\ldots N\}$ represents number of end-users n. $E_{t,n}$ and $e_{t,n}$ indicate the energy demand and energy consumption of end-users n at time slot t respectively.

Curtailable load: apart from critical loads, electricity demands such as heating, ventilation, and air conditioning (HVAC) of end-users usually decrease as the electricity and gas prices increase. The consumed energy of the curtailable load for end-users n at time slot t is defined as:

$$e_{t,n}^{curt} = E_{t,n}^{curt}\left(1 + \xi_t \cdot \frac{\lambda_{t,n} - \pi_t}{\pi_t}\right) \tag{3.2}$$

subject to the constraints:

$$\xi_t < 0$$

$$\lambda_{t,n} \geqslant \pi_t$$

Where ξ_t is the elasticity coefficient at time slot t. $t, \lambda_{t,n}$represents the retail energy price for end-users n are time slot t, π_t denotes the ISO energy price at time slot t.

The elasticity ξ_t expresses the change in energy demand for a 1 % change in price. Elasticity is commonly negative, indicating an inverse relationship between energy demand and energy price.

Once end-users n consumes energy $e_{t,n}^{curt}$ at time slot t, the corresponding energy amount $e_{t,n}^{curt}$ of end-users load demand is satisfied and the remainder of the load demand $E_{t,n}^{curt} - e_{t,n}^{curt}$ is not satisfied. This reduced energy causes dissatisfaction of end-users n at time slot t, which is denoted by a dissatisfaction cost function:

$$\varphi_{t,n} = \frac{\alpha_n}{2}\left(E_{t,n}^{curt} - e_{t,n}^{curt}\right)^2 + \beta_n\left(E_{t,n}^{curt} - e_{t,n}^{curt}\right) \tag{3.3}$$

$$\alpha_n > 0$$

$$\beta_n > 0$$

$$D_{\min} < E_{t,n}^{curt} - e_{t,n}^{curt} < D_{\max}$$

Where α_n is customer preference value varying between different customers, it shows the attitude of a customer with respect to energy demand reduction: a greater value of α_n indicates that the customer prefers less demand reduction to improve their satisfaction level, and vice versa, β_n is a predetermined constant, $D_{\min}$ and $D_{\max}$ are the ranges of demand reduction when the retail electricity and gas prices are in effect.

From the end-users' side, the objective function is to minimise its costs as described below:

$$\min \sum_{t=1}^{T} \left[\lambda_{t,n} \cdot \left(e_{t,n}^{curt} + e_{t,n}^{critic}\right) + \varphi_{t,n}\right] \tag{3.4}$$

The cost incurred by end-users includes their dissatisfaction cost and energy payment:

$$\text{Cost}_{\text{End-Users}} = \sum_{n=1}^{N} \left[\lambda_{t,n} \cdot e_{t,n}^{\text{total}} + \mu_{t,n} \cdot g_{t,n}^{\text{total}} + \varphi_{t,n}\right] \tag{3.5}$$

3.2.3 DSOs Model

The DSOs participate in the wholesale electricity market, organised by the ISO. In this market, the DSOs purchase energy at wholesale prices set by the ISO and subsequently sell it to end-users at retail prices determined by their pricing strategies. The primary objective of the DSOs is to implement dynamic retail pricing policies that maximise their profit, as formulated below:

$$\max \sum_{n=1}^{N} \sum_{t=1}^{T} (\lambda_{t,n} - \pi_t) \cdot \left(e_{t,n}^{curt} + e_{t,n}^{critic}\right) \tag{3.6}$$

At any rate, $\lambda_{t,n}$ exceeds π_t; however, it is essential to ensure that the price difference remains reasonable. This can be viewed as a result of regulatory requirements or mutual agreements between Distribution System Operators (DSOs) and end-users to maintain fair pricing and safeguard their profits. Therefore, it should be constrained as follows:

$$\mathscr{K}_1 \pi_{t,\min} \leqslant \lambda_{t,n} \leqslant \mathscr{K}_2 \pi_{t,\max} \tag{3.7}$$

Here, $\mathscr{K}_1$ and $\mathscr{K}_2$ represent predetermined coefficients for the bounds of the retail prices.

The DSO's profit is derived from the difference between retail and wholesale prices for both electricity and gas:

$$\text{Profit}_{\text{DSO}} = \sum_{n=1}^{N} \left[(\lambda_{t,n} - \pi_t) \cdot e_{t,n}^{\text{total}} + (\mu_{t,n} - \psi_t) \cdot g_{t,n}^{\text{total}} \right] \tag{3.8}$$

3.2.4 Energy Balances

Electricity: The total electricity supplied must equal the sum of electricity demanded, including electricity used in gas compression (e.g., Power-to-Gas conversion)

$$P_t^{supply} = \sum_n P_{t,n}^{demand} + P_t^{compressor} \tag{3.9}$$

Gas: Similarly, the gas supply must balance with demand, including gas-fired electricity generation:

$$G_t^{supply} = \sum_n G_{t,n}^{demand} + G_t^{generation} \tag{3.10}$$

3.2.5 Coupled Energy Pricing

This study employs marginal pricing to accurately represent the interdependency of the systems:

$$\lambda_t = \frac{\partial \mathscr{L}}{\partial P_t}, \quad \mu_t = \frac{\partial \mathscr{L}}{\partial G_t} \tag{3.11}$$

Where λ_t is the locational marginal price of electricity at time t. And μ_t Locational marginal price of gas at time t. $\mathscr{L}$ is the system's Lagrangian function encompassing all constraints and objectives.

Optimisation Objective: Maximise the overall system efficiency by minimising the combined cost of electricity and gas usage:

$$\min \sum_{t=1}^{T} \left(\lambda_t P_t^{supply} + \mu_t G_t^{supply} \right) \tag{3.12}$$

3.2.6 System Constraints

In order to ensure the feasibility and efficiency of the proposed energy trading framework, the following constraints have been established:

Energy Balance Constraint: The total energy procured by the Distribution System Operators (DSOs) from the wholesale market must satisfy the overall energy demand of end-users:

$$\sum_{n=1}^{N} \left(e_{t,n}^{\text{curt}} + e_{t,n}^{\text{crit}} \right) \leq P_{t,\text{proc}} \tag{3.13}$$

Where $P_{t,\text{proc}}$ represents the total energy acquired by the DSOs from the Independent System Operator (ISO) at time slot t.

Price Bound Constraint: The retail prices $\lambda_{t,n}$ set by the DSOs must comply with regulatory or contractual price bounds:

$$\mathscr{K}_1\pi_{t,\min} \leq \lambda_{t,n} \leq \mathscr{K}_2\pi_{t,\max} \tag{3.14}$$

In this equation, $\mathscr{K}_1$ and $\mathscr{K}_2$ are predefined coefficients, while $\pi_{t,\min}$ and $\pi_{t,\max}$ represent the minimum and maximum wholesale prices at time slot t.

Curtailable Load Constraint: The energy consumption of curtailable loads must adhere to user-defined elasticity and pricing dynamics:

$$e_{t,n}^{\text{curt}} = E_{t,n}^{\text{curt}} \cdot \left(1 + \xi_t \cdot \frac{\lambda_{t,n} - \pi_t}{\pi_t}\right) \tag{3.15}$$

With

$$\xi_t < 0, \quad \lambda_{t,n} \geq \pi_t \tag{3.16}$$

Here, ξ_t denotes the elasticity coefficient at time period t.

End-user Satisfaction Constraint

End-user satisfaction must be maintained within allowable bounds for unmet curtailable loads:

$$D_{\min} \leq E_{t,n}^{\text{curt}} - e_{t,n}^{\text{curt}} \leq D_{\max} \tag{3.17}$$

Where $D_{\min}$ and $D_{\max}$ define the range of tolerable unmet energy demand for curtailable loads.

Profitability Constraint: The DSO must ensure profitability by maintaining a positive margin between the retail and wholesale prices:

$$\lambda_{t,n} > \pi_t \tag{3.18}$$

Capacity Constraint: The total energy supplied to end-users must not exceed the maximum capacity of the DSO's infrastructure:

$$\sum_{n=1}^{N} \left(e_{t,n}^{\text{curt}} + e_{t,n}^{\text{crit}} \right) \leq C_{\text{DSO}} \tag{3.19}$$

Where C_{DSO} represents the maximum capacity of the DSO's system.

Operational Limits:

$$0 \leq P_t^{generation} \leq P_t^{max}, \quad 0 \leq G_t^{generation} \leq G_t^{max} \tag{3.20}$$

Dynamic Pricing Constraint: The dynamic pricing policy must reflect changes in energy demand and supply conditions:

$$\lambda_{t,n} = f(P_{t,\text{proc}}, D_t, \pi_t) \tag{3.21}$$

Where D_t is the total demand at time slot t. f represents the functional relationship defining dynamic pricing.

Interdependency Between Electricity and Gas:

To incorporate the interdependency between electricity and gas systems, the following coupled constraints are added:

$$P_t^{\text{supply}} = \sum_n P_{t,n}^{\text{demand}} + P_t^{\text{compressor}} \tag{3.22}$$

$$G_t^{\text{supply}} = \sum_n G_{t,n}^{\text{demand}} + G_t^{\text{generation}} \tag{3.23}$$

Where P_t^{supply} and G_t^{supply} denote electricity and gas supply, $P_t^{\text{compressor}}$ is the electricity used for gas compression, and $G_t^{\text{generation}}$ is the gas used for electricity generation. These equations ensure the integrated balance of energy flows.

3.3 IDR Dynamic Pricing Framework

In this section, we provide a detailed overview of our model. We introduce the dynamical response functions for electricity and gas prices, along with the associated pricing policies. Next, we formulate the dynamic pricing IDR program problem encountered by the DSO as a Markov decision process (MDP). Figure 3.2 depicts the interaction between the DSO (acting as the agent) and the end-users (environment).

3.3.1 Dynamic Pricing Functions for Electricity and Gas

The DSO encounters the challenge of developing dynamic pricing strategies for electricity and gas over a specified time interval t. The interaction between the DSO and end-users is modelled through the following electricity response function $P_{t,n}$ and gas response function $G_{t,n}$:

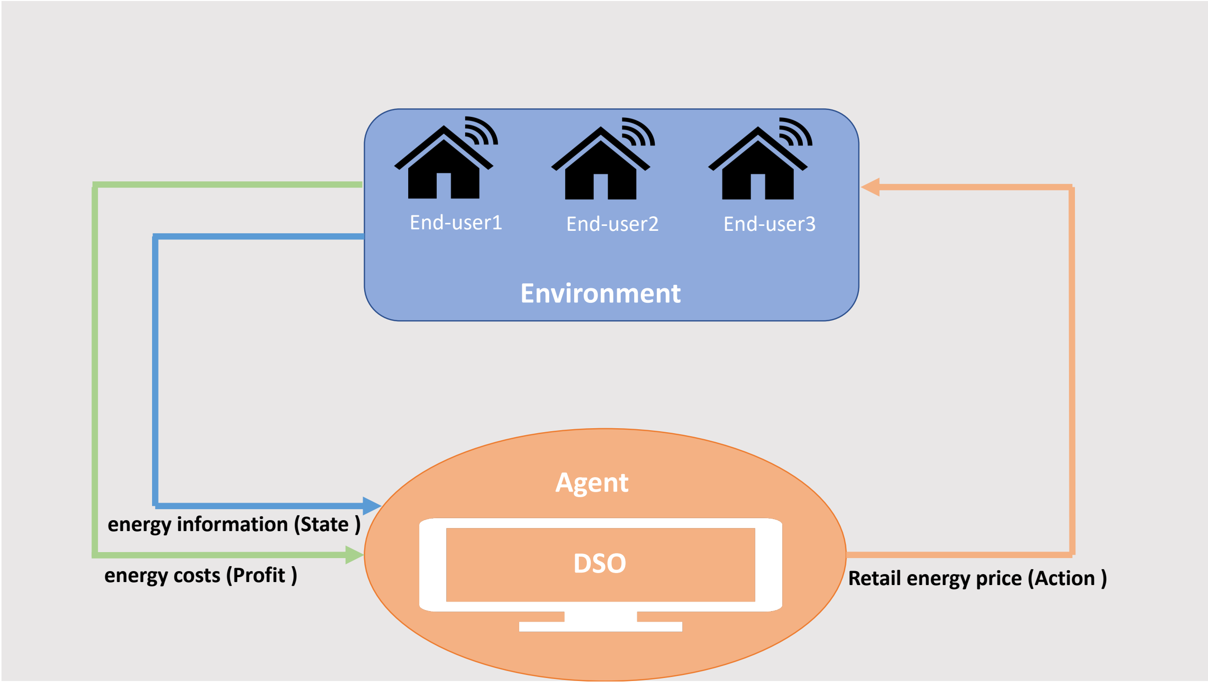

Figure 3.2: Interaction between the DSO agent and the end-users environment in the dynamic pricing framework.

$$\begin{aligned} \max \sum_{n=1}^{N} \sum_{t=1}^{T} \big[& \rho (P_{t,n} - \pi_t)(e_{t,n}^{\text{critic}} + E_{t,n}^{\text{critic}}) \\ & -(1-\rho) \left(P_{t,n}(e_{t,n}^{\text{critic}} + E_{t,n}^{\text{critic}}) + \varphi_{t,n} \right) \big] \end{aligned} \tag{3.24}$$

With

$$e_{t,n} = e_{t,n}^{\text{critic}} + E_{t,n}^{\text{critic}}$$

$$\lambda_{t,n} = P_{t,n}$$

Here, $\rho \in [0,1]$ is the weighting factor.

Analogously, we can articulate the gas pricing problem function as follows:

$$\max \sum_{n=1}^{N}\sum_{t=1}^{T}\Big[\rho(G_{t,n}-\pi_t)(e_{t,n}^{\text{critic}}+E_{t,n}^{\text{critic}}) \\ -(1-\rho)\big(G_{t,n}(e_{t,n}^{\text{critic}}+E_{t,n}^{\text{critic}})+\varphi_{t,n}\big)\Big] \tag{3.25}$$

With

$$e_{t,n}=e_{t,n}^{\text{critic}}+E_{t,n}^{\text{critic}}$$
$$\lambda_{t,n}=G_{t,n}$$

3.3.2 Dynamic Pricing Policies for Electricity and Gas

The objective of the DSO is to determine dynamic pricing policies for electricity and gas based on the available system state. These policies are expressed as:

$$P_t=\pi\left(\mathscr{I}_{t-1}\right) \tag{3.26}$$

$$G_t=\pi\left(\mathscr{I}_{t-1}\right) \tag{3.27}$$

Where $\mathscr{I}_{t-1}$ denotes the information available at time $t-1$.

$$\mathscr{I}_{t-1}=\{\pi_t,\lambda_{t,n},e_{t,n}^{\text{critic}},e_{t,n}^{\text{curt}},E_{t,n}^{\text{critic}},E_{t,n}^{\text{curt}}\} \tag{3.28}$$

3.3.3 MDP-Based Framework for Dynamic Pricing

The MDP characterises a system with a finite set of states, actions, and rewards. An MDP comprises a state space, an action space, a reward function, and a transition probability function that adheres to the Markov property [76]. Specifically, given the current state and action, the next state is independent of all preceding states and actions.

In the context of the IDR price-based problem, the IDR problem is modelled as an MDP, which comprises:

- **State Space:** The system state at time t represented as $s_t = (E_{t,n}, e_{t,n})$.
- **Action Space:** The DSO's actions, represented by electricity and gas prices $a_t = (p_t, g_t)$, are continuous.
- **Reward Function:** The reward r_t includes both DSO profit and end-user cost components, ensuring holistic optimisation:

$$
\begin{aligned}
r_t = & \rho \cdot \sum_{n=1}^{N} \left[(\lambda_{t,n} - \pi_t) \cdot e_{t,n}^{\text{total}} + (\mu_{t,n} - \psi_t) \cdot g_{t,n}^{\text{total}} \right] \\
& - (1-\rho) \cdot \sum_{n=1}^{N} \left[\lambda_{t,n} \cdot e_{t,n}^{\text{total}} + \mu_{t,n} \cdot g_{t,n}^{\text{total}} + \varphi_{t,n} \right]
\end{aligned}
\tag{3.29}
$$

The cumulative discounted reward R_t over a horizon T is expressed as:

$$
R_t = \sum_{T=t}^{T} \gamma^{T-t} r_T \tag{3.30}
$$

Where $\gamma \in [0,1]$ is the discount factor, and the transition between states follows the Bellman optimality principle.

3.3.4 Learning Response Functions for Pricing Optimisation

Reinforcement learning (RL) is employed to efficiently learn the optimal pricing policies. Transition samples (s_t, a_t, r_t, s_{t+1}) are used to approximate the Q -function, representing the expected reward:

$$Q(s_t, a_t) = \mathbb{E}[R_t | s_t, a_t]. \tag{3.31}$$

Using neural networks, the RL algorithm minimises the error between predicted and actual outcomes, allowing generalisation to unseen states. Policy gradient methods are employed to update the pricing policy parameters θ^π by maximising the expected return $J(\pi)$:

$$\nabla_{\theta^\pi} J = \mathbb{E}\left[\nabla_a Q(s, a) \nabla_{\theta^\pi} \pi(s)\right]. \tag{3.32}$$

This formulation enables the DSO to dynamically optimise pricing strategies while considering operational profitability and end-user satisfaction.

3.4 Reinforcement Learning Approach for Dynamic Pricing IDR

This section details the implementation of the DDPG algorithm, whose workflow is illustrated in Figure 3.3. Initially proposed by Lillicrap et al. [77], the DDPG algorithm employs an actor-critic architecture where:

- **The critic network** estimates the Q function, representing the expected return for a given state-action pair.
- **The actor network** approximates the optimal policy for selecting actions based on the current state.

3.4. Reinforcement Learning Approach for Dynamic Pricing IDR

Neural networks are utilised to approximate both the actor and critic functions. In the context of the IDR dynamic pricing problem, the Q-function is represented by a neural network referred to as the critic network, parametrised by θ^Q. The electricity pricing policy is modelled by the electricity policy network parametrised by θ^{P_t}, and the gas pricing policy is modelled by the gas policy network, parametrised by θ^{g_t}. These two policy networks collectively constitute the actor networks.

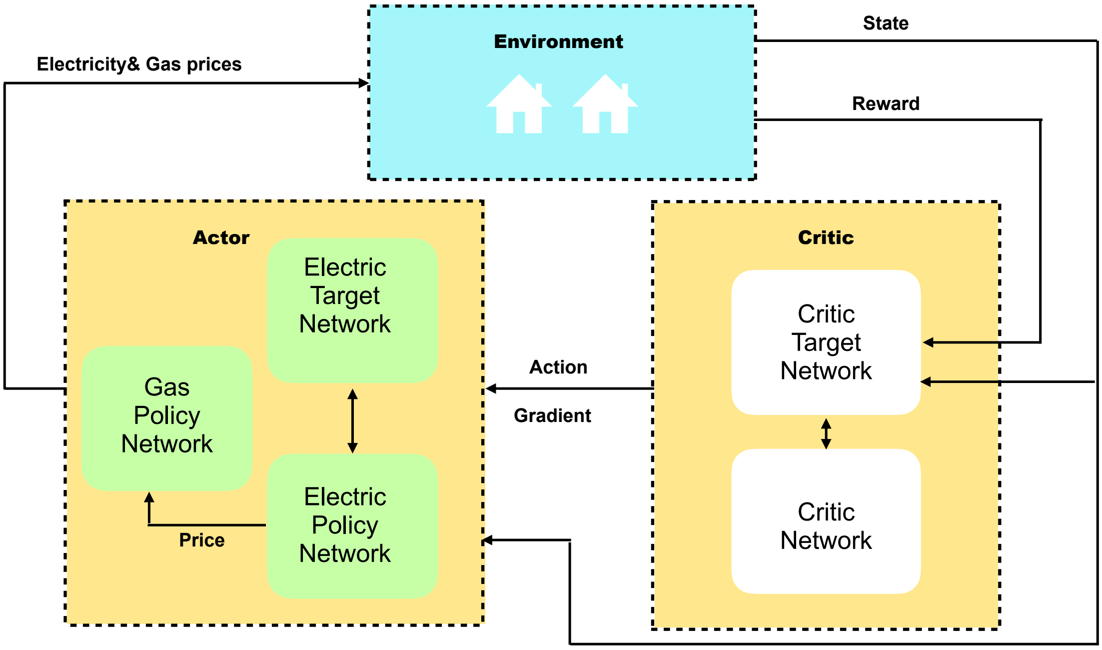

Figure 3.3: Deep Deterministic Policy Gradient (DDPG) algorithm workflow for Integrated Demand Response (IDR) dynamic pricing.

3.4.1 Key Features of the DDPG Algorithm

3.4.1.1 Target Networks

The critic network parameters θ^Q are updated to minimise the following loss function, which ensures that the network satisfies the Bellman optimality equation:

$$\ell = \frac{1}{m}\sum_i \left(r_i + \gamma Q'\left(s_{i+1}, \pi'\left(s_{i+1}\right)\right) - Q\left(s_i, a_i\right)\right)^2 \tag{3.33}$$

Here r_i is the reward observed for the i-th transition. γ is the discount factor. Q' and π' represent the target critic and actor networks, respectively.

3.4.1.2 Actor Network Update

The actor networks (electricity and gas policy networks) are updated to maximise the expected return $J(\pi)$. The update follows the direction of the action gradient, which is approximated as follows:

$$\nabla_{\theta^{\pi}} J \approx \frac{1}{m} \sum_{i} \nabla_{a} Q\left(s_i, \pi\left(s_i\right)\right) \nabla_{\theta^{\pi}} \pi\left(s_i\right) \tag{3.34}$$

Here $\nabla_a Q(s_i, \pi(s_i))$ represents the gradient of the Q-function with respect to the action. $\nabla_{\theta^{\pi}} \pi(s_i)$ is the gradient of the policy concerning its parameters.

3.4.1.3 Workflow Summary

The workflow of the DDPG algorithm is summarised in Figure 3.3. During training, response functions replace real-time end-user interactions to reduce exploration costs. The algorithm iteratively updates the critic and actor networks to converge toward optimal electricity and gas pricing policies.

Algorithm 1: DDPG-Based Dynamic Pricing IDR Program for Integrated Electricity and Gas System

Input: Initial state s_0, episodes M, steps T, replay buffer size m.

Output: Optimised electricity and gas pricing policies.

Initialize critic network $Q(s,a)$ and actor network $\pi(s)$ with weights θ^Q and θ^π;

Initialize target networks Q' and π' with weights $\theta^{Q'} \leftarrow \theta^Q$ and $\theta^{\pi'} \leftarrow \theta^\pi$;

Initialize replay buffer $\mathfrak{R}$;

for *episode* $= 1,\dots,M$ **do**

  Initialize a random process ζ for price exploration;

  Receive initial state s_0;

  for $t = 0,\dots,T-1$ **do**

    Select electricity and gas prices $a_t = (p_t, g_t)$: $a_t = \pi(s_t) + \zeta_t$;

    Execute action a_t, observe reward r_t, and transition to the next state s_{t+1};

    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathfrak{R}$;

    if $|\mathfrak{R}| > m$ **then**

      Sample a mini-batch of m transitions (s_i, a_i, r_i, s_{i+1}) from $\mathfrak{R}$;

      Update the critic network by minimising the loss in Eq 3.33;

      Update the actor network using the policy gradient in Eq 3.34;

      Update target networks:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$

$$\theta^{\pi'} \leftarrow \tau\theta^\pi + (1-\tau)\theta^{\pi'}$$

    end

  end

end

3.5 Case Studies

3.5.1 Simulation Set-up

This study simulates a DSO managing electricity and natural gas for about 150 customers. The DSO procures energy from the ISO via designated load zones and external connections. Demand profiles and market price forecasts are sourced from established references [78].

Parameter values in the simulation are specific to this study and may vary with different market structures and user characteristics, yet this does not affect the validity of the results. Over 24 hours, the DSO receives wholesale prices and load data, using these to apply the algorithm 1 for calculating Q-values that indicate optimal retail prices. The maximum Q-value identifies the optimal price for the next 24 hours.

The simulation employs neural networks developed with Tensorflow [79]. Hyperparameters, including two hidden layers with 256 neurons each, are optimised based on deep learning best practices [80]. An L2 regularisation technique (0.01) mitigates overfitting, while the ReLU activation function is used in both hidden and output layers. The Adam optimiser, with a learning rate of 0.001, trains the network for 1500 iterations.

3.5.2 Performance evaluation

The learning curve in Figure 3.4 highlights the Q-value convergence across episodes, demonstrating the reinforcement learning algorithm's effectiveness in optimising retail energy pricing strategies. As episodes progress, the average Q-value increases and then plateaus, indicating the system has learned an effective pricing policy. This convergence suggests that the DSO consistently identifies optimal electricity and natural gas retail prices, balancing profitability with customer satisfaction. The results validate the robustness of the reinforcement learning framework, showing its adaptability to dynamic market conditions and stable performance over time.

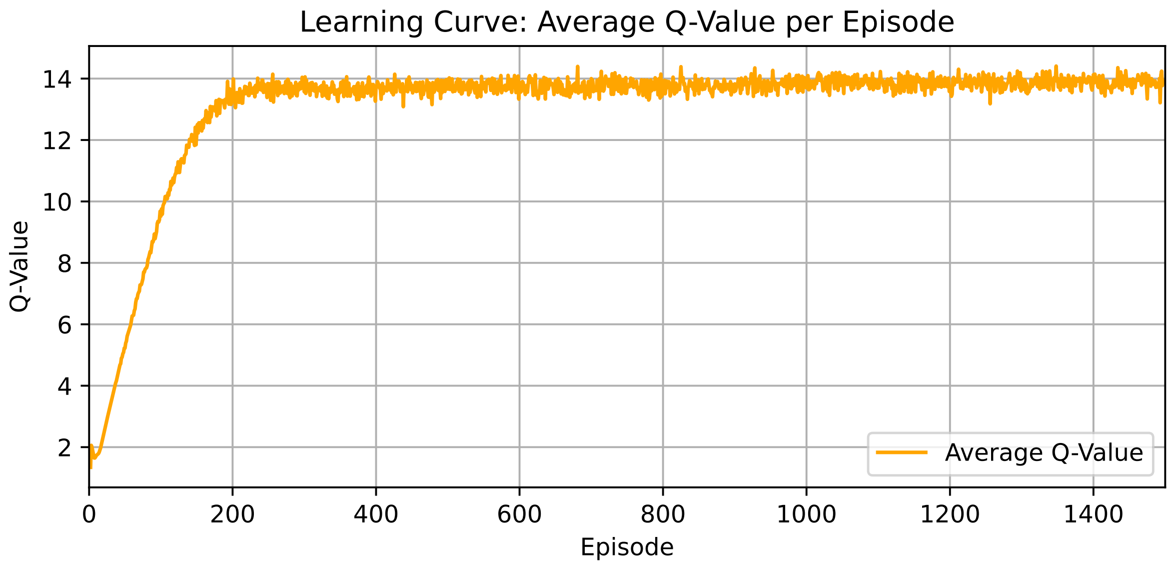

Figure 3.4: Learning curve of the DDPG algorithm, showing convergence of the Q-value across episodes.

