Artificial Intelligence-Based Control Schemes for Robust and Sustainable Wind Energy Conversion system

by
Yazdan H. Tabrizi

A thesis submitted in partial fulfilment of the requirement for the degree of
Doctor of Philosophy at
Department of Electrical Engineering
Lakehead University
Thunder Bay, Ontario, Canada
September 2024

Abstract

To reduce fossil fuel consumption, which causes carbon dioxide emissions and global warming, renewable energy is gaining popularity. Among various renewable energy sources, wind energy is one of the most cost-effective ways to generate electricity. Numerous studies have been conducted to improve the performance of wind energy conversion systems (WECS) in various aspects. However, traditional control strategies employed in WECS often lead to lower efficiency, complicated implementation, complex system modeling, sophisticated drive circuit design, and suboptimal responses. This PhD thesis presents a comprehensive exploration of cutting-edge techniques for optimizing wind energy conversion systems, unified by the application of a proposed multi-agent reinforcement learning (MARL) method. The research is structured around three primary objectives, each contributing to the advancement of renewable energy technologies through the innovative use of MARL. Firstly, the thesis delves into the control of a neutral point clamped (NPC) power converter employed in a direct-drive permanent magnet synchronous generator (PMSG)-based WECS. The focus is on enhancing power quality and meeting grid code requirements for total harmonic distortion (THD). Traditional controllers like PI often struggle with parameter tuning and adaptability to varying operating conditions, resulting in suboptimal performance under dynamic and unbalanced scenarios. AI-based approaches, while more adaptive, typically require extensive offline training and detailed system modeling, making them less practical for real-time applications. The proposed approach eliminates the need for offline training and extensive system modeling, distinguishing itself from traditional machine learning (ML), neural network-based techniques, and PI-based methods. Through simulations and comparative analysis, the effectiveness of the MARL strategy is validated, particularly in handling unbalanced voltage sag scenarios. The integration of meta-learning to optimize the discount factor (DF), a vital hyperparameter in RL-based approaches, further enhances the adaptability and convergence rate of the control system, ensuring power quality. Afterwards, the research addresses the challenges in maximum power point tracking (MPPT) for the wind energy conversion systems. Traditional methods like Perturb and Observe (P&O) and incremental conductance are known for their slow dynamic response and susceptibility to steady-state oscillations around the maximum power point, especially under rapidly changing wind conditions. The proposed customized MARL approach overcomes these limitations by employing multiple agents that work collaboratively, resulting in improved energy output and responsiveness to wind speed variations. The use of a meta-learned

discount factor optimizes the MARL algorithm, reducing learning duration and enhancing convergence. Extensive simulations and also a 1000W prototype implementation demonstrate the MARL strategy's superiority over traditional MPPT methods, confirming its practical benefits and reliability in real-world applications. Finally, the thesis explores power prediction and management, as well as energy scheduling, in a microgrid (MG) environment. The MG integrates renewable energy sources such as wind turbines (WT), photovoltaic (PV) systems, and battery energy storage systems (BESS), along with combined cooling, heating, and power (CCHP) units. Traditional forecasting methods, such as ARIMA models and simple neural networks, often fail to capture the complex temporal dependencies and variability in renewable energy sources, leading to inaccurate predictions. A multi-layer recurrent neural network (MLRNN) is developed for accurate 24-hour forecasting of renewable energy generation. A grid-search method is proposed to optimally tune the number of RNN layers and the optimizer learning rate. This model leverages historical wind and solar data to capture complex temporal dependencies and patterns. The predicted values are then used as the maximum WT and PV output capacity to optimize power management within the MG using the proposed MARL method. This approach minimizes fuel and CO_2 emissions costs, enhances coordination among MG components, and ensures efficient power distribution, resource utilization, and BESS scheduling. Traditional centralized control methods can be computationally intensive and less responsive to real-time changes, whereas the decentralized control provided by MARL will reduce computational burden and improves response quality, demonstrating its effectiveness in maintaining optimal MG performance. Comparative analysis validates the effectiveness of the approach.

Overall, this thesis provides a robust and innovative framework for enhancing wind energy conversion, renewable energy forecasting, and microgrid power management through the application of a unified MARL-based approach. For the coding and simulation Python programming and Simulink MATLAB are used, respectively. The findings underscore the potential of MARL application to significantly improve the efficiency, reliability, and environmental sustainability of renewable energy systems and WECS

Table of Contents

List of Figures

List of Tables

List of Acronyms

Wind energy conversion system	WECS
carbon dioxide	CO_2
greenhouse gas	GHG
International Energy Agency	IEA
Global wind energy council	GWEC
international renewable energy agency	IRENA
Latin America	LATAM
internal revenue service	IRS
original equipment manufacturers	OEM
Middle East and North Africa	MENA
asynchronous squirrel cage induction generator	SCIG
Doubly Fed Induction Generator	DFIG
wound rotor induction generator	WRIG
Permanent magnet synchronous generator	PMSG
back-to-back	BTB
voltage source inverter	VSI
flying capacitor	FC
neutral point clamped converter	NPC
modular multilevel converters	MMC
total harmonic distortion	THD
half-bridge	HB

full-bridge	FB
cascaded H-bridge	CHB
pulse width modulation	PWM
selective harmonic elimination	SHE
field-oriented control	FOC
voltage-oriented control	VOC
model predictive control	MPC
Direct torque control	DTC
direct power control	DPC
sliding mode control	SMC
machine side converters	MSC
grid side converters	GSC
genetic algorithms	GA
fuzzy logic controllers	FL
reinforcement learning	RL
artificial neural networks	ANN
machine learning	ML
maximum power point tracking	MPPT
micro-grid	MG
battery energy storage system	BESS
Photovoltaic	PV
Wind turbine	WT
Combined Cooling, Heating, and Power	CCHP

mean squared error	MSE
stochastic gradient descent	SGD
Mini-Batch Gradient Descent	MSG
Broyden–Fletcher–Goldfarb–Shanno	BFGS
Limited-memory BFGS	L-BFGS
Deep Q-Networks	DQN
Markov decision process	MDP
multi-agent reinforcement learning	MARL
Discount factor	DF
in-phase disposition	IPD
Proportional-Integral	PI
low voltage ride through	LVRT
indirect power control	IPC
tip speed ratio	TSR
optimal torque	OT
power signal feedback	PSF
incremental conductance	INC
optimum relation based	ORB
Perturb and observe	P&O
Digital signal processor	DSP
multi-layer recurrent neural network	MLRNN
distributed generation	DG
energy management strategy	EMS

Stochastic optimization	SO
system advisor model	SAM
Normalized root means square error	nRMSE
cost function	CF
electric chiller	EC
absorption chiller	AC

Chapter 1
Introduction

1.1 Overview on Global Wind Power Capacity

Traditionally, electric power generations depend on the nonrenewable resources such as fossil fuels, which lead to carbon dioxide (CO_2) emissions resulting in air pollution and global warming [1]. Moreover, some technical issues like power supply and transmission efficiency can bring us the milestone of looking for a suitable replacement. Consequently, researchers' focus in this area has shifted from reducing the detrimental effects of climate change to embracing clean energies as a safe and effective approach to offset the progression of these problems [2]. Currently, renewable energy technologies, also known as clean energy, such as wind, solar, and hydro-power, are commonly used in electrical power generations, resulting in a decline in fossil fuel usage and carbon emissions. The CO_2 emission is a crucial issue that prompted the United Nations to reiterate its call for immediate action to achieve net zero greenhouse gas (GHG) emissions by 2050, a call which has since been echoed by more than 120 countries, as well as thousands of businesses, investors, cities, regions, and universities. The Europe, Japan, South Korea, Canada, and South Africa all pledged to achieve net zero carbon emissions by 2050 in 2020, marking a turning point in climate change policy. Countries that have adopted or are considering net zero objectives currently account for two-thirds of the world economy and 63 % of global GHG emissions, when combined with China's net zero by 2060 aim and the United States' plan to attain net zero emission by 2050 under the Biden administration.

The era of fossil fuels is ended, and the global energy transformation is here to stay, more than ever. Wind energy is the dominant renewable resource due to its abundant availability, cost-effectiveness, and technological advancements. Wind is a naturally occurring and inexhaustible source of energy, harnessed efficiently through modern wind turbines. According to the prediction annual wind installations must increase dramatically to reach net zero emission by 2050. New global wind installations (in GW) to reach the target are illustrated in Fig. 1.1 from 2020 to the year 2030. International Energy Agency (IEA) Outlook has estimated the volume of required wind

capacity installation in 2022-2024 and 2026-2029 periods, so far, and it is demonstrated in this figure.

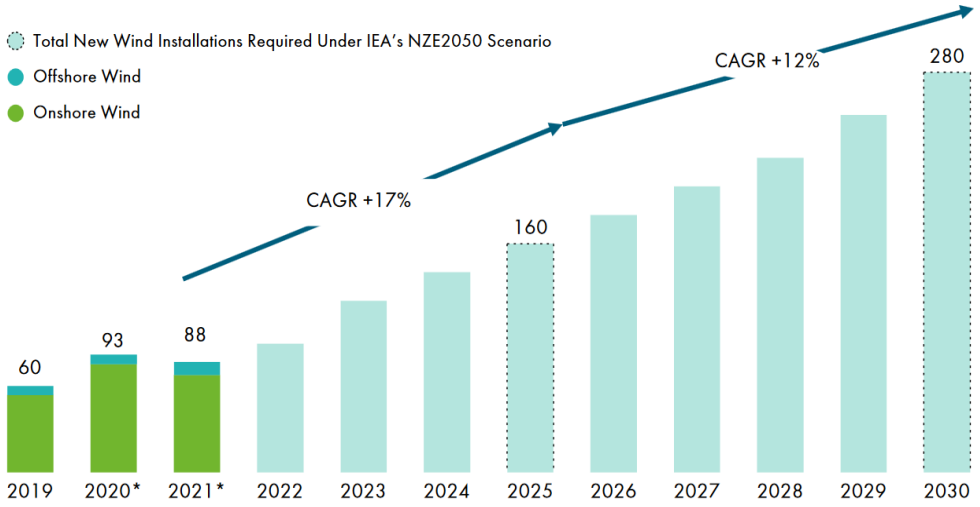


Fig. 1.1: New required global wind installations (in GW) to reach the net zero GHG [3].

The extended period of high energy prices following the COVID-19 pandemic, worsened by the Russian invasion of Ukraine, has revealed a deep energy crisis, as well. This crisis is a result of a fragmented and delayed transition to alternative energy sources, leaving energy markets vulnerable to unstable fossil fuel supplies influenced by political motives and unfair competition, even more than ever. The crisis has not only exposed consumers and industries to high energy costs but has also significantly contributed to the resurgence of inflation as a major global economic challenge. Despite the increasing clarity of the impacts of accelerated global warming, many countries are still postponing necessary actions outlined in the agreement to reduce emissions with the aim of reaching net zero greenhouse gas emission by the year 2050. Nevertheless, this "poly-crisis" has spurred action from society and policymakers, with a growing momentum toward emission reduction and sustainability. Governments worldwide have taken unprecedented measures over the past year to accelerate the energy transition and reduce reliance on fossil fuels. This has led to the adoption of ambitious targets for renewable energy and the phasing out of fossil-based technologies in various sectors. Intergovernmental energy agencies predict that renewable energy, particularly wind and solar power, will dominate electricity demand growth in the coming years. Wind energy offers several advantages over solar energy. Wind turbines can generate electricity day and night, unlike solar systems that are dependent on sunlight.

This continuous generation capability allows wind energy to provide a more consistent and reliable power supply. Furthermore, wind turbines have a smaller land footprint compared to solar PV systems, making them more suitable for areas with limited space availability. Additionally, wind energy is often more cost-effective than solar PV, especially in regions with strong and consistent wind resources. Given these advantages, the focus of this work will be on wind power, exploring its technical aspects and potential for sustainable energy generation.

Increased wind energy will need immediate infrastructure and grid expansion, along with investment in storage technologies and demand-side management. For the world to attain an energy route that is compliant, there is a clear and unequivocal consensus that oil and gas usage must be consistently and substantially decreased during the next three decades. According to the international energy agency, renewable sources will account for 98% of the 2,518 TWh of electricity generation to be added between 2022 and 2025. Global wind energy council (GWEC) expects 680 GW of wind capacity to be added globally between 2023 and 2027, with significant contributions from onshore wind installations. Wind energy is on track to achieve 1 terawatt of installed capacity by mid-2024, and Bloomberg forecasts that wind power will add nearly another terawatt by 2030, after taking 33 years to reach the first terawatt. The 2 TW milestone which is expected to be achieved in just seven years is provided in Fig. 1.2. The information provided is sourced from GWEC Market Intelligence, IEA Net Zero by 2050 Roadmap (2021), and projected new wind capacity from 2023-2030, assuming a ~7.2% CAGR based on GWEC's 2023 Global Outlook. Capacity gap figures are estimations based on the IEA roadmap milestone for 2030. Cumulative global installations for wind energy align roughly with the international renewable energy agency (IRENA) world energy transitions outlook: 1.5°C Pathway (2021). It's important to note that this data represents new and cumulative capacity and does not account for decommissioned projects.

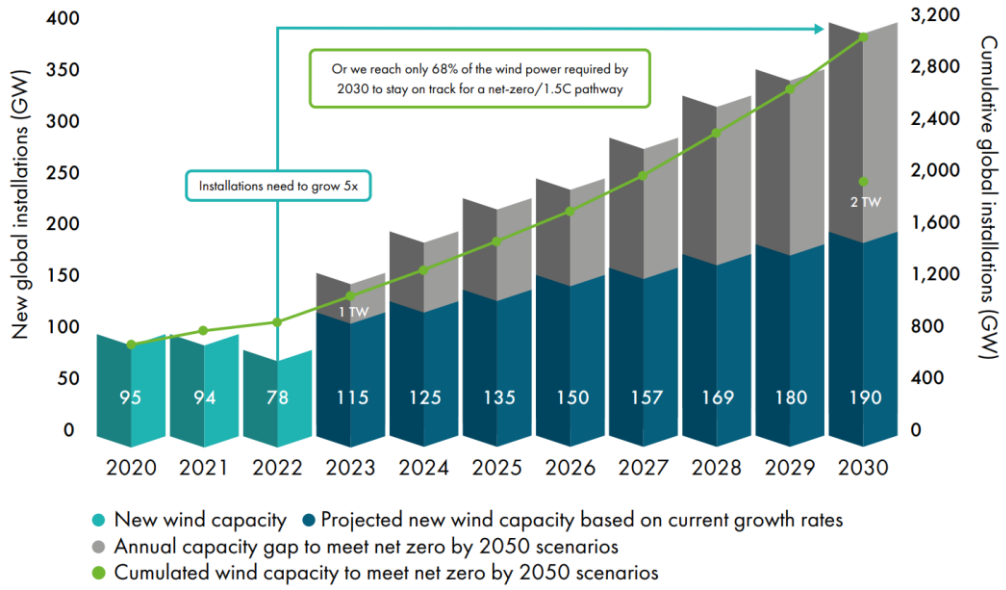


Fig. 1.2: The 2 TW new global wind capacity installation milestone [3].

In 2022, meticulously, the global wind power industry saw the addition of 77.6 GW of new capacity connected to power grids, resulting in a total installed wind capacity of 906 GW. This represents a 9% growth compared to 2021. Despite a 5% year-over-year decline in new onshore installations, 2022 marked the third-highest year in history for additions. Following a record year in 2021 with over 21 GW of new offshore wind capacity commissioned, 2022 saw a drop to 8.8 GW, making it the second-highest year on record. The new installation in onshore and offshore sections are demonstrated in Fig. 1.3 from 2018 to 2022.

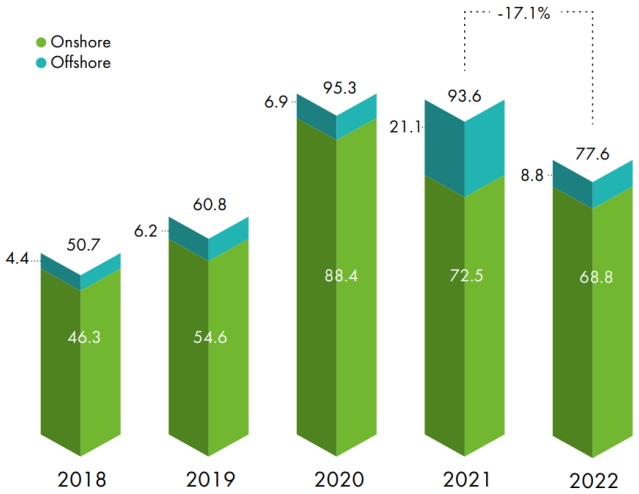


Fig. 1.3: The new wind capacity installation from 2018 to 2022 [3].

The Asia-Pacific region experienced a 3% decrease in market share compared to 2021 but remains the world's largest wind market, with China contributing 87% of its 2022 additions. Europe, as the second-largest market, witnessed record onshore wind installations in 2022, boosting the region's market share from 19% in 2021 to 25%. North America (mainly USA and Canada) retained third place but lost 2% in market share due to slower growth in the US. Latin America increased its market share in 2022 by 1%, driven by a record year for installations in Brazil. Despite a record year in new installations in 2021, Africa & the Middle East connected only 453 MW of wind power in 2022, the lowest since 2013. The top five markets for new installations in 2022 were China, the US, Brazil, Germany, and Sweden, collectively making up 71% of global installations last year, which was 3.7% lower than in 2021. This decline was primarily due to China and the US, the world's two largest markets, losing a combined 5% market share compared to the previous year. This marks the second consecutive year that both countries have lost market share.

In terms of cumulative installations, the top five markets as of the end of 2022 remained unchanged, with China, the US, Germany, India, and Spain together accounting for 72% of the world's total installed wind power capacity, consistent with 2021. New wind power capacity in 2022 by region and also top 10 markets are shown in Fig. 1.4, as well.

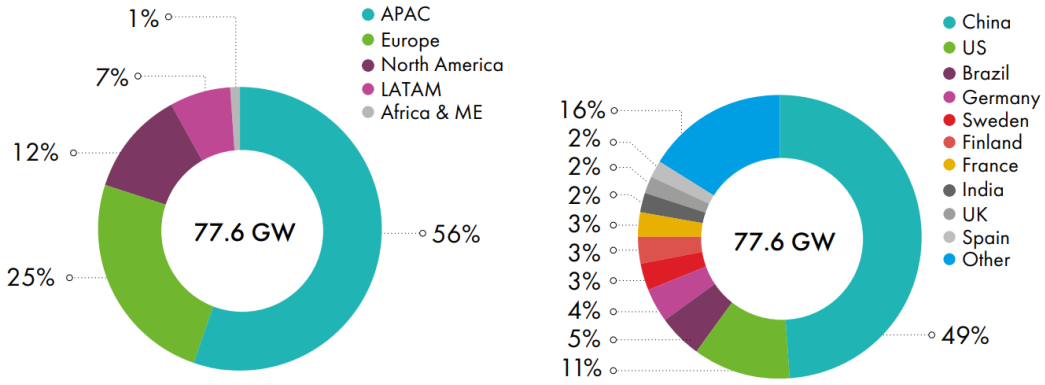


Fig. 1.4: New wind power capacity by region on the left and top 10 markets on the right for the year 2022 [3].

Although, Canada is almost new to exploring its vast wind energy resources, the country has grown fast in the sector, and it is now the ninth largest producer of wind energy in the world with installed capacity of 15.3 GW by the end of 2022. The provincial leaders in wind power capacity are Ontario, Quebec, and Alberta. According to the Canadian Renewable Energy Association,

Canada has 317 wind energy projects producing power. The largest wind farm is Black Spring Ridge, located in Vulcan County, Alberta, with 166 wind turbines. Thirty-nine onshore and offshore wind power projects are planned for the next ten years, with a total project value of $16 billion.

1.2 Onshore and Offshore Wind Turbines

Onshore WECSs are known for their lower installation costs, primarily because they are more accessible and require less logistical effort. They are easier to maintain and repair, leading to reduced operational costs over the system's lifespan. Furthermore, onshore projects typically have shorter permitting and project development timelines, allowing for quicker deployment. Land availability is another key advantage, as onshore sites are abundant and do not necessitate significant land reclamation or marine engineering. However, onshore wind turbines may experience less consistent and weaker winds compared to offshore locations, affecting energy output. Space limitations can also be a concern, particularly in regions with limited available land. Moreover, the visual and noise impact of onshore wind turbines on nearby communities may present challenges. In the year 2022, the global onshore wind sector witnessed the connection of 68.8 GW of new grid-connected capacity, resulting in a cumulative global onshore capacity of 842 GW, reflecting a year-over-year growth of 8.8%. Despite global additions in 2022 being 5% lower than the previous year, Europe stood out by adding a record 16.7 GW of onshore wind capacity, thanks to record installations in Sweden, Finland, Poland, and recovering installations in Germany.

The slowdown in Latin America (LATAM), Africa & the Middle East contributed to the overall decline, but the primary reason was the deceleration of onshore installations in the US. In China, onshore wind installations had declined in 2021 as the market transitioned to ‘grid parity’, where onshore wind electricity receives the same regulated price as coal power. However, it was forecasted that Chinese installations would rebound, reaching 46 GW in 2022. While the Chinese Wind Energy Association reported 44.7 GW of onshore wind capacity installed in 2022, the National Energy Administration reported only 32.6 GW of new grid-connected capacity. In the US, stable onshore wind growth was anticipated for 2022, with the internal revenue service (IRS) granting a one-year extension for projects that began construction in 2016 or 2017, allowing them to qualify for the full production tax credit (PTC) rate if operational by the end of 2022.

Nevertheless, the US wind industry only commissioned 8.6 GW of onshore wind capacity in 2022, the slowest year since 2018, mainly due to supply chain constraints and grid interconnection issues.

Apart from China and the US, other top onshore wind markets in 2022 included Brazil (4.1 GW), Sweden (2.4 GW), and Finland (2.4 GW). Market support mechanisms such as ‘grid parity’, auctions/tenders, and the PTC remained crucial for onshore wind capacity additions in 2022, collectively accounting for a 91% market share. However, excluding China, the global awarded onshore wind capacity through auctions saw a 30% decrease compared to 2021, with Europe experiencing a 29% drop. In China, although only 11 GW of onshore wind capacity was approved under the ‘grid parity’ mechanism in 2022, provincial governments announced more than 50 GW of capacity under the same support mechanism by January 2023, aligning with China's '30-60' targets.

On the other hand, Canada has a robust onshore wind energy sector, with over 13 GW of installed capacity across the country as of 2024. This capacity is spread across more than 300 wind farms, predominantly located in provinces such as Ontario, Quebec, and Alberta. Ontario leads with over 5 GW of installed capacity, followed by Quebec with approximately 4 GW and Alberta with about 1.7 GW. The onshore wind industry has seen significant growth due to supportive government policies, advances in turbine technology, and decreasing costs. Projects like the Henvey Inlet Wind Project in Ontario, which has an installed capacity of 300 MW, illustrate the scale and impact of onshore wind farms. Future onshore projects are expected to add several GW of capacity by 2030, driven by provincial renewable energy targets and increased investment in sustainable infrastructure.

Besides, offshore type has its own set of advantages. Offshore sites generally experience stronger and more consistent winds, resulting in higher energy production. Additionally, offshore wind turbines have minimal visual impact on land, making them more suitable for densely populated coastal regions. The scalability of offshore projects is a significant benefit, with vast offshore areas allowing for the installation of a larger number of turbines and the potential for large-scale wind farms. Moreover, offshore projects typically have fewer land-use conflicts compared to onshore installations. However, offshore wind projects are more expensive due to complex foundation structures, offshore installation, and more challenging maintenance. Logistics can be major challenge, as transporting equipment and personnel to offshore sites can be

logistically complex and costly. Furthermore, offshore projects may have environmental concerns, such as seabed disturbance and potential impacts on marine ecosystems.

In 2022, the global offshore wind sector added 8.8 GW of new capacity to the grid, bringing the total global offshore wind capacity to 64.3 GW by the end of the year. While this represents a 58% decrease from the record-breaking year of 2021, it still marks 2022 as the second-highest year in offshore wind history. China maintained its leadership in global offshore wind development, despite new installations in 2022 being 70% lower than in 2021. This decline was due to the end of the feed-in tariff (FiT), transitioning China's offshore wind market to the era of 'grid parity' starting January 1, 2022. Despite these challenges, China's offshore wind industry demonstrated resilience by commissioning more than 5 GW of new capacity in 2022, surpassing a cumulative installation milestone of 30 GW by the year's end, a feat that took Europe over three decades to achieve.

Europe accounted for the majority of the remaining new offshore wind capacity in 2022, with 2.5 GW connected to the grid across six countries. The UK solidified its leading position in the European offshore wind market by completing the commissioning of wind turbines at the 1.4 GW Hornsea Project 2, now the world's largest operational offshore wind farm. Additionally, the UK grid-connected 27 wind turbines at the 1.1 GW Sea green Project in 2022. France emerged as Europe's second-largest offshore wind market in new additions in 2022, following the full commissioning of its first commercial offshore wind project, the 480 MW Saint-Nazaire wind farm, in November. The Netherlands and Germany also contributed significant new capacity, adding 369 MW and 342 MW, respectively. Italy entered the offshore wind market with its first commercial project, the 30 MW Beleolico offshore wind farm, featuring Chinese wind turbines and becoming the first offshore wind project commissioned in the Mediterranean Sea. Norway's 94.6 MW Hywind Tampen floating wind project, featuring Siemens Gamesa turbines on a concrete floating foundation, faced delays due to supply chain issues, with only seven wind turbines (60.2 MW) operational by the end of 2022.

Overall, 66.4 MW of floating wind capacity was commissioned in 2022, including the Hywind Tampen project in Norway and a 6.2 MW floating wind turbine installed in China on a prototype floater called 'Fuyao'. New offshore installation globally from 2018 to 2022 is provided in Fig. 1.5. The offshore wind market has doubled in size, increasing from 4.4 GW in 2018 to 8.8 GW in 2022. This growth has raised its market share in global new installations from 9% to 11%. However, this

figure is 11 % lower than in 2021, mainly because new installations in China slowed down after an incentive-driven installation rush. GWEC Market Intelligence predicts that the global offshore wind market will maintain its rapid growth rate.

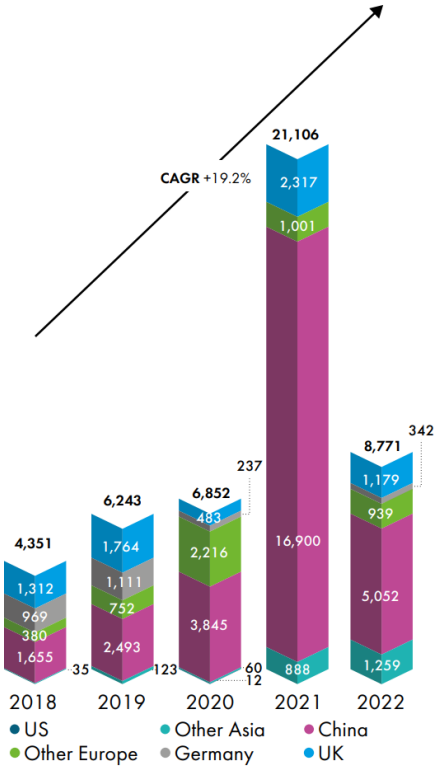


Fig. 1.5: New offshore wind capacity (MW) installation globally from 2018 to 2022 [3].

Canada's offshore wind energy sector is in its early stages, with considerable potential for future development. As of 2024, there are no operational offshore wind farms, but several projects are in the planning and development phases. The most notable among these is the proposed Atlantic Offshore Wind Project, which aims to capitalize on the strong and consistent winds off the coast of Nova Scotia. This project is expected to contribute significantly to Canada's renewable energy capacity, with initial plans targeting an installed capacity of up to 500 MW. Additionally, the federal and provincial governments are investing in research and development to overcome technical and environmental challenges associated with offshore installations. By 2030, Canada aims to establish a competitive offshore wind market, aligning with global trends and enhancing the country's renewable energy portfolio.

This section provides a comprehensive summary of the global wind power capacity installation trends from 2001 to 2022, as depicted in two distinct figures. The Fig. 1.6 illustrates

the new wind power capacity installations annually, while the Fig. 1.7 showcases the cumulative globally installed wind power capacity over the same period. The overall trend indicates a remarkable exponential growth in installed wind power capacity, reflecting the increasing global adoption of wind energy as a sustainable energy source. However, the new installation graph reveals certain fluctuations, particularly during worldwide crises, which have influenced the pace of new capacity additions. Both figures encompass data from both onshore and offshore wind sectors, highlighting the expansion of wind energy technologies in diverse geographical regions.

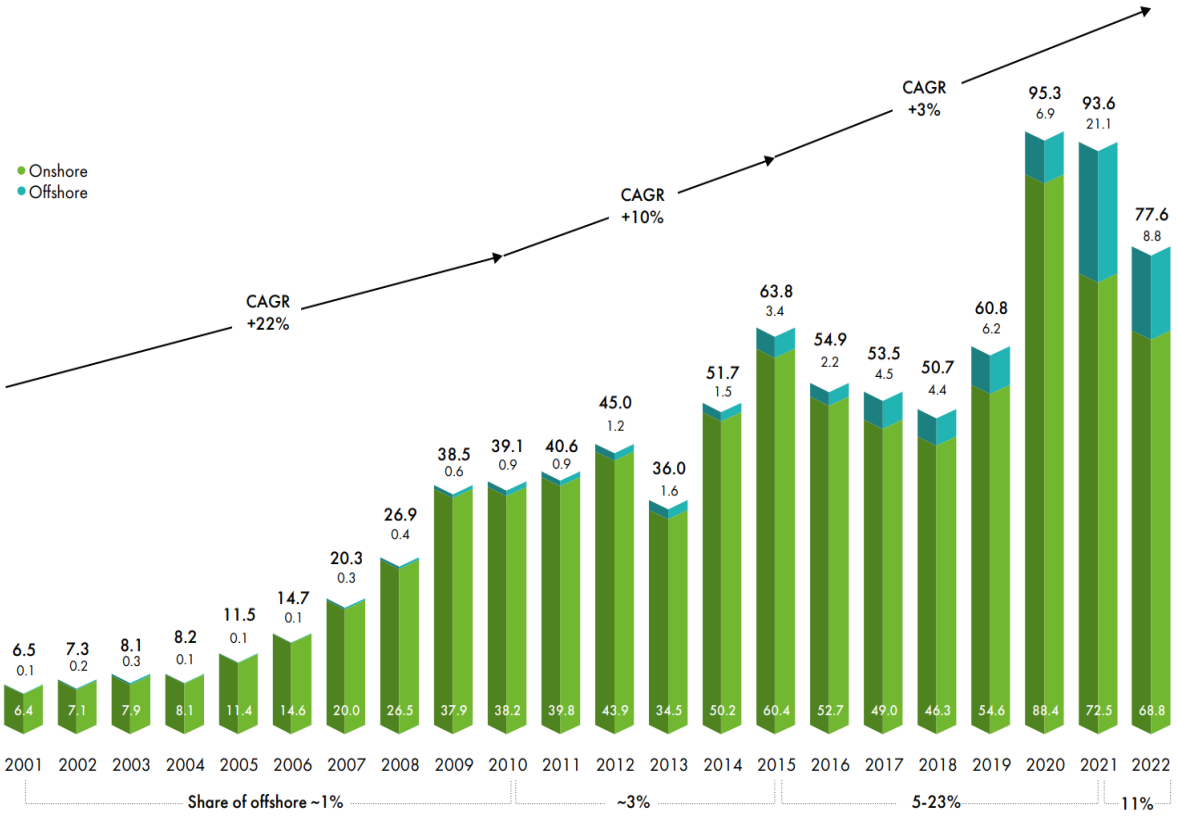


Fig. 1.6: New wind power capacity (GW) installation globally from 2001 to 2022 [3].

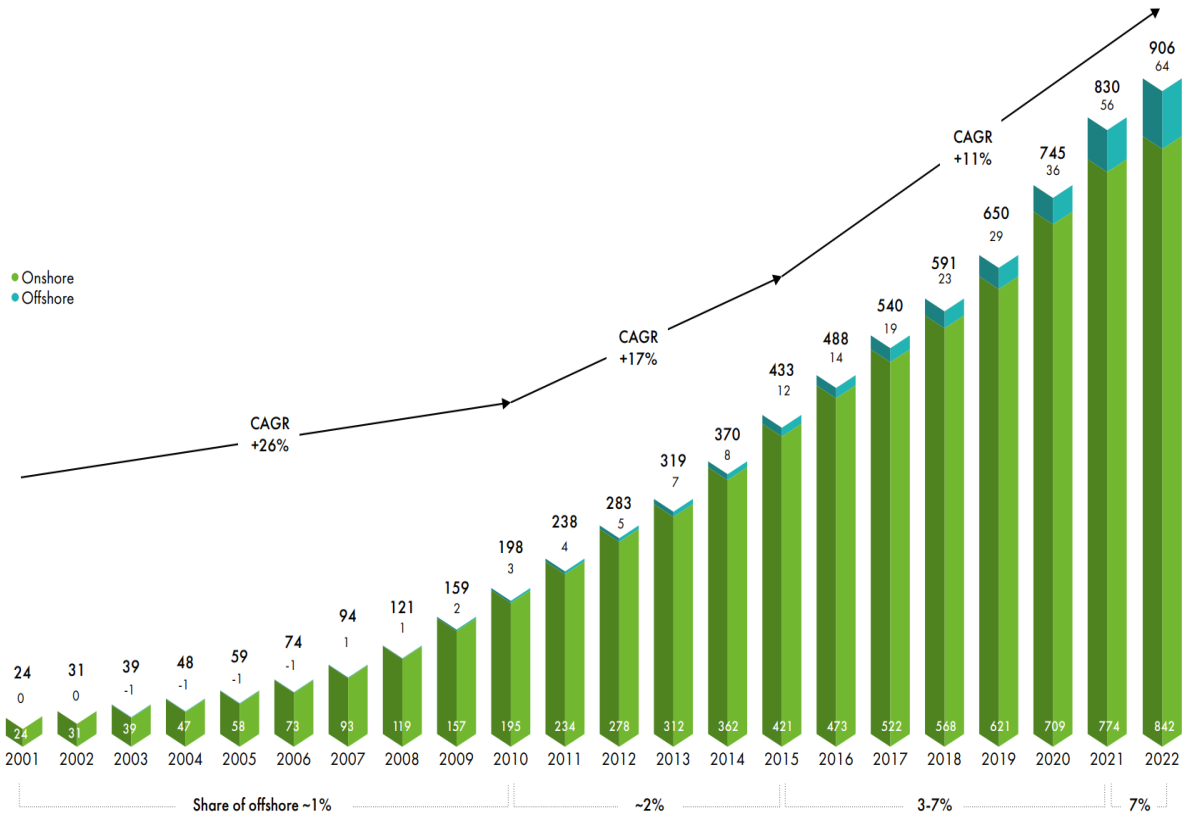


Fig. 1.7: Total wind power capacity (GW) installation globally from 2001 to 2022 [3].

Additionally, Table 1.1 is included to detail the leading countries in both the onshore and offshore wind sectors in terms of new wind power capacity installations and total installed capacity for the years 2021 and 2022. This table provides a detailed breakdown of each country's contribution to global wind power capacity expansion, offering insights into regional trends and developments. The inclusion of onshore and offshore sectors separately allows for a more nuanced analysis of the evolving landscape of wind energy deployment, highlighting the growing significance of offshore wind farms in certain regions.

Table 1.1: Leading countries in onshore and offshore wind power capacity for years 2021 and 2022 [3].

MW, onshore	New installations 2021	Total installations 2021	New installations 2022	Total installations 2022
Total onshore	**72499**	**773818**	**68816**	**841898**
Americas	**19243**	**189582**	**14829**	**204134**
USA	12747	135848	8612	144184
Canada	677	14255	1006	15261
Brazil	3830	21567	4065	25632
Mexico	473	7159	158	7317
Argentina	669	3291	18	3309
Chile	615	3444	824	4268
Other Americas	232	4018	146	4165
Africa, Middle East	**1809**	**9359**	**349**	**9708**
Egypt	237	1702	0	1702
Kenya	102	435	0	435
South Africa	668	3442	0	3442
Morocco	197	1512	276	1788
Saudi Arabia	416	422	**0**	422
Other Africa	189	1846	73	1919
Asia-Pacific	**37352**	**365887**	**36970**	**402852**
China	30670	301419	32579	333998
India	1459	40083	1847	41930
Australia	1746	9125	1412	10537
Pakistan	229	1516	301	1817
Japan	211	4523	149	4668
South Korea	64	1562	96	1658
Vietnam	2717	3102	0	3102
Philippines	0	443	0	443
Kazakhstan	88	337	418	755
Other APAC	168	3776	169	3945
Europe	**14095**	**208991**	**16667**	**225204**
Germany	1925	56814	2403	58951
France	1192	19079	1590	20653
Sweden	2104	11952	2441	14393
United Kingdom	328	14074	502	14575
Spain	750	28134	1659	29793
Finland	671	3186	2430	5607
Netherlands	952	5370	933	6223
Turkey	1400	11102	867	11969
Other Europe	4773	59280	3842	63040

MW, offshore	New installations 2021	Total installations 2021	New installations 2022	Total installations 2022
Total offshore	**21106**	**55549**	**8771**	**64320**
Americas	**0**	**42**	**0**	**42**
USA	0	42	0	42
Asia-Pacific	**17788**	**27695**	**6311**	**34006**
China	16900	26390	5052	31442
Japan	0	52	84	136
South Korea	0	142	0	142
Vietnam	779	874	0	874
Taiwan	109	237	1175	1412
Europe	**3317**	**27812**	**2460**	**30272**
United Kingdom	2317	12739	1179	13918
Germany	0	7713	342	8055
France	0	2	480	482
Netherlands	392	2460	369	2829
Denmark	605	2308	0	2308
Belgium	0	2262	0	2262
Other Europe	4	328	90	418

1.3 Wind Turbine Supply Chain

In terms of the global wind supply chain, Europe has a well-developed and mature supply chain, covering turbine nacelles, key components, and raw materials. However, since establishing its local wind supply chain in 2008-2010, China has emerged as the world's leading wind turbine manufacturing base and the largest production hub for key components and raw materials. European and American wind turbine original equipment manufacturers (OEMs) opted to broaden their supply chain to ensure a secure supply in the aftermath of the COVID-19 pandemic. India, the second-largest hub in the Asia-Pacific (APAC) region for turbine assembly and key components production, has emerged as a more prominent player in the global wind supply chain. While the majority of wind industry suppliers are still located in APAC, Europe, and the Americas, new players have also emerged in the Middle East and North Africa (MENA) region. As for the

adequacy of the supply chain capacity to support growth, globally, there are currently 153 turbine assembly plants in operation, with an additional 74 facilities either under construction or in the planning stages. China leads the way with over 100 nacelle assembly facilities in operation and 64 under construction, boasting a turbine nacelle production capacity of 98 GW per year, which accounts for 60% of the global market share. Europe ranks as the second-largest turbine nacelle production base, with assembly facilities mainly located in countries such as Germany, Denmark, Spain, France, Portugal, and Turkey. The US follows as the world's third-largest wind nacelle manufacturing hub, trailed by India and LATAM, primarily Brazil. Overview of global wind turbine (WT) nacelle facilities are provided in Table 1.2. Globally, there is a nacelle production capacity of 163 GW in 2023. While initially, the wind industry seems to have enough nacelle assembly capacity to meet the projected global demand up to 2027.