3.5.3 DSO Profitability and End-User Cost Reduction Analysis

The analysis of DSO profits provides substantial insights into the effects of dynamic pricing IDR mechanisms. By comparing the DSO profits attained under IDR conditions with the baseline profits determined by fixed electricity and gas prices, we can assess the effectiveness of the implemented strategies.

3.5. Case Studies

The DSO profits, driven by electricity and gas markets, outpaced baseline profits, highlighting the effectiveness of IDR strategies in boosting DSO profitability. The ability to adjust retail prices in response to wholesale fluctuations significantly contributed to these gains. Figure 3.5 shows that electricity profits were a significant part of both DSO and baseline profits, with gas profits also improving. This underscores the complementary relationship between electricity and gas markets.

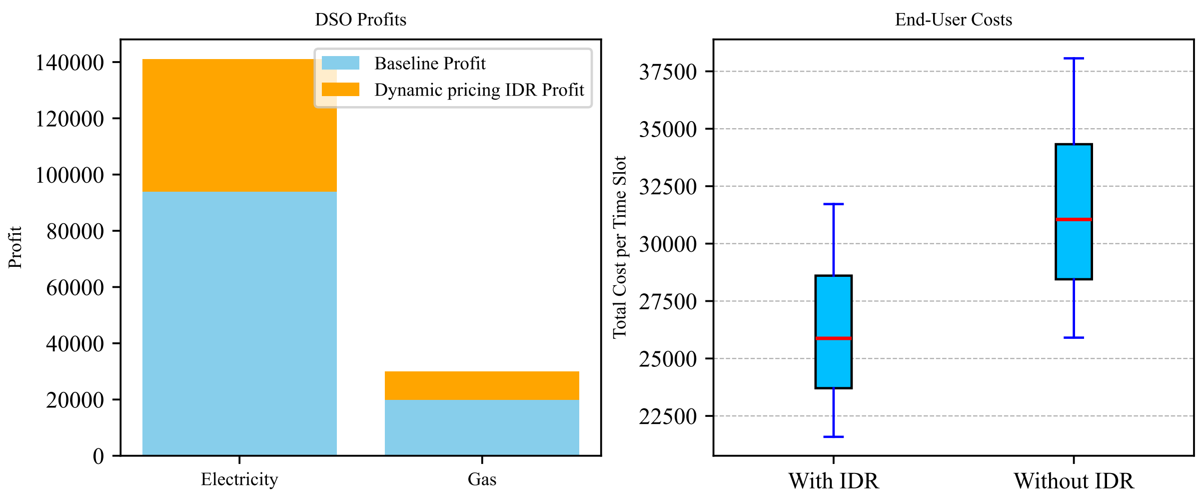

Figure 3.5: Comparison of Distribution System Operator (DSO) profits and end-user costs with and without Integrated Demand Response (IDR).

The evaluation of end-users' costs per episode with and without IDR shows the significant impact of dynamic pricing on energy expenditures. Costs under IDR were based on real-time market-adjusted retail prices, while the baseline scenario used static prices.

The findings reveal that end-users' costs with IDR were consistently lower than in the baseline, highlighting the effectiveness of dynamic pricing in encouraging consumers to optimise their energy usage and achieve substantial cost savings. Box plot visualisations further emphasise the stark differences in cost distributions between the two scenarios, showcasing IDR's potential to improve economic efficiency and consumer satisfaction in energy systems.

3.5.4 Supply and Demand Analysis

The correlation analysis between energy prices and supply-demand dynamics in Figure 3.6 offers valuable insights into the system's effectiveness in facilitating load shifting and enhancing energy efficiency. A negative correlation was observed between electricity prices and demand, suggesting that higher prices result in decreased consumption during peak periods. This finding underscores the effectiveness of dynamic pricing in motivating consumers to reduce energy usage.

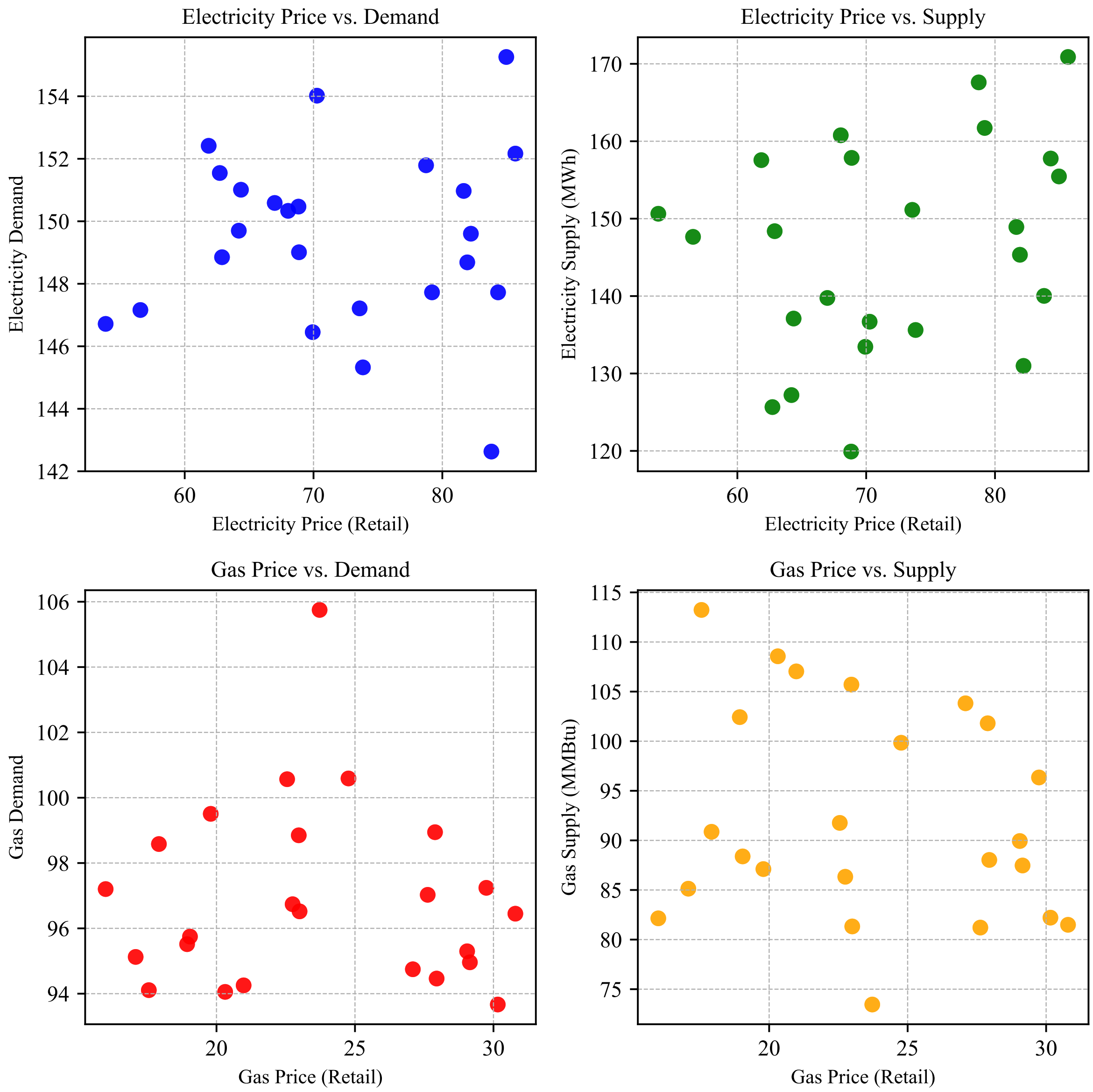

Figure 3.6: Correlation analysis between energy prices and supply-demand dynamics.

Conversely, a positive correlation between electricity prices and supply shows that higher prices incentivise suppliers to increase output, ensuring availability during high demand. This alignment of supply with demand highlights the role of price signals.

Similarly, gas prices showed a negative correlation with demand, as consumers reduced usage in response to price hikes. This demonstrates the impact of pricing on sustainable energy practices. The positive correlation between gas prices and supply indicates that higher prices encourage suppliers to boost output, maintaining equilibrium in the market.

These findings affirm the effectiveness of dynamic pricing mechanisms in managing energy systems, optimising utilisation, and supporting sustainability goals. Understanding the interplay of energy prices, demand, and supply is crucial for building resilient energy markets.

3.5.5 Supply Constraints

This analysis explores the effects of sudden supply constraints on an integrated energy system, specifically focusing on a 40 per cent reduction in electricity supply (such as grid complications) and a 50 per cent reduction in gas supply (pipeline interruptions) during designated time periods. The study assesses the system's capacity to prioritise critical loads while minimising dissatisfaction costs by implementing a dynamic pricing IDR mechanism, as illustrated in Figure 3.7. The enforced reductions in electricity supply that occur between the hours of 08:00 and 16:00 demonstrate the system's capacity to prioritise critical loads despite facing significant constraints. To maintain operational integrity, the system opts to curtail non-essential loads. Similarly, during gas supply reductions from 12:00 to 20:00, there is a noticeable decrease in the system's ability to meet non-essential demand. However, critical needs are effectively prioritised within the limitations of available resources, ensuring the continuity of essential services.

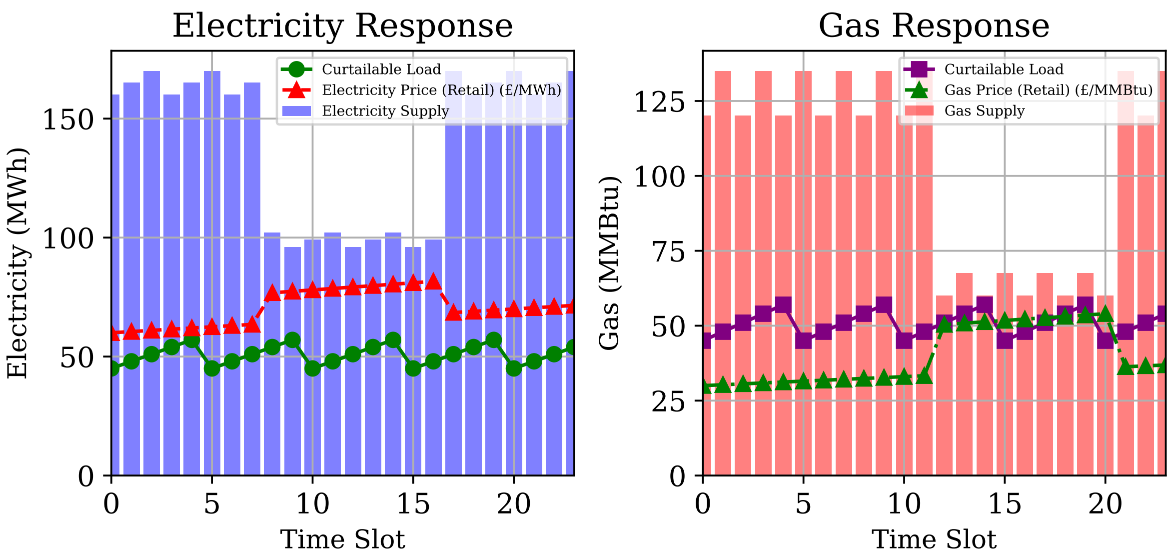

Figure 3.7: Electricity and gas response to supply constraints.

Dynamic pricing plays a crucial role in managing supply scarcity. Retail electricity and gas prices increase during constrained periods, discouraging non-essential consumption and reallocating resources to critical loads. These price adjustments align with economic principles, wherein price signals influence demand behaviour during shortages. The system's response underscores the effectiveness of dynamic pricing in optimising resource allocation under stressful conditions.

Periods characterised by constrained supply observe an increase in dissatisfaction costs, particularly for electricity during time slots 8 to 16 and gas between time slots 12 and 20. Figure 3.8 demonstrates the End-user dissatisfaction cost. Despite these escalations, the costs are effectively contained, reflecting the system's efficiency in navigating the trade-off between limited supply and user satisfaction. This containment emphasises the system's capacity to sustain economic and operational stability amid adverse circumstances.

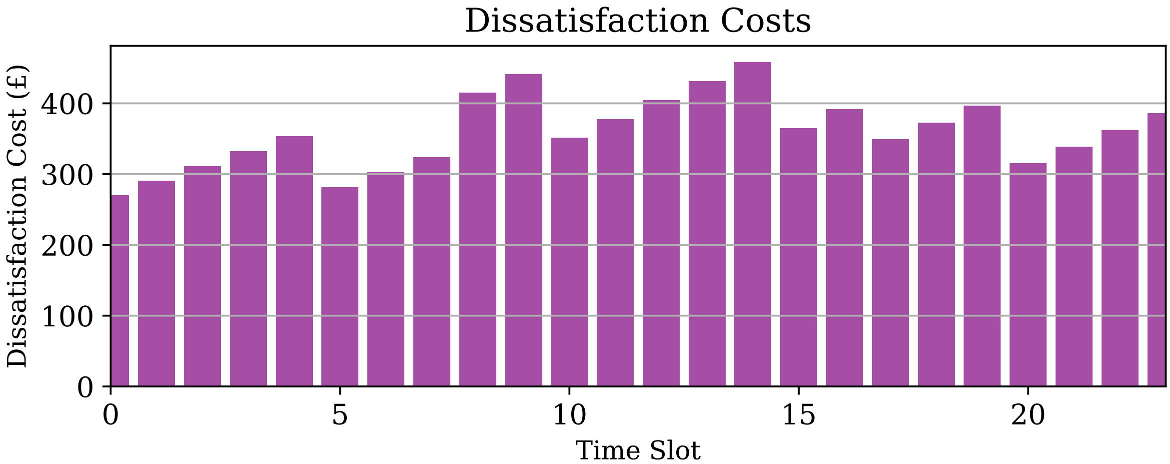

Figure 3.8: End-user's dissatisfaction costs.

The resilience of the IDR system is emphasised by its capacity to satisfy essential energy requirements even amidst challenging constraints. By utilising dynamic pricing, the system effectively mitigates pressure on resources and allocates them according to the elasticity of demand. This adaptability underscores the system's potential for broader applicability within the energy management domain.

The findings support the implementation of dynamic pricing mechanisms within integrated energy systems. These mechanisms provide an effective approach to balancing supply limitations while ensuring that essential demands are adequately met. Policymakers should explore the incorporation of such strategies to strengthen system resilience.

Educating end-users about demand response and price elasticity during periods of supply constraint can encourage greater collaboration between consumers and system operators. Increased awareness can lead to more efficient energy consumption and greater acceptance of dynamic pricing strategies.

3.6 Chapter Summary

This chapter presented a novel dynamic pricing framework for integrated electricity and gas systems using the DDPG algorithm. The key contribution was demonstrating that this DRL-based approach can simultaneously enhance DSO profitability and reduce end-user costs. Furthermore, the model proved its resilience and adaptability by effectively managing supply-demand imbalances during simulated supply constraints.

Chapter 4

Dynamic Pricing IDR for Multiple Energy Carriers with DRL

The traditional scope of demand response has evolved to encompass IDR, which leverages advancements in energy integration technologies. In this chapter, we explore the relationship between service providers (SPs) and end-users within the IDR program. The aim of the IDR initiative is to maximise profits for natural gas and electricity utility companies while simultaneously reducing customers' consumption costs and ensuring system stability.

We illustrate a framework of hierarchical decision-making using DRL. To tackle this challenge, we characterise the high-dimensional state and action spaces through the Deep Deterministic Policy Gradient (DDPG) technique, which utilises deep neural networks to estimate the state and generate actions. Service providers can adaptively adjust retail energy pricing during the online learning process, taking into account the uncertainties in end-user load demand profiles and the variability of wholesale energy prices.

Experimental results indicate that our proposed approach demonstrates high performance. The findings suggest that the IDR program can create mutual benefits for both customers and suppliers by reducing electricity and natural gas consumption costs and lessening peak load demand in the associated load profiles.

4.1 Introduction

Given the recent developments in energy resources, it is increasingly difficult to overlook the urgent need for a more complex, robust, efficient, and sustainable energy system. Environmental concerns and economic factors must be considered [50]. In this context, the smart grid is being actively developed to facilitate the significant integration of renewable generation [51], [52]. One of the key advantages of the smart grid is its ability to enhance the integration of variable and uncertain RES through the use of various forms of energy storage, especially when compared to traditional energy systems, which operate their sectors independently of one another [53]. Nevertheless, due to the limited capacity of smart grids, renewable energy curtailment remains a challenge in existing energy systems.

A recently developed method for energy systems is both simpler and more effective. DR has been introduced to enhance grid reliability and reduce energy costs. It improves the capacity to quickly respond to supply-demand mismatches by adjusting flexible loads on the demand side, aided by modern advanced information and communication technologies in smart grid systems [54]. DR can be defined as a structured program or tariff designed to provide incentive payments that encourage reduced electricity usage during peak times, when market prices are high, or when grid reliability is at risk. It may also motivate changes in electricity pricing over time [55]. DR programs can be categorised into two main types: price-based and incentive-based [56]. In price-based DR, participants are encouraged to alter their energy consumption habits in response to fluctuating electricity prices. In contrast, incentive-based DR offers fixed or time-varying incentives to participants for decreasing their energy usage during periods of power system stress [57].

Recent advancements in energy co-generation and integration technologies have prompted a transition from DR in smart grids to IDR within integrated energy systems [58]. The primary objective of IDR is to fully leverage the DR capabilities of all users, thereby enhancing the economic efficiency and reliability of multi-energy systems. By participating in IDR programs, energy users can shift their energy consumption patterns and even switch the sources of the energy they consume [66]. From the perspective of system operators, IDR enables the maximisation of social welfare within a broader optimisation framework. Furthermore, IDR dismantles the barriers among different energy forms, allowing users to adapt their energy sources based on varying energy prices [81]. This flexibility facilitates the integration of significant amounts of renewable energy, enabling the conversion of electricity into gas and thermal energy. Consequently, this approach can lead to a substantial reduction in overall operational costs [71].

Although IDR represents a significant area of interest in the energy systems field, there have been limited efforts to develop a well-designed pricing mechanism for the scheduling strategies of multi-energy systems incorporating IDR. Integrated demand response often leads to variations in multiple factors, such as load and energy prices, and these variations can yield differing outcomes. Consequently, the response of causal factors to the market-oriented environment remains inadequately understood [52]. The deregulation of energy markets and the active engagement of consumers add complexity to the search for solutions that facilitate the integration of distributed energy resources [72]. As a result, future energy markets necessitate systems capable of monitoring, forecasting, scheduling, learning, and making real-time decisions regarding energy consumption and production. This requirement underscores the need for more efficient and intelligent solutions, such as deep reinforcement learning. In this paper, we demonstrate that integrating various types of energy resources can effectively enhance the performance of existing demand response programs. In fact, the interconnection of different energy carriers allows customers to engage in demand response not only through load shifting but also by altering their energy source.

Reinforcement learning is a machine learning technique that has garnered significant interest across various fields due to its effectiveness in tackling complex sequential decision-making problems [76]. Notably, substantial advancements have been made by integrating RL with deep learning, resulting in the development of deep reinforcement learning [82]. Recently, there has been a growing focus on applying DRL in power systems, prompting numerous researchers to address a wide array of decision-making, control, and optimisation challenges within the energy sector. These challenges encompass energy management, demand response, electricity markets, operational control, and more.

To enhance grid reliability and manage peak demand, integrated demand response systems must incorporate consumer feedback and consumption data into the control loop. A significant advantage of DRL lies in its ability to provide an effective optimal control approach, supported by data-driven models to tackle such issues [68], [57].

For example, Ghasemkhani and Yang proposed an optimal pricing strategy for a demand response program utilising reinforcement learning (RL), designed to enhance the performance of a load-serving entity (LSE) by effectively balancing exploration and exploitation during the learning process [73]. Incorporating considerations such as consumer satisfaction, stochastic renewable energy, and associated costs, Remani et al. introduced an RL-based optimal model for residential load scheduling [74]. Additionally, Lu et al. presented a dynamic pricing demand response framework that integrates the service provider's profit and customer expenses, where retail electricity prices are dynamically determined through RL based on electricity demand and wholesale market prices [75].

4.1. Introduction

Despite the considerable potential of IDR in power systems, research exploring the full range of its advantages remains limited. For example, the market mechanisms that govern IDR have not been thoroughly examined. Furthermore, optimal bidding strategies for IDR resources in both energy and ancillary service markets—particularly regarding their role as price-takers—are not well-defined. There is also significant uncertainty about the market equilibrium involving multiple IDR resources. The underlying mechanisms within the energy market that could effectively address asymmetric information on the demand side are not yet fully understood. By leveraging the complementarity of Modular Energy Systems (MESs), the capabilities of IDR can be maximised without sacrificing consumer comfort. As a result, IDR has emerged as a promising approach to enhance the future interaction between demand-side resources and renewable generation.

4.2 Proposed Framework

4.2.1 Market Structure and Key Assumptions

In this model, we assume that the hierarchical energy market consists of three sectors: energy supply, the trading centre, and energy consumption, as illustrated in Fig. 4.1. The electricity system and natural gas system are recognised as a wholesale energy market. The trading centre comprises an SP and a network of energy hubs. Energy consumption is represented by a diverse group of end-users. The proposed IDR program is primarily divided into two stages and is supported by three types of participants aimed at mitigating supply-demand mismatches.

The framework of the proposed program can be outlined as follows:

4.2. Proposed Framework

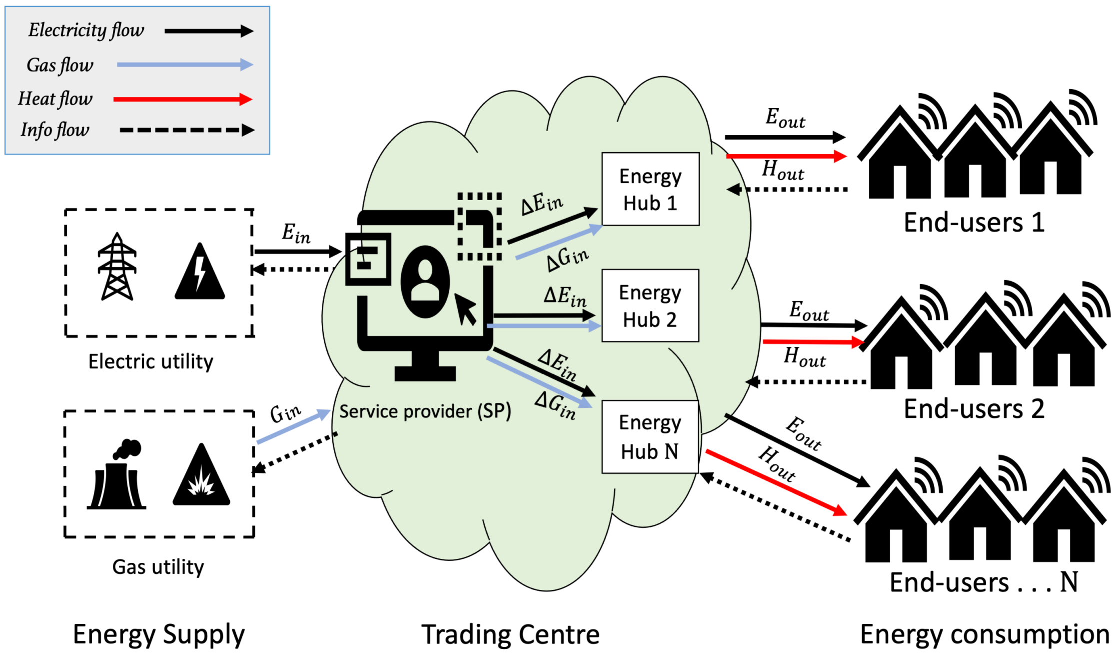

Figure 4.1: An illustration of the hierarchical market structure, including energy supply, trading centres, and energy consumption sectors.

1. **Energy Supply:** Comprising electricity and natural gas utility companies operating in wholesale markets. These utilities supply primary energy inputs to the system.
2. **Trading Centre:** Managed by the SP, this layer facilitates energy trading and pricing. The SP interacts with energy hubs and coordinates balancing strategies.
3. **Energy Consumption:** Representing end-users who consume energy through EHs. These users actively participate in the IDR program by responding to SP's dynamic pricing signals.

In the initial stage, as energy supply actors, the electricity and natural gas utilities assess their projected power needs (E_{in} and G_{in}) for an upcoming short-term period. This assessment includes identifying any surplus energy that cannot be spontaneously consumed by end-users, as well as any deficits that cannot be supplied. Once this imbalance in power is determined, the information is communicated to the trading centre.

4.2. Proposed Framework

Upon receiving the details regarding the unbalanced power, the SP submits their bidding strategies to the trading market, aiming to secure compensation for providing a specific amount of balancing power within a designated time frame. Subsequently, the trading centre discloses the outcomes of the SP's bidding strategies. By the conclusion of the first stage, the SP is aware of the required balancing power they need to deliver, along with the potential revenue they could earn if they successfully meet this requirement.

The IDR program follows a two-stage process:

- **Supply-Demand Balancing** Utility companies predict short-term energy imbalances and communicate surplus or deficit information to the SP. The SP formulates bidding strategies for the trading market and determines balancing power requirements.
- **Consumer Incentivising** The SP sets dynamic electricity and gas prices to incentivise end-users. These prices are optimised to align consumer energy consumption with system balancing needs, minimising incentive costs while maximising SP's revenue.

To achieve the necessary balancing power with minimal incentive costs, SP optimises the dispatch factors of its EH along with the incentive prices offered to end-users. The EH integrates electricity and natural gas infrastructures to fulfil customers' demands for both electricity and heating. The following section will provide a detailed explanation of the EH mechanism. Once SP has optimised and published the incentive prices for electricity and gas, consumers can select their actual balancing power in response to these prices, taking into account potential reductions in their energy bills and associated dissatisfaction costs. When SP receives highly accurate data on consumer-dependent parameters, the total actual power supplied by consumers can align with the required balancing power that SP must provide, considering the energy conversion capabilities of the EH.

4.2.2 Energy Hub Model

An energy hub can be defined as a concept that incorporates multiple energy carriers, including electricity, gas, heat, and others, which can be stored, converted, and transmitted. This paper examines the impact of integrating electricity and natural gas utilities within an energy hub on the dynamic pricing program for IDR, providing end-users with electricity and heating services. Figure 4.2 illustrates an energy hub that integrates electricity and natural gas infrastructures to deliver power and heating to customers.