Table 1.2: Overview of global WT nacelle facilities [3].

	China	Europe	India	USA	LATAM	Asia Pacific	Africa	Total
Number of nacelle assembly facilities (onshore)	77	16	13	4	6	3	1	123
Number of nacelle assembly facilities (offshore)	20	5	0	0	0	4	0	30
Number of announced nacelle assembly facilities (onshore)	17	0	2	0	0	0	0	19
Number of announced nacelle assembly facilities (offshore)	47	1	0	3	0	4	0	55

The provided information underscores the critical importance of studying wind energy conversion systems in the context of achieving net-zero greenhouse gas emissions by 2050. As the global community strives to mitigate climate change and transition to renewable energy sources, a deep understanding of WECS becomes paramount. Increased knowledge and research in various fields related to wind power systems, including aerodynamics, structural engineering, control systems, and grid integration, are crucial. Advancements in these areas are essential for enhancing the efficiency, reliability, and cost-effectiveness of WECS. Moreover, as the world looks to scale up wind energy capacity significantly, innovative solutions and interdisciplinary collaboration will play key roles in accelerating the adoption of wind power and achieving sustainability goals. The upcoming sections will delve into the technical concepts of WECS, providing a detailed exploration of the principles and technologies driving this vital component of the renewable energy landscape, apart from general numerical information which has been discussed, so far.

1.4 Wind Energy Conversion System

Typically, a WECS is a structure consists of both mechanical and electrical components to transform wind energy into electrical power. The WECS should be configured to harvest the maximum possible amount of power from wind turbine generators under various operating conditions in order to meet the grid criteria [4]. A standard configuration of a wind energy conversion system, can be observed in Fig. 1.8, which primarily consist of a wind turbine rotor, gearbox, electrical generator, and power electronic converter as an interface, and a transformer for electrical power grid integration purpose.

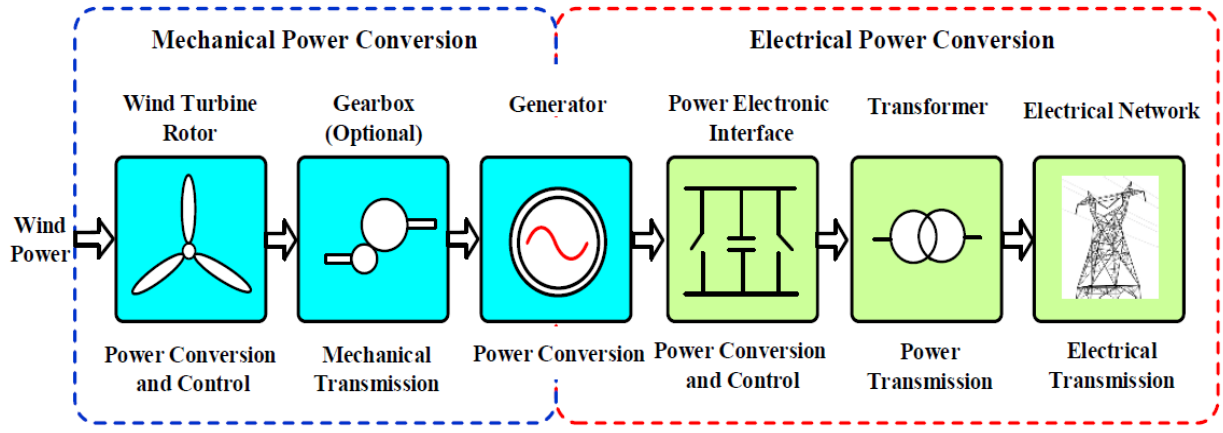


Fig. 1.8: Standard configuration of a wind energy conversion system [4].

Where the wind turbine's shaped blades can aerodynamically absorb the wind energy, it transforms to rotating mechanical power, which is then converted to electrical power by the generator unit. Gearboxes are used in WECS to optimize generator speed, enable the use of smaller generators, enhance torque production at lower wind speeds, maintain a constant generator rotational speed for stable AC output, adapt to variable wind speeds, reduce mechanical stress, lower noise levels, improve safety, and facilitate maintenance and repairs, making the overall wind energy system more efficient and cost-effective. Power converters are essential in such systems to ensure grid compatibility, optimize power capture from variable wind speeds, control voltage and frequency, provide reactive power support, enhance grid reliability, enable fault ride-through, and facilitate energy storage integration, all of which contribute to efficient wind energy generation and seamless integration with the electrical grid [4], [5], and [6]. And the last but not the least, transformers are employed in a WECS for grid connection to step up the generated electricity's voltage to match grid voltage levels, reducing transmission losses and enabling efficient power transfer over long distances.

Based on the above-mentioned components, various types of WECSs have been developed to optimize wind energy conversion and grid integration. These systems can be categorized as follows: Fixed-Speed WECS, traditional systems employing gearboxes and power converters for grid compatibility; Variable-Speed WECS, which utilize power converters to adapt to changing wind conditions and improve efficiency; Direct-Drive WECS, eliminating gearboxes for increased reliability and cost-effectiveness. These diverse WECS types cater to specific operational requirements and environmental contexts, and they will be explained in more detail in the following sections.

1.4.1 Feasible Configurations of WECS

This section contains a literature review on practical solutions for wind energy conversion systems. Fixed speed wind energy and variable speed wind energy systems are two commonly known types of technological solutions that can meet all generator and grid side specifications [7]. In the 1980s, a fixed-speed wind energy system was suggested [8]. An asynchronous squirrel cage induction generator (SCIG), which is also known as type 1 WECS, with a soft starter using thyristors makes up a configuration (Fig. 1.9). The wind turbine generators are connected to the

power grid directly via step-up transformers. They can therefore be operated with a rotor speed variation of less than 1%. To achieve a unity power factor, a bank of shunt capacitors is required to provide reactive power compensation. SCIGs may have appealing alternatives, such as simple construction and low cost. The soft starter is connected to the grid without the use of synchronization devices, resulting in a smooth grid integration. The key disadvantage is lacking a speed control mechanism. For reliability and stability, due to the high mechanical stress induced by wind torque pulsation, it needs a stiff grid. To have almost 10% speed variation resistors can be added to the rotor of the SCIG and it is considered as the 2nd type of WECS. However, the efficiency is still low, power factor is low, and capacitor bank is still a must. Mechanical stress on turbine also remains high so soft starter should be placed in such systems, as the first type.

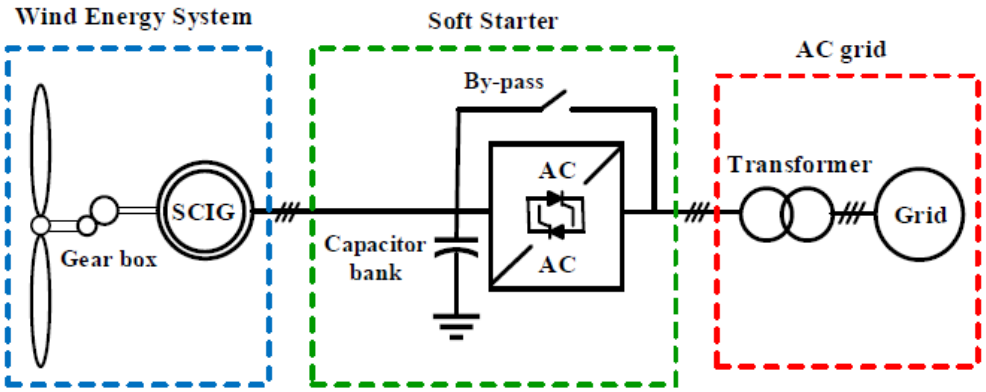


Fig. 1.9: Fixed speed WECS converter equipped with soft starter.

Due to the instability in wind speed and torque, the higher variable speed wind energy system has many advantages over the fixed speed WECS, including lower mechanical tension on components, such as the gearbox and the shaft. Furthermore, the system can be infused with full power output with less fluctuated power. A portion-scale and full-scale power converters are presented in this literature review as the two major configurations of the variable speed wind energy system. This variable speed configurations are used in modern WECSs to maximize the power energy captured from the wind. They can also maintain active and reactive power management under all operational conditions. Reference [7] has defined a Doubly Fed Induction Generator (DFIG) based WECSs (type 3). With a wound rotor induction generator (WRIG), it can provide partial variable speed system. The power grid is directly connected to the generator stator

in this configuration. As seen in Fig. 1.10, a back-to-back (BTB) indirect ac/dc/ac power electronic converters integrate the rotor into the grid via slip rings.

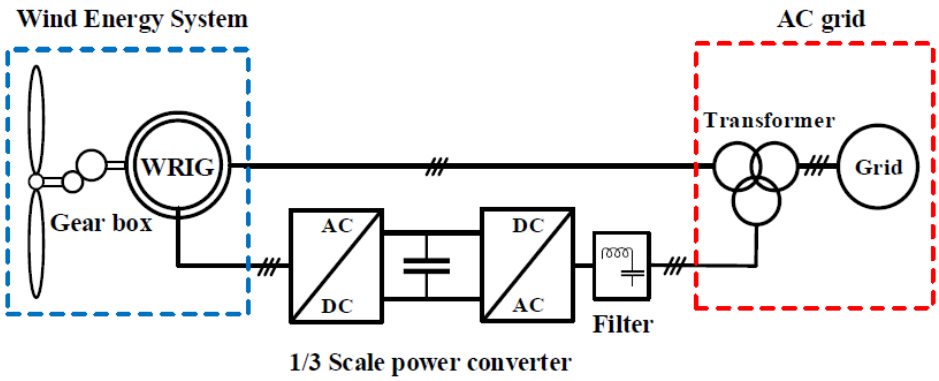


Fig. 1.10: Variable speed WECS based on the partial scale power converter.

Aside from the inclusion of slip rings, another major disadvantage of this arrangement is the stator direct connection to the grid, which limits controllability with a safety mechanism during a power grid fault.

However, as seen in Fig. 1.11, another variable speed WECS configuration based on the full-scale power converter can be employed (WECS type 4). Squirrel cage induction generator (SCIG), WRSG, and PMSG are examples of generators that are validated to be used in this configuration. Although, PMSG has a few advantages over other generators, such as; it regulates the grid voltage by supplying further reactive power, since it is interfaced with the power system via a full-scale back-to-back converter. It also eliminates the requirement for a gearbox, reducing the weight and size of nacelle hardware, as well as mechanical losses and maintenance demands. Via full-scale BTB power converters and transformers, they are smoothly linked to the grid. The ac-dc converter on the generator side is used to control generator output power over a large speed range and maintain unity power factor, while the dc-ac converter on the grid side is used to provide active and reactive power conversion to keep the dc link voltage of the capacitor fixed. The wind turbine can be directly coupled to the BTB power converter without using a gearbox for a variable speed WECS that uses a multi-pole PMSG, or the gearbox size can be decreased, as mentioned above. For example, working at low speeds with no gears will improve the WECS reliability, as well [9].

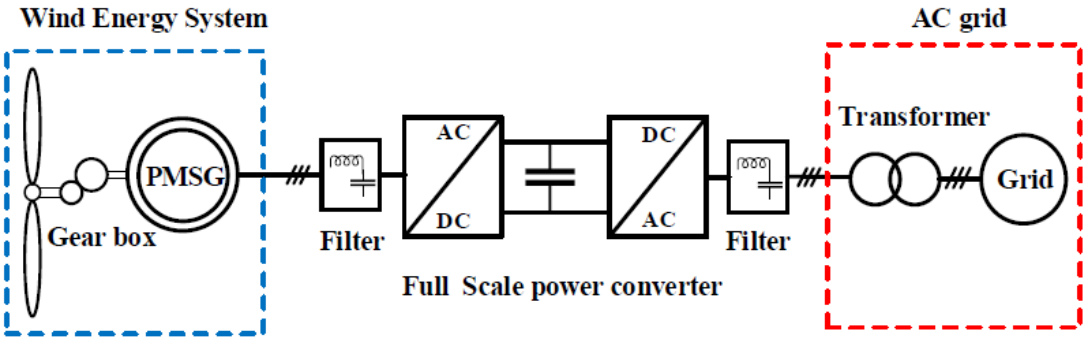


Fig. 1.11: Variable speed WECS based on the full-scale power converter.

While PMSG based WECS offer several advantages, as stated, there are some drawbacks associated with this system. One drawback is its relatively higher initial cost due to the use of permanent magnets and full-rated power converter. Additionally, PMSG-based systems may face challenges in grid integration, as they require almost complex power electronics and control systems to ensure proper grid compatibility. However, it's important to emphasize that these drawbacks can be seen as trade-offs for the benefits of increased efficiency, reliability, and other pros which would be offered by employing this configuration.

Overall, considering the potential and positive attributes of PMSG-based WECS, and DFIG-based ones are promising candidate for industry applications. However, in this work the main concentration is on direct-drive PMSG-based WECS.

As previously mentioned, a power converter is a must in a WECS. To link a wind turbine equipped with PMSG to the grid system, a variety of power converter topologies can be used. The key goal of these topologies is to meet the specifications of both the wind generator and the power grid in terms of optimum performance, low cost, and reliability with a straightforward maintenance approach. They should also be able to monitor active power and compensate reactive power, so that grid codes are met independent of wind speed variations [9]. Hence, the following sections study and address some of the possible power converter topologies which can be employed within the WECSs. The power converter classification is also provided; however, the concentration is on the 4th type of WECS which is aimed in this work.

1.5 Power Converter Topologies

The power converters can be classified based on many aspects such as; direct / indirect conversion. In direct conversion the constant AC voltage and frequency directly converted to variable AC voltage and variable frequency. However, in an indirect one there are two stages; AC to DC, then DC to AC [10]. Direct and indirect conversion methods in power electronics play crucial roles in converting electrical energy efficiently. In the context of direct conversion (AC/AC), it involves a single-stage process of converting AC power to the desired output voltage or frequency, which can reduce overall power loss and increase efficiency. However, direct conversion systems often require sophisticated semiconductor technology and control algorithms, leading to higher initial costs and potential reliability issues. In contrast, indirect conversion (AC/DC/AC) involves multiple stages, which may introduce additional energy losses due to the intermediate conversion steps, reducing efficiency. Yet, indirect conversion methods are typically more versatile, offer better fault tolerance, and are more cost-effective for certain applications. Overall, despite the inherent energy losses in indirect conversion, the flexibility, reliability, and cost-effectiveness it offers often make it the preferred choice for providing effective and practical solutions in many power-electronic applications. However, by optimizing the control of multilevel power converters, enhancing semiconductor devices, and developing more efficient switching techniques, it is possible to further enhance the energy efficiency and overall performance of these systems. As a result, ongoing efforts are directed toward refining these control strategies and technologies to continually improve the efficiency and reliability of indirect conversion methods in power electronics.

Another critical categorization in power electronics pertains to the distinction between current source and voltage source power converters. In the context of WECS, voltage source power converters are frequently favored. These converters are well-suited for applications like WECS due to their ability to provide stable voltage output regardless of fluctuations in load or disturbances in the grid. This is vital for maintaining grid compatibility and for ensuring that the electrical energy generated by the wind turbines is seamlessly integrated into the existing power grid, as it allows the WECS to generate power at a constant voltage and frequency that aligns with grid requirements. Voltage source power converters also offer a high degree of control and are adaptable to varying wind conditions, making them an ideal choice for optimizing power capture and energy production in WECS installations.

To summarize and have a clearer overview, the most common available topologies for medium-voltage high-power converter can be shown as Fig. 1.12. As mentioned, according to this figure, power converters are broadly classified into two main categories: direct and indirect. The indirect class further divides into unclassified types, including voltage source and current source converters. In contrast, the direct class comprises only two types: matrix converters and cycloconverters. These classifications are crucial for understanding the diverse range of power conversion technologies available and their specific applications in various industries, especially in the context of wind energy where converters play a vital role in transforming and managing electrical power efficiently.

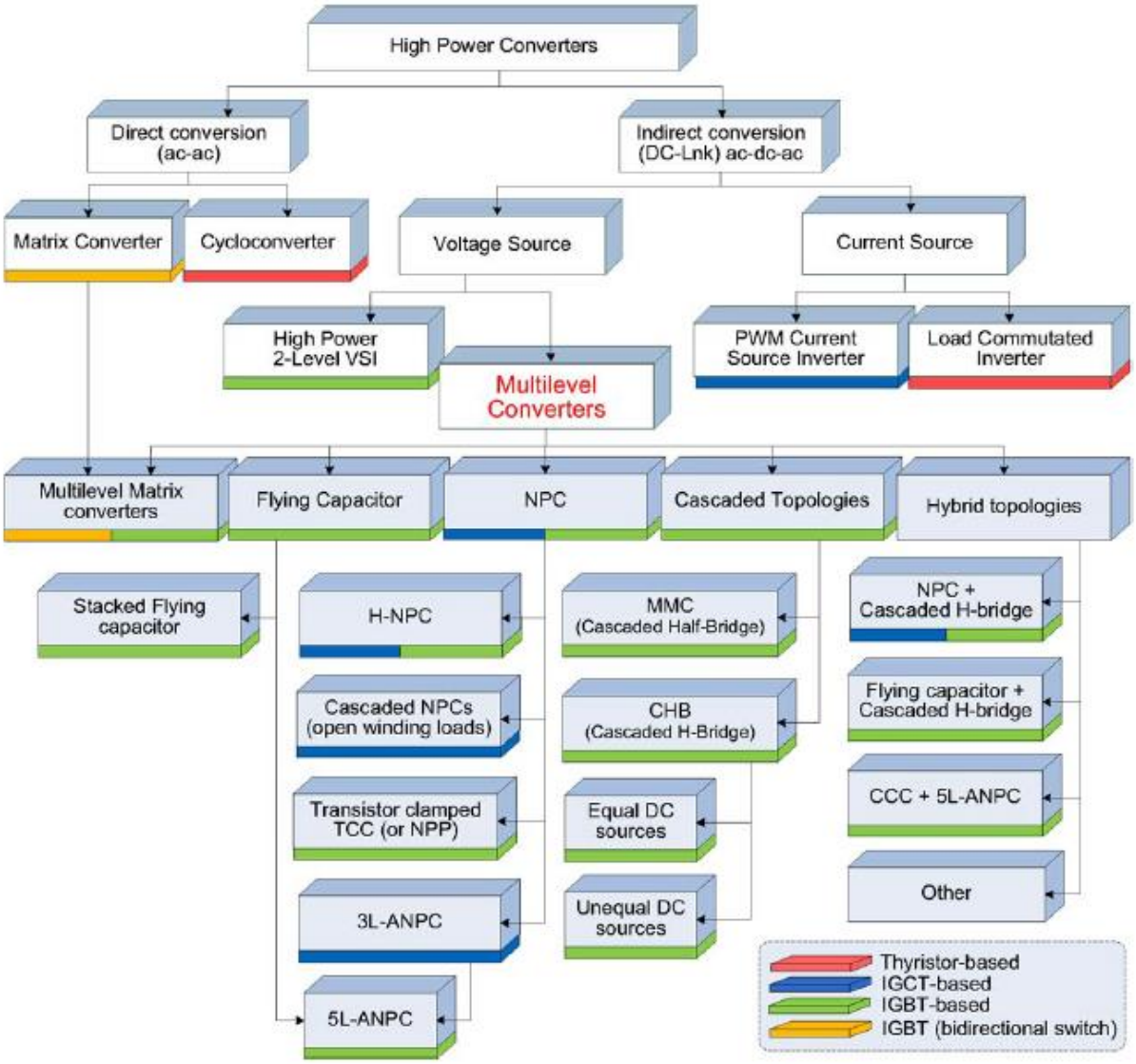


Fig. 1.12: Multilevel converter classification [10].

Based on the pros and cons of the concepts discussed thus far, the fourth type of WECS, which contains PMSG, utilizing indirect power conversion has emerged as a promising configuration. The indirect power conversion method offers enhanced flexibility and versatility, making it well-suited for optimizing wind energy conversion in varying conditions. Consequently, this approach has garnered attention as a viable option for WECS. In light of this, the next section will delve into the details of a suitable and well-established power converters within this system, shedding light on its crucial role in efficiently harnessing wind energy, and their advantages and disadvantages over each other. It contains two-level and multi-level power converters for PMSG-based WECS.

1.5.1 Two-Level Converter for PMSG Based WECS

The back-to-back two-level voltage sourced converter (VSC), as seen in Fig 1.13, is the most general topology used to interface direct drive of PMSG based WECS with the gird. The voltage sourced rectifier (VSR) will absorb the most wind energy on the generator side. The voltage source inverter (VSI) can regulate active and reactive power while maintaining a steady dc link voltage. Despite the fact that this converter has a simple structure with six switches on both sides, it can experience high voltage stress dv/dt in high power medium voltage applications. Furthermore, large size filters may be needed to meet gird device specifications for lower total harmonic distortion [11], [12].

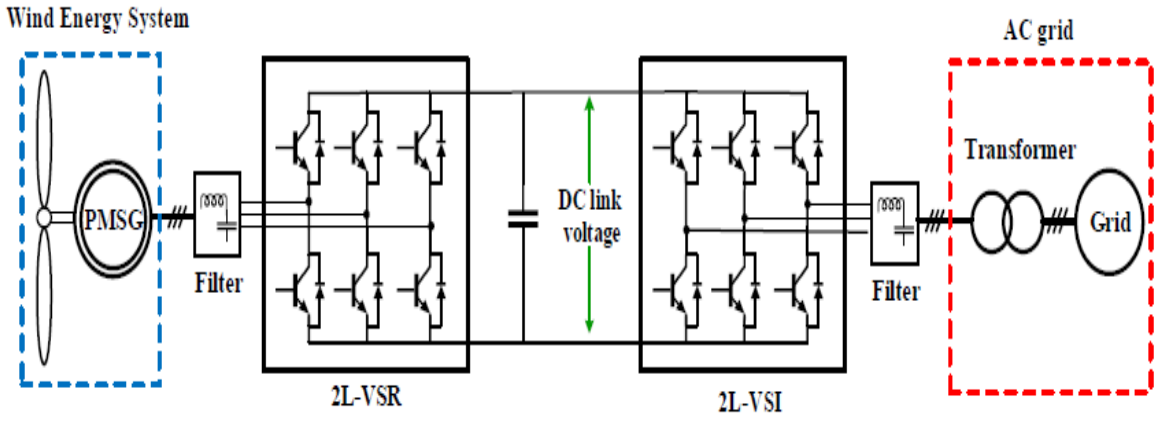


Fig. 1.13: Configuration of direct drive PMSG for WECS using two-level BTB converter.

Figure 1.14 demonstrates a semi-controlled system with the BTB power converter. By replacing three-phase, six diodes instead of a two-level VSR, a low cost-effective solution with high efficiency can be achieved as compared to the previous power converter topology. Diodes rectifier, dc-dc boost converter, and two-level voltage source inverter make up this topology. Since reactive power is not available in such a PMSG-based system, converter topology uses the voltage source inverter (VSI) to pass active power unidirectionally to the power grid system [13]. However, the diode rectifier output is small, so boost converter is required to compensate the grid side converter low output. In this regard, the number of elements, cost, and control complexity will increase. Additionally, the use of a diode rectifier may lead to higher total harmonic distortion, potentially requiring larger filters to meet grid specifications.

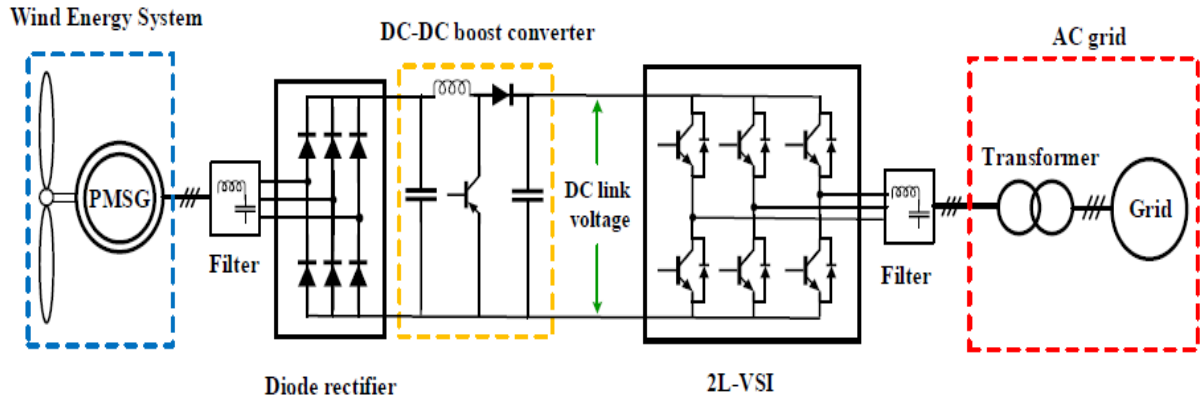


Fig. 1.14: PMSG for WECS configuration with direct drive using diodes rectifier, dc-dc boost converter and two-level voltage source inverter.

1.5.2 Multi-Level Converter for PMSG Based WECS

Multilevel converter topologies can reach a several number of output voltage levels as compared to traditional two-level converters. As the output voltage waveform quality improves, a smaller filter size is needed. The voltage tension around the converter electronic devices is also diminished. The topology of the flying capacitor (FC) multilevel BTB converter, is considered as one of the typical structures, can be combined with PMSG direct drive for WECS. Fig. 1.15 depicts a 3-level FC converter. The capacitors are used to synthesize the output voltage levels in terms of switching states. Furthermore, the FC converter stage can be easily expanded. The key drawbacks

of flying capacitor converters in WECS include their complex control and management requirements, voltage stress concerns, limited scalability, a higher component count leading to increased costs and maintenance complexity, limited fault tolerance, and increased size and weight. While these converters offer advantages as mentioned, their limitations make them less preferable in certain WECS configurations, particularly in larger installations or those where scalability, reliability, and cost-effectiveness are paramount, driving the choice of alternative converter topologies [15].

In WECS, a stacked flying capacitor power converter can also be used to connect variable-speed wind turbines to the electrical grid. It comprises multiple flying capacitor cells stacked in series, each containing a capacitor, a switch, and a diode. This configuration enables the conversion of variable frequency and voltage AC from the wind turbine to fixed frequency and voltage AC suitable for grid connection. However, compared to a regular flying capacitor converter, the stacked configuration offers several advantages. It allows for increased voltage levels, improved voltage balance across capacitors, and potentially higher efficiency due to reduced switching losses and improved voltage balancing. However, this configuration also introduces complexity in control and operation, requiring more sophisticated control algorithms. Additionally, the increased component count and complexity may lead to higher costs compared to a regular flying capacitor converter. So, regular FC option is more common in wind energy industry.

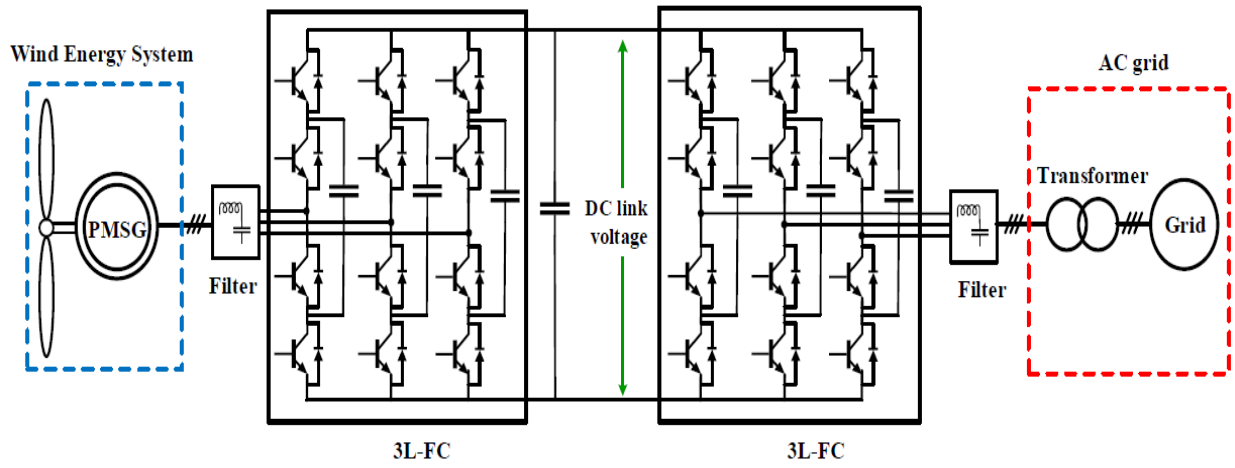


Fig. 1.15: Configuration of direct drive PMSG based WECS using BTB regular flying capacitor multi-level converter.

As seen in [14], and [15], neutral point clamped (NPC) multilevel converter is used for direct-drive PMSG based WECS. To satisfy the specifications of the gird integration, this topology is developed by the BTB structure. In this case, a zero-voltage level can be achieved by clamping diodes connecting the midpoint of the IGBT switches to the converter neutral point. Though, in order to balance the dc link capacitor voltage in this topology, a control circuit is needed. The utilization of NPC power converters in PMSG-based WECS comes with a range of noteworthy advantages. Notably, NPC converters are known for their inherent simplicity in control and operation compared to the advanced features they offer. This simplicity can translate to easier maintenance and lower control system costs. Additionally, NPC converters excel in delivering high-quality voltage waveforms, effectively reducing harmonics and enhancing power quality, which is crucial when integrating renewable energy sources into the grid. Their ability to efficiently operate at high switching frequencies can lead to downsized filtering components, resulting in a more compact and cost-effective system. Moreover, NPC converters offer enhanced fault tolerance and high reliability, making them well-suited for applications where uninterrupted energy supply is essential.

However, like any technology, NPC converters also present some drawbacks. The increased semiconductor count required by the converter configuration compared to the 2-level voltage source converter, while offering advantages in fault tolerance, leads to higher complexity in control algorithms and increased manufacturing costs. This higher number of semiconductor devices can also contribute to potential reliability concerns, as each component represents a point of failure. Moreover, NPC converters are not as suitable for applications requiring very high voltage levels, which may necessitate the use of alternative converter topologies. Although, the maximum voltage that can be used with a NPC power converter depends on several factors, including the specific design of the converter, and the voltage rating of the semiconductor devices used. But, in typical wind turbine s, for very higher voltage applications, such as grid-scale wind farms, other converter topologies like modular multilevel converters (MMC) may be more appropriate due to their higher voltage capabilities.

As a result, for high-power, medium voltage wind energy applications, a direct-drive PMSG wind turbine system with a 3L-NPC BTB power converter, which is illustrated in Fig. 1.16, seems to be a promising system, and will be comprehensively analysis along with the work.

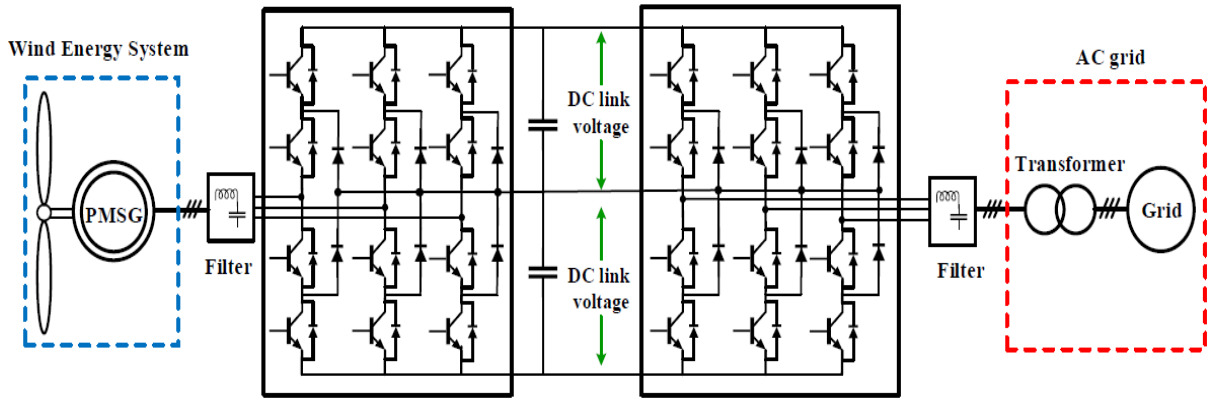


Fig. 1.16: Configuration of direct drive PMSG based WECS using BTB neutral point diode clamped multilevel converter.

Marquardt and Lesnicar proposed the MMC for the first time in 2004. Then, [16] discussed the state of the art of MMC topologies. The MMC, has emerged as a candidate in WECS and HVDC applications due to its more appealing technological approach. The MMC is made up of many submodules connected in series. In contrast to a 2-level VSC, this structure can cause a system switching frequency in the MMC to be decreased. As a result of the low voltage tension dv/dt through the IGBT system, the total harmonic distortion (THD) of the output voltage will be dropped at the same switching frequency [17]. The MMC, nowadays, are aimed in many literatures. However, it has no much propensity to employ MMC in medium voltage WECS due to the major drawbacks such as requirement of large number of semi-conductors, other power electronic devices, and capacitors. Furthermore, since individual capacitor voltage balancing control is needed, the MMC control circuits can become more complicated. As can be seen in Fig. 1.17, the MMC can be developed with several number of half-bridge (HB) or full-bridge (FB) submodules. Based on the above-mentioned flaws, the MMC will not be investigated in this research.

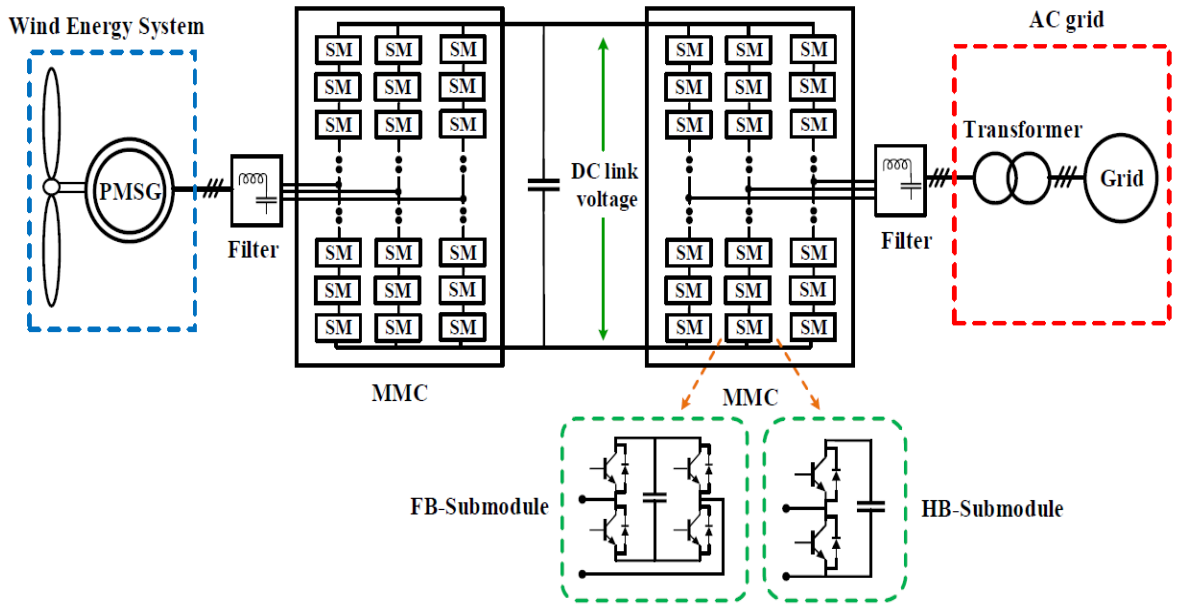


Fig. 1.17: Configuration of direct drive PMSG based WECS using BTB modular multi-level converter.

Apart from the above-mentioned multilevel converter, the cascaded H-bridge (CHB) converter stands as another viable choice for WECS in industry. Cascaded H-bridge converters offer certain advantages and disadvantages. One of the key advantages is their modularity, which allows for easy scalability and maintenance. They also have a relatively high level of voltage and current control, making them suitable for various wind turbine configurations. However, they do come with some drawbacks, including increased component count, which can result in higher initial costs and increased susceptibility to component failures. In contrast as stated, the NPC multilevel converter, offers more efficient power conversion and reduced harmonic distortion proportional to its simplicity. Its lower component count and proven reliability make it a preferable choice for long-term sustainability and cost-effectiveness.

Comparing the analyses of four distinct popular multilevel inverter topologies are included in Table 1.3. Each topology contains advantages and disadvantages. Overall, the implementation of power converters varies depending on their benefits and downsides. CHB or MMC are commonly employed in the high-voltage and high-power applications. The NPC, on the other hand, is used in medium-voltage, high-power applications. Hence, the output of the WECSs mostly is in medium range (less than 1 kV), the NPC can be considered as a promising architecture in wind energy industrial applications based on literature.

Table 1.3: Compares the benefits and downsides of several multilayer converters.

Topology	Advantages	Disadvantages	Ref.
CHB	• Good for applications that require higher fault tolerance. • It is possible to use in an asymmetric source arrangement. • Using a modular layout, makes it able to operate in a high-power rating. • Just unidirectional switches are needed. • Reliable and almost simple structure	• There is a need for more gate drivers. • To increase the output voltage, a large number of individual dc sources are necessary. • Asymmetric configuration causes a loss of modularity. • Implementation costs are high. • Semiconductor with different voltage ratings.	[18] [19]
FC	• Ideal for applications involves high fault tolerance. • The quantity of DC sources is reduced. • Lower output filter is required. • Switches are under less voltage stress.	• A significant number of electrolytic capacitors are required. • Switching efficiency is poor. • The expense of installation is considerable.	[20] [21]
MMC	• Modularity at its finest. • Low output harmonics. Good for various voltage ratings. • Transformer is optional in some cases. • The number of semi-conductor power switches does not increase as voltage levels rise.	• A significant number of independent DC voltage sources is required. • Various phases of the inverter have different voltage imbalances. • Through the capacitor, high voltage ripple can be found. • High thermal and conduction losses.	[22] [23]
NPC	• It's excellent in EMI and thermal management. • The number of DC sources is reduced. • Ideal for applications which require high fault tolerance. • The design is simple. • Superior dynamic responsiveness and harmonic spectrum performance. • Neutral clamping switches help solve voltage balancing and the unequal sharing of losses amongst switches.	• The voltage balancing circuit can be complicated. • Inner and outer switches have an unequal share of voltage stress and losses.	[24] [25]

1.5.3 Overview of Industrial Power Converters

In the realm of industrial wind energy conversion systems, the selection of converter topologies is pivotal for the efficient conversion of variable wind turbine output into usable electrical power for the grid. Various converter types are employed based on factors such as power rating, grid requirements, and cost considerations. The CHB converter is a prominent choice, offered by manufacturers like ABB, Siemens, and GE Renewable Energy. Typically, CHB converters for WECS applications exhibit rated power levels ranging from 1 MW to 10 MW per turbine, with voltage specifications varying from 690 V to 34.5 kV or higher for grid connectivity.

Another prevalent converter topology in the WECS sector is the NPC converter, manufactured by companies such as Schneider Electric, Toshiba, and Hitachi. NPC converters share similar rated power and voltage characteristics with CHB converters, making them suitable alternatives for medium to high-power WECS applications. For instance, Schneider Electric's NPC converters can handle power ratings ranging from 1 MW to 10 MW, with voltage levels from 690 V to 34.5 kV, ensuring compatibility with various grid integration requirements. However, NPC converters have lower switching losses due to their reduced number of switching devices per phase, leading to higher efficiency and reduced heat generation. They also exhibit lower harmonic distortion in the output voltage waveform, which helps in maintaining a cleaner output voltage and reducing stress on connected equipment. Additionally, NPC converters offer improved voltage balance, higher modularity, scalability, better overload capability, and improved fault tolerance compared to CHB converters in industry.