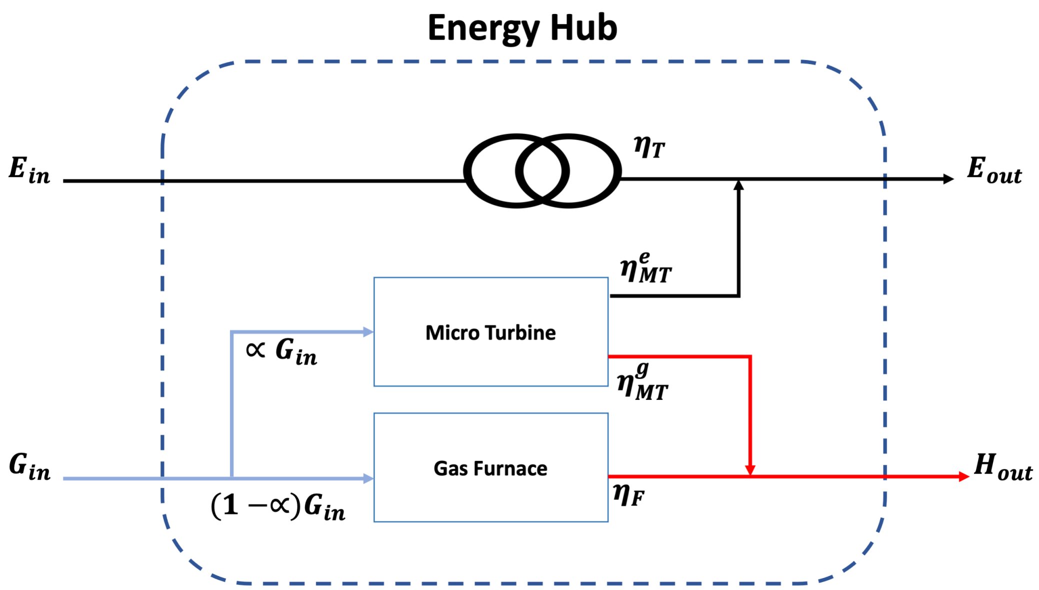

Figure 4.2: Schematic representation of an energy hub, demonstrating the coupling between electricity and natural gas systems for energy conversion and distribution.

Let E_{in} and G_{in} represent the input electricity and natural gas powers purchased from utility companies, respectively. In contrast, E_{out} and G_{out} denote the output electricity and heating powers. The converter devices include a transformer, a gas boiler, and a micro-turbine, with their efficiencies represented by η_T, η_F, η^e_{MT} (the electrical efficiency of the micro-turbine), and η^g_{MT} (the thermal efficiency of the micro-turbine). The dispatch factor is indicated by $\alpha \in [0,1]$.

The following matrix equation conveys the relationship between the inputs and outputs of various energy carriers within an energy hub:

$$\begin{bmatrix} E^{\text{out}} \\ H^{\text{out}} \end{bmatrix} = \begin{bmatrix} \eta_T & \alpha\eta_{MT}^{e} \\ 0 & \eta_F(1-\alpha)+\alpha\eta_{MT}^{g} \end{bmatrix} \begin{bmatrix} E^{in} \\ G^{in} \end{bmatrix} \tag{4.1}$$

In this study, we examine how utility providers are implementing the IDR program to motivate end users in energy hubs to either reduce their energy consumption or switch to alternative energy sources during peak hours. As depicted in Fig. 4.1, there are N energy hubs supplied by a single electric utility and a single natural gas utility. The strategy profile for energy hub $I \in N$ is defined as:

$$\mathbf{x}_i = (E_{i,1}^{\text{in}}, \ldots, E_{i,T}^{\text{in}}, E_{i,1}^{\text{out}}, \ldots, E_{i,T}^{\text{out}}, G_{i,1}^{\text{in}}, \ldots, G_{i,T}^{\text{in}}, H_{i,1}^{\text{out}}, \ldots, H_{i,T}^{\text{out}}) \tag{4.2}$$

4.2.3 Service Provider Optimisation

In the proposed IDR, the SP functions as a price-maker within the IDR programs, incentivizing consumers to contribute to balancing power. End-users act as price-takers, determining the quantity of balancing power they are willing to provide. As a profit-driven organisation, the SP aims to maximise revenue derived from trading energy resources with electricity and natural gas utility companies while minimising incentive payments to its end-users.

4.2. Proposed Framework

On the supplier side, we consider that the electricity and natural gas utility companies are generating E_t^{total} and G_t^{total} at time slot $t \in \{1,2,3,\ldots,T\}$ to meet the demands of the SP. The electricity and natural gas input for the EH, indexed by $I \in N$, at time $t \in T$ are represented as $E_{i,t}^{in}$ and $G_{i,t}^{in}$ respectively.

$$E_t^{\text{total}} = \sum_{i \in N} E_{i,t}^{in} \tag{4.3a}$$

$$G_t^{\text{total}} = \sum_{i \in N} G_{i,t}^{in} \tag{4.3b}$$

The electricity utility company operates with a time-dependent generation cost function, denoted as $C_e(E_t^{\text{total}})$. It is assumed that the primary objective of the utility company is to maximise its profit. Given the electricity price $p_e(t)$ as a function of time t, the utility company seeks to supply the SP with a certain amount of electricity that resolves the following profit optimisation problem:

$$\underset{E_t^{\text{total}} \geq 0}{\text{maximize}} \sum_{t \in T} \left(E_t^{\text{total}} p_e(t) - c_e\left(E_t^{\text{total}}\right)\right) \tag{4.4}$$

The solution to this optimisation problem is given by:

$$c_e'\left(E_t^{\text{total}}\right) = p_e(t) \tag{4.5}$$

In the natural gas network, prices are primarily influenced by supply and demand fundamentals. Additionally, natural gas prices may be linked to crude oil prices [83]. Various pricing models can be employed, including fixed-rate pricing and real-time pricing schemes. In this study, we assume that the gas price can be estimated as an increasing linear function of the total gas demand G_t^{total}. Let $p_g(G_t^{\text{total}})$ represent the natural gas price at time t. We can express this relationship as:

$$p_g\left(G_t^{\text{total}}\right) = \theta_1 G_t^{\text{total}} + \theta_0 \tag{4.6}$$

Where θ_1 and θ_0 are positive coefficients of the linear function known to the natural gas utility company.

4.2.4 End-Users Model

End-users participate actively in the IDR program by leveraging advanced energy management strategies, including detailed monitoring and control of their energy usage patterns. This active participation enables them to optimise their energy consumption schedules, shift usage to off-peak times, and switch energy sources dynamically based on price signals. Through smart energy management systems integrated into Energy Hubs (EHs), end-users can respond effectively to real-time price changes, enhancing both cost savings and overall system efficiency. By participating in load-balancing efforts, they contribute to a more stable and sustainable energy ecosystem.

End-users can respond to dynamic prices by either load shifting, modifying their consumption schedules, or energy switching, where they utilise energy hubs (EHs) to replace electricity with natural gas during peak hours. These strategies effectively reduce peak electricity demand without diminishing overall energy consumption, thereby aligning user convenience with system objectives.

In this context, $E_{i,t}^{out}$ and $H_{i,t}^{out}$ represent the electrical and heating loads at the energy hub's output ports, respectively. Moreover, E_i^d and H_i^d denote the energy hub's daily electricity and heat demand, respectively. As a result, we derive:

$$E_i^d = \sum_{t \in T} E_{i,t}^{out} \tag{4.7a}$$

$$H_i^d = \sum_{t \in T} H_{i,t}^{out}. \tag{4.7b}$$

The IDR program adjusts the timing of electrical and heating demands, shifting them to different time slots. Consequently, the total daily energy consumption for each energy hub remains unchanged. As a result, both E_i^d and H_i^d are fixed throughout the day, as indicated in 4.7.

As mentioned earlier, energy hubs receive pricing information for electricity and natural gas from utility companies via the SP. This means that the SP can track the impact of its actions on market prices. According to Equation 4.1, the input and output powers in EH $I \in N$ are proportional. By inverting the matrix presented in 4.1, we can express $E_{i,t}^{in}$ and $G_{i,t}^{in}$ in terms of $E_{i,t}^{out}$ and $H_{i,t}^{out}$ Thus:

$$\begin{bmatrix} E_{i,t}^{\text{in}} \\ G_{i,t}^{\text{in}} \end{bmatrix} = \begin{bmatrix} A_{i,t} & B_{i,t} \\ 0 & C_{i,t} \end{bmatrix} \begin{bmatrix} E_{i,t}^{\text{out}} \\ H_{i,t}^{\text{out}} \end{bmatrix} \tag{4.8}$$

where

$$A_{i,t} = \frac{1}{\eta_T} \tag{4.9a}$$

$$B_{i,t} = \frac{-\alpha_{i,t}\eta_{MT}^{e}}{\eta_T\left((1-\alpha_{i,t})\,\eta_F + \alpha_{i,t}\eta_{MT}^{g}\right)} \tag{4.9b}$$

$$C_{i,t} = \frac{1}{(1-\alpha_{i,t})\,\eta_F + \alpha_{i,t}\eta_T\eta_{MT}^{g}}. \tag{4.9c}$$

From 4.9a - 4.9c, we get

$$E_{i,t}^{in} = A_{i,t}E_{i,t}^{\text{out}} + B_{i,t}H_{i,t}^{\text{out}} \tag{4.10a}$$

$$G_{i,t}^{in} = C_{i,t}H_{i,t}^{\text{out}} \tag{4.10b}$$

$$0 \leq \alpha_{i,t} \leq 1 \tag{4.10c}$$

The framework established by the deep reinforcement learning algorithm facilitates the analysis and development of an IDR program for price forecasting SP. In this proposed IDR algorithm, service providers act as agents. They are profit-maximising entities that purchase gas and electrical energy from utility companies and sell these resources to end users. Furthermore, they determine which dynamic retail pricing policies to implement, aiming to promote more efficient energy usage while maximising profits. The objective function of the SP is defined as follows:

$$u_i(\mathbf{x}_i, \mathbf{x}_{-i}) = \sum_{t \in T} \left(U_i^e \left(E_{i,t}^{out} \right) - p_e(t) E_{i,t}^{in} \right) + \sum_{t \in T} \left(U_i^g \left(H_{i,t}^{out} \right) - p_g(t) G_{i,t}^{in} \right) \tag{4.11}$$

In this context, $p_e(t)E_{i,t}^{in}$ and $p_g(t)G_{i,t}^{in}$ represent the electricity and gas prices that the SP must pay to the electricity and natural gas utility companies, respectively. Moreover, $U_i^e(E_{i,t}^{out})$ and $U_i^g(H_{i,t}^{out})$ denote the value received by the customer from the SP. The SP, anticipating price fluctuations, acknowledges that electricity and natural gas prices are computed based on equations 4.5 and 4.6. By substituting these into equation 4.11, we derive the following optimization problem:

$$\underset{\mathbf{x}_i}{\text{maximize}} \sum_{t \in T} \left(U_i^e \left(E_{i,t}^{\text{out}} \right) - c_e' \left(E_t^{\text{total}} \right) E_{i,t}^{in} \right) + \sum_{t \in T} \left(U_i^g \left(H_{i,t}^{\text{out}} \right) - \left(\theta_1 G_t^{\text{total}} + \theta_0 \right) G_{i,t}^{in} \right). \tag{4.12}$$

We derive the objective function by integrating equations 4.10a and 4.10b into equation 4.12.

$$\begin{aligned} \underset{\mathbf{y}_i}{\text{maximize}} \sum_{t \in T} & \left[U_i^e \left(E_{i,t}^{\text{out}} \right) - \left(A_{i,t} E_{i,t}^{\text{out}} + B_{i,t} H_{i,t}^{\text{out}} \right) c_e' \left(\sum_{i \in r} A_{i,t} E_{i,t}^{\text{out}} + B_{i,t} H_{i,t}^{\text{out}} \right) \right] \\ + \sum_{t \in T} & \left[U_i^g \left(H_{i,t}^{\text{out}} \right) - C_{i,t} H_{i,t}^{\text{out}} \left(\theta_1 \left(\sum_{i \in r} C_{i,t} H_{i,t}^{\text{out}} \right) + \theta_0 \right) \right]. \end{aligned} \tag{4.13}$$

4.3 Proposed DRL approach

IDR programs are pivotal for optimising multi-energy systems. DRL provides a promising framework for dynamic pricing by enabling agents to make sequential decisions in complex environments. In this context, we propose using the Deep Deterministic Policy Gradient (DDPG) algorithm to develop optimal pricing strategies for SP. These strategies balance cost efficiency for consumers and profitability for SP while maintaining system stability.

4.3.1 MDP Formulation for Dynamic Pricing IDR

In this study, we propose a discrete finite-horizon Markov Decision Process (MDP) to model the dynamic retail pricing problem, which presents a decision-making challenge in a stochastic environment. Specifically, within the context of the dynamic pricing Integrated Demand Response (IDR) problem, the electricity and heat energy demand and consumption are defined as the state s_t at a given time interval t. The electricity and heat energy prices are represented as the action a_t for that same time interval. Consequently, the agent's profit, denoted as the Selling Price (SP) in our example, is represented as the reward r Figure 4.3 depicts the interaction between the Service Provider (agent) and end-users (environment), showcasing the reinforcement learning process within the IDR dynamic pricing framework.

In summary, the key components of the MDP that need to be modelled in the IDR dynamic pricing problem include:

- **State:** $s_t = (E_{i,t}^{\mathrm{in}}, E_{i,t}^{\mathrm{out}}, G_{i,t}^{\mathrm{in}}, H_{i,t}^{\mathrm{out}})$.
- **Action:** $a_t = U_i^e(E_{i,t}^{\mathrm{out}}), U_i^g(H_{i,t}^{\mathrm{out}})$.
- **Reward:** $r(s_t, a_t) = (\Delta E_{i,t}, U_i^e, \Delta G_{i,t}^{\mathrm{in}}, H_{i,t}^{\mathrm{out}}, U_i^g)$.

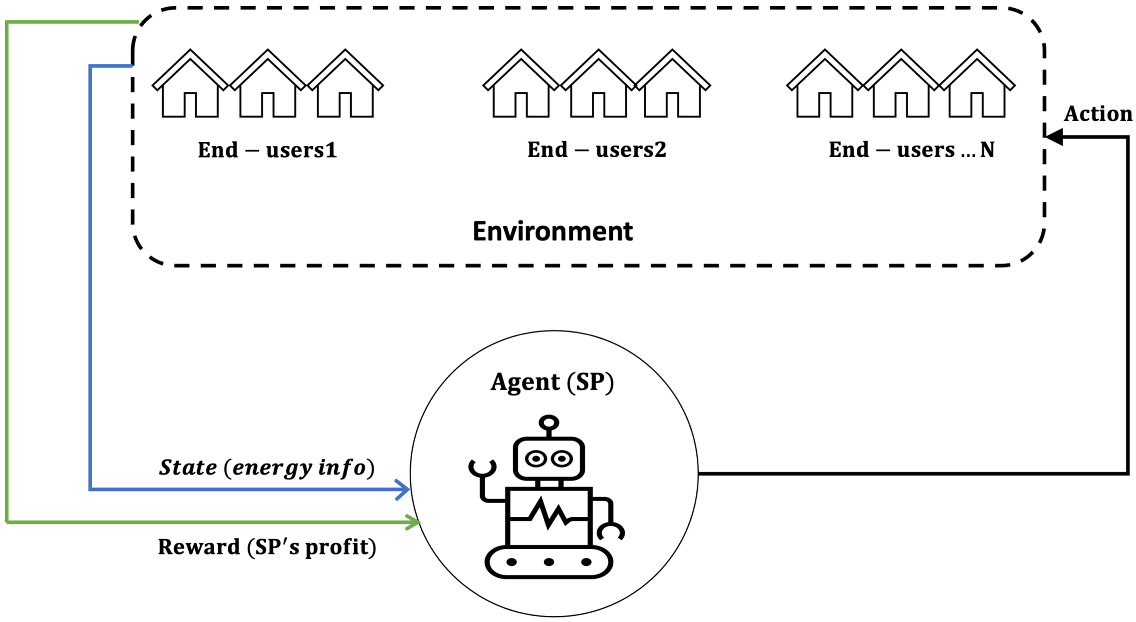

Figure 4.3: Agent and Environment Interaction.

Neither the state nor the action spaces are discrete. Given s_t, a_t, and s_{t+1} determined by equation (4.13), the Markov property is satisfied. The IDR dynamic pricing aims to maximise the service provider's profit (SP). Therefore, the reward for time interval t is defined as the SP's profit:

$$r_t = U_i^e(E_{i,t}^{\text{out}}) - p_e(t)E_{i,t}^{\text{in}} + U_i^g(H_{i,t}^{\text{out}}) - p_g(t)G_{i,t}^{\text{in}} \tag{4.14}$$

The agent must consider current and future returns when calculating long-term rewards, as potential reductions may also apply to future incentives. Consequently, future rewards are multiplied by a discount factor γ. The cumulative discounted reward for interval t can be expressed as:

$$R_t = \sum_{T=t}^{T} \gamma^{T-t} r_T \tag{4.15}$$

where $\gamma \in [0,1]$ serves as the discount factor. As such, future rewards are scaled by this discount factor.

ν can be used to express the policy that maps states to actions.

$$Q^{\nu}(s,a) = \mathbb{E}\left[R_t \mid s_t, a_t; \nu\right] \tag{4.16}$$

The objective of the dynamic pricing problem is to establish an optimal policy ν^* that always chooses an action (energy price) that maximises the expected discounted reward. The Q function under optimal policy ν^* denoted by $Q^*(s_t, a_{t)}$ satisfies the Bellman optimally equation:

$$Q^*\left(s_t, a_t\right) = \mathbb{E}\left[r_t\right] + \gamma \int_{\mathscr{S}} \mathbb{P}\left\{s_{t+1} \mid s_t, a_t\right\} \max_a Q^*\left(s_{t+1}, a\right) \tag{4.17}$$

Where $\mathbb{P}\{s_{t+1} \mid s_t, a_t\}$ is the probability of $S_t + 1$ given s_t , a_t

4.3.2 Proposed DDPG Solutions

Once the objective problem has been formulated as an MDP, reinforcement learning methods can be employed to identify the optimal policy. The DDPG algorithm, which integrates an actor-critic framework with neural networks, allows for the parametrisation of both the Q-value function and the policy [77]. The DDPG algorithm is implemented for end-users, as illustrated in Fig. 4.4. Both the Actor and Critic networks consist of deep neural networks (DNNs). Initially, utilising the current policy, the actor-network makes joint decisions based on the compacted state of the end users. Subsequently, the Critic network computes the approximate Q-value, drawing input from the environment state, the Actor's output, and the immediate reward. These two networks are alternately updated until the training process reaches completion.

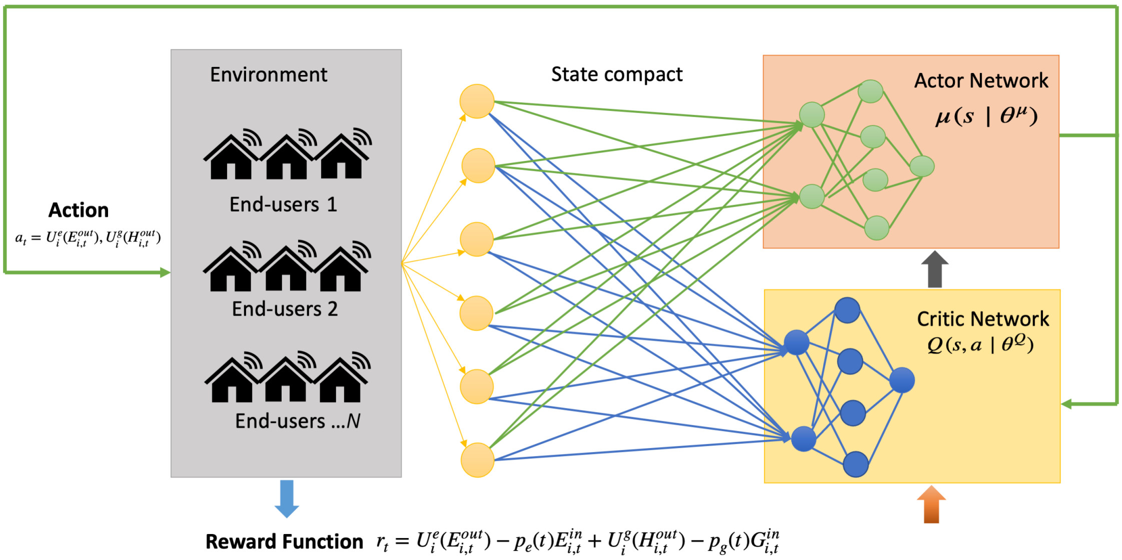

Figure 4.4: An overview of the DDPG framework applied to the IDR dynamic pricing problem, featuring Actor and Critic networks for policy optimisation and Q-value approximation.

The *Actor* and *Critic* networks are defined as $\mu(s \mid \theta^{\mu})$ and $Q(s,a \mid \theta^{Q})$, utilizing parameters θ^{μ} and θ^{Q}, respectively. The transition samples (s_t, a_t, r_t, s_{t+1}) are stored in a replay buffer $\mathscr{R}$, collected according to the current policy θ^{μ}. Once the replay buffer reaches its maximum capacity, the *Actor* and *Critic* networks are updated by drawing a mini-batch of transition samples from the buffer.

Monte Carlo estimators are utilised to derive estimated values from mini-batches of state transitions (s_t, a_t, r_t, s_{t+1}) of size m, sampled from $\mathscr{R}$, to estimate the gradients. At each step, the parameters θ^{Q} can be updated by minimising the Mean Square Error (MSE) loss:

$$L^m = \frac{1}{m} \sum_{i=1}^{m} \left(Q \left(s_i, a_i \mid \theta^Q \right) - y_i \right)^2 \tag{4.18}$$

The parameters of the Actor network policy, denoted as θ^μ, are updated in the direction of the Q-value gradient as follows:

$$\nabla_{\theta^\mu} J^m = \frac{1}{m} \sum_{i=1}^{m} \nabla_a Q \left(s, a \mid \theta^Q \right)_{s=s_i, a=\mu(s_i|\theta^\mu)} \times \nabla_{\theta^\mu} \mu \left(s \mid \theta^\mu \right)_{s=s_i} \tag{4.19}$$

Given that even a minor update to θ^Q or θ^μ can lead to substantial changes in the action-value and policy, the training process employs target networks $Q'(s, a \mid \theta^{Q'})$ and $\mu'(s \mid \theta^{\mu'})$ to provide consistent targets. The weights of these target networks are gradually updated in response to the learning progress of the primary networks as follows:

$$\theta^{Q'} \leftarrow \tau^Q \theta + (1 - \tau) \theta^{Q'} \tag{4.20}$$

$$\theta^{\mu'} \leftarrow \tau^\mu \theta + (1 - \tau^\mu) \theta^{\mu'} \tag{4.21}$$

Since $\tau^Q \ll 1$ and $\tau^\mu \ll 1$, the target values change slowly. Algorithm 2 outlines a detailed DDPG-based DRL approach for addressing dynamic pricing in the IDR context.

4.3. Proposed DRL approach

Algorithm 2: DDPG-Based Dynamic Pricing IDR Program for multi-energy carriers

Initialization: Randomly initialize *Actor* $\mu(s \mid \theta^{\mu})$ and *Critic* networks $Q(s,a \mid \theta^{Q})$, and target networks and $Q'(s,a \mid \theta^{Q'})$ and $\mu'(s \mid \theta^{\mu'})$

Input: Replay buffer $\mathscr{R}$, batch size m, number of episodes N, number of time steps in each episode T, learning rate r^{Q} and r^{μ} for *Critic* and *Actor* networks, update rate τ^{Q} and τ^{μ} for target *Critic* and *Actor* networks.

Output: The optimal action a

for *episode = 1, ..., N* **do**

- Initialise a random process for action exploration;
- receive initial state:
- $s_1 = (E_1^{in}, E_1^{out}, G_1^{in})$
- **for** $t = 1$ *to* T **do**
  - Select action required according to 4.13
  - Compute reward r_t according to 4.14 and observe the next state $(S_t + 1)$
  - Store transition $(s_t, a_t, r_t, s_t + 1)$ into replay buffer $\mathscr{R}$
  - **if** *Stored transition > Replay buffer capacity* **then**
    - Discard the oldest transition samples
  - **end**
  - Sample a random batch from $\mathscr{R}$ of a mini-batch m
  - Update *Critic* using the gradient descent in 4.18.
  - Update *Actor* using the policy gradient descent in 4.19.
  - Update target networks Q' and μ' in 4.20 and 4.21
- **end**

end

4.4 Case Studies

4.4.1 Simulation Setup

For illustrative purposes, simulations have been conducted using a single SP and three distinct end-user groups. Additionally, three EHs are connected to the SP, with each EH serving a specific group of end-users. An example of the load demand profiles for these three end-user clusters at each time slot was derived from [84]. Specifically, we selected data on electricity load (MW), gas load (kcf), electricity price (kWh/h), and gas price (Kcf/h) spanning from January 10 to January 14, 2021, to represent the wholesale energy market from which the SP procures energy.

4.4.2 Performance Evaluation

To showcase the performance of the proposed IDR scheme using DRL, Table 4.1 provides a summary of the network architecture and details regarding the training process parameters.

Table 4.1: DDPG Algorithm Parameters

Parameter	**Description**
Number of episodes	500
Learning Rate (Actor)	0.001
Learning Rate (Critic)	0.005
Batch Size	64
Replay Buffer Size	10^6
Discount Factor (γ)	0.99

4.4. Case Studies

Figure 4.5 illustrates the performance and convergence of the DDPG algorithm over 500 episodes, showcasing the learning curve. On average, this curve offers a detailed overview of the agent's training progress. At the beginning of each episode, we reset the environment to a new initial state, denoted as $s_{(1)}$. We calculate the cumulative reward and update the model parameters for the DDPG method every 1,000 steps. In the initial five episodes, the actor network makes the decision to collect data for the memory buffer randomly, leading to lower rewards. After this initial phase, both the Critic and Actor networks are updated using the transitions stored in the memory buffer. This visualisation serves as a valuable tool for analysing the proposed model and its effectiveness in achieving optimal performance in dynamic environments.