The Two-Level Full Converter is commonly employed in large-scale wind energy projects, with manufacturers like Vestas, Nordex, and Goldwind offering solutions. These converters exhibit rated power and voltage ranges akin to CHB and NPC converters, catering to diverse grid integration requirements. For example, Vestas' 2LFC converters can handle power ratings up to 3 MW with voltage levels from 690 V to 34.5 kV, ensuring seamless integration into the grid.

Matrix converters, offered by ABB, Siemens, and Danfoss, are also prevalent in medium to high-power WECS setups, eliminating the need for bulky DC-link capacitors and enhancing power density and efficiency. Full-Bridge and Half-Bridge converters, supplied by Infineon Technologies, ON Semiconductor, and STMicroelectronics, cater to medium to high-power and low to medium-power applications, respectively, each with its own voltage and power specifications. For instance, Infineon Technologies' Full-Bridge converters can handle power

ratings up to 10 MW, while their Half-Bridge converters are suitable for power ratings ranging from 100 kW to 1 MW, ensuring compatibility with various WECS requirements.

Flying Capacitor Converters, manufactured by companies such as ABB, Siemens, and Fuji Electric, are also utilized in medium-power WECS setups up to 5 MW, offering advantages in terms of efficiency and performance. However, this type of power converter come with several disadvantages in the WECS industry. These include complex control requirements due to the need to manage flying capacitors for voltage balancing, which can increase system cost and complexity. Additionally, despite efforts to balance the flying capacitors, voltage imbalances can occur, leading to potential issues like overvoltage or undervoltage conditions, reducing overall converter efficiency and reliability. These factors make FC converters less favorable in certain WECS applications compared to other converter topologies, contributing to their declining popularity in the industry.

In addition to the converter topologies, voltage and power ratings, an important consideration in industrial power converters is the switching frequency range, which has a significant impact on performance, efficiency, and reliability. The switching frequency of industrial power converters in WECS varies depending on the converter topology, power rating, and specific application requirements. NPC converters typically operate within the range of 1 kHz to 3 kHz for high-power applications, as higher frequencies can lead to significant switching losses, reducing efficiency. CHB converters generally work within the 1 kHz to 5 kHz range, particularly for medium-power applications, balancing efficiency and harmonic reduction [26]. Two-Level Converters which are more common in lower power setups, can operate at slightly higher frequencies, typically between 2 kHz and 6 kHz, due to their simpler design and lower power demands. Matrix converters, which eliminate the need for bulky DC-link capacitors, often operate at higher switching frequencies, ranging from 5 kHz to 20 kHz, making them suitable for compact, high-efficiency applications. Meanwhile, Flying Capacitor Converters usually switch within the 1 kHz to 3 kHz range, similar to NPC converters, due to the need for voltage balancing across capacitors. Therefore, the choice of switching frequency is crucial, balancing power efficiency, heat management, and the quality of the output waveform [27].

Afterwards, beyond the choice of converter topologies, it is essential to recognize the critical role that control strategies play in the effective operation of these systems. Control strategies are pivotal in ensuring efficient power conversion, grid integration, and the optimization of energy

output. They determine the system's response to varying wind conditions and grid requirements, affecting its overall performance and reliability. In the next section, we will provide an in-depth review of available, commonly-used, and modern control strategies employed in conjunction with the mentioned converters. This analysis will shed light on the importance of control strategies in the context of WECS, offering valuable insights for engineers and researchers striving to enhance the efficiency and effectiveness of wind energy technologies.

1.6 Control Techniques

Power converter control strategies play vital roles in ensuring efficient and reliable operation of power conversion systems. The control of a power converter is crucial for maintaining stable voltage and current levels, regulating power flow, and protecting the converter and connected devices from faults and overloads. Additionally, the control strategy determines the converter's response to dynamic changes in the input or output conditions, such as load variations or disturbances in the power supply. In the following section, various applicable control strategies for power converters will be discussed in detail, highlighting their benefits and disadvantages.

1.6.1 Linear Control

A modulation stage is almost always incorporated in the linear controllers of power converters. This modulator device linearizes the nonlinear converter control approach by providing control signals for the switches. The most well-known linear controller is the proportional-integral (PI) controller. To generate a pulse width modulation (PWM) signal for the switches, a reference of a sinusoidal signal is compared to a triangular carrier one in this approach. When the carrier value is less than the reference signals, for example, the switch state is altered to increase the output signal, and vice versa. Other modulation approaches, such as space vector modulation (SVM) and selective harmonic elimination (SHE), are also accessible in the literature. A linear controller for current regulation of the system has been provided in Fig. 1.18 using SVM/PWM, where the reference load currents are compared with the measured one, and the error identified between them is processed by typical PI modulators.

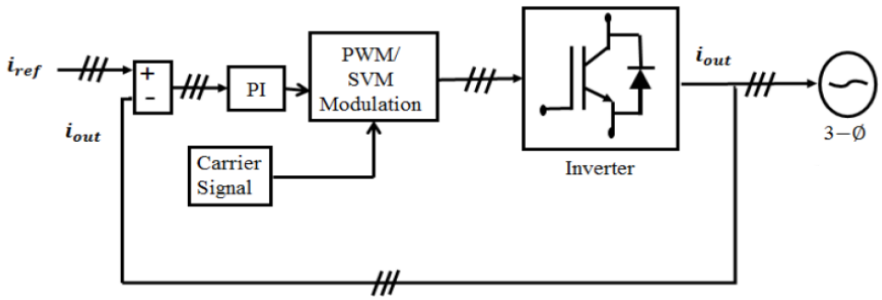


Fig. 1.18: PI controller based linear current controller using PWM/SVM for a three-phase load.

Linear controllers are widely used in the fields of industrial drives and power generation systems, such as; field-oriented control (FOC) which is used to manage the flux and torque of the generator in the machine-side converter. On the other hand, PI controllers in voltage-oriented control (VOC) is used to control the grid-side inverter in a similar way.

The linear control approach for NPC power converters employed in WECS offers several advantages and disadvantages. On the positive side, linear control strategies are well-established and relatively easy to implement. They provide stable and predictable system responses under varying wind conditions, ensuring consistent power generation. Linear control is also effective in grid synchronization, allowing WECS to seamlessly integrate with the electrical grid. However, it does have its drawbacks. Linear control methods may not fully exploit the converter's capabilities, potentially leading to suboptimal energy capture and reduced efficiency, particularly in the presence of rapid wind fluctuations. To address these limitations, modern control strategies like model predictive control (MPC) and hysteresis control have been developed, aiming to enhance the performance of power converters in WECS by providing more sophisticated and adaptive control. The choice between linear and modern control approaches depends on specific project requirements, emphasizing the importance of a nuanced evaluation when designing WECS.

1.6.2 Hysteresis Control

As illustrated in Fig. 1.19, a hysteresis error threshold is used in hysteresis control to determine switching operating modes by comparing the measured quantity with the reference signal [28]. When the regulated current reaches the limit, the switch status is altered. The scheme's implementation is as straightforward as the current control. Direct torque control (DTC) [29] and [30] are two examples of higher-level multidimensional nature applications. When using a digital

control technique, a substantial inspection recurrence is required to maintain the controlled parameters constantly within the hysteresis range.

Hysteresis control is a well-established control approach for power converters in WECS, offering its own set of pros and cons. One of the key advantages of hysteresis control is its simplicity and robustness. It provides fast and accurate control, making it effective for rapid switching in power converters. However, hysteresis control has its limitations. It can lead to increased switching losses and higher stress on converter components due to frequent switching, potentially reducing the converter's overall efficiency and longevity. Additionally, it may generate higher harmonic distortions, which could impact the quality of power injected into the grid. In contrast to more advanced control strategies, hysteresis control may not fully optimize the converter's performance in scenarios where precise control and efficiency are paramount.

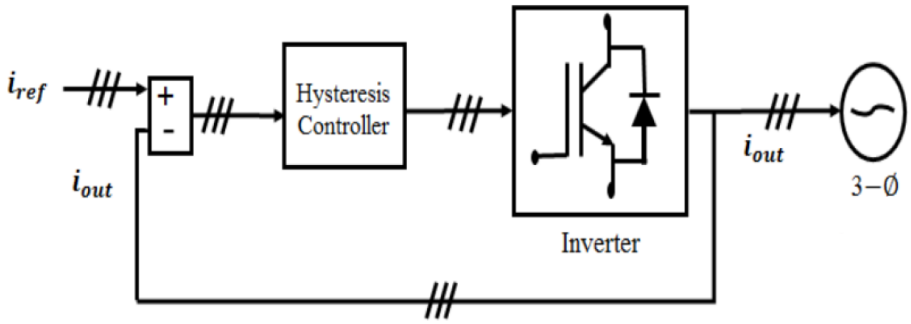


Fig. 1.19: A three-phase hysteresis current control for a three-phase system.

1.6.3 Sliding Mode Control

The sliding mode control (SMC) methodology is a propulsive converter control method which belongs to variable structure control and flexible adaptive control families. This non-linear control approach is nicely related to both linear and nonlinear architectures. Figure 1.20 shows a sliding mode control approach in addition to the PWM/SVM block. Load voltage references are generated by the controller. The control variable is forced to follow or slide in the pre-determined direction [31]. This strategy involves changing the structure of the controller over time in order to provide a strong and stable functioning in the face of parameter and load variations [32].

The sliding mode control approach for power converter presents distinct advantages and disadvantages. One of its key advantages is its robustness in the face of uncertainties and disturbances. Sliding mode control excels in tracking reference signals, even when the wind conditions are highly variable. It offers excellent dynamic response and disturbance rejection

capabilities, making it suitable for enhancing the overall system's reliability and grid integration. However, there are notable drawbacks to consider. Sliding mode control can generate high-frequency switching, leading to increased switching losses and potential stress on the converter components, which may reduce overall efficiency and longevity. Furthermore, it may exhibit chattering behavior, characterized by rapid switching between control modes, which can lead to increased electrical and mechanical wear and tear. While it offers compelling benefits in terms of robustness, sliding mode control may not be the ideal choice when minimizing switching losses and harmonics is crucial.

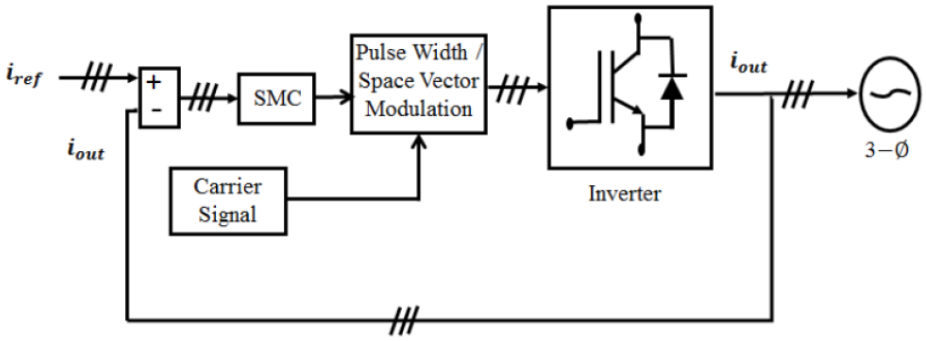


Fig. 1.20: A sliding mode control strategy alongside the PWM/SVM for a three-phase load.

1.6.4 Model Predictive Control

Model predictive controllers, as illustrated in Fig. 1.21, allow for the use of any algorithm which employs a model to predict behavioral patterns and selects the most appropriate control operation based on a cost/loss function. Predictive control necessitates a considerable number of computations than traditional controllers. Because high-speed microprocessors are readily available, this large number of computations may be completed in a shorter period of time, of course [34]. MPC is an optimization approach where a cost function is optimized over a pre-determined time horizon while taking into account system limitations and the model [35].

Model predictive control of power converters has attracted many researchers. Authors in [36] developed a ubiquitous and moderate-complexity MPC technique for DFIG-based wind energy conversion systems, which could perform successfully in both balanced and unbalanced circumstances. To regulate the three-level boost converter and grid-side NPC inverter, a model predictive procedure was provided in [37]. The predictions were assessed using two autonomous

cost functions, besides diminishing the switching states. The switching signals, then, directly applied to the machine (MSC) and grid side converters (GSC). Since the MPC controller needs to investigate extensively for all potential switching signals to obtain the appropriate switching signal in each controller time span, it had a high computing cost. In [38], a MPC was employed on both sides of a back-to-back NPC converter, which linked a permanent magnet synchronous based WECS to the grid. The predictive control technique, which took into account the redundancy of the switching states of the NPC, was applied to accomplish DC-link neutral-point balancing. With the predictive control, the count of switching commutations was lessened, the low-voltage ride-through criterion was met, and the current spectrum was controlled, as well. However, the required computations were still high. In [39], an effective model predictive control approach for torque and power regulation of a back-to-back NPC converters employed in WECS with PMSG was presented, based on hexagon and triangle candidate regions. The variety of feasible switching states was dramatically decreased owing to proper candidate region selection, which diminished calculation time. Reference [40] provided the predictive voltage source converter model with a cascaded control structure. A mathematical technique was used to compute the sensitivity coefficients for power injections and slack bus voltage. A rigorous finite-control set MPC approach with modified predictions for a PMSG based WECS with a three-level back-to-back NPC power converter system was suggested and verified in [41]. The technique considered the power converter nonlinearity and switching nature, combining the control objective optimization and switching state determination procedures into a single computation phase. Not only was the system resilience against parameter fluctuations increased, the control variable ripples were also significantly decreased. The MPC, however, has a number of shortcomings. The downside of MPC is the complicated algorithm, which necessitates a proper process model and a high number of model coefficients. Furthermore, if there is an actual disturbance at the plant input, it may not work well, no matter how much it is improved. Moreover, frequency variation is another main disadvantage of MPC.

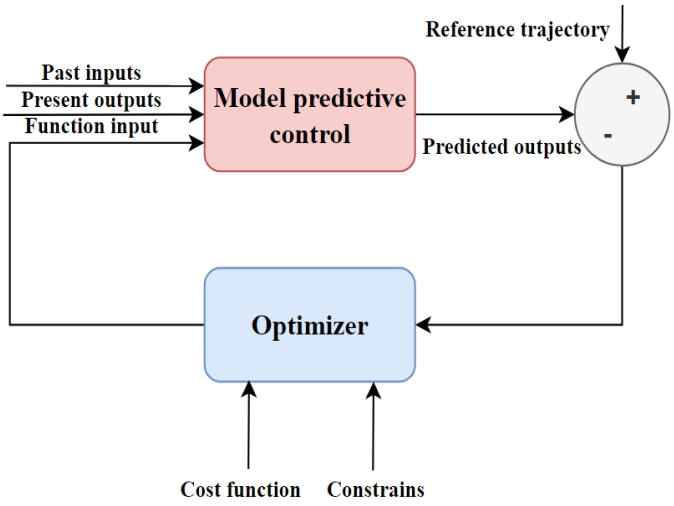


Fig. 1.21: Fundamental concept of predictive control.

1.6.5 Artificial Intelligence Based Control Approaches

The shift toward using AI-based control techniques for power converters in WECS is driven by the need for more efficient, adaptive, and robust solutions to manage the variability and complexity of wind energy. Traditional control methods, as mentioned, often struggle to optimize performance under changing environmental conditions and system uncertainties. AI-based approaches offer the ability to learn and adapt to real-time data, enhancing the ability to predict and respond to the situations. Machine learning [42], artificial neural networks (ANN) [41], genetic algorithms (GA) [43], and fuzzy logic controllers (FL), reinforcement learning (RL), etc., have all been grouped together as intelligent control approaches [44]. Fig. 1.22 depicts an FLC technique in which the PI controller is replaced by the FL. The FL input is the derivative and error of the reference current. For membership functions, this controller adds the converter administrator/experience, planner's learning, and instinct. Because power converters are non-linear in essence, using the FL without understanding the suitable converter model might improve the system's strength when parameter variations occur [45]. Fig. 1.23, on the other hand, depicts the ANN-based controller. Through an acceptable gain (K), the load current tracking error is sent into the ANN, and the ANN generates changing signals for power converter. A constant switching frequency may be accomplished using this manner. However, their implementation can be challenging due to their inherent complexity. Designing and training ANNs requires specialized knowledge in both control systems and machine learning, which can increase development time

and costs. Additionally, ANNs require large amounts of training data to learn the control behavior effectively, and ensuring their generalization to new, unseen scenarios can be difficult. ANNs are also considered "black-box" models, making their decision-making process less interpretable, which can be a drawback in safety-critical applications. ANNs offer great potential for enhancing the control of power converters in WECS, but careful consideration and expertise are necessary to mitigate their disadvantages and ensure their effective implementation. Although, the advantages of FLC and ANN may be combined as shown in Fig. 1.24 to improve control performance and mitigate above-mentioned shortcomings.

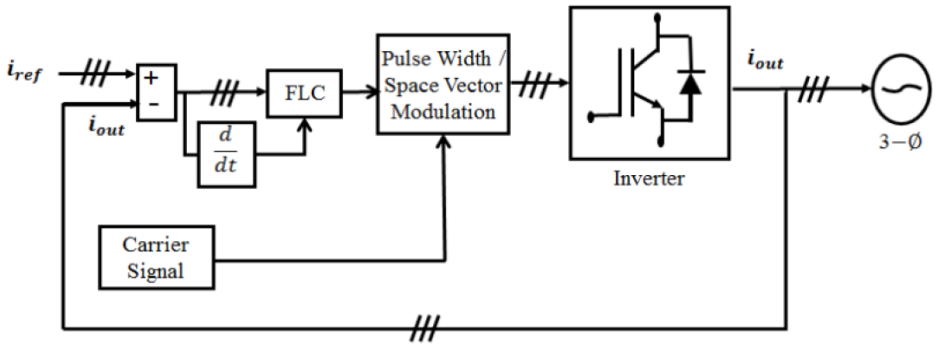


Fig. 1.22: FLC strategy utilizing pulse width/space vector modulation for a three-phase system.

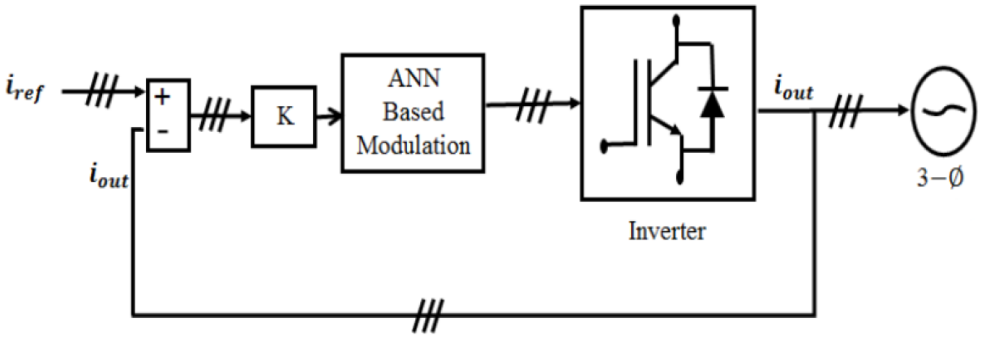


Fig. 1.23: An ANN based controller for a three-phase system.

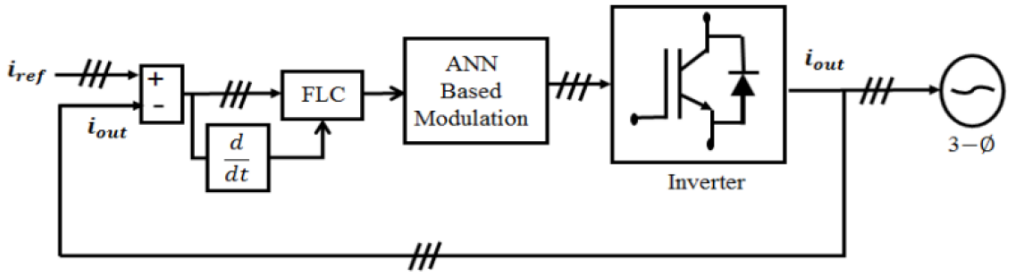


Fig. 1.24: A Neuro-fuzzy controller for a three-phase system.

Learning-based control approaches represent a pivotal and highly influential component of AI-based methods. In the realm of control systems and automation, these approaches have gained significant prominence for their adaptability and capacity to address complex and dynamic real-world scenarios. By leveraging machine learning (ML) or RL techniques, learning-based control methods enable systems to acquire knowledge from data and adapt their behavior accordingly, often surpassing the performance of traditional control strategies. These approaches are particularly beneficial when dealing with systems that are challenging to model accurately or when encountering varying environmental conditions. By continuously learning and evolving, they empower systems to optimize their control strategies, making them invaluable tools in the pursuit of enhanced efficiency, autonomy, and adaptability in a wide array of applications across industries. Consequently, learning-based control approaches stand as a leading frontier in the ongoing integration of AI into control systems, offering remarkable potential for the future of automation and robotics [46]. In this regard, power electronic systems may be imbued with consciousness by adopting ML, and thus the system's autonomy would be strengthened. Reference [47] offered a MPC trained methodology for modular multilevel converter based on a ML and ANN models. The converter was controlled by a highly accurate artificial neural network model. So, data from the standard rapid MPC approach was collected to train a ML algorithm. The suggested ML-based method had the same performance as a traditional MPC, but in a much more computation-effective way. Though, ML solutions were mostly conducted as tools for condition monitoring and fault diagnostic of converters in researches [48], [49], or control of a single-phase power converter, not three-phase ones [50].

Overall, Table 1.4 lists some of the pros and downsides of the mentioned strategies, which are popular for power converter employed in WECS, to summarized what have been discussed.

Table 1.4: Compares the benefits and downsides of the control strategies.

Strategy	Advantages	Disadvantages	Ref.
Linear	• It is less expensive. • Simple and straightforward to implement. • Less complicated	• Switching losses are higher. • Lower order harmonics are present. • Power quality is poor. • Less reliable. • Error in the steady state is high.	[51]
Hysteresis	• Simple to put into practice. • There's no need for complex technology. • There is no need for a modulator.	• Switching losses make it less efficient for low-power purposes. • The width of the hysteresis band, load factors, non-linearity, and working circumstances all affect the switching frequency. • It necessitates the use of costly filters to eliminate the spectral component.	[52]
Sliding Mode	• Offers a fast response. • During load disturbances and parameter changes, it remains steady.	• It's complicated in implementation. • High-frequency oscillations are generated. • Unable to cope with unforeseen uncertainty. • The choice of sliding surface has an effect on performance.	[53]
Model Predictive	• Allows for a faster dynamic reaction. • The design is more straightforward. • Increased tracking precision. • It is possible to include non-linear characteristics and constraint. • Less sensitive to the system model.	• Sophistication of computation. • Switching frequency can be varied. • High maintenance cost. • Lack of flexibility	[54]
Artificial Intelligence based	• can adapt to changing environmental conditions. • Get rid of human error. • Costs of computation and time are being reduced by means of system modeling elimination. • Understanding of high-Dimensional Data.	• It is necessary to have enough training data for training. • may not always generalize well to unseen conditions. • AI controllers can be vulnerable to cyber-attacks if not properly secured.	[54]

Based on the comprehensive investigation and the information presented in the provided table, it is evident that AI-based techniques offer a compelling and promising approach for controlling power converters in WECS and can be a proper candidate for the counterparts. Given the focus of this work on RL and ML based approaches for controlling the NPC power converters within WECS, these methodologies will be subjected to thorough investigation in the upcoming chapters. RL and ML bring about a paradigm shift in the way control strategies are conceived and implemented, offering the potential for greater adaptability, improved performance, and robustness in the face of uncertain environmental conditions and system dynamics. By delving deep into these cutting-edge techniques, this study aims to provide valuable insights into their application and effectiveness within the context of WECS. Understanding and harnessing the power of RL and ML within the control of NPC converters is pivotal to unlocking the full potential of wind energy systems, making this research an important contribution to the field of renewable energy and automation, in terms of power quality and maximum power point tracking (MPPT). In Chapter 2, the AI controller's working concept, particularly the ML and RL-based as well as the proposed control strategy, are detailed.

1.7 Motivation of the Thesis Work

Based on the previous discussions, wind energy conversion systems are found to be instrumental, in the global transition toward sustainable energy sources to achieve net zero GHG emission by 2050. Among the critical components of WECS, power converters play a pivotal role in optimizing energy extraction and grid integration. Traditionally, PI-based controllers have been used for control tasks, but these may fall short in managing the dynamic and complex nature of wind energy systems. AI-based control strategies hold immense potential to revolutionize the control of power converters for power quality at the grid side and maximum power point tracking at the machine side in a three-phase system. This part seeks to provide an overview of the motivations and contributions of applying the cutting-edge methodology to enhance the performance and efficiency of WECS, which will be investigated in this research.

This study is structured into three distinct but interrelated parts, each addressing critical aspects of WECS. The first part focuses on the control of the grid side NPC power converter with a primary objective of enhancing power quality. Grid side control strategies are essential for ensuring that the power supplied to the electrical grid is stable and high quality, ultimately

contributing to grid stability and the seamless integration of renewable energy. The second part of the study delves into the control of the machine side NPC power converter, with a specific aim to implement advanced MPPT strategies. MPPT is pivotal in wind energy systems, as it enables wind turbines to capture the maximum possible energy from varying wind conditions. Both objectives address critical control challenges in WECS, collectively contributing to the efficient and reliable operation of wind energy systems in the pursuit of sustainable and clean energy production. Finally, the micro-grid (MG) power management will be addressed, in this thesis work. In MGs utilizing wind energy systems and other uncertain technologies, effective power management and battery energy storage system (BESS) scheduling are essential, instantly, due to the inherent output fluctuations associated with these sources. This is not only a must, but also it enhances grid stability and reliability and maximizes the utilization of renewable energy, reduces dependency on fossil fuels, in order to minimize the fuel cost.

As the motivation for the grid Side NPC Control to maintain the power quality, the followings are considered:

1. **Challenges in Conventional Control**: Traditional control methods, such as PI controllers, often struggle to handle the dynamic and non-linear behavior of wind turbines and grid-connected systems, which are essential for maintaining power quality. AI-based control approaches can offer the capability to adapt to varying grid conditions resulting in improved grid-side power quality.
2. **Limited Application of AI in wind industry**: An analysis of the literature reveals that AI, specifically RL-based controller in the context of power converters has received limited attention, with most studies concentrating on AI models applied to single-phase or DC/DC systems. This motivates the author to explore the application of RL in controlling three-phase NPC power converters.
3. **Detailed system modeling requirement:** Conventional PI controllers in WECS demand detailed and accurate system modeling that is often complex and difficult to achieve. In contrast, RL-based controllers significantly reduce the need for intricate system modeling by learning optimal control strategies through interactions with the environment. This capability of RL-based approaches not only simplifies the control design process but also can enhance the system THD performance.

The impact of this research on power converter power quality are summarized, as follows:

1. **Eliminating PI Controller Tuning**: With the elimination of PI controllers, the need for parameter tuning and bandwidth adjustments becomes obsolete. This simplifies the control process and reduces maintenance complexities.
2. **Mitigation of Power Quality Issues**: This study seeks to significantly reduce power quality issues by effectively managing voltage sags, swells, and harmonic distortions. This contribution is vital for preserving the integrity of the electrical grid and minimizing the risk of equipment damage.
3. **Considering Unbalance Power System Condition**: The study contributes to solving unbalanced condition problems, single-phase and double-phase voltage drop, that other AI solution models might not work properly.
4. **Innovative Application**: The research introduces MARL to control power converters. MARL extends the capabilities of RL by allowing multiple agents to interact, making it well-suited for handling complex, multi-dimensional scenarios. Ans it is also an innovative application of MARL in wind energy systems represents a pioneering contribution to the field of control strategies.
5. **Optimal MARL hyperparameter tuning:** a novel approach is proposed to optimize the MARL hyperparameter known as the discount factor (DF), which significantly enhances the performance and response of the system compared to traditional tuning methods which often rely on manual adjustments.

As the motivation for machine side NPC power converter control for MPPT purpose, the followings are considered:

1. **Dependency on Accurate System Modeling:** Techniques like Perturb and Observe (P&O) and Incremental Conductance (IC) require precise modeling of the wind turbine and power electronics to function effectively. Inaccurate models can lead to suboptimal performance and incorrect MPP tracking.
2. **Slow Dynamic Response:** Traditional methods often exhibit slow response times to rapidly changing wind conditions, which can result in significant energy losses. The lag in adapting to new maximum power points reduces the overall efficiency of the WECS.

3. **Oscillations Around MPP:** Many conventional MPPT algorithms suffer from oscillations around the maximum power point, particularly under steady-state conditions. These oscillations can cause wear and tear on the system components and reduce the lifespan of the turbine and converters.

The impact of this research on highly-efficient MPPT are summarized, as follows:

1. **Improved MPPT:** The development of an advanced MPPT algorithm enhance the energy capture capabilities of wind turbines, ensuring that the system operates at their peak efficiency under varying wind conditions, and provide fast response to the wind speed changes.
2. **Novelty in Multi-Agent RL for MPPT:** The work extends the novelty by applying multi-agent RL not only to power converter control, but also to MPPT purposes, further expanding the horizon of research in this domain. Also, a highly-efficient approach is proposed to optimize the MARL discount factor hyperparameter, which significantly enhances the performance and response of the system compared to methods rely on manual adjustments.
3. **Lab Implementation of MPPT:** As an additional contribution, the study emphasizes the practicality of the proposed MARL ta tackle MPPT by implementing it in a laboratory environment. This validation helps bridge the gap between theoretical RL-based model and real-world applications in power systems.

The first two objectives of this thesis focus on enhancing power quality and maximizing energy extraction in WECS through advanced NPC converter control and MPPT techniques. These improvements set a strong foundation for the third objective, which involves effective power management in a microgrid comprising wind turbines, photovoltaic systems, and combined cooling, heat, and power units. Given the inherent uncertainties associated with WT power generation, robust power management is crucial. Hence, the research motivation for power management and energy scheduling of a MG, are as follows:

1. **Optimizing Renewable Integration:** With the increasing emphasis on renewable energy integration into the MG, managing the intermittent nature of sources like solar and wind becomes critical.
2. **Economic Benefits:** Effective power management can result in significant economic benefits. By optimizing energy use and storage systems, the operational costs of running a microgrid can be substantially lowered. Additionally, improved efficiency and reliability can enhance the economic attractiveness of microgrid projects, encouraging further investments and adoption.
3. **CO_2 emissions concern:** One of the primary motivations for this study is the significant potential of microgrids to reduce CO_2 emissions, by an optimal power management strategy. By integrating renewable energy sources such as wind turbines and solar panels, as well as the properly-managed units' energy generation, microgrids can diminish fossil fuel burning, directly lowering carbon footprints.

The impact of the research on this environmentally-friendly power management and BESS energy scheduling are summarized, as follows:

1. **Multi-Objective Optimization in A Comprehensive Microgrids:** A novel MARL approach is proposed for power management in a microgrid with wind turbines, photovoltaic panels, and Combined Cooling, Heating, and Power (CCHP) units, including chillers, micro-turbines, etc., aiming to optimize a multi-objective function to reduce both fuel consumption and CO2 emissions, considering the penalty factors for PV and WT generations.
2. **Optimizing Agent Number for Efficient MG energy scheduling:** Determined the optimal number of agents in the MARL system, a previously unexplored aspect, ensuring the right balance between complexity and effectiveness in managing the MG, leading to improved performance and efficiency**.**
3. **Solving fluctuations in the training phase:** The discrepancy in co-player policies is an acute issue as it makes restoring them intrinsically challenging and causes undesired fluctuations in the training phase. The matter is resolved here employing the Q-based

Bellman equation. The Q-based Bellman equation helps resolve this issue by explicitly modeling the interactions and dependencies between agents.

1.8 Objective of the Thesis

This thesis aims to address the existing challenges in power converter control strategy, MPPT, and power management within MGs containing wind energy systems by proposing a novel MARL method. The main objectives of this research are as follows:

1. **Improving Power Quality of WECS**:
    - **Develop a ML approach:** Design and develop an improved ML approach to control grid-side power converters employed in a direct-drive PMSG-based WECS for ensured power quality.
    - **Develop an Advanced MARL Approach**: Design and implement a MARL-based control strategy for grid-side and machine-side power converters in a direct-drive PMSG-based WECS. This approach aims to enhance the power quality by effectively managing voltage sags, swells, and reduction in total harmonic distortions.
    - **Reduce System Modeling Complexity**: By utilizing MARL, minimize the need for detailed and complex system modeling required by traditional PI controllers, thus simplifying the control design process.
2. **Enhancing MPPT Efficiency**:
    - **Develop an Advanced MPPT Algorithm**: Create a highly-efficient MPPT algorithm based on MARL to maximize energy extraction from wind turbines. This algorithm ensures rapid adaptation to varying wind conditions and reduce oscillations around the maximum power point.
    - **Optimize MARL Hyperparameters**: Implement a meta-learning approach to optimally adjust the MARL discount factor, significantly enhancing the system's response time and overall efficiency.

- **Practical Implementation and Validation**: Validate the proposed MARL-based MPPT algorithm in a laboratory environment to bridge the gap between theoretical models and real-world applications, ensuring practical feasibility and robustness.

3. **Optimizing Power Management in Microgrids**:
    - **Develop Accurate Power Predictions**: Utilize a recurrent multi-layer ANN to accurately predict the output power of WT and PV systems over a 24-hour period, incorporating a stochastic gradient-based optimizer to minimize the RMSE cost function.
    - **Model a Comprehensive Microgrid and develop Multi-Objective Cost Function**: Design and mathematically model a microgrid (MG) that includes WT, PV, CCHP units, micro-turbines, and BESS. Formulate the system's constraints and equality equations to balance loads and generations effectively. Also, develop a multi-objective cost function considering fuel costs, CO2 emission costs, and penalty factors for WT and PV due to their inherent uncertainties.
    - **Employ MARL for Optimal Management**: Apply the proposed MARL method to manage power among generation units and optimize BESS scheduling within a 24-hour period. Focus on achieving cost-effective and efficient power management.
    - **Determine Optimal Agent Configuration and DF value**: Identify the optimal number of agents in the MARL system, as well as the DF value to balance complexity and effectiveness, ensuring improved performance and efficiency in managing the microgrid.

1.9 Thesis Organization

In this thesis, the subsequent chapters are meticulously structured to address the key components of AI-based control strategies and their applications in wind energy systems. Chapter-2 delves into the fundamentals of AI-based control strategies, with a specific focus on the proposed MARL method. This chapter presents formulation of the MARL approach, detailing its innovative aspects and theoretical underpinnings. Chapter-3 is dedicated to controlling the grid-side NPC power converter using the developed MARL method. And validates its applicability and superiority over conventional methods through extensive simulation results. In Chapter-4, the focus shifts to the MPPT problem. This chapter presents a series of extensive scenarios and simulation results that substantiate the efficacy of the proposed method. Additionally, it includes the implementation of the MARL-based MPPT on a laboratory setup, providing empirical evidence of its practical viability and effectiveness. Afterwards, Chapter-5 addresses the power management problem, encompassing the modeling, development of mathematical equations, and constraints of the proposed microgrid. This chapter also presents the results of energy scheduling, supported by detailed discussions on the outcomes and their implications. Finally, Chapter-6 provides a succinct conclusion of the research findings and suggests directions for future work. This chapter encapsulates the contributions of the thesis, highlighting the advancements made in the field of wind energy systems through the innovative use of MARL-based control strategies. It also identifies potential areas for further research, ensuring that the groundwork laid by this thesis can be built upon by future studies.

Chapter 2
Literature Review, and Development of the Proposed Power Converter Control Strategy

2.1 Introduction

Power converters are fundamental components in electrical systems, responsible for converting and regulating electrical energy across a wide spectrum of applications. Traditionally, the control of power converters has been grounded in heuristic and model-based methods, which, while effective, come with their inherent limitations. With the emergence of artificial intelligence, particularly machine learning and reinforcement learning methods, the landscape of power converter control is undergoing a transformative shift. These AI-based approaches offer numerous advantages, ranging from enhanced efficiency and adaptability to the potential for autonomous optimization. However, it is worth noting that, despite the immense potential, relatively less attention has been directed toward applying AI-based control strategies to power converters. While AI has made substantial inroads in various domains, the application of AI to power converters has primarily centered around single-phase DC/DC converters. This has garnered noteworthy success in improving control performance, efficiency, and adaptability. Yet, the vast domain of three-phase AC converters, which play a critical role in industrial and commercial electrical systems, remains relatively unexplored, especially in WECS research area. The advantages that AI-based control strategies offer in the context of single-phase converters set the stage for a compelling exploration of their potential in three-phase AC converters.

This chapter serves as an introduction to the use of AI, specifically ML and RL. Moreover, the chapter draws attention to the relative scarcity of research and implementation in AI-based control of power converters, particularly in the context of three-phase AC converters, employed in WECSs. By bridging this gap in research, it is aimed to shed light on the untapped potential and opportunities for enhancing power converter control in various applications. So, the proposed strategy will be detailed here, and will be served for WECS power quality and MPPT purposes, in the following chapters.

2.2 Machine Learning Models

Without an explicit prior programming task, AI algorithms can learn directly from data. The motif omits the complexity of the system modeling (here, WECS). Machine learning, with its capacity to learn from data, is particularly well suited to issues that are either too complicated to properly characterize or whose description is inaccurate. Several ML algorithms can be applied to the diverse problems which originate in data networks, such as; regression, logistic regression, polynomial logistic regression [55], support vector machine [56], K-nearest neighbors [57], K-Means [58], random forest [59], gradient boosting machine [60], multilayer perceptron [61], Naïve Bayes [62], and many others. These algorithms can be sub-divided into four main categories:

Supervised learning: The ML endeavor of learning a function that maps inputs to an output according to the input-output pairs is known as supervised learning. It utilizes labeled training data and a collection of training examples to infer a function. Each sample in supervised learning is a pair of an input and an output usually known as the target. A supervised learning algorithm evaluates the training dataset and generates a highly-accurate estimated function that can be applied to new unseen cases. The algorithm must be able to predict a correct output for the mentioned unseen data. It is noteworthy that based on the output type, the supervised learning would be either regression or classification problems [63]. As one of the WECS application for this type of ML-based solution, [64] introduced a multivariate estimation for the power curve of the wind turbine considering the weather conditions such as wind speed, air density, wind turbulence, and wind share. Or, [65] compared various supervised techniques in a data-centric approach to wind turbine component failure detection.

Unsupervised learning: Unsupervised Learning is a ML approach in which the model does not require the users' supervision. Alternatively, it empowers the model to operate independently to identify previously unnoticed patterns and trends. It is mostly concerned with unlabeled data. In comparison to supervised learning, unsupervised learning algorithms allow users to execute more sophisticated data processing. Unsupervised learning, on the other hand, might be more unpredictable than other learning approaches. Clustering, and anomaly detection are two examples of such category [66]. In WECS application, [67] carried out the demand analysis of operation mode extraction of high proportion renewable energy power system. Then, an automatic mode extraction algorithm based on K-mean algorithm and improved unsupervised-based cluster validity index was proposed.