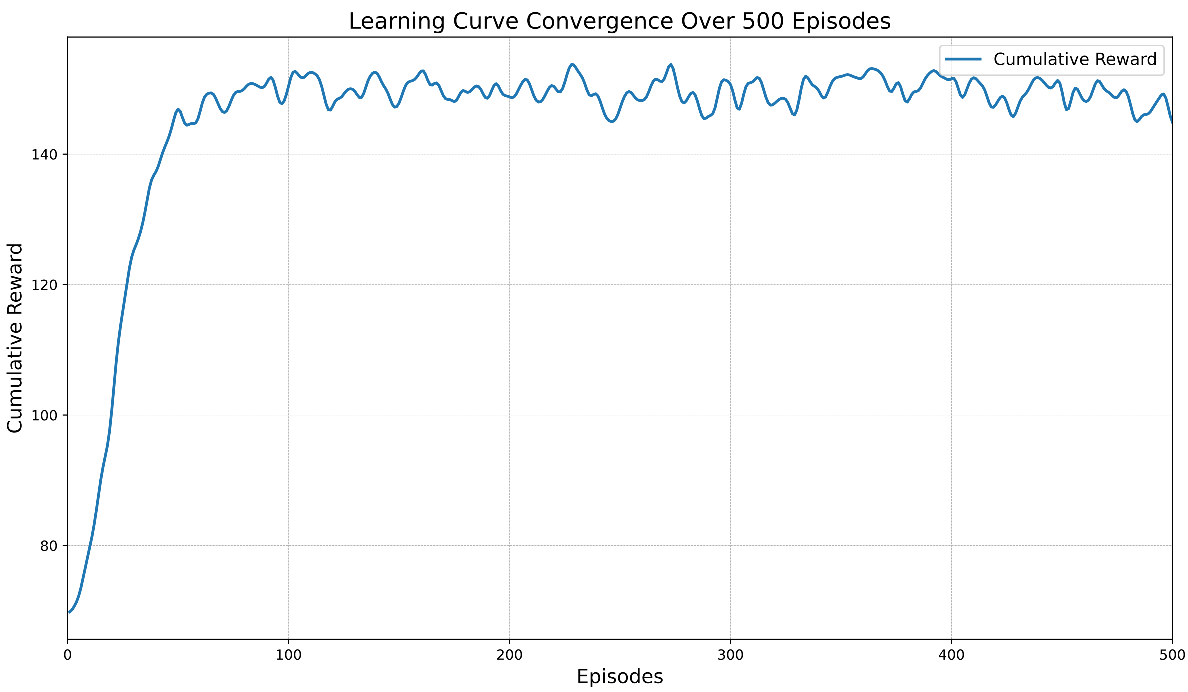

Figure 4.5: Training performance of the proposed DDPG algorithm over 500 episodes, demonstrating convergence to higher rewards and improved policy performance.

4.4.3 Impact of the proposed IDR program on the energy market

The SP purchases energy from suppliers based on wholesale market rates. All simulation parameters are hypothetical and may vary due to market design and user characteristics, but these differences won't affect the results. The analysis spans one day with 24 samples represented as $T = [1, 2, 3, \ldots, 24]$.

As shown in Fig. 4.6, there is a strong correlation between wholesale and retail electricity prices before and after IDR implementation, indicating that SP's pricing strategies reflect market trends while offering stable rates to end-users. Conversely, the gas system shows different pricing patterns, with retail prices less affected by wholesale trends, particularly between hours 14 to 20 when higher retail gas prices are noted due to increased natural gas purchases for electricity generation.

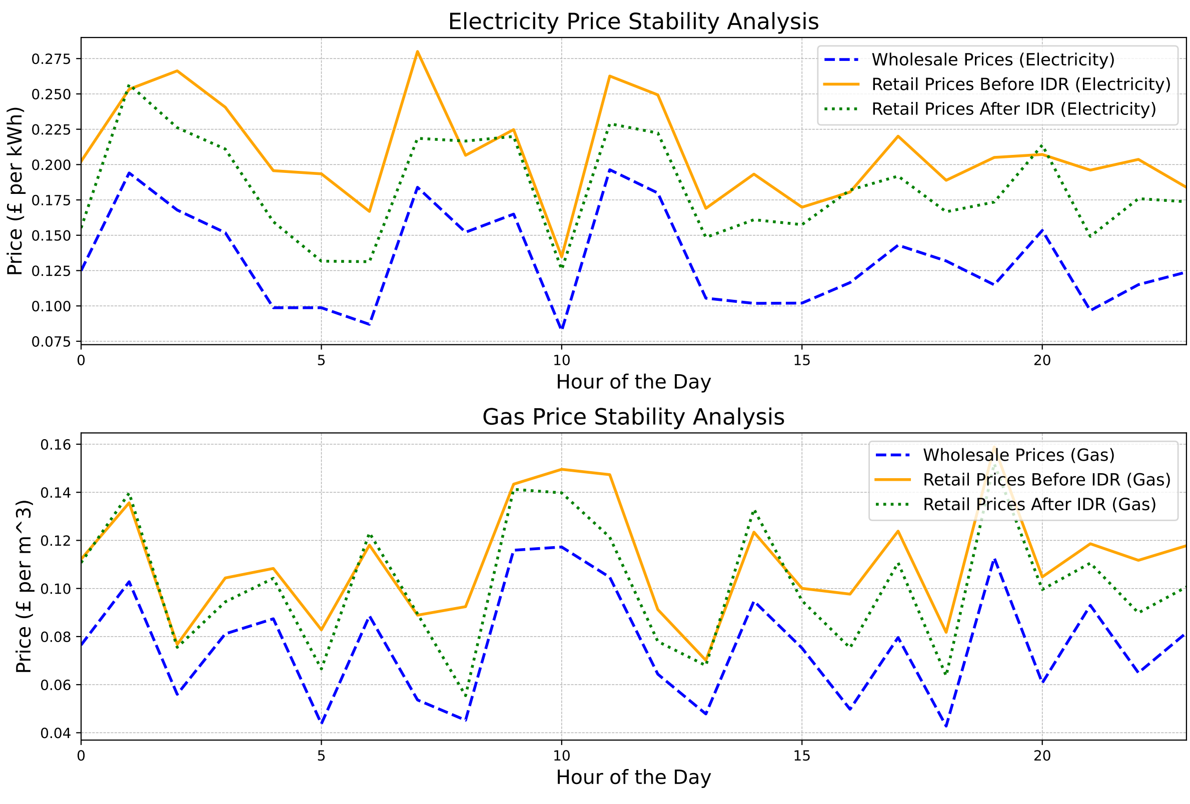

Figure 4.6: Comparison of wholesale and retail electric energy prices under the DRL policy, illustrating trends and their implications for SP and end-user interactions.

4.4. Case Studies

The IDR mechanism reduces peak electricity demand by using natural gas for power generation. This load-shifting strategy balances grid stability and, despite higher gas costs, leads to lower overall peak electricity costs and improved SP profitability. By integrating electricity and gas systems, the SP enhances operational efficiency and ensures reliable energy for end-users.

4.4.4 Profit and Cost Analysis

The figures 4.7 and 4.8 compare profit margins and consumer costs before and after implementing IDR.

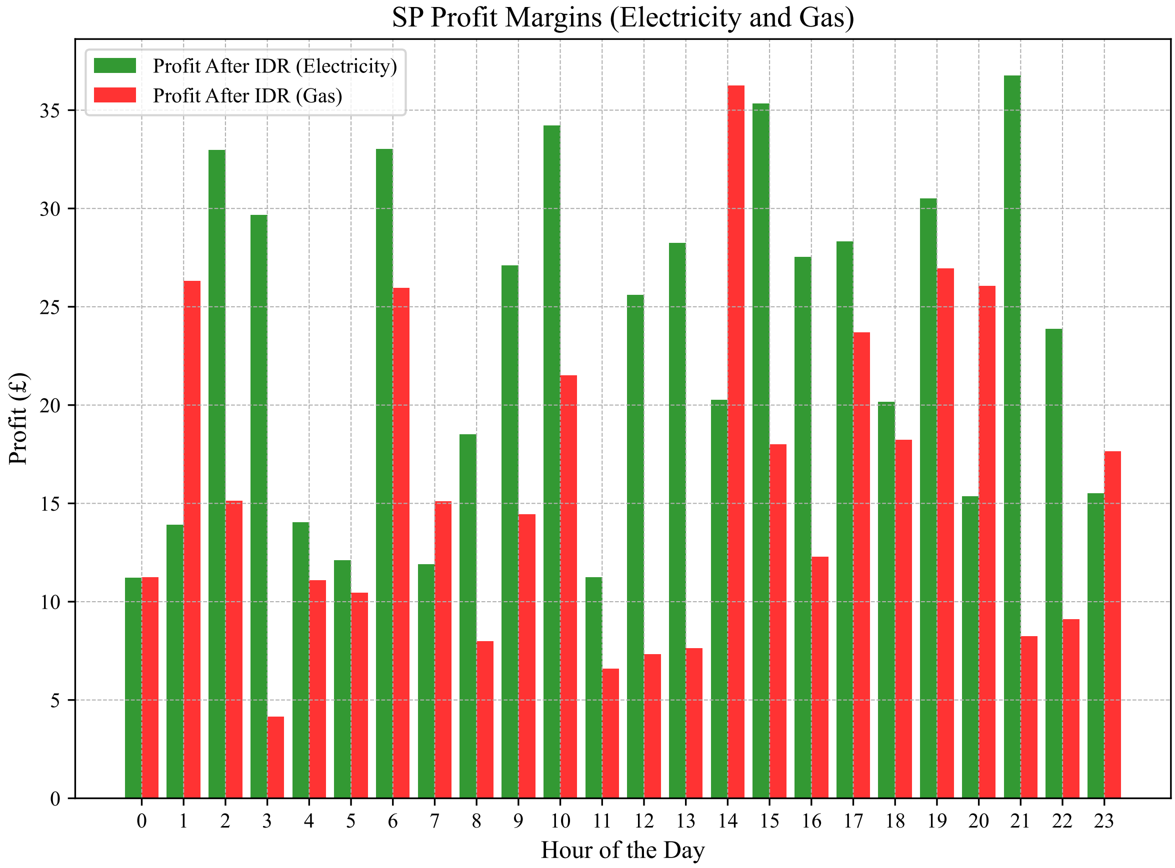

Figure 4.7: SP Profit margins before and after implementing Dynamic Pricing IDR

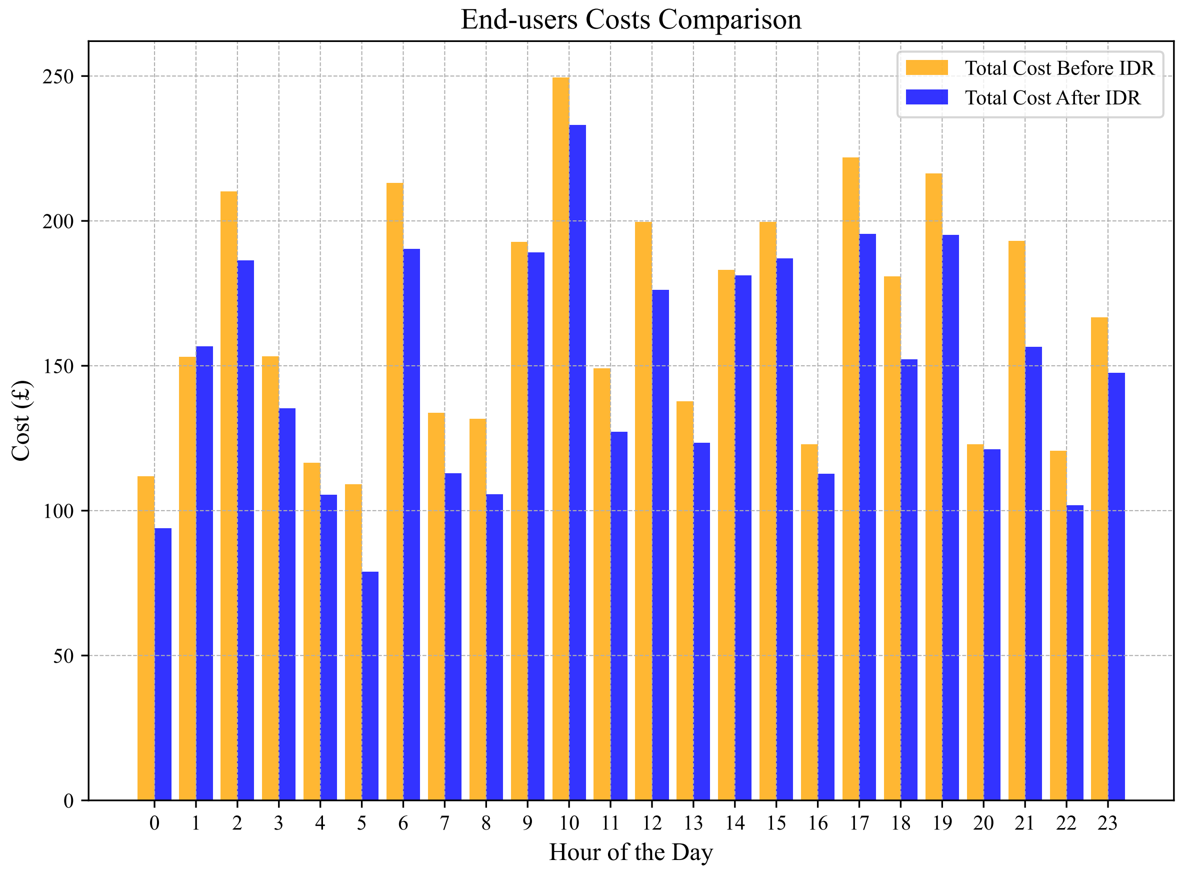

Figure 4.8: End-users Cost Analysis before and after implementing Dynamic Pricing IDR

Service provider profit margins showed a clear trend in the Figure. 4.7: before IDR, profits peaked at £60. Still, they decreased post-IDR due to reduced energy consumption during peak hours, a result of IDR's load redistribution strategy. Although profit margins fell, IDR led to a more stable energy trading environment, easing operational stress and long-term costs for providers.

End-users' costs also significantly improved as shown in the Figure. 4.8. Before IDR, costs soared to nearly £200 during high-demand periods but have since dropped. Dynamic pricing mechanisms encouraged reduced peak-hour usage, promoting a more balanced demand distribution. Overall, IDR has made energy consumption more economical by shifting usage patterns and leveraging alternative energy sources.

4.4. Case Studies

The analysis illustrates dynamic pricing IDR's transformative impact on the energy market. By reducing consumer costs and promoting balanced trading, IDR emerges as a vital tool for sustainable energy management.

4.4.5 Peak Demand and Cost Distribution Analysis

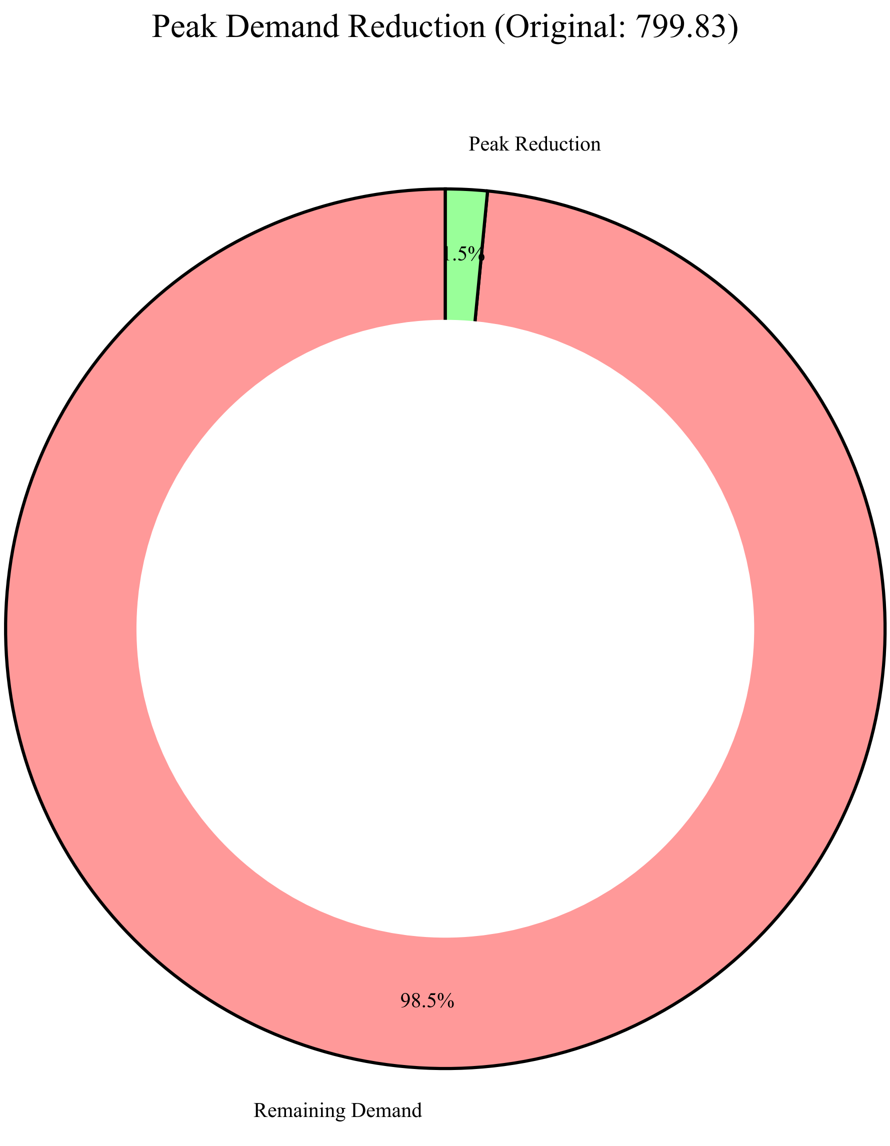

Figure 4.9: Illustrate the distribution of peak demand in the energy system before and after the implementation of Dynamic Pricing IDR

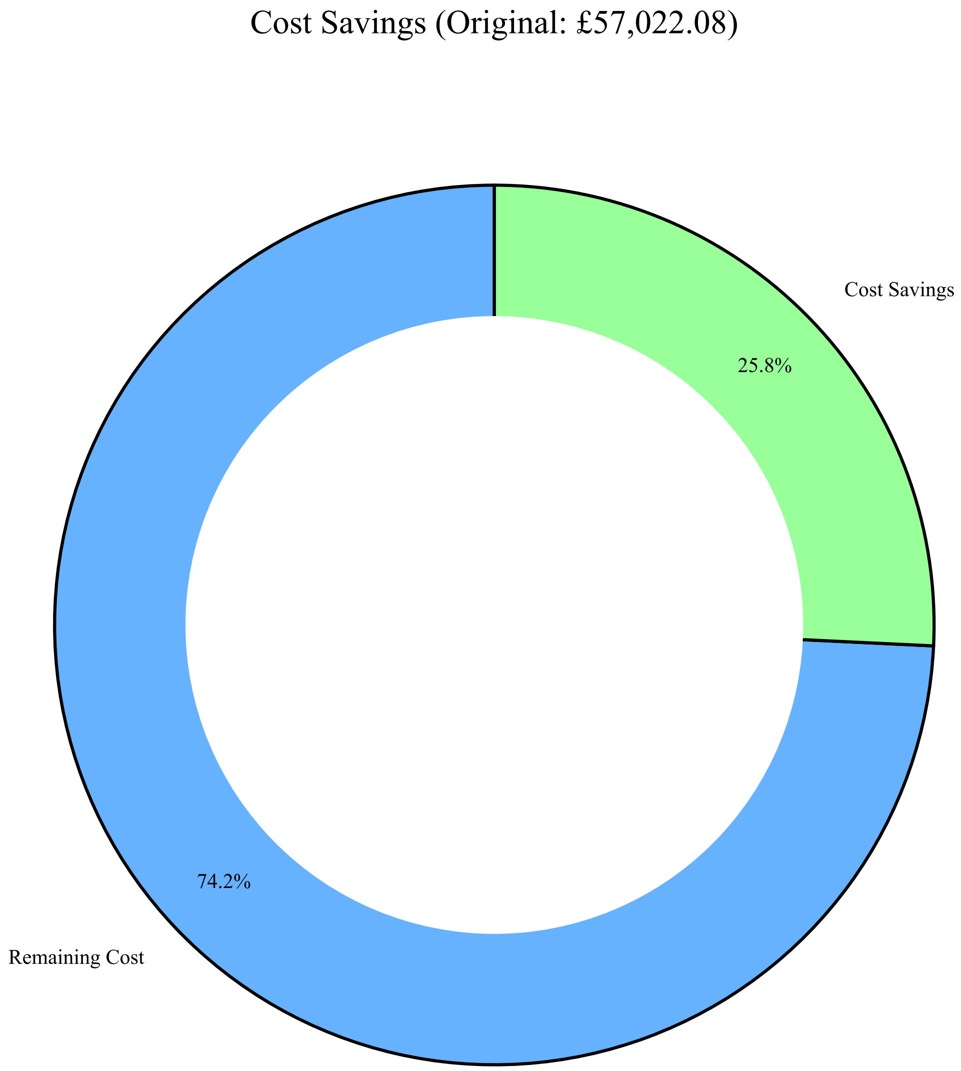

Figure 4.10: Illustrate the distribution of costs in the energy system before and after the implementation of Dynamic Pricing IDR

An analysis of the IDR program's impact reveals significant improvements in both grid demand and cost efficiency, as illustrated by the accompanying figures 4.9 and 4.10.

The "Peak Demand Reduction" chart indicates that from an original peak demand of 799.83 MWh, the IDR program successfully achieved a reduction of 1.5%. While a modest percentage, this decrease in peak consumption is crucial for enhancing grid stability and mitigating strain on energy infrastructure during critical periods.

The financial benefits are more pronounced. The "Cost Savings" chart shows that from an original total cost of £57,022.08, the IDR implementation generated substantial savings of 25.8%. This represents a significant financial advantage, highlighting the program's effectiveness in optimising energy expenditure.

The data demonstrate that the IDR program is a powerful tool, effectively curbing peak energy demand and delivering considerable cost savings for consumers and providers alike.

4.4.6 Energy Shift Analysis Based on IDR Implementation

The figure 4.11 illustrates the changes in energy source utilisation before and after the implementation of dynamic pricing IDR mechanisms. It reveals a significant reduction in electricity usage from 80% to 60%, demonstrating IDR's success in reducing peak demand. In contrast, the reliance on natural gas increased from 15% to 30%, indicating consumers shifted to natural gas in response to dynamic pricing signals. EH usage also rose from 5% to 10%, reflecting greater adoption of EH solutions.

These shifts imply improved load management, reducing grid strain and enhancing efficiency. The increased use of natural gas and energy hubs highlights economic benefits as consumers optimise costs, while the adoption of EH solutions supports renewable energy integration.

The observed energy shifts validate IDR's role in changing consumption patterns and improving grid stability. Continued refinement and investment in related technologies are essential for maximising the long-term impact on energy systems.

4.5. Chapter Summary

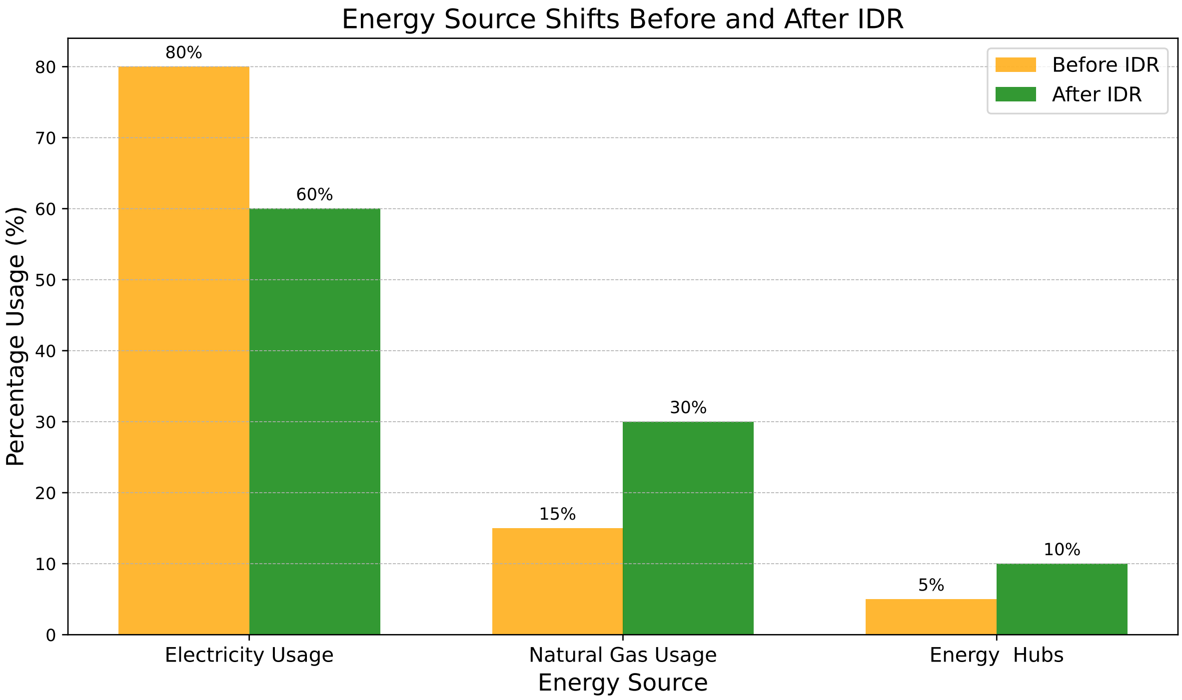

Figure 4.11: demonstrates the shifts in energy source utilisation before and after the implementation of Dynamic Pricing IDR

4.5 Chapter Summary

This chapter extended the DRL-based framework to a multi-energy system incorporating electricity, gas, and heat, with energy hubs as a central enabling component. The key contribution was demonstrating that this integrated approach, optimised by the DDPG algorithm, facilitates significant energy source substitution and achieves notable peak demand reductions. The findings highlight how EHs provide the necessary flexibility for a truly integrated and efficient demand response strategy.

Chapter 5

Dynamic Pricing IDR in P2P Multi-Energy Trading system

The integration of Peer-to-Peer (P2P) energy trading with integrated demand response presents an effective solution for decentralised energy systems. This chapter introduces a new method for managing energy in interconnected multi-energy microgrids, combining dynamic pricing with IDR strategies in a P2P trading framework. We utilise a modified Double-Actors Regularised Critics (DARC) algorithm to optimise energy resource allocation and improve system resilience. This algorithm enables the SP to implement dynamic pricing and IDR strategies that maximise overall welfare for both the SP and the participants in P2P energy trading. Numerical simulations with real-world data show that the DARC model outperforms traditional fixed pricing and other reinforcement learning methods. It enhances renewable energy use, reduces deficits, and increases profitability for service providers and traders. Thus, it offers a sustainable solution for energy resource management in decentralised markets.

5.1 Introduction

The integration of renewable energy sources and distributed energy systems has increased interest in innovative energy management strategies. P2P energy trading and IDR are effective methods for addressing the challenges of decentralised energy systems. [85] [86].

P2P energy trading enables direct transactions between consumers and prosumers in a decentralised market.[87]. This approach can lower transaction costs, improve the integration of renewable energy, and empower consumers to participate actively. [88]. As the market becomes more complex, prosumers are increasingly important.[89].

P2P energy markets can help maximise household incomes by allowing individuals and small groups to buy and sell excess electricity generated from renewable sources. However, the lack of clear trading strategies and intense competition may result in significant losses if irrational behaviour occurs. [90] [91].

IDR encourages consumers to adjust their energy consumption based on price signals and system conditions [7]. Its main goal is to improve the efficiency and reliability of energy systems while lowering costs for consumers and utilities [8]. IDR helps reduce peak demand, supports renewable energy integration, and aligns with real-time market prices by utilising the interdependencies among energy sources.

Integrating P2P energy trading with dynamic pricing can greatly enhance efficiency. Dynamic pricing, which adjusts based on real-time demand and supply, encourages consumer participation in P2P trading. [92] [93]. This combination can create a more flexible and reliable energy system that is better suited to handle the variability of renewable energy sources and diverse consumer needs. [94].

5.1. Introduction

The literature on P2P energy trading has garnered attention, leading to various proposed pricing mechanisms. Alfaverh et al. [24] introduced a pricing model based on dynamic supply-demand ratios to facilitate energy sharing in community microgrids. Qiu et al. [95] developed a double auction framework for multi-energy microgrid coordination using multi-agent reinforcement learning. Yin et al. [96] applied a Stackelberg game approach to creating win-win scenarios in energy pricing and sharing systems.

Dynamic pricing in P2P energy systems has been explored using cooperative Stackelberg game models [97] and evolutionary game theory to lower costs for prosumers and boost system efficiency.[98]. Tushar et al. [97] examined a hybrid incentive approach combining feed-in tariffs with flexible grid access to enhance P2P energy marketplace viability. Liu et al. [94] proposed an optimal bidding strategy using a double-auction framework to manage uncertainties in load demand and renewable energy production in P2P contexts.

IDR has been extensively researched, with various optimisation methods for multi-energy management, including load scheduling and energy storage. [99–101]. Aghamohamadi et al. [102] proposed a robust optimisation model for energy hub systems using a block-coordinate-descent approach. Techniques like linear programming, mixed-integer programming, and metaheuristic algorithms address complex optimisation challenges in IDR settings [103–105]. Additionally, forecasting methods, such as time-series analysis and machine learning, help accurately estimate load demand and renewable generation, improving energy resource planning and coordination. [106, 107].

5.1. Introduction

Extensive research on game theory and market mechanisms has focused on designing effective IDR frameworks, considering the interactions among stakeholders [108–111]. Techniques like double-auction mechanisms and Nash bargaining solutions have been used to create decentralised IDR models that ensure fair and efficient energy resource distribution [112–114]. Wang et al. [115] explored IDR programs that categorize consumers by their responsiveness to incentives, proposing an incentive pricing mechanism to enhance program efficiency.

In their paper "Dynamic Electricity Price Adjustment in Trading Markets Using Reinforcement Learning," Hu et al. [116] investigate using reinforcement learning to adjust electricity prices in real-time. This method aims to balance supply and demand while benefiting both suppliers and consumers, showcasing machine learning's potential in complex markets.

Zhang et al. [117] introduces a multi-objective optimisation model that integrates demand response (DR) and dynamic pricing in smart integrated energy systems (IES). This approach optimises system operations and enhances energy efficiency by considering the demand's spatio-temporal characteristics and stakeholder preferences.

Das et al. [118] propose a Q-learning-based dynamic pricing algorithm to optimise retail energy prices, reducing curtailable load and enhancing power system stability, benefiting both consumers and service providers.

Almannouny et al. [93] develop the integrated demand response field by using a deep reinforcement learning framework, specifically the DDPG method, to optimise retail energy pricing in a dynamic multi-energy market. This approach allows service providers to adjust energy prices in response to fluctuations in end-user demand and wholesale prices, benefiting both utility companies and consumers through cost savings and better peak load management.

5.1. Introduction

The existing literature has not fully explored the potential of these techniques for optimal energy trading and demand response strategies, especially considering the dynamic nature of renewable energy sources and diverse consumer needs. There is also a lack of clarity on the key factors affecting model performance, including pricing schemes, energy storage systems, and consumer responsiveness to dynamic price signals.

Despite the progress made, the combination of P2P energy trading and dynamic pricing for IDR, especially within the framework of interconnected multi-energy microgrids, has not been thoroughly investigated. This study intends to fill the current research voids by introducing an innovative model incorporating dynamic pricing IDR into a P2P energy trading system, utilising advanced optimisation methods to facilitate efficient management of energy resources. The primary contributions of this study are summarised as follows:

- To the best of our knowledge, this is the first work to consider P2P energy trading, energy conversion, and dynamic pricing IDR together holistically. A new P2P energy trading platform for interconnected residential, commercial, and industrial microgrids has been established.
- Utilising the modified Double Actors Regularised Critics (DARC) Algorithm, a sophisticated DRL approach is employed to address the complex optimisation problem of energy trading and integrated demand response in a multi-energy P2P setting. DARC improves value estimation and exploration by employing double actors and regularised critics, reducing biases and facilitating optimal energy trading strategies.

The rest of this chapter is structured as follows: Section II outlines the system architecture and develops the mathematical model for the dynamic pricing IDR. Section III introduces the proposed method for dynamic pricing IDR within P2P multi-energy trading across interconnected microgrids. Section IV details case studies to assess the proposed method's efficacy. Finally, Section V concludes the chapter.

5.2 SYSTEM MODEL

This section outlines the proposed system model, which includes energy producers, interconnected microgrids, a peer-to-peer energy trading platform, and SP. It emphasises peer-to-peer multi-energy trading in interconnected microgrids, using a dynamic pricing model based on DRL.

5.2.1 System Overview

The proposed model for dynamic pricing IDR)within a peer-to-peer energy trading framework builds on recent studies. It adapts existing designs to incorporate demand response mechanisms effectively. Influenced by Chen et al. (2021), which uses multi-agent deep reinforcement learning for P2P trading and energy conversion [119], our model introduces a dynamic pricing IDR mechanism, offering a novel contribution by integrating these strategies into P2P energy trading for multi-energy microgrids.

In our framework, as illustrated in Fig. 5.1, energy hubs play a crucial role in overseeing the distribution and transformation of different forms of energy within the system. Based on the concept by Yang et al. (2022), these hubs serve as key nodes in P2P networks that convert, store, and distribute energy from multiple sources, including renewables [120]. Each EH utilises advanced technology to handle solar, wind, and conventional energy inputs, enhancing the microgrids' flexibility and sustainability.

Residential microgrids (RES MG) use solar panels to generate electricity. Excess energy is stored as hydrogen through water electrolysis. This hydrogen can be stored indefinitely and later converted back into electricity and heat, with natural gas as a backup fuel.

Commercial microgrids (COM MG) also harness solar energy but primarily use electric heat pumps to heat water. They employ thermal storage systems to reserve surplus heat.

Industrial microgrids (IND MG) focus on energy efficiency. They use combined heat and power (CHP) generators to produce electricity and heat. Wind turbines support electricity generation, and storage systems manage energy fluctuations.

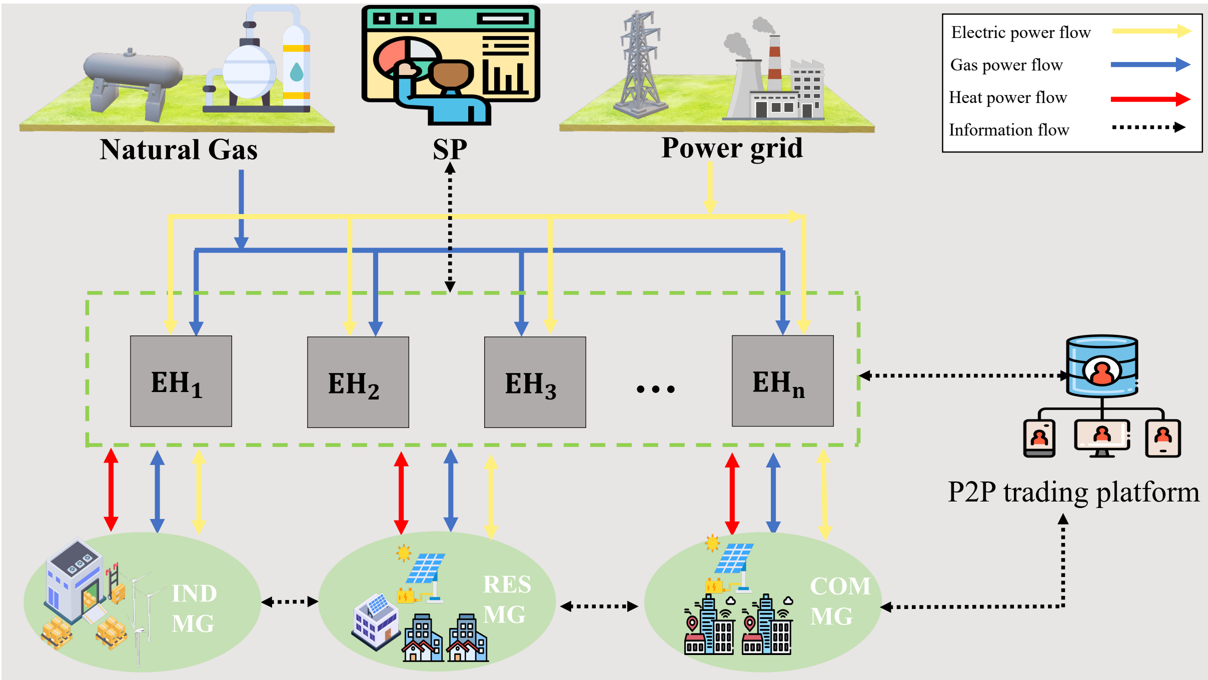

Figure 5.1: representing the integration of dynamic pricing IDR into a P2P multi-energy trading system.

The dynamic pricing IDR mechanism in our P2P platform adapts to real-time changes in energy demand and supply, allowing for more efficient trades between consumers and prosumers. By integrating dynamic pricing into the IDR framework, our model aligns prices with market conditions and encourages energy-saving behaviours and optimal consumption patterns.

The P2P trading platform is central to this system, enabling energy trades between hubs based on dynamic pricing. It promotes a competitive yet cooperative environment, helping microgrids optimise their energy portfolios for a more resilient and economically viable network.

This architecture, inspired by effective models, incorporates integrated dynamic pricing IDR strategies to meet the needs of interconnected multi-energy microgrids. This integration aims to enhance system resilience, optimise resource allocation, and improve overall efficiency.

5.2.2 Problem Formulation and System Dynamics

This subsection presents the mathematical formulation of the dynamic pricing IDR model for P2P multi-energy trading in interconnected microgrids. The model adjusts pricing based on changing conditions to ensure prosumers' profitability and system stability while accurately representing energy flow between electricity and gas systems, including energy conversion and storage.

5.2.2.1 *State Space*

The state space captures the information available to agents at each time step t, represented by the vector s_t, which includes energy demand, supply, storage levels, renewable output, and pricing.

$$s_t = \left[D_t, S_t, E_{\text{storage }, t}, G_{\text{storage }, t}, E_{\text{renewable }, t}, P_{\text{elec }, t}, P_{\text{gas }, t}\right] \tag{5.1}$$

5.2.2.2 *Action Space (A)*

At time t, the action a_t involves adjustments to dynamic electricity and gas prices, along with changes in energy consumption patterns in response to IDR signals.

$$a_t = \left[P^{\mathrm{d}}_{\mathrm{elec}\ ,t}, P^{\mathrm{d}}_{\mathrm{gas}\ ,t}, IDR_t^{\mathrm{adjust}}\right] \tag{5.2}$$

where $P^{d}_{\mathrm{elec,t}}$ and $P^{\mathrm{d}}_{\mathrm{gas,t}}$ represent the adjustments in electricity and gas prices, respectively, and IDR_t^{adjust} indicates changes in demand response actions (e.g., energy conservation or shifting consumption to off-peak times).

5.2.2.3 *Reward function*

The reward function aims to maximise SP profitability while ensuring the user's net income. It can be mathematically represented as:

$$R(s_t, a_t) = \lambda_1 \cdot \pi_{SP,t} + \lambda_2 \cdot \pi_{p2p,i} + \lambda_3 \cdot U_{IDR} - \lambda_4 \cdot C_{Op,t} \tag{5.3}$$

In the given equations: $R(s_t, a_t)$ is the reward at time t for state s_t and action a_t. $\pi_{SP,t}$ represents the SP's profit at time t. $\pi_{i,t}$ represents the user's net income at time t, considering energy costs and participation incentives. $C_{Op,t}$ represents the operational costs at time t, including penalties for not meeting regulatory requirements.

The SP's revenue is calculated as the difference between revenue from selling energy and the cost of procuring/generating this energy. This can be expressed as follows:

$$\pi_{SP,t} = \text{Maximize} \sum_{t \in T} \left(P_t \cdot S_{\mathrm{sold}\ ,t} - C_t \cdot B_{\mathrm{bought}\ ,t}\right) \tag{5.4}$$

The profit for any microgrid i participating in P2P trading during time step t can be determined by considering both the revenue generated from selling energy to other microgrids and the costs incurred from purchasing energy. The profit $\pi_{p2p,i}$ for microgrid i at time t can be formulated as follows:

$$\pi_{p2p,i} = \sum_{j:(i,j)\in\mathscr{E}} P_{ij,t}E_{ij,t} - \sum_{k:(k,i)\in\mathscr{E}} P_{ki,t}E_{ki,t} \tag{5.5}$$

Here, the first term represents the total revenue from selling energy to neighbouring microgrids via P2P platform j, where $(i,j) \in \mathscr{E}$ (the set of microgrids that i can trade with directly). The second term represents the total cost of purchasing energy from neighbouring microgrids through the P2P platform k.

Operational costs at time t can encompass several factors, including energy procurement costs, maintenance, regulatory compliance penalties, and the costs associated with demand response actions. This can be defined as:

$$CO_{p,t} = CO_{\text{proc },t} + CO_{\text{maint },t} + CO_{\text{reg},t} + CO_{\text{IDR},t} \tag{5.6}$$

Where: $CO_{\text{proc },t}$ energy procurement costs at time t, covering expenses for purchasing energy and production costs. $CO_{\text{maint },t}$ maintenance costs at time t, including routine and unscheduled maintenance. $CO_{\text{reg},t}$ regulatory compliance costs or penalties incurred at time t, including fines for emissions or delivery schedule deviations. $CO_{\text{IDR},t}$ costs for integrated demand response actions at time t, including consumer incentives for demand reduction or usage adjustments.

5.2.3 Three-stage System Process

The three-stage system process includes P2P energy trading, energy conversion, and dynamic pricing IDR. Each stage informs the next: outcomes from P2P trading influence energy conversion decisions, and the system status post-conversion shapes pricing strategies. This integrated approach seeks to create a more flexible, efficient, and sustainable energy system that adapts to renewable energy fluctuations and meets diverse consumer needs.

5.2.3.1 P2P Energy Trading

Prosumers submit buying or selling bids to the P2P platform based on their energy generation and consumption forecasts. The platform matches bids, enabling direct transactions at agreed prices. At time t, microgrid i decides on its trading actions, represented by a vector x_t^i. Therefore,

$$x_t^i = [x_t^{ij}]_{-}\{1 \leq j \neq i \leq N\} \tag{5.7}$$

x_t^{ij} denotes the intended amount of energy to be traded between microgrid i and microgrid j at the time t. If $x_t^{ij} > 0$, microgrid i intends to buy energy from microgrid j. And if $x_t^{ij} < 0$, microgrid i intends to sell energy from microgrid j. After negotiations, the actual energy trading is finalised and represented by

$$z_t^i = [z_t^{ij}]_{-}\{1 \leq j \neq i \leq N\} \tag{5.8}$$

Where z_t^i denotes the actual amount of energy traded between microgrid i and microgrid j at the time t, the actual trading may differ from the intended trading due to negotiations and constraints.

5.2.3.2 Energy Conversion Stage

Each energy hub h at time t decides on its energy conversion and storage actions, represented by a vector y_t^h. It includes conversion technologies (Combined heat and power (CHP) units, power-to-gas (P2G) units, electric heat pumps (EHPs)) and storage technologies (electricity storage and gas storage). Therefore,

$$y_t^h = E_{\text{elec},CHP}(t), E_{\text{gas},CHP}(t), E_{\text{heat},CHP}(t), E_{\text{gas},P2G}(t), E_{\text{heat},EHP}(t), S_{\text{ele}}(t), S_{\text{gas}}(t) \tag{5.9}$$

Energy conversion and storage occur within each EH based on the outcomes of the P2P trading stage z_t^i and the EH actions y_t^h.

5.2.3.3 Dynamic Pricing IDR Stage

In this stage, SP monitors the system state s_t and determines the dynamic pricing for electricity P_{ele,t^d} and P_{gas,t^d}. The SP also issue IDR signals, IDR_adjust,t, to influence the energy consumption patterns of prosumers and consumers. These signals promote energy conservation during peak demand periods, facilitate the integration of renewable energy, and enhance overall system stability and cost-effectiveness. The adjusted demand represented as $D_{\text{adjusted },t}$ is a function of the original demand $D_t - \Delta$, the dynamic prices indicated by the IDR signals.

$$D_{\text{adjusted },t} = D_t - \Delta D\left(P_t, IDR_t\right) \tag{5.10}$$

The three interconnected stages influence each other. The outcomes of the P2P energy trading stage z_t^i affect the energy resources available for each microgrid during the energy conversion stage. In turn, the energy conversion actions y_t^h impact the overall system state (s_t), which influences the dynamic pricing and IDR decisions made by the SP in the next time step. Additionally, the dynamic prices P_{ele,t^d}, P_{gas,t^d} and IDR signals IDR_adjust,t shape the energy demand and P2P trading behaviour of prosumers and consumers in subsequent time steps.

5.2.3.4 Objective Function

The overall system objective is to maximise the following function:

$$\text{Maximize: } \lambda_1 \cdot \pi_{SP,t} + \lambda_2 \cdot \pi_{p2p,i} + \lambda_3 \cdot U_{IDR} - \lambda_4 \cdot C_{Op,t} \tag{5.11}$$

This objective function aims to achieve a balanced and efficient energy management strategy that benefits all P2P multi-energy trading system stakeholders.

5.2.4 Constraints

The constraints define the feasible operating region for the P2P energy trading system, which encompasses dynamic pricing, IDR, and energy hubs. The optimisation problem aims to identify the most effective actions, including P2P trading decisions, energy hub operations, dynamic pricing adjustments, and IDR signals, that will maximise the system's overall objectives while ensuring compliance with these physical constraints.

5.2.4.1 P2P Trading Constraints

- *Trading Limits:* The amount of energy traded between microgrids is subject to limitations based on network capacity, generation capabilities, and demand constraints.

$$x_{min}^{ij} \leq x_t^{ij} \leq x_{max}^{ij} \tag{5.12}$$

  Where x_{min}^{ij} and x_{max}^{ij} represent the minimum and maximum allowable energy trading amounts between microgrids i and j, respectively.

- *Energy Balance:* The total energy bought and sold by a microgrid in the P2P market should be balanced with its net energy production/consumption and any interactions with the external grid.

$$\sum_{j \neq i} z_t^{ij} + z_t^{ii} = G_t^i - D_t^i \tag{5.13}$$

G_t^i is the net energy generation of microgrid i at time t (considering renewable generation and energy hub output). D_t^i is the total energy demand of microgrid i at time t.

5.2.4.2 Energy Hub Constraints

- *Conversion Efficiency:* The energy conversion processes within the energy hub are subject to efficiency limitations.

$$0 \leq \eta_{conv} \leq 1 \tag{5.14}$$

  Where η_{conv} represents the conversion efficiency of a specific energy conversion technology within the energy hub.

- *Storage Capacity:* The energy storage systems within the energy hub have limited capacity.

$$0 \leq S_t \leq S_{max} \tag{5.15}$$

Where S_t is the energy stored in the storage system at time t. And S_{max} is the maximum storage capacity

- *Storage Dynamics:* The energy levels in the storage systems evolve over time based on charging/discharging rates and self-discharge characteristics.

$$S_{t+1} = S_t + \eta_{ch} \cdot E_{in}(t) - \frac{1}{\eta_{disch}} \cdot E_{out}(t) - \eta_{sd} \cdot S_t \tag{5.16}$$

Where η_{ch} and η_{disch} are the charging and discharging efficiencies, respectively. $E_{in}(t)$ and $E_{out}(t)$ are the energy charged into and discharged from the storage system at time t, respectively. η_{sd} is the self-discharge rate.

- *Energy Balance:* The energy hub must maintain the balance between the energy inflow (from P2P trading and external grid) and the energy outflow (to meet local demand and for storage).

$$\sum E_{in}(t) = \sum E_{out}(t) + \sum E_{stored}(t) \tag{5.17}$$

5.2.4.3 Dynamic Pricing IDR Constraints

- *Price Bounds:* The dynamic prices set by the ESP should lie within reasonable bounds to avoid market distortions and ensure affordability for consumers.

$$P_{min} \leq P_{t^d} \leq P_{max} \tag{5.18}$$

Where P_{min} and P_{max} represent the minimum and maximum allowable dynamic prices, respectively.

- *Price Adjustment:* The dynamic price adjustments should be smooth and gradual to avoid sudden price shocks and promote market stability.

$$|P_t^t - P_{t-1}^d| \leq \Delta P_{max} \tag{5.19}$$

Where ΔP_{max} is the maximum allowable price change between two consecutive time steps.

- *IDR Signal Limits:* The IDR signals issued by the SP should be within technically feasible and acceptable ranges for the prosumers/consumers.

$$IDR_{min} \leq IDR_{adjust,t} \leq IDR_{max} \tag{5.20}$$

Where $IDR_{adjust,t}$ and IDR_{max} represent the minimum and maximum allowable IDR signal levels, respectively.

5.3 Proposed Enhanced DARC Algorithm

This research presents an enhanced DARC algorithm designed to tackle the challenges of dynamic pricing IDR in P2P energy trading across interconnected multi-energy microgrids. It incorporates double-actor networks for improved exploration and utilises regularised critic networks to increase the stability of value estimation.

The adjusted DARC algorithm acts as the SP's decision-making framework within our proposed system. Its primary goal is to identify the optimal dynamic pricing IDR strategies that maximise overall system welfare, factoring in both the SP's profit and the collective benefits of prosumers and consumers engaged in P2P energy trading.

5.3. Proposed Enhanced DARC Algorithm

An effective enhancement in our approach is regularising the critic networks, which reduces the risk of overestimation bias that can arise from using multiple critics. The DARC algorithm seeks to maximise the overall system objective by balancing the interests of the SP with those of prosumers and consumers while also fostering efficient energy utilisation and demand-side flexibility.

5.3.1 Modified DARC algorithm implementation

This research implements the DARC algorithm, which comprises two key neural networks: the actor and critic networks. The actor network's main function is to produce optimal actions, specifically by adjusting dynamic pricing for electricity and gas, as well as altering energy consumption patterns in response to IDR signals. In contrast, the critic network assesses the value of these actions by predicting the expected future rewards linked to them. The approach is illustrated in Figure 5.2, which shows the information flow.

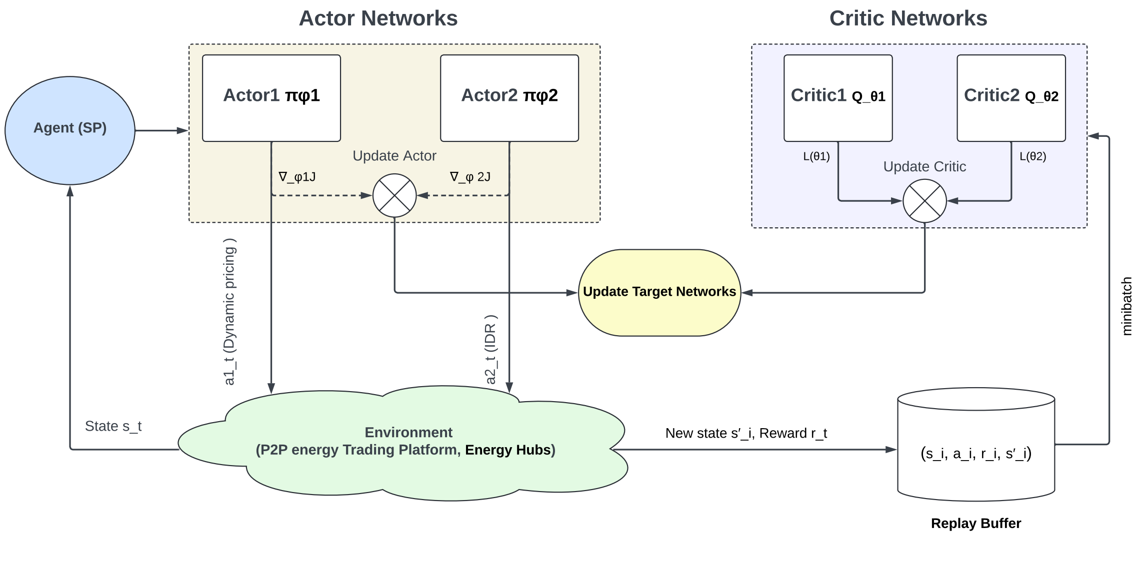

Figure 5.2: illustrates the information flow of the modified DARC algorithm.

5.3.1.1 Double Actor Networks

The actor-network develops a policy that translates the current system state, which encompasses factors such as energy demand, supply levels, storage capacities, and renewable energy output, into the most beneficial action. The network's design captures the intricate relationships among these state variables and potential actions, allowing it to make well-informed decisions aimed at maximising overall system welfare.

The DARC algorithm enhances this decision-making process through the use of double actors, which introduces an element of exploration and helps to mitigate biases in action selection. Mathematically, this can be expressed as:

$$a_t = \pi_\phi(s_t) = [P^d_{elec,t}, P^d_{gas,t}, IDR_{adjust,t}] \tag{5.21}$$

Where π_ϕ represents the policy function learned by the actor-network.

Policy update: The actor network's policy is updated using the policy gradient method to maximise the expected future rewards. The gradient of the policy network is calculated as follows:

$$\nabla_\phi J(\phi) = \mathbb{E}_{s_t}[\nabla_\phi \pi_\phi(s_t) \nabla_a Q_\theta(s_t, a_t)|_{a_t=\pi_\phi(s_t)}] \tag{5.22}$$

Where $J(\phi)$ is the objective function representing the expected future rewards. $\nabla_\phi \pi_\phi(s_t)$ is the gradient of the policy function for the network parameters ϕ. $\nabla_a Q_\theta(s_t, a_t)$ is the gradient of the Q-value function (critic network) for the actions a_t. The gradient is evaluated at the actions suggested by the actor-network itself $a_t = \pi_\phi(s_t)$.