Semi-supervised learning: A learning issue with a small number of labeled instances and a large number of unlabeled cases is known as semi-supervised learning. This sort of learning issue is difficult to address since neither supervised nor unsupervised learning algorithms can effectively employ a mixture of labeled-unlabeled data. As a result, specialized semi-supervised learning methods are necessary [68]. In WECS application, [69] proposed a workflow to detect faults based on automatic feature selection, and semi-supervised classification for fault detection.

Amongst mentioned sub-methods the adoption of supervised ML for power converter control in a WECS is underpinned by several compelling reasons:

1. Adaptability: Power converters often operate in dynamic environments where conditions, such as load characteristics and input voltages, can vary substantially. Supervised ML models excel in adapting to such changes. They learn from historical data, automatically adjusting control parameters to maintain optimal performance without the need for manual tuning.

2. Nonlinear System Handling: Power converters and wind generators are inherently nonlinear systems, exhibiting complex and intricate dynamics. Traditional control methods frequently resort to linearization, which can compromise control accuracy, especially in dynamic applications. In contrast, supervised ML models can inherently handle these nonlinearities, capturing intricate system dynamics with precision.

3. Data-Driven Control: Power converters generate substantial data through sensors and feedback mechanisms. Supervised ML leverages this rich data environment, making it particularly well-suited for applications, where real-time information drives decision-making is required. By learning from past data, these models make informed decisions based on the most recent system conditions.

4. Fault Tolerance: Supervised ML models can detect and respond to faults in real-time. This enhances the overall reliability of power converters by identifying anomalies and mitigating potential damage, a vital consideration in applications where system safety is paramount.

These underscore the importance of the supervised learning approach in power converter control strategy employed in a WECS. The following section focuses on the development of supervised machine learning solutions and the underlying mathematics are also provided to support the concept explanation.

2.3 Development of Supervised Machine Learning

In the simplest terms, linear regression is a supervised machine learning model that identifies the best fit linear line seen between independent and dependent variables, i.e., it discovers the linear connection between the two variables. There are two forms of linear regression: simple and multiple. Only one independent variable is included in simple linear regression, and the model must identify a linear connection between the only parameter and the dependent variable. Multiple linear regression, on the other hand which is the type of this study problems, uses more than one independent feature variable to identify a link. Equation (2.1) describes the multiple linear regression. And b_0 is the bias, $w_1, w_2, \dots, w_n$ are coefficients of the independent variables of $x_1, x_2, \dots, x_n$, and $h_w(x)$ is the hypothesis function or predicted target.

$$h_w(x) = b_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n \tag{2.1}$$

In the context of supervised machine learning, nonlinear regression, on the other hand, is used to find a nonlinear function that best fits the data. The general mathematical representation of nonlinear regression can be expressed as follows:

$$h_\omega(x) = f(X, w) + \epsilon \tag{2.2}$$

Where $f(X, w)$ is the nonlinear function that relates the inputs to the target output. And ϵ represents the error term, accounting for the difference between the predicted values and the actual observed values. The goal of nonlinear regression is to determine the best-fitting function $f(X, w)$ and estimate the parameters ω in such a way that the error term ϵ is minimized, indicating a close alignment between the predicted and actual values. The specific form of the non-linear function will depend on the nature of the problem and the chosen model. Common types of nonlinear models include:

Polynomial Regression: This is a type of nonlinear regression where the function $f(X, w)$ takes the form of a polynomial equation. It should be noted that using a high degree of polynomial tries to overfit the data and for smaller values of degree, the model tries to underfit so it is required to find the optimum value of a degree. The best degree of the nonlinear function can be attained by trial and error in the model [70].

$$h_w(x) = b_0 + w_1 x_1 + w_2 x_1^2 + \cdots + w_n x_1^n + w_{n+1} x_2 + \cdots \quad (2.3)$$

Exponential Regression: In this case, the function follows an exponential equation, like:

$$h_w(x) = b_0 + w_1 . e^{w_2 X} \quad (2.4)$$

Logistic Regression: Logistic regression models the probability of a binary outcome, and the function is the logistic function, given as:

$$h_w(x) = \frac{1}{1+e^{-(b_0+wx_1+w_2x_1^2+\cdots+w_nx_1^n)}} \quad (2.5)$$

However, when it comes to the specific domain of power converter control, the choice of the most suitable nonlinear regression model is a critical decision. Given the complex and often nonlinear relationships between input variables and the desired output in power converter control, polynomial regression emerges as a compelling choice. Polynomial regression allows for a more flexible representation of the data, accommodating intricate dynamics. It can capture nonlinear relationships by introducing higher-order terms, offering a better fit for the dynamic behavior of power converters. In this context, the versatility and adaptability of polynomial regression make it a strong candidate for effectively modeling and controlling power converters, enhancing their performance, efficiency, and adaptability.

2.3.1 Regularized Polynomial Locally Weighted Regression

The prediction accuracy of the commonly-used polynomial regression may not be enough. So, locally weighted regression can assist in enhancing the overall performance of polynomial regression algorithms. The primary idea underlying locally weighted regression is to create a local model in which data points in close vicinity of the current query point are assigned more weight over data points further away, as follows:

$$W = e^{-\frac{\left(x^{(j)}-x\right)^2}{2\tau^2}} \quad (2.6)$$

In this equation, the weights are determined by the x point around which the attempt is given to evaluate the regression. The weight $w^{(j)}$ gets near the 1 if $|x^{(j)} - x|$ is low, and it would be low if $|x^{(j)} - x|$ is high. The τ is also known as the bandwidth parameter. In this instance, increasing the τ of the bell shape curve equation will lead to raise of the further points weight. In this problem the bandwidth parameter could be attained by trial and error, as well.

Moreover, developing a ML model is more than just loading data and getting desirable results. There are several shortcomings which thus impact the accuracy of any model. Overfitting in ML is one such defect that degrades the model accuracy. When a model learns the details and noise in the training data to the point that it degrades the model performance on new data, this is referred to overfitting [71]. To solve the issue, many regularization solutions have been proposed by data analysts. Ridge regularization (L1 regularization), and Lasso regularization (L2 regularization) strategies are two well-known regularization methods which are provided to prevent parameters from being too large. Reference [72] proposed a wind speed forecasting method based on regularized ML method, as an effective computing technique to deal with control and operation of WECS. In this regard, overfitting will be restrained. To achieve this, a weighted summation of the parameters (L1) or the squared summation of them (L2), as $R(w)$, will be concatenated to the cost function, which will be defined in following section.

In which:

$$R(\omega) = \begin{cases} \frac{1}{2}\sum_{j=1}^{n} w_j = \|w\| & \rightarrow \quad if: L1 \\ \frac{1}{2}\sum_{j=1}^{n} w_j^2 = \|w\|_2^2 & \rightarrow \quad if: L2 \end{cases} \tag{2.7}$$

2.3.2 Cost Function

As mentioned, the main goal of a regression model is to determine the best-fit line and the appropriate interception and coefficient values so that the error is minimized. To assess the model performance a cost/loss function is vital to demonstrate how well it operates. Several cost functions are represented in literature. The mean squared error (MSE) stands out as the most suitable cost function for controlling power converters in machine learning applications due to its inherent properties and relevance to the task. MSE measures the average squared difference between

predicted and actual values, making it ideal for tasks where the goal is to minimize the difference between the output of the power converter and the desired target. In the context of power converter control, MSE provides a clear, quantitative measure of performance that directly reflects the system's accuracy. Furthermore, MSE is convex, which means it has a single global minimum, making it easier to optimize using gradient-based methods. This property ensures that the optimization process converges reliably to a solution, enhancing the stability and efficiency of the control algorithm. Additionally, MSE is well-suited for continuous and differentiable optimization, aligning with the requirements of many power converter control algorithms that rely on iterative optimization techniques. To eliminate any potential of negative errors, a square of the difference between the actual and forecasted value is computed. The average of the sum of squared discrepancies between forecasts and actual observations is used to calculate the MSE, as follows. For better understanding the equation, the *m* demonstrates the number of data and $y^{(j)}$ is the actual output values.

$$MSE = J(w) = \frac{1}{2m}\sum_{j=1}^{m}\left(h_w\left(x^{(j)}\right) - y^{(j)}\right)^2 \tag{2.8}$$

After applying the local weights and regularization terms to the cost function based on (8), the overall cost function would be (2.9), in which, λ is regularization strength, which balances the minimization importance, between the two terms in the cost function.

$$J(w) = \frac{1}{2}W^{(j)}\sum_{j=1}^{m}\left(h_w\left(x^{(j)}\right) - y^{(j)}\right)^2 + \lambda R(w) \tag{2.9}$$

2.3.3 Cost Function Optimizers

In the realm of supervised machine learning, the optimization of the cost function is a pivotal step. This process involves finding the set of model parameters that minimizes the difference between predicted values and actual observations. Numerous optimization algorithms have been developed for this purpose, each with its unique strengths and characteristics. Some well-known optimizers include stochastic gradient descent (SGD), Adam, RMSprop, and Adagrad, to name a few. While the choice of optimizer may depend on the specific problem and the nature of the data, several optimizers have gained widespread use due to their effectiveness. In the following sections,

we will delve into some of the most commonly used optimization algorithms, providing a detailed exploration of their principles and practical applications in the context of power converter control.

Stochastic Gradient Descent: The iterative approach of SGD is often used to optimize an objective function with sufficient smoothness criteria. Stochastic gradient descent algorithms are a modification of gradient descent, in terms of computational volume. In SGD, the gradient is calculated using just one of the randomly-selected observations instead of all of them. In some cases, this approach can reduce computation time. SGD is commonly used to avoid local optimum and saddle-points, which would otherwise trap deterministic processes. In SGD, the true gradient of $J(\omega)$ is approximated by a gradient at the chosen sample, and the parameters will be updated, as regards [73]. where, α is the hyperparameter of the step size which is obtained by trial and error, in this work.

$$w^{updated} = w^{old} - \alpha(\nabla J(w)) \tag{2.10}$$

Mini-Batch Gradient Descent (MSG): Mini batched updates are a more accurate computing procedure than SGD. As a result in MGD, in each updating iteration, a batch of data is picked instead of only one sample [74]. To update the parameters, here, commonly a batch of 8, 16, 32, 64, 128 or 256 data is involved in each iteration. In this scenario, the model mostly outperforms SGD in terms of learning rate and accuracy. Hence, a higher step size can be used to have lower number of iterations required to achieve the convergent.

$$w^{updated} = w^{old} - \alpha\left(\frac{1}{Batch\ number}\sum_{n=1}^{Batch\ number}\nabla J(w)\right) \tag{2.11}$$

Broyden–Fletcher–Goldfarb–Shanno (BFGS): The BFGS optimizer is basically an approach for addressing nonlinear optimization problems with no constraints. It preconditions the gradient with curvature data to identify the descent direction. It performs by progressively refining an approximation to the loss function. Hessian matrix (H) derived only from gradient assessments using the extended secant approach. It is computationally cost-effective to update the BFGS curvature matrix and it is usually more stable than gradient descent approaches [75]. The iterative process begins with an initial guess for parameter matrix ω_0, and Hessian matrix H_0. Then:

$$H_t P_t = -\nabla J(w_t) \tag{2.12}$$

$$s_t = \alpha_t P_t \tag{2.13}$$

$$w_{t+1} = w_t + s_t \tag{2.14}$$

$$y_t = \nabla J(w_{t+1}) - \nabla J(w_t) \tag{2.15}$$

$$H_{t+1} = H_t + \frac{y_t y_t^T}{y_t^T s_t} - \frac{H_t s_t s_t^T H_t^T}{s_t^T B_t s_t} \tag{2.16}$$

where, P_t represents the search direction, α_t is the stepsize which can be found through a one-dimensional linear optimization, and the s_t illustrates the update step.

Limited-memory BFGS: L-BFGS, such as above-mentioned BFGS, navigates its search across variable space through an estimation of the inverse Hessian matrix, but instead of storing all the problem variables, L-BFGS just maintains a few vectors that reflect the estimation implicitly. In other words, L-BFGS uses the m most recent iterations for Hessian matrix approximation. The L-BFGS approach is particularly well suited for optimization problems with multiple variables due to the linear memory demand [76].

Adagrad: Adagrad is an optimizer frequently used in machine learning, particularly for minimizing the MSE cost function. This algorithm stands out for its adaptive learning rates, which automatically adjust the step size for each parameter during training. The mathematics behind Adagrad involves updating the learning rate for each parameter based on the historical gradient information. Specifically, for a parameter θ, the update rule is as follows:

$$\Delta w_t = (\eta \,/\, sqrt(G_t + \varepsilon)) * \nabla\theta\, L(w_t) \tag{2.17}$$

Here, $\Delta\omega_t$ represents the change to the parameter ω at time step t, η is the learning rate, G_t accumulates the squared gradients of w up to time t, and ε is a small constant added to avoid division by zero. Adagrad is effective in scenarios where some parameters require larger updates than others, as it adapts the learning rates individually, ensuring faster convergence for infrequent features and slower updates for frequent ones, making it a valuable tool for optimizing MSE cost functions in various machine learning tasks.

Adam: Adam (Adaptive Moment Estimation) is a popular optimization algorithm that combines the advantages of both the Adagrad and root mean square propagation algorithms. It's known for its effectiveness in training deep neural networks. Adam uses adaptive learning rates for each parameter and maintains two moving averages of past gradients to adjust these learning rates.

While there are numerous optimization algorithms available, several well-known ones were briefly explained for ML cost function optimization tasks. However, to boost the efficiency of the ML model, apart from the regularizing or choosing proper optimizer, data pre-processing plays a crucial role by ensuring that the data is clean, consistent, and ready for analysis. It involves a series of steps such as cleaning, transforming, and aggregating data to improve the quality and usability for modeling. In the following part, the subject of data pre-processing will be discussed.

2.3.4 Data Pre-Processing

Data preprocessing is a fundamental and indispensable stage in the machine learning pipeline. It involves a series of techniques applied to raw data to make it suitable for analysis and modeling. Proper data preprocessing is critical for several reasons. First and foremost, it helps in cleaning and refining the dataset, ensuring that it is free from missing values, outliers, and errors. By doing so, it reduces the potential for introducing bias or inaccuracies into the model, thereby enhancing the robustness and reliability of the results. Secondly, data preprocessing aids in the transformation of data into a format that can be effectively used by machine learning algorithms. This includes encoding categorical variables into numerical representations, scaling features, and addressing imbalanced datasets. Without these transformations, algorithms may struggle to extract meaningful patterns and relationships within the data. Lastly, data preprocessing often contributes to dimensionality reduction, which can lead to more efficient model training and interpretation, as well as reduced computational overhead. Among the myriad data preprocessing methods, the most well-known ones are:

Feature Scaling: Feature scaling is essential because it ensures that features are on a similar scale, preventing some from dominating others and skewing the learning process of many machine learning algorithms that rely on distance measures. The most common techniques for feature scaling include Min-Max scaling and Standardization (Z-score scaling).

$$x_i^{(j)} = \frac{\left(x_i^{(j)} - \mu_i^{(j)}\right)}{\sigma_i^{(j)}} \tag{2.18}$$

wherein, $x_i^{(j)}$ is the i^{th} input feature of the j^{th} sample data, $\mu_i^{(j)}$ is mean, and $\sigma_i^{(j)}$ is the standard deviation, respectively. Relying on the scaled labeled data the model will be trained, and a regression algorithm seeks to attribute inputs into an output in lower iterations.

Feature Encoding: Categorical data, cannot be directly used in many machine learning algorithms. Feature encoding is vital for converting these categories into a numerical format that models can understand. Two common methods are One-Hot Encoding and Label Encoding.

Handling Missing Data: Missing data is a common issue in real-world datasets and can lead to erroneous results. Proper handling ensures that the model's training and predictions are not compromised. For imputing missing values, common approaches include mean imputation which is replace missing values with the mean of the column.

These three data preprocessing methods are foundational in ensuring data quality and compatibility for machine learning tasks. Feature scaling normalizes the data to a consistent range, feature encoding translates categorical data, and handling missing data maintains the integrity of the dataset. Their effective application enhances the accuracy and generalizability of machine learning models and is pivotal in the quest for meaningful insights and reliable predictions.

2.4 Limitations of ML technique for Power Convert Control

Machine learning solutions for controlling power converters in WECS face several significant challenges and limitations. One of the primary issues is the complexity of the WECS environment, which includes highly nonlinear dynamics, uncertainties in wind speed and direction, and varying load demands. These complexities make it challenging to develop accurate ML models that can effectively control the power converters under all operating conditions. Additionally, the performance of ML models heavily depends on the quality and quantity of training data, which can be limited or difficult to obtain for WECS, especially for rare or extreme operating conditions.

Another limitation of ML technique is their lack of interpretability, particularly in deep learning models. This lack of transparency can be problematic in critical applications such as

WECS control, where understanding the reasoning behind control decisions is essential for ensuring safe and reliable operation. Moreover, ML models may suffer from the problem of overfitting, where they perform well on the training data but fail to generalize to unseen data, leading to poor performance in real-world applications. These would probably disable the ML model to tackle with unbalance and faulty conditions in wind energy systems.

Reinforcement learning offers a promising alternative to traditional ML approaches for power converter control in WECS. RL algorithms, such as Q-learning and Deep Q-Networks (DQN), learn optimal control policies through trial and error, without requiring explicit models of the system dynamics. This makes RL well-suited for controlling complex systems like WECS, where the underlying dynamics are not fully understood or are difficult to model accurately. Furthermore, RL algorithms can adapt to changing environmental conditions and learn optimal control strategies over time, potentially outperforming traditional ML approaches in terms of control performance and efficiency.

Due to the afore-mentioned limitations of ML method for power converter control, it is required to move toward the principles of RL and explore how it can be applied to address these issues. Also, the advantages of RL in learning optimal control policies in complex and uncertain environments will be explained, and proposed novel RL-based method for controlling power converters in WECS for this study will be developed, step by step. This method will aim to overcome the challenges faced by traditional ML approaches and provide a more effective and efficient solution for controlling power converters in wind energy systems for power quality and MPPT purposes.

2.5 Reinforcement learning

Reinforcement Learning is a fascinating subfield of AI learning that centers on training agents to make sequential decisions in an environment. Unlike supervised learning, where labeled data guides the model, and unsupervised learning, which uncovers patterns in unlabeled data, RL relies on trial-and-error learning. Reinforcement Learning is based on the idea of learning by interacting with an environment. An RL agent takes actions in the environment, and in return, the environment provides feedback in the form of rewards or punishments. The agent's objective is to learn a policy, a strategy that maximizes cumulative rewards over time. This process is analogous to training a pet where you reinforce desirable behaviors with treats. Key Components of RL are as follows:

1. Agent: The agent is the core decision-maker in the RL framework. It can be a software agent or a physical entity (e.g., a robot). The agent's role is to interact with the environment, make decisions, and learn from the consequences of those decisions. The agent's goal is typically to maximize the cumulative reward it receives over time by selecting actions that lead to desirable outcomes. The agent in a WECS could be a control system or algorithm that determines the optimal control actions for the system.

2. Environment: The environment is the external system that the agent interacts with. It encompasses everything outside the agent itself that can influence or be influenced by the agent's actions. The environment is usually characterized by its states, actions, transition dynamics, and reward structure. In RL, the environment is the entity that provides feedback to the agent in the form of rewards or punishments.

3. State (s): A state represents the current situation or configuration of the environment at a specific time. States capture all relevant information needed to make decisions in the context of the RL problem. States can be as simple as the position of a chessboard or as complex as the sensory data from a self-driving car. They serve as input to the agent's decision-making process and influence its choice of actions. States in a WECS problem could include wind speed, turbine speed, power output, blade pitch angle, and other relevant parameters that describe the current operating conditions of the wind turbine system

4. Action (a): An action represents the choices made by the agent to interact with the environment. Actions can be discrete, or continuous. The agent selects actions from a predefined set of possibilities. The chosen action affects the state of the environment and can lead to different consequences, including receiving rewards or penalties. An action in a WECS problem corresponds to the control actions which the agent can take to influence the system's behavior, such as adjusting the rotor speed, blade pitch angle or the generator torque.