Double Actor Exploration: In the modified DARC algorithm, two actor networks ($\pi_\phi 1$ and $\pi_\phi 2$) are employed to enhance exploration. For each state s_t, both actor networks propose actions, and the action leading to a higher Q-value is selected:

$$a_t = argmax_{a \in \{\pi_{\phi_1}(s_t), \pi_{\phi_2}(s_t)\}} Q_\theta(s_t, a) \tag{5.23}$$

Both actor networks are updated independently using their respective policy gradients.

$$\begin{aligned} \nabla_\theta J &\approx \frac{1}{N} \sum \nabla_{a_1} Q_{\theta_1}(s, a_1)\big|_{a_1 = \pi_\theta(s)} \nabla_\theta \pi_\theta(s) \\ \nabla_\phi J &\approx \frac{1}{N} \sum \nabla_{a_2} Q_{\theta_2}(s, a_2)\big|_{a_2 = \pi_\phi(s)} \nabla_\phi \pi_\phi(s) \end{aligned} \tag{5.24}$$

5.3.1.2 Regularised Critic Networks

The Critic network functions as the evaluative counterpart to the Actor-network, employing its neural network architecture to estimate the value function. This function quantifies the expected future rewards associated with a specific state-action pair. This estimation is essential for guiding the Actor network's learning and ensuring that it converges toward an optimal policy. Furthermore, the regularisation mechanism incorporated into the Critic network enhances the stability of the learning process and mitigates the risk of overestimating the value function, thereby leading to more robust and reliable performance.

Critic Network Update: The critic network, parameterised by θ, estimates the Q-value function $Q_\theta(s_t, a_t)$, representing the expected future rewards for taking action a_t in state s_t. The network is updated by minimising the mean squared error between its estimated Q-values and the target values y_t.

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} [(Q_\theta(s_i, a_i) - y_i)^2] \tag{5.25}$$

Where N is the batch size. y_i is the target value for the $i-$ th sample, calculated as:

$$y_i = r_i + \gamma(1-d_i)\hat{V}(s_i';\nu) \tag{5.26}$$

where r_i is the immediate reward received after taking action a_i in state s_i. γ is the discount factor. d_i is a binary flag indicating whether the episode terminated after taking action a_i in state s_i. $\hat{V}(s_i';\nu)$ is the soft target value for the next state s_i', calculated using a convex combination of the Q-values from both critics and actors.

Regularisation of Critic Networks: To mitigate the potential overestimation bias and reduce variance in value estimation, a regularisation term is added to the critic loss function:

$$\begin{aligned} L(\theta_1) &= \frac{1}{N}\sum(y_i - Q_{\theta_1}(s_i,a_{1i}))^2 \quad + \lambda\left(Q_{\theta_1}(s_i,a_{1i}) - Q_{\theta_2}(s_i,a_{1i})\right)^2 \\ L(\theta_2) &= \frac{1}{N}\sum(y_i - Q_{\theta_2}(s_i,a_{2i}))^2 \quad + \lambda\left(Q_{\theta_1}(s_i,a_{1i}) - Q_{\theta_2}(s_i,a_{2i})\right)^2 \end{aligned} \tag{5.27}$$

Where λ is the regularisation coefficient.

5.3.1.3 Target Network Update:

The target critic and actor networks are updated slowly to stabilise the learning process using a soft update rule with a parameter:

$$\begin{aligned} \pi'_{\theta_1} &\leftarrow \tau \cdot \pi_{\theta_1} + (1-\tau)\cdot \pi'_{\theta_1} \\ \pi'_{\theta_2} &\leftarrow \tau \cdot \pi_{\theta_2} + (1-\tau)\cdot \pi'_{\theta_2} \\ Q'_{\theta_1} &\leftarrow \tau \cdot Q_{\theta_1} + (1-\tau)\cdot Q'_{\theta_1} \\ Q'_{\theta_2} &\leftarrow \tau \cdot Q_{\theta_2} + (1-\tau)\cdot Q'_{\theta_2} \end{aligned} \tag{5.28}$$

5.3.2 Training Process of the Modified DARC Algorithm

The modified DARC algorithm is employed for dynamic pricing IDR in P2P energy trading among interconnected multi-energy microgrids. The training process involves iterative steps that optimise the policies of actor networks and the value estimations of critic networks for effective energy management.

Training begins with initialising actor and critic networks using random parameters and setting up a replay buffer to store experiences. This buffer is essential for agents to learn from past interactions.

At each time interval, agents assess the system's existing state and determine actions in accordance with their established policies. They employ dual actor networks to optimise the Q-value. These actions result in subsequent states and rewards contingent upon performance outcomes.

The replay buffer records transitions, including the current state, selected action, reward, next state, and an indication of whether the episode has terminated. This enables agents to enhance their decision-making by revisiting previous experiences.

The network updates involve sampling batches of transitions from the replay buffer. The critic networks are updated by minimising the mean squared error between their estimated Q-values and the target values, with a regularisation term for consistency. The actor networks are updated using the policy gradient method, informed by the critic networks' value estimations. The target critic and actor networks are updated periodically to ensure stable learning and gradual adjustments, which help stabilise the training process.

This process includes action selection, experience storage, and network updates and is repeated until convergence or a predefined stopping criterion is met. Through this, policies and value estimations are refined to achieve an optimal solution that balances individual microgrid objectives with user benefits.

The modified DARC algorithm employs a multi-agent framework with double actors for improved exploration and regularised critic networks for better value estimation. By incorporating dynamic pricing IDR mechanisms, it effectively addresses the complexities of P2P multi-energy trading, leading to adaptive strategies that optimise energy management and enhance system efficiency.

Algorithm 3: Modified DARC Algorithm for Dynamic Pricing IDR in P2P Energy Trading

---

Input: Initialize actor networks π_{ϕ_1}, π_{ϕ_2} with random parameters ϕ_1, ϕ_2

Input: Initialize critic networks Q_{θ_1}, Q_{θ_2} with random parameters θ_1, θ_2

Input: Initialize target networks π'_{ϕ_1}, π'_{ϕ_2}, Q'_{θ_1}, Q'_{θ_2} with parameters $\phi'_1 \leftarrow \phi_1$, $\phi'_2 \leftarrow \phi_2$, $\theta'_1 \leftarrow \theta_1$, $\theta'_2 \leftarrow \theta_2$

Input: Initialize replay buffer $\mathscr{D}$

for *episode = 1 to M* **do**
- Initialise environment and state s
- **for** t = *1 to* T **do**
  - Select action a based on Eq 5.23 using current state s and both actor networks
  - Execute action a and observe reward r and new state s'
  - Store transition (s,a,r,s') in replay buffer $\mathscr{D}$
  - Sample a random minibatch of N transitions (s_i,a_i,r_i,s'_i) from $\mathscr{D}$
  - Compute target values y_i for each transition in the minibatch using Eq 5.26
  - Update critic networks θ_1, θ_2 by minimizing the loss functions in Eq 5.27
  - Update actor networks ϕ_1, ϕ_2 using the policy gradients in Eq 5.24
  - Update target networks using Eq 5.28 with a soft update rate τ
  - $s \leftarrow s'$
- **end**

end

5.3.3 Execution Steps

The DARC algorithm implementation in Algorithm 3 begins by initialising actor and critic networks with random parameters and a replay buffer designed to store experiences. At each time step, the algorithm evaluates the system's state, which includes energy demand, supply, storage levels, and renewable energy output. Actions are determined based on the policies of both actor networks, prioritising those that yield a higher Q-value, which reflects expected future rewards. After selecting an action, it is executed, and the resulting reward and updated system state are observed.

This transition, comprising the current state, the action taken, the reward received, the next state, and an episode termination indicator, is stored in the replay buffer. Batches of transitions are then sampled from this buffer to update the critic networks by minimising the mean squared error between the estimated Q-values and the target Q-values, with a regularisation term added for consistency. The actor networks are updated through the policy gradient method, guided by the value estimations from the critic networks. Furthermore, the target critic and actor networks are periodically refreshed to ensure stable learning. This iterative process continues until convergence is achieved or a predetermined stopping criterion is met.

5.4 Numerical Simulation and Analysis

This section evaluates the DARC algorithm's performance within a P2P multi-energy trading framework designed for dynamic pricing and IDR in interconnected microgrids.

5.4.1 Simulation Setup

Our simulation utilised datasets from the Open Energy Data Initiative (OEDI) [121] and Energy Data and Research [122]. Including energy pricing, generation, demand, supply, and renewable generation.

For the configuration of microgrids, we utilized real-world data from three different settings: residential microgrids (RES MG) in Mueller, Austin, Texas [123], commercial microgrids (COM MG) also in Mueller [123], and industrial microgrids (IND MG) from a trial site in Aachen/Cologne, Germany. Each dataset provided high-resolution (hourly) data for components such as Photovoltaic (PV) systems, Electrical Energy Storage (EES), Thermal Energy Storage (TES), Fuel Cells (FC), Gas Boilers (GB), and interactions with the main grid.

5.4.2 Performance Evaluation

This study conducts a comparison between a modified DARC-based scheme and a baseline method that employs traditional fixed pricing, as well as three DRL approaches: DDPG, MADDPG, and MATD3. A consistent Python simulation environment has been utilised to ensure the comparability of results, with hyperparameters for all models detailed in Table 5.1. Each model was trained over 200 episodes, with each episode consisting of 50 steps, in order to facilitate a fair comparison.

The average hourly costs of various DRL models MDARC, DDPG, MADDPG, and MATD3 are compared across three microgrid types: Residential, Commercial, and Industrial MEMGs, as shown in Fig. 5.3.

Table 5.1: Hyper-parameters for Different RL Models

Hyperparameter	**DARC**	**DDPG**	**MADDPG**	**MATD3**
Max Steps per Episode	50	50	50	50
Episodes	200	200	200	200
Replay Buffer Size	10^3	10^3	10^3	10^3
Batch Size	128	32	64	64
Learning Rate (Actor)	0.001	0.001	0.001	0.001
Learning Rate (Critic)	0.002	0.002	0.002	0.002
Discount Factor γ	0.99	0.99	0.99	0.99
Soft Update Rate τ	0.01	0.005	0.005	0.005

Residential MEMG: The MDARC model demonstrates the lowest and most stable costs, reflecting effective management practices. In contrast, the MATD3 model incurs significantly higher costs due to inefficient energy trading.

Commercial MEMG: Similarly, MDARC showcases lower average costs, while both MATD3 and MADDPG exhibit higher and more variable costs, indicating challenges with efficiency.

Industrial MEMG: Once again, MDARC stands out with the lowest costs, whereas MATD3 experiences elevated costs with considerable fluctuations, underscoring difficulties in cost control.

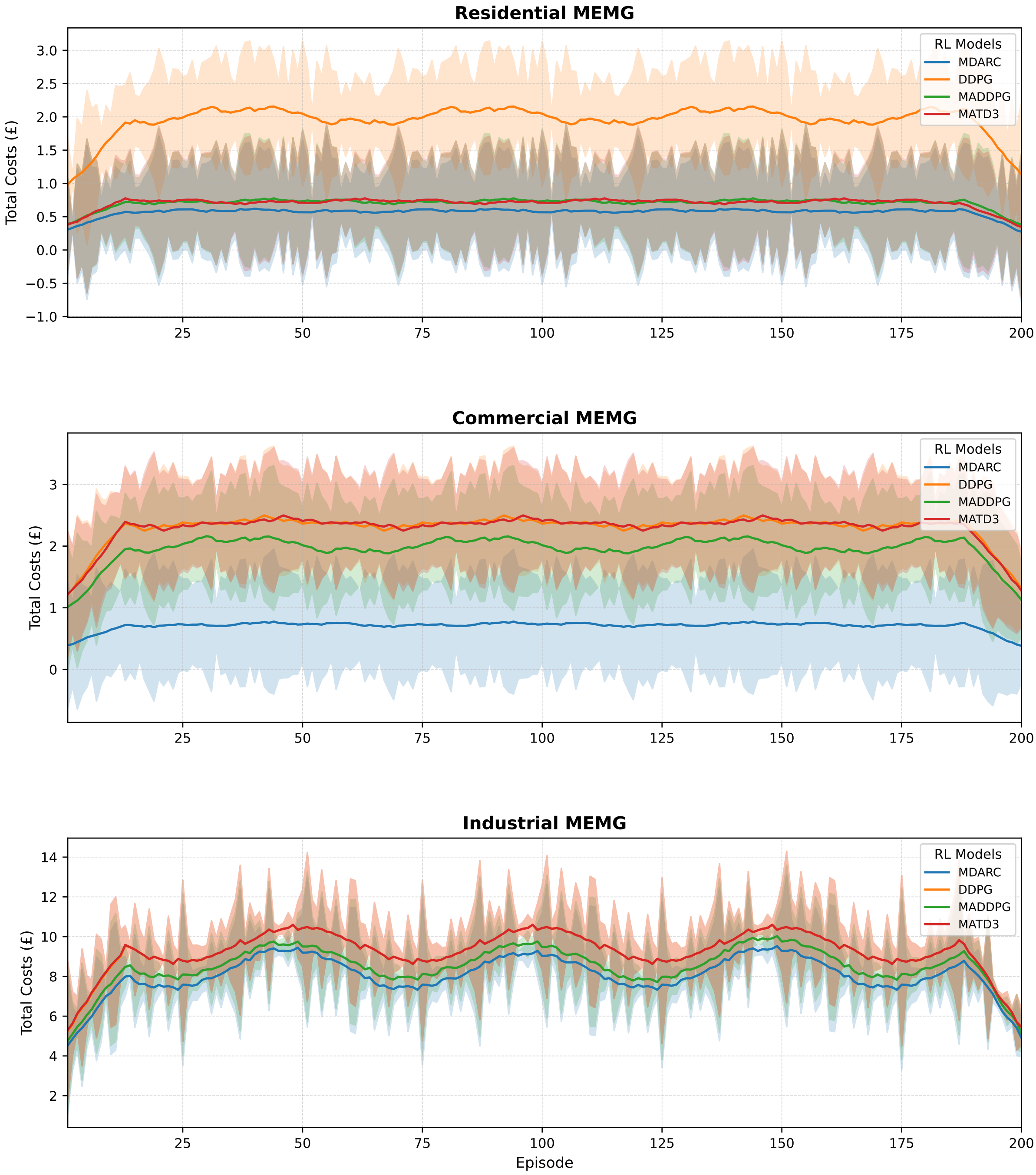

Figure 5.3: Shows Comparison of Average Hourly Costs across Reinforcement Learning Models for Residential, Commercial, and Industrial MEMGs.

5.4.3 Profitability Analysis

The profitability analysis provides a comprehensive comparison of the financial performance of various DRL models, specifically focusing on their impacts on both SP and participants within the P2P energy trading network. This detailed evaluation plays a crucial role in identifying the most economically viable model that not only maximises profits but also fosters community engagement and benefits.

Among the analysed models, the MDARC model stands out due to its exceptional capability to generate higher profits for service providers operating within various MEMGs. In particular, this model produces substantial earnings, yielding £78,226 from the residential micro grid, £82,621 from the Commercial micro grids, and £81,486 from the industrial micro grids. These figures, which are visually represented in Figure 5.4, highlight the MDARC model's effectiveness across different sectors, underscoring its potential to enhance financial stability and promote sustainable energy practices within the community.

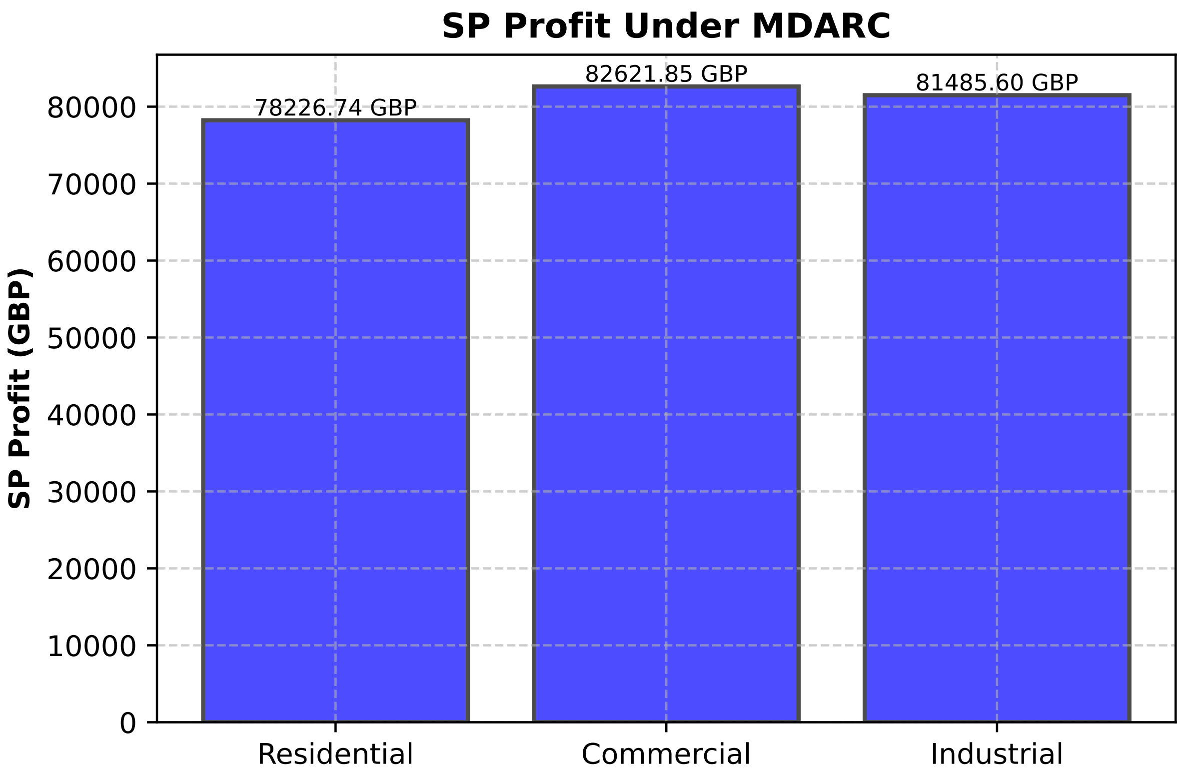

Figure 5.4: show SP profitability analysis under MDARC across three MEMGs.

5.4. Numerical Simulation and Analysis

The MDARC algorithm demonstrates superior performance in enhancing net income for participants within P2P energy trading frameworks, as evidenced by sector-specific financial outcomes Figure. 5.5. Residential MEMG achieve a notable net income of £4,136, reflecting the algorithm's capacity to optimise decentralised energy exchanges in smaller-scale, demand-flexible environments. Commercial MEMGs exhibit even greater financial gains, with earnings reaching £10,243.08, underscoring MDARC's effectiveness in balancing high-energy consumption patterns with dynamic pricing incentives. Industrial MEMGs, while slightly lower at £9,357.49, still showcase robust profitability, indicative of the model's adaptability to complex, high-demand operational contexts. These disparities highlight MDARC's ability to tailor energy allocation strategies across diverse sectors, fostering enhanced economic viability and equitable market participation. The results emphasise the algorithm's role in advancing sustainable energy ecosystems by aligning stakeholder profitability with efficient resource utilisation.

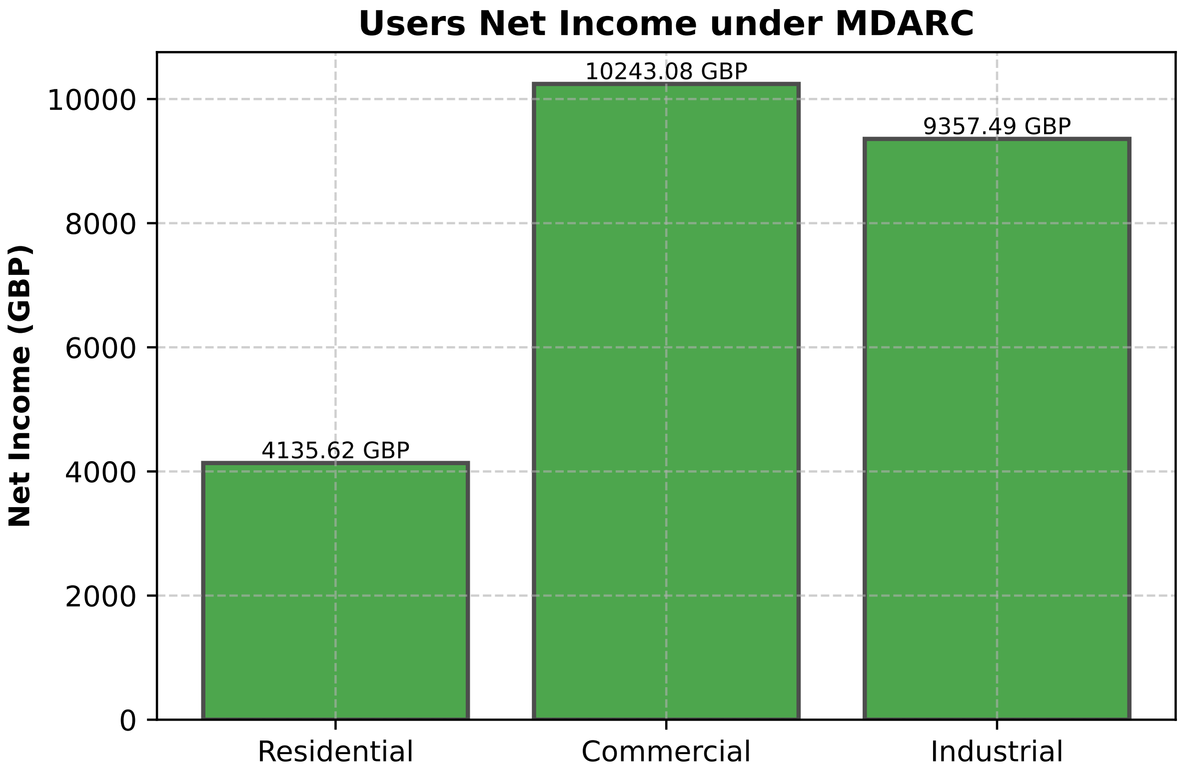

Figure 5.5: Demonstrate total users' net income under MADARC algorithm for participants within peer-to-peer (P2P) energy trading frameworks

5.4. Numerical Simulation and Analysis

The figure 5.6 provides a comparative analysis of profit generation across three distinct MEMG residential, commercial, and industrial evaluated under two frameworks: Service Provider (SP) and Peer-to-Peer (P2P) energy trading. This analysis is conducted using four RL models: MDARC, DDPG, MADDPG, and MATD3. Profits in each microgrid are divided into SP-driven revenues and P2P-derived earnings, offering valuable insights into the economic viability of decentralised energy markets.

MDARC consistently outperforms the other models across all MEMGs, demonstrating a superior ability to balance SP profitability with equitable P2P trading gains. In residential microgrids, MDARC's dual emphasis on dynamic pricing and demand flexibility aligns well with lower energy requirements, resulting in stable returns for both SPs and participants. In the commercial sector, MDARC reveals enhanced profitability, attributed to its capacity to manage complex consumption patterns and incentivise participation from demand-side players. Although industrial applications yield slightly lower P2P profits compared to the commercial sector, they still showcase MDARC's robustness in optimising high-volume, multi-energy transactions.

The comparatively lower performance of DDPG, MADDPG, and MATD3 highlights the challenges associated with scalability and real-time adaptability, particularly in balancing centralised revenue goals with the dynamics of decentralised markets. These findings underscore MDARC's pivotal role in fostering sustainable energy ecosystems by harmonising stakeholder profitability with efficient resource allocation, thereby offering a blueprint for future advancements in intelligent energy management systems.

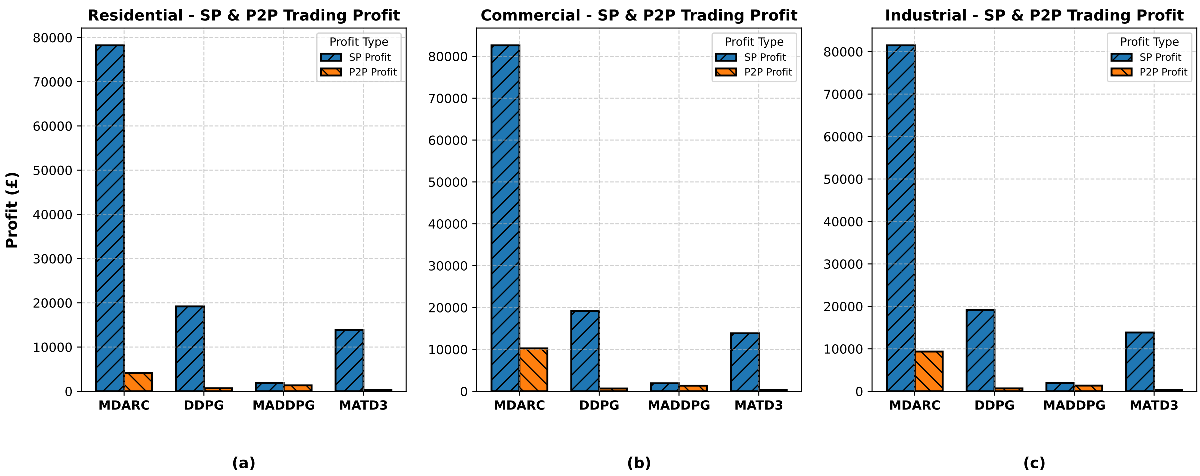

Figure 5.6: (a)-(c) showcasing SP profit and P2P profit for different MEMG types (Residential, Commercial, Industrial) and across RL models (MDARC, DDPG, MADDPG, MATD3).

5.4.4 Dynamic Price IDR Effectiveness

The analysis covers the integration of dynamic pricing and IDR in a P2P multi-energy trading system. It focuses on the flow of electricity, gas, and heat among various microgrids and highlights how energy hubs (EHs) manage energy distribution to improve flexibility and sustainability. Compared to much lower profits under DDPG, MADDPG, and MATD3, this showcases DARC's ability to create a cooperative trading environment that benefits both prosumers and consumers.

Figure 5.7 shows the impact of dynamic pricing IDR on energy demand over 24 hours across residential, commercial, and industrial MEMGs. The bar plots present average electricity demand before and after dynamic pricing IDR implementation, while line graphs depict percentage reductions in demand. Residential MEMGs exhibit significant reductions during peak hours due to their responsiveness to IDR signals, whereas commercial and industrial MEMGs demonstrate more consistent demand shifts, indicating varying operational flexibility. These results highlight the potential of IDR to optimise energy consumption, reduce grid stress, and promote sustainability across sectors.

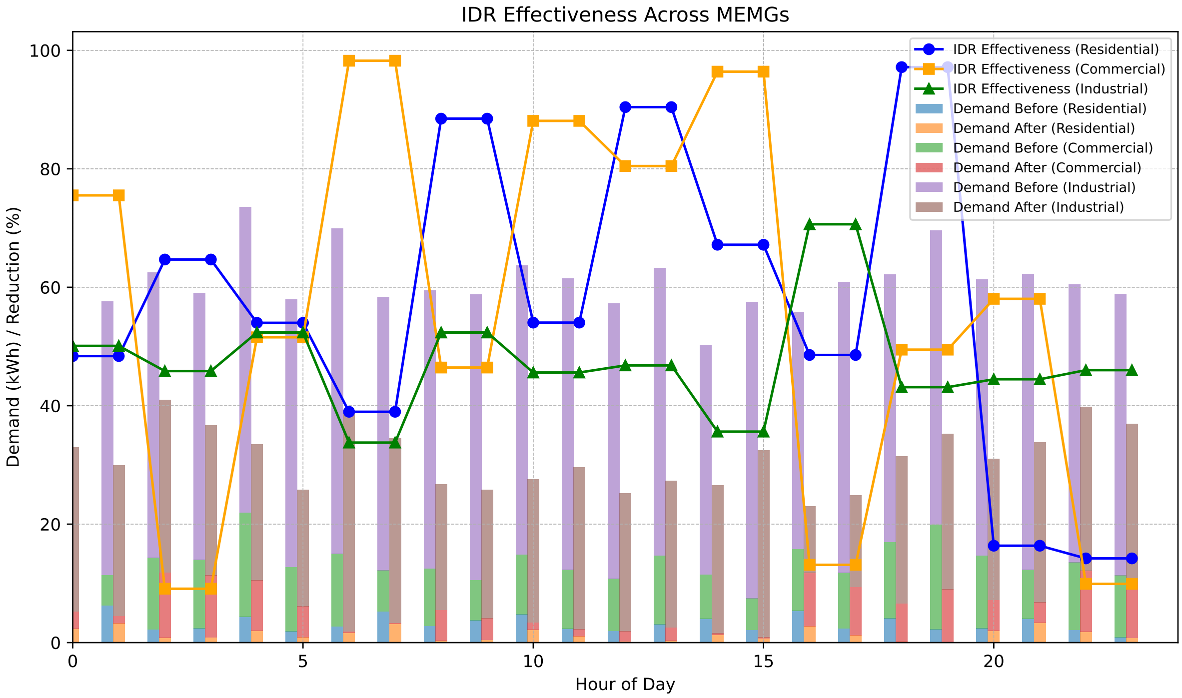

Figure 5.7: Depicts Visualisation of IDR effectiveness and energy demand shifts before and after IDR implementation across MEMG types.

Figure 5.8 also showcases the interactions between different microgrids (residential, commercial, and industrial) and the main power grid as an external energy resource. It reveals how different MEMGs depend on local versus external energy resources. The stacked bar chart indicates that residential MEMGs rely heavily on local generation and storage. In contrast, commercial and industrial MEMGs maintain a balance between local and external sources, reflecting their operational needs. This underscores the importance of local resources for enhancing system self-sufficiency.

The impact of dynamic pricing IDR on peak shifting within MEMGs is noteworthy. This strategy assesses the effectiveness of adjusting energy consumption during peak times. Industrial MEMG demonstrates the highest capacity for peak shifting, followed by commercial MEMG, which exhibits moderate shifts. Residential MEMG contributes smaller but consistent shifts, playing a crucial role in stabilising peak demand, as shown in Figure 5.9.

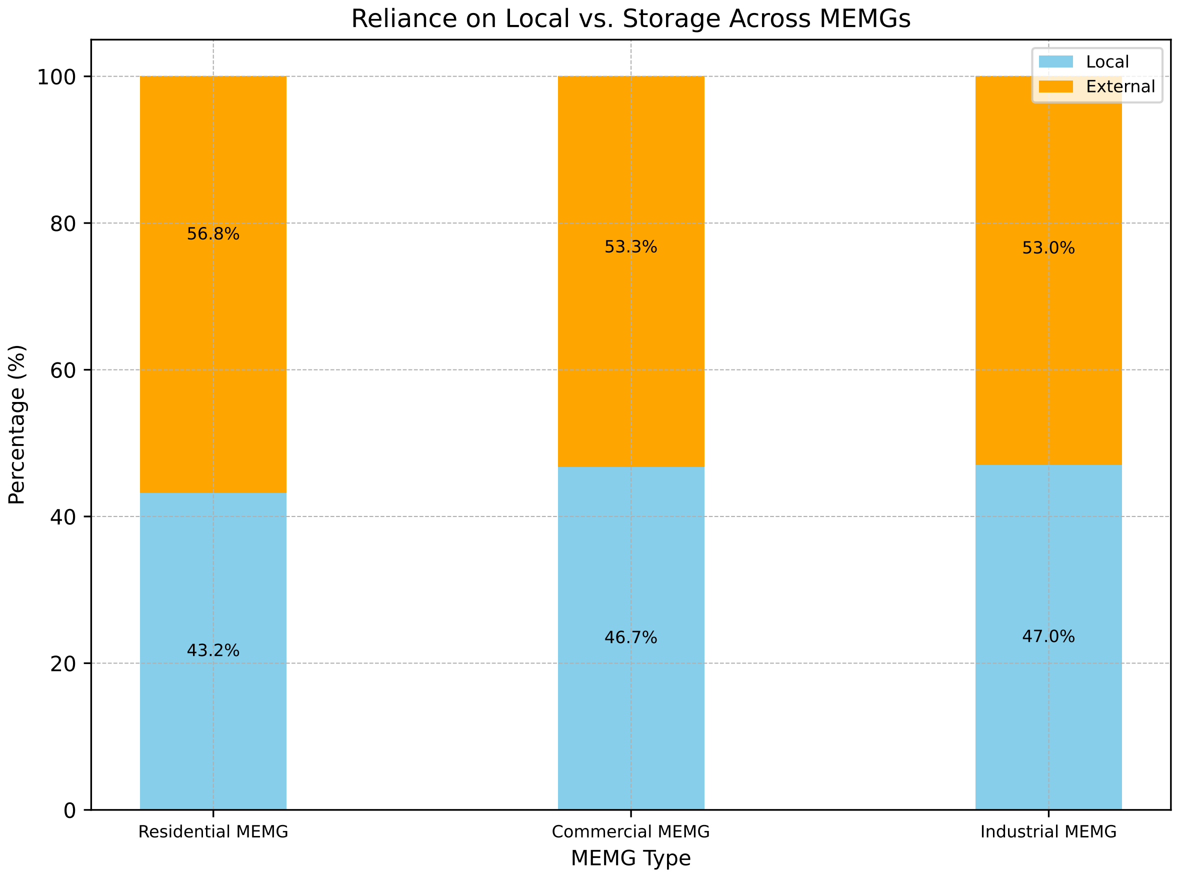

Figure 5.8: Depicts comparison of energy supply sources (local vs. external) across Residential, Commercial, and Industrial MEMGs. Energy shifts in electricity and heat across MEMG types

This analysis highlights how our proposed system can enhance energy efficiency and grid stability through effective peak demand management.

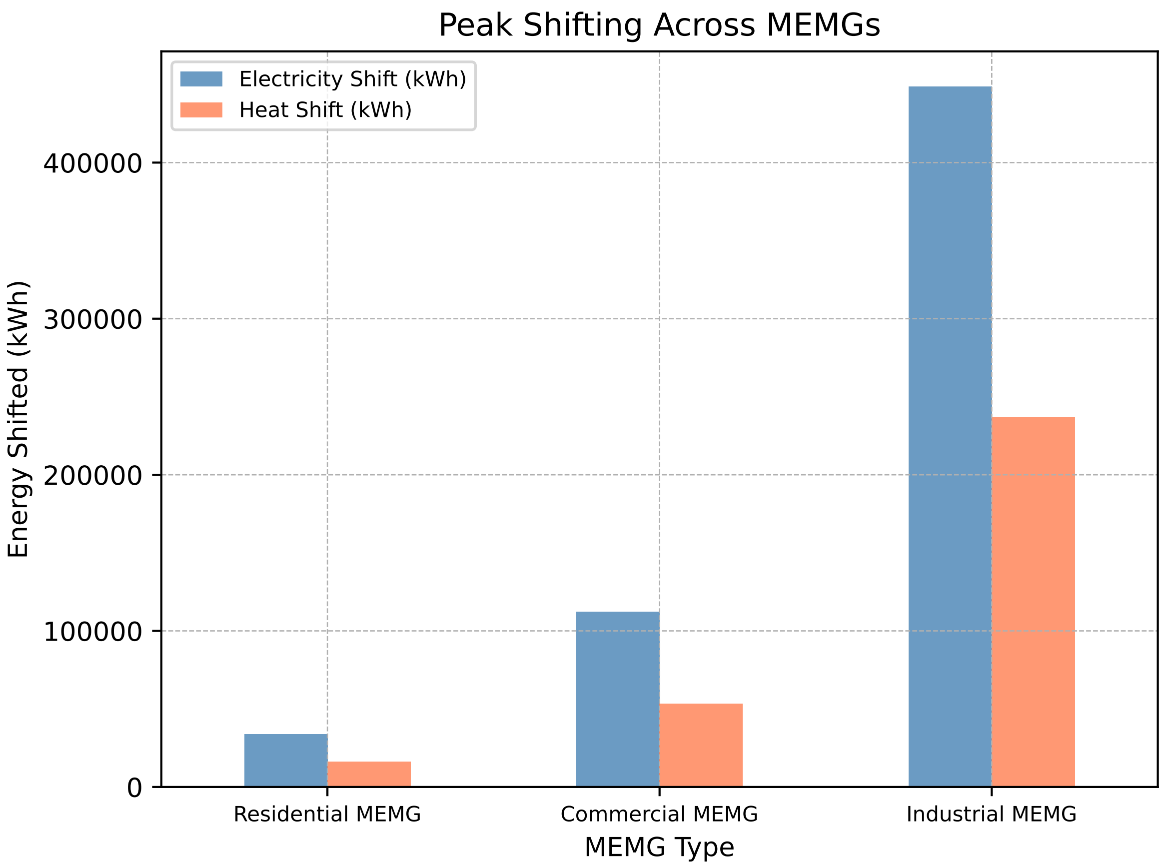

Figure 5.9: Depicts energy shifts in electricity and heat across MEMG types

5.4.5 Energy balance

This study examines the effectiveness of diverse reinforcement learning models in managing renewable energy utilisation, energy deficiencies, and energy storage systems within residential, commercial, and industrial MEMGs. Figure 5.10 compares the average renewable energy generation (kWh) across MEMG types.

The findings indicate that MADARC and MATD3 consistently outperform other models in terms of renewable energy utilisation, particularly within industrial MEMGs. For example, industrial MEMGs utilising MADARC generate approximately 20% more renewable energy compared to the Baseline model, showcasing their effectiveness in optimising

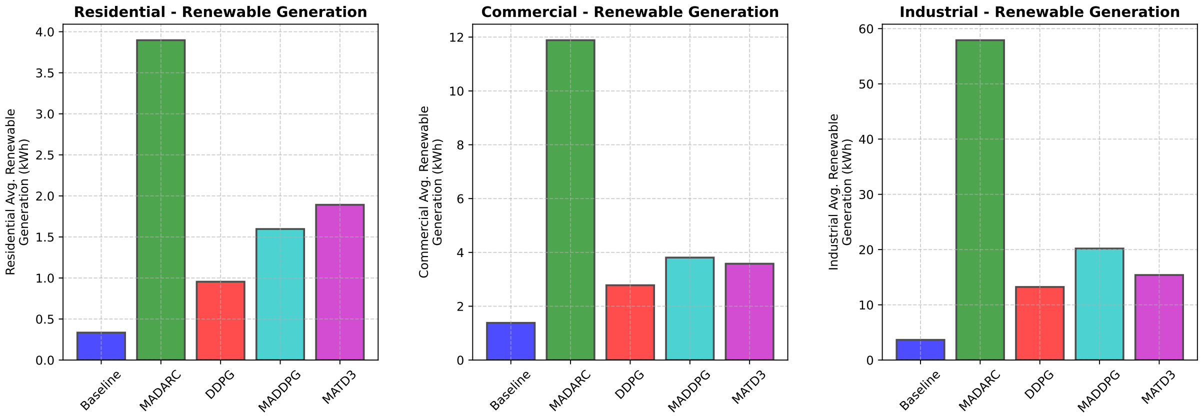

Figure 5.10: Shows a comparison of the average renewable energy production (kWh) among different MEMG types.

renewable integration. In contrast, residential MEMGs demonstrate moderate improvements, while commercial MEMGs exhibit balanced performance across various models. The Baseline model significantly underperforms, emphasising the advantages of reinforcement learning-driven strategies in maximising the use of renewable resources.

Figure 5.11 presents data on electricity deficit counts, which represent instances where demand surpasses supply. The DDPG and MADARC models exhibit the lowest counts across all types of MEMGs, achieving reductions of approximately 40–60% when compared to the Baseline. Notably, industrial MEMGs utilising MADARC attain the fewest deficits, highlighting the effectiveness of their demand response coordination.

In contrast, the Baseline model, with its static control strategy, frequently encounters deficits, particularly within commercial MEMGs. These findings underscore the vital role of the MADARC model in enhancing the balance between electricity supply and demand dynamics.

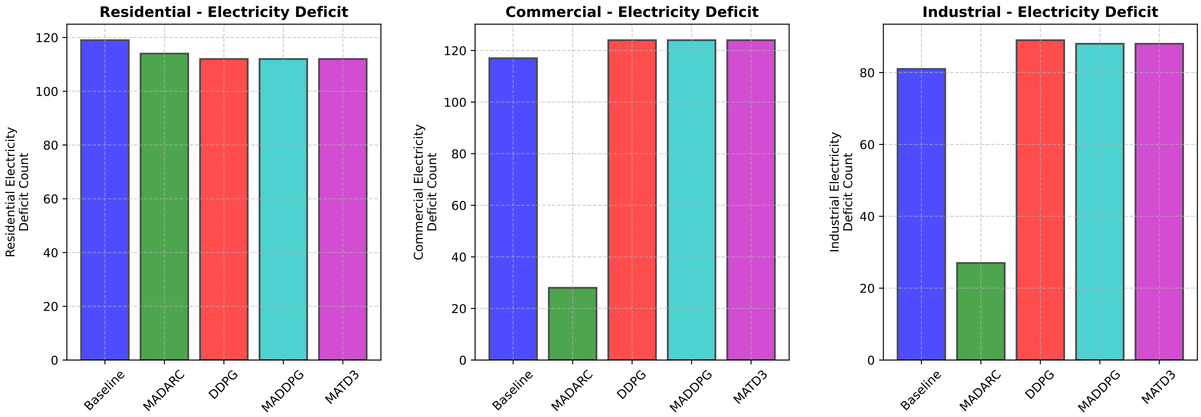

Figure 5.11: Illustrate the electricity deficit counts.

Furthermore, Figure 5.12 emphasises the heat deficit counts. MADARC outperforms other models in commercial and industrial MEMGs, reducing heat deficits by approximately 35–50% compared to the Baseline. In residential MEMGs, the differences among the models are less pronounced; however, MATD3 demonstrates slightly superior performance. These results underscore MADARC's effectiveness in managing thermal energy systems, likely attributable to its adaptive pricing mechanisms, which more effectively balance heat generation and consumption.

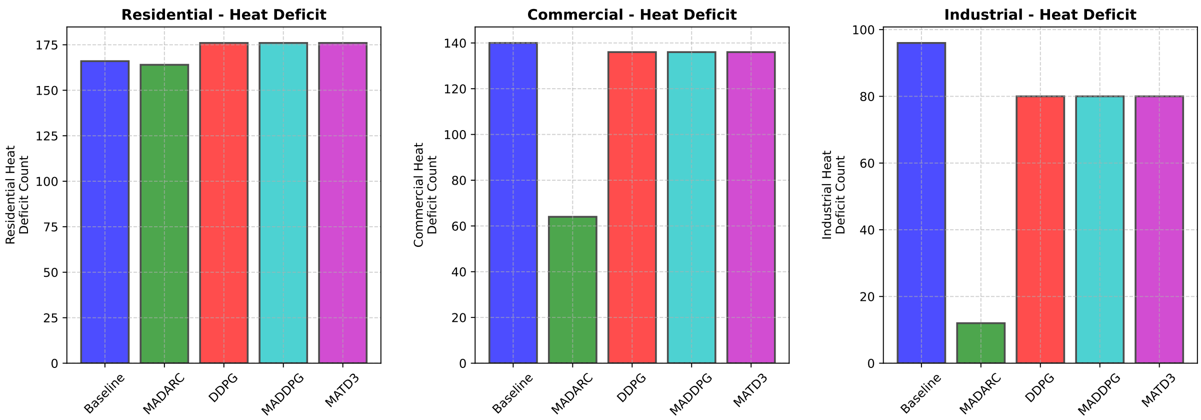

Figure 5.12: Depict the heat deficit counts.

Figure 5.13 analyses the average electricity storage levels (kWh). Both MADARC and MATD3 achieve optimal storage capacities, with industrial MEMGs under DARC storing approximately 25% more energy than the Baseline. This reflects effective energy buffering aimed at mitigating supply variability. In contrast, MADDPG and DDPG exhibit moderate levels of storage utilisation, while the Baseline significantly underutilises its storage infrastructure, resulting in inefficiencies. A higher storage capacity is associated with reduced deficits, which underscores the interdependence of storage and demand management within reinforcement learning frameworks.

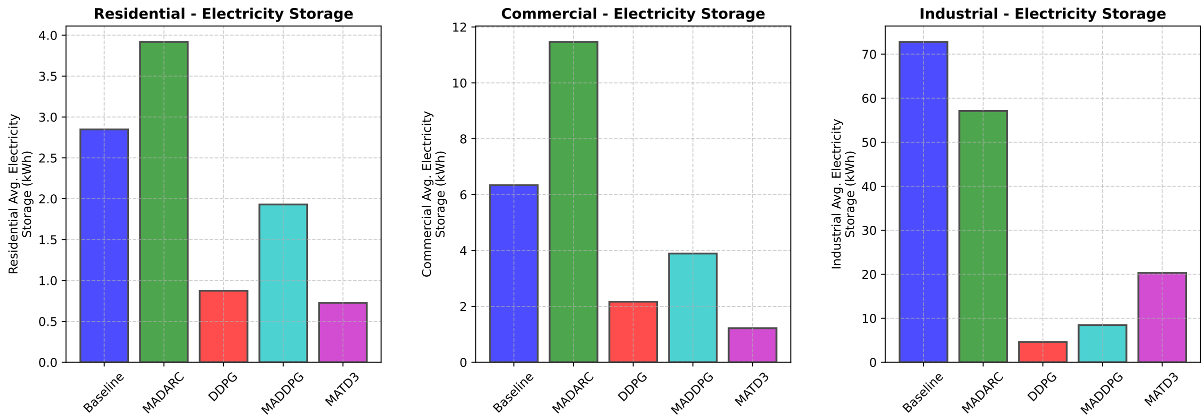

Figure 5.13: Depict the heat deficit counts.

In general, these results emphasise MADARC's reliable performance across different microgrid categories.

Residential Microgrids: MADARC consistently outperforms other methods in reducing heat and electricity deficits while achieving a higher average capacity for electricity storage. This demonstrates MADARC's effectiveness in balancing energy demand and supply, as well as its ability to utilise storage efficiently.

Commercial Microgrids: While all DRL methods significantly reduce heat deficits compared to the baseline, MADARC stands out with a slight advantage in minimising electricity deficits and maximising electricity storage. This underscores MADARC's potential to optimise energy distribution and storage in commercial settings.

Industrial Microgrids: MADARC's performance is closely matched by MADDPG and MATD3's efforts to reduce electricity deficits and maximise storage. However, MADARC clearly excels in minimising heat deficits, showcasing its adaptability to the specific energy demands of industrial environments.

Overall, these findings highlight MADARC's consistent effectiveness across various microgrid types. Its capability to minimise deficits and optimise storage indicates its potential to enhance the resilience and efficiency of microgrids operating under dynamic pricing and demand response conditions.

5.5 Chapter Summary

This chapter introduced a novel framework for P2P multi-energy trading by employing a modified DARC algorithm. The primary contribution was demonstrating that this advanced DRL approach effectively optimises the complex interactions in a decentralised market, outperforming other models. The results showed that the DARC-driven system significantly enhances profitability for both the service provider and P2P participants while minimising energy deficits and improving overall system resilience.

Chapter 6

Conclusion and Future Work

This thesis has embarked on a systematic investigation into the application of intelligent energy management strategies for modern, decentralised multi-energy systems. By leveraging advanced DRL algorithms, this research has developed and evaluated a series of frameworks for integrating dynamic pricing with IDR. The progressive journey from a dual-carrier system to a complex P2P multi-energy trading environment has provided a comprehensive exploration of the challenges and opportunities in this domain. This final chapter synthesises the key findings of the research in relation to the questions posed in Chapter 1, consolidates the limitations of the work, and outlines promising directions for future research.

6.1 Overall Conclusions

The research presented in this thesis successfully addressed its overarching aim and answered the specific research questions formulated in Section 1.3. The findings collectively demonstrate the significant potential of DRL-driven strategies to enhance the efficiency, stability, and economic viability of multi-energy systems.

6.1. Overall Conclusions

In response to Research Question 1 (RQ1), which queried the effectiveness of a DRL-based dynamic pricing IDR in an integrated electricity and gas system, Chapter 3 provided a definitive answer. The DDPG framework successfully learned to optimise retail pricing, leading to the dual benefit of increased DSO profitability and reduced end-user costs. Furthermore, the model demonstrated robust adaptability, effectively managing supply constraints by dynamically adjusting prices to prioritise critical loads and contain consumer dissatisfaction costs. This confirmed that even in a foundational dual-carrier system, an intelligent, model-free approach can create significant value for both providers and consumers

Addressing Research Question 2 (RQ2), which explored the extension of the DRL framework to multi-carrier systems incorporating EHs, Chapter 4 illustrated the profound impact of this integration. The introduction of EHs provided the necessary physical mechanism for true energy substitution, a cornerstone of advanced IDR. The DDPG-based agent learned to leverage this flexibility, resulting in significant shifts in energy source utilisation, specifically, a reduction in peak electricity demand by encouraging the use of natural gas for local heat and power generation. This study validated that DRL can effectively manage the increased complexity of multi-carrier interactions, leading to tangible improvements in overall system efficiency and significant peak load reductions

Finally, in answering Research Question 3 (RQ3), which investigated the performance of an advanced DRL algorithm in a P2P multi-energy trading environment, Chapter 5 highlighted the necessity of algorithmic adaptation for complex systems. The modified DARC algorithm was shown to be demonstrably superior to both traditional pricing and other DRL models like DDPG in this decentralised setting. The DARC-driven framework achieved higher profitability for both the SP and the P2P trading participants, effectively creating a larger and more equitably distributed economic benefit. Moreover, it proved more effective at improving the energy balance by minimising electricity and heat deficits and enhancing the utilisation of renewable energy resources. This confirmed that as system complexity grows, so too does the need for more sophisticated DRL solutions capable of robust exploration and stable value estimation.

In conclusion, this thesis makes a significant contribution by demonstrating a clear, progressive pathway for applying DRL to energy management. It establishes that DRL-driven dynamic pricing and IDR are not only feasible but highly effective, offering scalable and adaptive solutions that create mutual benefits for all stakeholders. The research validates the synergy between multiple energy carriers, especially when enabled by technologies like Energy Hubs, and provides a blueprint for designing intelligent, resilient, and economically efficient decentralised energy markets.

6.2 Limitations of the Research

A critical reflection on this work necessitates the acknowledgement of its limitations, which define the boundaries of its conclusions and provide context for future inquiry. These limitations can be categorised into model simplifications, algorithmic challenges, and the overall scope of the research.

- **Model Simplifications:** The frameworks developed in this thesis rely on several key abstractions. The modelling of consumer behaviour, while incorporating price elasticity and dissatisfaction costs, is a simplification of complex, often irrational, human decision-making. The assumption of perfect information and instantaneous communication networks does not capture the latencies, noise, and potential failures of real-world systems. Furthermore, while the economic interactions were modelled in detail, the underlying physical network constraints of electricity grids and gas pipelines were largely abstracted, which in reality would impose hard limits on operational decisions.

- **Algorithmic and Methodological Limitations:** Deep Reinforcement Learning, as a methodology, has inherent limitations. The algorithms used, DDPG and DARC, can be sample-inefficient, requiring a significant number of interactions to learn optimal policies, a challenge for real-world, non-simulated training. The ”black box” nature of deep neural networks also poses a challenge for interpretability, which can be a barrier to trust and adoption by system operators. While the research demonstrated scalability across progressively complex scenarios, applying these frameworks to city- or nation-wide systems with millions of agents would present significant computational and data management hurdles.
- **Scope of the Research:** The research was intentionally focused on operational timescales (e.g., day-ahead and real-time decision-making). Consequently, it did not address the equally important challenge of long-term investment and infrastructure planning based on these operational strategies. Moreover, while the thesis explored decentralised P2P trading, it did not delve deeply into the critical ancillary topics of cybersecurity vulnerabilities or the data privacy implications inherent in collecting and processing granular energy usage data.

6.3 Future Work

While this research provides a solid foundation, several avenues remain for further exploration. Future studies could concentrate on integrating emerging technologies, such as blockchain, to enable secure and transparent peer-to-peer energy trading. Additionally, examining the relationship between dynamic pricing in IDR and consumer behaviour models could yield valuable insights into user engagement strategies.

6.3. Future Work

Furthermore, expanding the framework to encompass large-scale, cross-regional energy systems could help validate its scalability and effectiveness across diverse market conditions. Real-world implementation studies are also crucial for assessing the practical challenges and benefits of deploying demand response learning DRL-driven IDR solutions. Finally, utilising advanced forecasting techniques that leverage real-time data could significantly improve the accuracy of demand-supply predictions, thereby enhancing the efficiency and reliability of multi-energy systems.

Appendices

This appendix provides supplementary materials to support the research presented in this thesis. It includes a list of publications arising from this work, the pseudocode for the core algorithms developed, and additional details regarding the simulation setups.

A Publications Arising from this Thesis

The following is a list of publications that have been accepted, submitted, or are in preparation based on the research conducted for this thesis.

Paper from Chapter 3:

- Almannouny, G., Bu, S., & Yang, J. (2025). Deep reinforcement learning for integrated demand response dynamic pricing of electricity and gas systems. [submitted to International Symposium on POWER ELECTRONICS Ee2025].

Paper from Chapter 4:

A. Publications Arising from this Thesis

- Almannouny, G., Bu, S., & Yang, J. (2022). Dynamic pricing integrated demand response for multiple energy carriers with deep reinforcement learning. In 2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe) (pp. 1–6). IEEE.

Paper from Chapter 5:

- Almannouny, G., Bu, S., & Yang, J. (2025). Dynamic pricing IDR in P2P multi-energy trading systems using a modified DARC algorithm. [Under preparation for submission to IEEE Transactions on Smart Grid].

B Supplementary Simulation Details

B.1 DDPG Hyperparameters (Chapter 3)

Table 6: DDPG Algorithm Parameters

Parameter	**Description**
Number of episodes	1500
Learning Rate (Actor)	0.001
Learning Rate (Critic)	0.005
Batch Size	64
Replay Buffer Size	10^6
Discount Factor (γ)	0.99

B.2 DDPG Hyperparameters (Chapter 4)

Table 7: DDPG Algorithm Parameters

Parameter	Description
Number of episodes	500
Learning Rate (Actor)	0.001
Learning Rate (Critic)	0.005
Batch Size	64
Replay Buffer Size	10^6
Discount Factor (γ)	0.99

B.3 Comparative RL Model Hyperparameters (Chapter 5)

Table 8: Hyper-parameters for Different RL Models

Hyperparameter	DARC	DDPG	MADDPG	MATD3
Max Steps per Episode	50	50	50	50
Episodes	200	200	200	200
Replay Buffer Size	10^3	10^3	10^3	10^3
Batch Size	128	32	64	64
Learning Rate (Actor)	0.001	0.001	0.001	0.001
Learning Rate (Critic)	0.002	0.002	0.002	0.002
Discount Factor γ	0.99	0.99	0.99	0.99
Soft Update Rate τ	0.01	0.005	0.005	0.005

Bibliography

[1] J. P. Barton and D. G. Infield, “Energy storage and its use with intermittent renewable energy,” *IEEE transactions on energy conversion*, vol. 19, no. 2, pp. 441–448, 2004.

[2] M. H. Albadi and E. F. El-Saadany, “A summary of demand response in electricity markets,” *Electric power systems research*, vol. 78, no. 11, pp. 1989–1996, 2008.

[3] Y. Chen, P. Xu, J. Gu, F. Schmidt, and W. Li, “Measures to improve energy demand flexibility in buildings for demand response (dr): A review,” *Energy and buildings*, vol. 177, pp. 125–139, 2018.

[4] N. G. Paterakis, O. Erdinç, and J. P. Catalão, “An overview of demand response: Key-elements and international experience,” *Renewable and Sustainable Energy Reviews*, vol. 69, pp. 871–891, 2017.

[5] L. Bai, F. Li, H. Cui, T. Jiang, H. Sun, and J. Zhu, “Interval optimization based operating strategy for gas-electricity integrated energy systems considering demand response and wind uncertainty,” *Applied energy*, vol. 167, pp. 270–279, 2016.

[6] Y. Lin and Z. Bie, “Study on the resilience of the integrated energy system,” *Energy Procedia*, vol. 103, pp. 171–176, 2016.

[7] J. Wang, H. Zhong, Z. Ma, Q. Xia, and C. Kang, “Review and prospect of integrated demand response in the multi-energy system,” *Applied Energy*, vol. 202, pp. 772–782, 2017.

[8] S. Bahrami and A. Sheikhi, “From demand response in smart grid toward integrated demand response in smart energy hub,” *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 650–658, 2015.

[9] W. Huang, N. Zhang, C. Kang, M. Li, and M. Huo, “From demand response to integrated demand response: Review and prospect of research and application,” *Protection and Control of Modern Power Systems*, vol. 4, no. 2, pp. 1–13, 2019.

[10] S. Bahrami, M. Toulabi, S. Ranjbar, M. Moeini-Aghtaie, and A. M. Ranjbar, “A decentralized energy management framework for energy hubs in dynamic pricing markets,” *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6780–6792, 2017.

[11] C. Zhang, J. Wu, Y. Zhou, M. Cheng, and C. Long, “Peer-to-peer energy trading in a microgrid,” *Applied energy*, vol. 220, pp. 1–12, 2018.

[12] K. Umer, Q. Huang, M. Khorasany, M. Afzal, and W. Amin, “A novel communication efficient peer-to-peer energy trading scheme for enhanced privacy in microgrids,” *Applied Energy*, vol. 296, p. 117075, 2021.

[13] C. Shao, Y. Ding, P. Siano, and Y. Song, “Optimal scheduling of the integrated electricity and natural gas systems considering the integrated demand response of energy hubs,” *IEEE Systems Journal*, vol. 15, no. 3, pp. 4545–4553, 2020.

[14] G. Dutta and K. Mitra, “A literature review on dynamic pricing of electricity,” *Journal of the Operational Research Society*, vol. 68, no. 10, pp. 1131–1145, 2017.

[15] D. T. Nguyen, H. T. Nguyen, and L. B. Le, “Dynamic pricing design for demand response integration in power distribution networks,” *IEEE Transactions on power systems*, vol. 31, no. 5, pp. 3457–3472, 2016.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, “Human-level control through deep reinforcement learning,” *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[17] A. Perera and P. Kamalaruban, “Applications of reinforcement learning in energy systems,” *Renewable and Sustainable Energy Reviews*, vol. 137, p. 110618, 2021.

[18] Q. Zhang and J. Li, “Demand response in electricity markets: A review,” *Electric Power Systems Research*, pp. 1–15, 2020.

[19] M. H. Albadi and E. F. El-Saadany, “A summary of demand response in electricity markets,” *Electric Power Systems Research*, vol. 78, no. 11, pp. 1989–1996, 2008.

[20] B. Dey, S. Misra, and F. P. Garcia Marquez, “Microgrid system energy management with demand response program for clean and economical operation,” *Applied Energy*, vol. 334, p. 120717, 2023.

[21] C. Yang, Z. Wu, X. Li, and A. Fars, “Risk-constrained stochastic scheduling for energy hub: Integrating renewables, demand response, and electric vehicles,” *Energy*, vol. 288, p. 129680, 2024.

[22] J. Xie, A. Ajagekar, and F. You, “Multi-agent attention-based deep reinforcement learning for demand response in grid-responsive buildings,” *Applied Energy*, vol. 342, p. 121162, 2023.

[23] J. Liu, H. Hu, S. S. Yu, and H. Trinh, “Virtual power plant with renewable energy sources and energy storage systems for sustainable power grid-formation, control techniques and demand response,” *Energies*, vol. 16, no. 3705, pp. 1–28, 2023.

[24] F. Alfaverh, M. Denai, and Y. Sun, “A dynamic peer-to-peer electricity market model for a community microgrid with price-based demand response,” *IEEE Transactions on Smart Grid*, vol. 14, no. 5, pp. 3976–3989, 2023.

[25] S. Bahrami, M. H. Amini, M. Shafie-khah, and J. P. S. Catalao, “A decentralized electricity market scheme enabling demand response deployment,” *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4218–4229, 2018.

[26] I. Kamwa, L. Bagherzadeh, and A. Delavari, “Integrated demand response programs in energy hubs: a review of applications, classifications, models and future directions,” *Energies*, vol. 16, no. 11, p. 4443, 2023.

[27] L. Chen, Y. Yang, and Q. Xu, “Retail dynamic pricing strategy design considering the fluctuations in day-ahead market using integrated demand response,” *International Journal of Electrical Power & Energy Systems*, vol. 130, p. 106983, 2021.

[28] S. A. Mansouri, S. Maroufi, and A. Ahmarinejad, “A tri-layer stochastic framework to manage electricity market within a smart community in the presence of energy storage systems,” *Journal of Energy Storage*, vol. 71, p. 108130, 2023.

[29] e. a. Mansouri, “Multi-stage joint planning for energy hubs with idr,” *International Journal of Electrical Power & Energy Systems*, vol. 140, p. 108103, 2022.

[30] e. a. Zheng, “Incentive-based idr and consumer behaviour coupling,” *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3231–3245, 2020.

[31] Bahrami and Sheikhi, “Demand switching between energy carriers in idr,” *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 650–658, 2015.

[32] e. a. Shao, “Integrated electricity and gas systems with idr,” *IEEE Systems Journal*, vol. 15, no. 3, pp. 4545–4553, 2020.

[33] B. Gao, C. Chen, Y. Qin, X. Liu, and Z. Zhu, “Evolutionary game-theoretic analysis for residential users considering integrated demand response,” *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 6, pp. 1500–1509, 2020.

[34] Z. Yang, M. Ni, and H. Liu, “Pricing strategy of multi-energy provider considering integrated demand response,” *Ieee Access*, vol. 8, pp. 149 041–149 051, 2020.

[35] Z. Li and Z. Zhang, “Day-ahead and intra-day optimal scheduling of integrated energy system considering uncertainty of source & load power forecasting,” *Energies*, vol. 14, no. 9, p. 2539, 2021.

[36] A. Esmat, M. de Vos, Y. Ghiassi-Farrokhfal, P. Palensky, and D. Epema, “A novel decentralized platform for peer-to-peer energy trading market with blockchain technology,” *Applied Energy*, vol. 282, p. 116123, 2021.

[37] T. Alskaif, J. L. Crespo-Vazquez, M. Sekuloski, G. van Leeuwen, and J. P. S. Catalão, “Blockchain-based fully peer-to-peer energy trading strategies for residential energy systems,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 231–242, 2022.

[38] E. A. Soto, L. B. Bosman, E. Wollega, and W. D. Leon-Salas, “Peer-to-peer energy trading: A review of the literature,” *Applied Energy*, vol. 283, p. 116268, 2021.

[39] T. Morstyn and M. D. McCulloch, “Multiclass energy management for peer-to-peer energy trading driven by prosumer preferences,” *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 4005–4016, 2019.

[40] A. Paudel, L. P. M. I. Sampath, J. Yang, and H. B. Gooi, “Peer-to-peer energy trading in smart grid considering power losses and network fees,” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4727–4737, 2020.

[41] Y. Wu, Y. Wu, H. Cimen, J. C. Vasquez, and J. M. Guerrero, “Towards collective energy community: Potential roles of microgrid and blockchain to go beyond p2p energy trading,” *Applied Energy*, vol. 314, p. 119003, 2022.

[42] J. Guerrero, A. C. Chapman, and G. Verbic, “Decentralized p2p energy trading under network constraints in a low-voltage network,” *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5163–5173, 2019.

[43] B. Han, M. Hu, S. Lv, Z. Bao, L. Lu, and M. Yu, “An integrated demand response-based energy management strategy for integrated energy system based on deep reinforcement learning,” *2023 6th International Conference on Renewable Energy and Power Engineering (REPE)*, pp. 407–412, 2023.

[44] R. Shen, S. Zhong, X. Wen, Q. An, R. Zheng, Y. Li, and J. Zhao, “Multi-agent deep reinforcement learning optimization framework for building energy systems with renewable energy,” *Applied Energy*, vol. 312, p. 118724, 2022.

[45] J. Lee and H. Choi, “Federated deep reinforcement learning for smart electric vehicle charging,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5551–5560, 2021.

[46] T. Chen, S. Bu, X. Liu, J. Kang, F. R. Yu, and Z. Han, “Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning,” *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 715–726, 2022.

[47] Y. Xu, L. Yu, G. Bi, M. Zhang, and C. Shen, “Deep reinforcement learning and blockchain for peer-to-peer energy trading among microgrids,” in *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*. IEEE, 2020, pp. 360–365.

[48] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, “Reinforcement learning and its applications in modern power and energy systems: A review,” *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1029–1042, 2020.

[49] M. M. Forootan, I. Larki, R. Zahedi, and A. Ahmadi, “Machine learning and deep learning in energy systems: A review,” *Sustainability*, vol. 14, p. 4832, 2022.

[50] J. Duan, D. Shi, R. Diao, H. Li, Z. Wang, B. Zhang, D. Bian, and Z. Yi, "Deep-reinforcement-learning-based autonomous voltage control for power grid operations," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 814–817, 2019.

[51] E. Mocanu, "Machine learning applied to smart grids," *Energy*, vol. 2, no. 3, p. 4, 2017.

[52] M. F. Zia, E. Elbouchikhi, and M. Benbouzid, "Microgrids energy management systems: A critical review on methods, solutions, and prospects," *Applied energy*, vol. 222, pp. 1033–1055, 2018.

[53] E. Dall'Anese, P. Mancarella, and A. Monti, "Unlocking flexibility: Integrated optimization and control of multienergy systems," *IEEE Power and Energy Magazine*, vol. 15, no. 1, pp. 43–52, 2017.

[54] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: Pricing methods and optimization algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152–178, 2014.

[55] Q. Qdr, "Benefits of demand response in electricity markets and recommendations for achieving them," *US Dept. Energy, Washington, DC, USA, Tech. Rep*, 2006.

[56] B. Shen, G. Ghatikar, Z. Lei, J. Li, G. Wikler, and P. Martin, "The role of regulatory reforms, market changes, and technology development to make demand response a viable resource in meeting energy challenges," *Applied Energy*, vol. 130, pp. 814–823, 2014.

[57] P. Siano, "Demand response and smart grids—a survey," *Renewable and sustainable energy reviews*, vol. 30, pp. 461–478, 2014.

[58] M. Qadrdan, M. Cheng, J. Wu, and N. Jenkins, "Benefits of demand-side response in combined gas and electricity networks," *Applied energy*, vol. 192, pp. 360–369, 2017.

[59] P. Mancarella, "Mes (multi-energy systems): An overview of concepts and evaluation models," *Energy*, vol. 65, pp. 1–17, 2014.

[60] A. Arteconi, D. Patteeuw, K. Bruninx, E. Delarue, W. D'haeseleer, and L. Helsen, "Active demand response with electric heating systems: Impact of market penetration," *Applied energy*, vol. 177, pp. 636–648, 2016.

[61] N. Liu, J. Wang, and L. Wang, “Distributed energy management for interconnected operation of combined heat and power-based microgrids with demand response,” *Journal of Modern Power Systems and Clean Energy*, vol. 5, no. 3, pp. 478–488, 2017.

[62] M. Shahidehpour, Y. Fu, and T. Wiedman, “Impact of natural gas infrastructure on electric power systems,” *Proceedings of the IEEE*, vol. 93, no. 5, pp. 1042–1056, 2005.

[63] Y. He, M. Shahidehpour, Z. Li, C. Guo, and B. Zhu, “Robust constrained operation of integrated electricity-natural gas system considering distributed natural gas storage,” *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1061–1071, 2017.

[64] C. Wang, W. Wei, J. Wang, L. Wu, and Y. Liang, “Equilibrium of interdependent gas and electricity markets with marginal price based bilateral energy trading,” *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4854–4867, 2018.

[65] S. Nolan and M. O’Malley, “Challenges and barriers to demand response deployment and evaluation,” *Applied Energy*, vol. 152, pp. 1–10, 2015.

[66] A. Alabdulwahab, A. Abusorrah, X. Zhang, and M. Shahidehpour, “Coordination of interdependent natural gas and electricity infrastructures for firming the variability of wind energy in stochastic day-ahead scheduling,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 606–615, 2015.

[67] N. Zhao, B. Wang, and M. Wang, “A model for multi-energy demand response with its application in optimal tou price,” *Energies*, vol. 12, no. 6, p. 994, 2019.

[68] J. R. Vázquez-Canteli and Z. Nagy, “Reinforcement learning for demand response: A review of algorithms and modeling techniques,” *Applied energy*, vol. 235, pp. 1072–1089, 2019.

[69] Q. Wei, D. Liu, and G. Shi, “A novel dual iterative q-learning method for optimal battery management in smart residential environments,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2509–2518, 2014.

[70] H. Wang, T. Huang, X. Liao, H. Abu-Rub, and G. Chen, “Reinforcement learning in energy trading game among smart microgrids,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 8, pp. 5109–5119, 2016.

[71] S. Behboodi, D. P. Chassin, C. Crawford, and N. Djilali, "Renewable resources portfolio optimization in the presence of demand response," *Applied Energy*, vol. 162, pp. 139–148, 2016.

[72] M. Joung and J. Kim, "Assessing demand response and smart metering impacts on long-term electricity market prices and system reliability," *Applied Energy*, vol. 101, pp. 441–448, 2013.

[73] A. Ghasemkhani and L. Yang, "Reinforcement learning based pricing for demand response," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–6.

[74] T. Remani, E. Jasmin, and T. I. Ahamed, "Residential load scheduling with renewable generation in the smart grid: A reinforcement learning approach," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3283–3294, 2018.

[75] R. Lu, S. H. Hong, and X. Zhang, "A dynamic pricing demand response algorithm for smart grid: reinforcement learning approach," *Applied Energy*, vol. 220, pp. 220–230, 2018.

[76] R. S. Sutton, A. G. Barto *et al.*, "Reinforcement learning: An introduction 2nd ed," *MIT press Cambridge*, vol. 1, no. 2, p. 25, 2018.

[77] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[78] ISO New England, "ISO New England ISO Express: Hourly Day-Ahead LMP Reports," https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/lmps-da-hourly, accessed: 2020-02-21.

[79] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.

[80] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[81] M. Mazidi, H. Monsef, and P. Siano, "Robust day-ahead scheduling of smart distribution networks considering demand response programs," *Applied Energy*, vol. 178, pp. 929–942, 2016.

[82] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *arXiv preprint arXiv:1811.12560*, 2018.

[83] H. Van Goor and B. Scholtens, "Modeling natural gas price volatility: The case of the uk gas market," *Energy*, vol. 72, pp. 126–134, 2014.

[84] Ofgem, "Energy data publications, Data portal," 2021, https://www.ofgem.gov.uk/energy-data-and-research/data-portal, Last accessed on 2021-01-14.

[85] Y. Tao, J. Qiu, S. Lai, and J. Zhao, "Integrated electricity and hydrogen energy sharing in coupled energy systems," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1149–1162, 2020.

[86] A. Hussain, V.-H. Bui, and H.-M. Kim, "Microgrids as a resilience resource and strategies used by microgrids for enhancing resilience," *Applied energy*, vol. 240, pp. 56–72, 2019.

[87] T. Sousa, T. Soares, P. Pinson, F. Moret, T. Baroche, and E. Sorin, "Peer-to-peer and community-based markets: A comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 104, pp. 367–378, 2019.

[88] E. Mengelkamp, D. Schlund, and C. Weinhardt, "Development and real-world application of a taxonomy for business models in local energy markets," *Applied Energy*, vol. 256, p. 113913, 2019.

[89] C. Schick, N. Klempp, and K. Hufendiek, "Role and impact of prosumers in a sector-integrated energy system with high renewable shares," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 3286–3298, 2020.

[90] P. Siano, G. De Marco, A. Rolán, and V. Loia, "A survey and evaluation of the potentials of distributed ledger technology for peer-to-peer transactive energy exchanges in local energy markets," *IEEE Systems Journal*, vol. 13, no. 3, pp. 3454–3466, 2019.

[91] L. Ableitner, V. Tiefenbeck, A. Meeuw, A. Wörner, E. Fleisch, and F. Wortmann, "User behavior in a real-world peer-to-peer electricity market," *Applied Energy*, vol. 270, p. 115061, 2020.

[92] Q. Wu, Z. Xie, H. Ren, Q. Li, and Y. Yang, "Optimal trading strategies for multi-energy microgrid cluster considering demand response under different trading modes: A comparison study," *Energy*, vol. 254, p. 124448, 2022.

[93] G. Almannouny, S. Bu, and J. Yang, "Dynamic pricing integrated demand response for multiple energy carriers with deep reinforcement learning," in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, 2022, pp. 1–6.

[94] W. Liu, D. Qi, and F. Wen, "Intraday residential demand response scheme based on peer-to-peer energy trading," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1823–1835, 2019.

[95] D. Qiu, T. Chen, G. Strbac, and S. Bu, "Coordination for multienergy microgrids using multiagent reinforcement learning," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 5689–5700, 2022.

[96] S. Yin, Q. Ai, J. Li, Z. Li, and S. Fan, "Energy pricing and sharing strategy based on hybrid stochastic robust game approach for a virtual energy station with energy cells," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 2, pp. 772–784, 2020.

[97] W. Tushar, T. K. Saha, C. Yuen, T. Morstyn, H. V. Poor, R. Bean *et al.*, "Grid influenced peer-to-peer energy trading," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1407–1418, 2019.

[98] A. Paudel, K. Chaudhari, C. Long, and H. B. Gooi, "Peer-to-peer energy trading in a prosumer-based community microgrid: A game-theoretic model," *IEEE Transactions on Industrial electronics*, vol. 66, no. 8, pp. 6087–6097, 2018.

[99] Z. Ma, Y. Zheng, C. Mu, T. Ding, and H. Zang, "Optimal trading strategy for integrated energy company based on integrated demand response considering load classifications," *International Journal of Electrical Power & Energy Systems*, vol. 128, p. 106673, 2021.

[100] L. Wang, Z. Tao, L. Zhu, X. Wang, C. Yin, H. Cong, R. Bi, and X. Qi, "Optimal dispatch of integrated energy system considering integrated demand response resource trading," *IET Generation, Transmission & Distribution*, vol. 16, no. 9, pp. 1727–1742, 2022.

[101] Y. Li, C. Wang, G. Li, and C. Chen, “Optimal scheduling of integrated demand response-enabled integrated energy systems with uncertain renewable generations: A stackelberg game approach,” *Energy Conversion and Management*, vol. 235, p. 113996, 2021.

[102] M. Aghamohamadi, A. Mahmoudi, J. K. Ward, M. H. Haque, and J. P. Catalão, “A block-coordinate-descent robust approach to incentive-based integrated demand response in managing multienergy hubs with must-run processes,” *IEEE Transactions on Industry Applications*, vol. 58, no. 2, pp. 2352–2368, 2022.

[103] L. Zhang, J. Kuang, B. Sun, F. Li, and C. Zhang, “A two-stage operation optimization method of integrated energy systems with demand response and energy storage,” *Energy*, vol. 208, p. 118423, 2020.

[104] T. M. Alabi, L. Lu, and Z. Yang, “A novel multi-objective stochastic risk co-optimization model of a zero-carbon multi-energy system (zcmes) incorporating energy storage aging model and integrated demand response,” *Energy*, vol. 226, p. 120258, 2021.

[105] H. Tan, W. Yan, Z. Ren, Q. Wang, and M. A. Mohamed, “A robust dispatch model for integrated electricity and heat networks considering price-based integrated demand response,” *Energy*, vol. 239, p. 121875, 2022.

[106] J. Chen, B. Qi, Z. Rong, K. Peng, Y. Zhao, and X. Zhang, “Multi-energy coordinated microgrid scheduling with integrated demand response for flexibility improvement,” *Energy*, vol. 217, p. 119387, 2021.

[107] Z. Zhang, Y. Su, M. Tan, and R. Cao, “Fusing domain knowledge and reinforcement learning for home integrated demand response online optimization,” *Engineering Applications of Artificial Intelligence*, vol. 121, p. 105995, 2023.

[108] Q. Lu, S. Lü, and Y. Leng, “A nash-stackelberg game approach in regional energy market considering users’ integrated demand response,” *Energy*, vol. 175, pp. 456–470, 2019.

[109] D. Wang, H. Jia, K. Hou, W. Du, N. Chen, X. Wang, M. Fan *et al.*, “Integrated demand response in district electricity-heating network considering double auction retail energy market based on demand-side energy stations,” *Applied Energy*, vol. 248, pp. 656–678, 2019.

[110] P. Liu, T. Ding, Z. Zou, and Y. Yang, "Integrated demand response for a load serving entity in multi-energy market considering network constraints," *Applied Energy*, vol. 250, pp. 512–529, 2019.

[111] L. Chen, H. Tang, J. Wu, C. Li, and Y. Wang, "A robust optimization framework for energy management of cchp users with integrated demand response in electricity market," *International Journal of Electrical Power & Energy Systems*, vol. 141, p. 108181, 2022.

[112] H. Aghamohammadloo, V. Talaeizadeh, K. Shahanaghi, J. Aghaei, H. Shayanfar, M. Shafie-khah, and J. P. Catalão, "Integrated demand response programs and energy hubs retail energy market modelling," *Energy*, vol. 234, p. 121239, 2021.

[113] K. Li, N. Ye, S. Li, H. Wang, and C. Zhang, "Distributed collaborative operation strategies in multi-agent integrated energy system considering integrated demand response based on game theory," *Energy*, p. 127137, 2023.

[114] W. Fan, Z. Tan, F. Li, A. Zhang, L. Ju, Y. Wang, and G. De, "A two-stage optimal scheduling model of integrated energy system based on cvar theory implementing integrated demand response," *Energy*, vol. 263, p. 125783, 2023.

[115] Z. Wang, X. Zhang, S. Zhu, and B. Yang, "An incentive pricing approach for integrated demand response in multi-energy system based on consumer classification," in *2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. IEEE, 2019, pp. 1–6.

[116] C. Hu and D. An, "Dynamic pricing strategy in electricity trading market based on reinforcement learning," in *2023 35th Chinese Control and Decision Conference (CCDC)*. IEEE, 2023, pp. 5485–5489.

[117] D. Zhang, H. Zhu, H. Zhang, H. H. Goh, H. Liu, and T. Wu, "Multi-objective optimization for smart integrated energy system considering demand responses and dynamic prices," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 1100–1112, 2021.

[118] B. Das, A. M. M. Reza, N. Mohammad, and S. H. Razo, "Smart grid dynamic pricing algorithm for benefit maximization of service providers and consumers," in *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2020, pp. 206–209.

[119] T. Chen, S. Bu, X. Liu, J. Kang, F. R. Yu, and Z. Han, "Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning," *IEEE transactions on smart grid*, vol. 13, no. 1, pp. 715–727, 2021.

[120] J. Yang, W. Xu, K. Ma, and C. Li, "A three-stage multi-energy trading strategy based on p2p trading mode," *IEEE Transactions on Sustainable Energy*, vol. 14, no. 1, pp. 233–241, 2022.

[121] U. E. I. Administration. (2023) Open data. Accessed: 2023-08-13. [Online]. Available: https://www.eia.gov/opendata/

[122] Ofgem. (2023) Wholesale market indicators data portal. Accessed: 2023-08-13. [Online]. Available: https://www.ofgem.gov.uk/energy-data-and-research/data-portal/wholesale-market-indicators

[123] S. Ong and N. Clark, "Commercial and residential hourly load profiles for all tmy3 locations in the united states," https://data.openei.org/submissions/153, 2014.