5. Policy (π): The policy defines the agent's strategy for selecting actions in various states. It is a mapping from states to actions that guides the agent's behavior. In essence, the policy represents the agent's decision-making algorithm. A good policy should aim to maximize the expected cumulative reward over time. Policies can be deterministic, meaning they always select the same action for a given state, or stochastic, introducing randomness into action selection.

6. Reward (r): The reward is a numerical value that provides feedback to the agent about the immediate consequence of taking a specific action in a particular state. Rewards can be positive, negative, or zero, reflecting whether an action was beneficial, detrimental, or had no impact. The agent's objective is to learn a policy that maximizes the cumulative reward over time. Reward functions are designed to guide the agent toward desirable outcomes. Rewards in a WECS problem could be defined based on the amount of energy generated by the wind turbine system in a given time period

7. Value Function (V): The value function, denoted as $V(s)$, estimates the expected cumulative reward that an agent can achieve starting from a specific state s while following a given policy π. In other words, it quantifies the goodness of being in a particular state and following a particular policy. It helps the agent assess the desirability of states and is essential for decision-making.

8. Q-Function: The Q-function, denoted as $Q(s,a)$, estimates the expected cumulative reward that an agent can achieve by taking a specific action a in a specific state s and then following a particular policy π. It is a fundamental concept in RL, particularly for value-based methods like Q-Learning. Q-values are crucial for determining which actions are the most rewarding in each state, thus guiding the agent's actions.

These components collectively form the foundation of RL, enabling the agent to learn from its interactions with the environment and make decisions to maximize cumulative rewards. In the realm of RL, a rich array of methods has emerged to address the challenge of training agents to make intelligent decisions in complex environments. These methods can be broadly categorized into model-based and model-free approaches.

Model-Based Methods: Model-based methods involve creating a model of the environment. This model includes a transition probability distribution that specifies how the environment changes from one state to another when an action is taken, and a reward function that assigns numerical values to state-action-state transitions. These learned models are then used for decision-making. The agent can simulate potential interactions with the environment, predict the outcomes, and determine the best course of action. Model-based methods are often sample-efficient, making them useful when collecting real-world data is expensive or risky.

Model-based reinforcement learning for WECS inherits some limitations from previously studied model-based control strategies, such as the difficulty and accurately modeling the complex dynamics of WECS, including uncertainties in wind conditions and turbine behavior. This modeling inaccuracy can lead to suboptimal control decisions and reduced performance.

Moreover, limited generalization of model-based RL algorithms considers as the most critical issues, specifically. These algorithms are often tailored to explicit WECS configurations and operating conditions, and they may struggle to generalize to new or unseen scenarios. This lack of generalization can necessitate retraining or adaptation of the RL algorithm when the system changes, adding complexity and computational burden.

Model-Free Methods: Model-free methods, in contrast to model-based ones, do not require an explicit model of the environment. Instead, they directly learn the optimal policy or value function through trial and error. They learn by continuously interacting with the environment, updating the agent's policy or value function based on the rewards and outcomes of their actions. Model-free methods are divided into two categories: policy-based and value-based approaches. These methods are often more robust and applicable to a wide range of problems but can be sample-inefficient, particularly in situations with sparse or delayed rewards.

- **Policy-Based Methods:** Policy-based methods focus on optimizing the policy itself, which is a mapping from states to actions. The goal is to find a policy that maximizes the expected cumulative reward. These methods often use gradient-based optimization techniques to update the policy, making it more likely to choose actions that lead to higher rewards. Policy-based methods are particularly useful when dealing with high-dimensional action spaces or scenarios where the optimal policy is complex and challenging to represent. They excel in stochastic environments and can handle both discrete and continuous action spaces. However, policy optimization can be computationally expensive, and finding the optimal policy can be challenging, with a trade-off between exploration and exploitation.
- **Value-Based Methods:** Value-based methods focus on estimating the value function or Q-function. The value function estimates the expected cumulative reward starting from a given state, while the Q-function estimates the expected cumulative reward of taking a specific action in a specific state. These methods estimate the value or Q-function and derive the policy from it. By selecting the action with the highest estimated value in a given

state, the agent can make decisions. Value-based reinforcement learning solutions offer distinct advantages over policy-based approaches for optimizing WECS. One key advantage is their simplicity, as value-based methods often feature simpler algorithms and implementations, facilitating easier understanding and deployment. Additionally, these methods exhibit scalability, effectively handling large state and action spaces inherent in complex WECS control tasks. Value-based approaches also excel in exploration, naturally navigating the state-action space to discover optimal policies, particularly beneficial in environments with uncertainties like WECS. Moreover, these methods tend to be more stable during training, focusing on learning the value of states rather than directly learning a policy. This stability, coupled with efficient value function optimization, typically leads to faster convergence to optimal policies. However, they may struggle with high-dimensional action spaces or problems where the optimal policy is stochastic and can suffer from overestimation biases that some algorithms aim to mitigate.

Due to the advantages of value-based RL methods, this technique will be used for optimizing WECS, in terms of power quality improvement, MPPT, and power management schemes. The mathematics and development are provided, as follows.

2.5.1 Mathematical Modeling of Reinforcement Learning

As mentioned, in RL, an agent sequentially interacts with an environment, E. The environment is characterized by the transition function, $P(s_{t+1} \mid s_t, a_t)$, the reward function $R(a_t, u_t)$, action space A, and discount factor, γ. The transition function then can be defined as:

$$P(s_{t+1} \mid s_t, a_t): S \times A \times S \to \mathbb{R} \tag{2.19}$$

This would specify the probability distribution of the next state, s_{t+1}, given the current state, s_t, and action, a_t. Here, S illustrates the state space, and A considers as the action space. While some RL-based methods can be extended to continuous action spaces, as well, the author focuses on discrete action spaces, which is proper for WECS units' control. This is due to the fact that as the likelihood of the subsequent state is solely determined by the preceding state and action, and

this would bring up the Markov decision process (MDP) concept. In RL, the Markov property is a foundational principle that streamlines the representation of environments and decision-making. It asserts that a system's future state is solely determined by its present state and is independent of the sequence of states that came before it. This principle can be mathematically formulated as:

$$P(s_{t+1} \mid s_t, a_t, s_{t-1}, a_{t-1}, \dots, s_0, a_0) = P(s_{t+1}|s_t, a_t) \tag{2.20}$$

According to this equation, independent of previous states or actions, the probability to shift from the present state of s_t to the following state, s_{t+1}, relies entirely on the present situation. As it enables agents to come to decisions based exclusively on the current state, this characteristic is essential to reinforcement learning as it streamlines the learning process. The state-action pairings, $\{s_t, a_t\}$, are mapped to rewards via the possibly stochastic reward function, $R(s_t, a_t): S \times A \to \mathbb{R}$. In fully observable circumstances, the agent picks an action $a_t \in A$ from their policy of $\pi(a_t \mid s_t): S \times A \to [0,1]$ after spotting the environment's Markov state, s_t. Following that, the environment transforms to the next state, $s_{t+1} \sim P(s_{t+1} \mid s_t, a_t)$, and offers the agent a reward, $r_t \sim R(s_t, a_t)$. Conventionally, it is assumed that the agent is unaware of the transition function or reward function; instead, it must acquire knowledge about these through interaction with the environment. A discount factor, γ, is also included in the task, indicating the relative value of future rewards. The agent's objective is to adjust the policy in order to maximize the total anticipated discounted return for each episode:

$$J = \mathbb{E}_{\tau \sim P(\tau)} R_0(\tau) \tag{2.21}$$

where τ is the trajectory of $\tau = \{s_0, a_0, r_0, \dots, s_{T+1}\}$ and the expected return from time step t onwards would be:

$$R_t(\tau) = \sum_{t'=t} \gamma^{t'} r(s_t, a_t) \tag{2.22}$$

employing the Markov assumption, it can be expressed that the probability of the trajectory is multiplication of probabilities across time steps:

$$P(\tau) = P(s_0) \prod_{t=0}^{t=T} P(r_t \mid s_{t+1}, s_t, a_t) P(a_t \mid s_t) P(s_{t+1} \mid s_t, a_t) \tag{2.23}$$

In (23), both the transition function, $P(s_{t+1} \mid s_t, a_t)$, and the policy $P(a_t \mid s_t) = \pi(a_t \mid s_t)$ can be defined.

A schematic for the high-level process of an RL agent is depicted in Fig 2-1. This diagram highlights the above-mentioned three signals that enable communication between the agent and environment at time steps t and $t + 1$: state, action, and reward

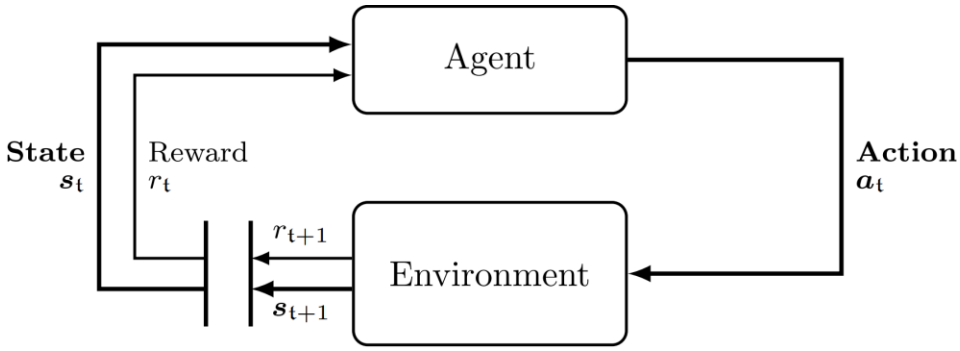


Fig. 2.1: The agent-environment process in RL

In value-based reinforcement learning, as mentioned, the value function is a crucial concept used to estimate the expected cumulative reward an agent can obtain from a given state (or state-action pair) following a specific policy. It helps the agent make decisions by quantifying the desirability of different states or actions. So, the state-value function of $V_\pi(s_t)$ would be calculated by (24). The output is a scalar which demonstrates the relative value of a particular state. Estimating the value of a state-action pair has comparable benefits. The action value function, represented by the notation $Q_\pi(s_t, a_t)$, evaluates the value $V_\pi(s_t)$ of acting at s_t and then adhering to the policy π. Equation (2.25), which calculates the state-value function, and the action-value function in (2.24) are closely connected.

$$V_\pi(s_t) = \mathbb{E}[R_t|s_t] \tag{2.24}$$

$$Q_\pi(s_t, a_t) = \mathbb{E}[R_t|s_t] \tag{2.25}$$

In this regard, the value function can be expressed as:

$$V_\pi(s_t) = \mathbb{E}[Q_\pi(s_t, a_t)] \tag{2.26}$$

Single-agent RL faces challenges when applied to WECSs due to the dynamic and interdependent nature of wind turbine operations in a wind farm. Single-agent RL mostly assumes a stationary environment, but wind conditions, which directly impact turbine performance, are non-stationary. In a wind farm, the wake effect caused by one turbine can significantly affect neighboring turbines, introducing further non-stationarity. Credit assignment, attributing rewards to specific turbine actions, is complex in a wind farm. Turbines' actions can influence each other's rewards, such as one turbine's operation affecting the wind speed and direction experienced by others. Balancing exploration and exploitation are critical in turbine control. Single-agent RL may struggle to balance these concepts in a WECS, where turbines must adapt to changing wind conditions while maximizing power output, potentially requiring coordination with neighboring turbines. Also, turbines in a wind farm may need to coordinate to avoid wake effects and maximize overall power output. However, they also compete for wind resources. Single-agent RL is not designed to handle the coordination and competition inherent in a multi-turbine wind farm or other sophisticated WECS structures.

While single-agent approaches have paved the way for advancements in reinforcement learning (RL), they are not without limitations, though. To address these shortcomings, multi-agent reinforcement learning (MARL) has been developed. In the next section, MARL will be explored and its application in enhancing the capabilities of RL algorithms will be discussed.

2.5.2 Multi-agent Reinforcement Learning

Multi-agent reinforcement learning can address the above-mentioned challenges in WECS by enabling system to learn to cooperate, compete, and coordinate. MARL algorithms allow turbines

to adapt to changing wind conditions, mitigate wake effects, and optimize overall farm performance by learning from each other's experiences. MARL is a subfield of reinforcement learning that focuses on scenarios where multiple agents interact in a shared environment, and their actions collectively determine the outcome.

Formally, MARL may be expressed as a stochastic game, which is denoted by the tuple contains; $G = < S, A, P, r, O, N, \gamma >$, where agents, demonstrated by $i = \{1, \ldots, N\}$, select successive actions from $a^i \in A$. There is also genuine state $s \in S$ in the environment, just like in the single agent setting. Every agent initiates a collective action at every time step, causing a change in the surroundings based on the state transition function $P(s'|s, a): S \times A \times S \rightarrow \mathbb{R}$. In this case, γ is the discount factor, a value between [0, 1] which balance the future rewards, and as before, $r(s, a, i): S \times A \times N$ defines an agent ith reward.

In the following sections, the various types of control strategies based on MARL, will be explored. These strategies can be broadly categorized into centralized and decentralized approaches, each offering distinct advantages and challenges. Additionally, different environmental settings, such as cooperative environments where agents work together to achieve a common goal, and zero-sum environments where agents' goals are in direct conflict, leading to competitive interactions, will be discussed, as well. Understanding these different strategies and environments is crucial for designing effective MARL systems for various applications.

2.5.3 Centralized vs Decentralized Control

Centralized control RL involves singles decision-making entity that coordinates actions for all agents based on the global state of the system. This approach offers advantages such as enhanced coordination among agents, leading to potentially more efficient and synchronized behavior. Additionally, centralized control can leverage global information about the environment, enabling more informed decision-making. However, it also has its drawbacks. For instance, centralized control introduces a single point of failure; if the central controller malfunctions, the entire system may fail. Moreover, the need for constant communication between the central controller and agents can result in communication overhead, potentially slowing down decision-

making processes. In the fully observable set-up, the centralized controller of $C^{\pi}(a|s_t)$, which maps the states into a probability distribution over joint action a would be:

$$C^{\pi}(a|s_t) = A \times S \rightarrow [0,1] \tag{2.27}$$

Decentralized control, on the other hand, allows each agent to make decisions independently based on its local state information. This approach offers scalability benefits, as it does not require coordination among all agents, making it more suitable for large-scale system of WECS. Decentralized control also tends to be more robust, as the failure of one agent does not necessarily affect the entire system. However, decentralized control may suffer from a lack of coordination among agents, which can lead to suboptimal or conflicting actions, especially in complex environments. Additionally, decentralized control may have limited access to global information, which can constrain decision-making capabilities. Each agent in this context has a local policy, denoted by $\pi^i(a^i \mid s_t)$, that maps states to a probability distribution across actions relevant to the particular agent. It would be observed that the probability distribution over the joint action is factorized by this:

$$P(a \mid s_t) = \prod \pi^i(a^i \mid s_t) \tag{2.28}$$

2.5.4 Cooperative, Zero-sum, and General-Sum Environment

In MARL, the environment refers to the external system in which agents operate and interact. It includes all elements that the agents can perceive and affect, such as other agents, objects, and the rules governing their interactions. The environment provides feedback to the agents in the form of rewards based on their actions, shaping their learning process. The different types of environments are summarized, as follows:

Cooperative: In a cooperative environment, agents collaborate to achieve a common goal. They share information and resources to maximize the collective reward. This setting promotes teamwork and can lead to solutions that are beneficial for all agents involved. And it can be formulated as:

$$r(s,a,i) = r(s,a,i'), \quad \forall i, i' \tag{2.29}$$

Pros include encouraging collaboration and teamwork, potentially leading to optimal solutions that benefit all agents. However, cons include the requirement for effective communication and coordination between agents, as well as the risk of one agent exploiting others' cooperation for its own gain.

Zero-sum: In a zero-sum environment, the total reward is constant, meaning one agent's gain comes at the expense of another agent's loss. This competitive setting often leads to strategic behavior as agents try to outperform each other. And the rewards can be defined as:

$$\sum_i r(s, \boldsymbol{a}, i) = 0, \quad \forall s, a \tag{2.30}$$

Pros include well-defined competitive dynamics that can lead to strategic decision-making, along with established game theory principles and algorithms for analysis. However, cons include constant conflict between agents, which can limit cooperation and joint learning, as well as the potential for unstable learning dynamics and suboptimal solutions.

General-Sum: In a general-sum environment, the total reward is not fixed and can vary based on the actions of all agents. This setting represents a wide range of real-world scenarios where agents' actions can collectively impact the overall reward. Pros include a more realistic representation of many real-world scenarios and encouraging agents to consider the global impact of their actions. However, cons include increased complexity in learning due to the varying nature of rewards and the requirement for sophisticated algorithms to handle the interactions between agents.

For the application of power quality and MPPT in WECS using MARL, the cooperative environment is often used. In a cooperative environment, agents work together to achieve a common goal, which in this case could be maximizing the power output of the WECS, and maintaining good power quality. Cooperation among agents in this context can involve sharing information about wind conditions, turbine performance, and power quality measurements to collectively make decisions that benefit the overall system performance. Using a cooperative

environment in this scenario allows the agents to coordinate their actions to ensure that the WECS operates efficiently and effectively. This can lead to better overall power output and improved power quality, as agents can collaborate to mitigate issues such as voltage fluctuations or harmonics in the system. Additionally, cooperation can help agents explore the state space more effectively, leading to faster convergence to optimal solutions for MPPT. In contrast, using a zero-sum or general-sum environment may not be as suitable for this application. In a zero-sum environment, agents' goals are directly opposed, which could lead to suboptimal behavior as agents compete against each other rather than cooperating. And general-sum environment, while is more flexible, may not provide the necessary structure for agents to effectively coordinate their actions towards a common goal.

While MARL offers promising capabilities for addressing complex problems in WECS, there are several shortcomings of general MARL approaches when applied to this domain. One key limitation is the scalability of traditional MARL algorithms as the number of agents increases. The complexity of coordination and communication among agents can become prohibitively high, leading to inefficient learning and decision-making processes. Another challenge is the non-stationarity of the WECS environment, where factors such as wind speed change. Traditional MARL algorithms struggle to adapt quickly to these dynamic environments, requiring frequent retraining and recalibration, which can be computationally expensive and time-consuming. Additionally, the discrepancy in such co-player policies among the agents is an acute issue as it makes restoring them intrinsically challenging and causes undesired fluctuations in the training phase. As the other challenges in the realm of MARL, the intricacies of inter-agent dependencies give rise noise and variance in rewards, leading to training instability.

Moreover, usually, fixed values are used for the discount factor in MARL. While it is straightforward, it would carry some limitations, especially in dynamic WECS environments. Using a fixed discount factor can lead to suboptimal policies, as it may not appropriately balance the importance of short-term and long-term rewards. Additionally, a fixed discount factor may not adapt well to changes in the environment or system dynamics, limiting the ability of RL algorithms to learn optimal control policies in complex and uncertain environments.

To address these limitations, a novel approach, based on MARL, is proposed for WECS applications for power quality and MPPT purposes. The details are provided in following sections and the above-mentioned issues are going to be fulfilled, step by step.

2.6 Proposed Multi-Agent Reinforcement Learning

The MARL represents a paradigm shift in artificial intelligence, extending the principles of RL to environments with multiple interacting agents, as mentioned. Moreover, MARL provide higher capacity to capture emergent behaviors and interactions in sophisticated systems, compared to the single agent RL. Here, a model-free based MARL will be employed to tackle the WECS problems, aiming to boost the power quality and MPPT efficiency. The model-free nature of the proposed MARL eliminates the need for an explicit model of the environment, allowing agents to learn optimal policies directly from experience [77], [78], [79].

Tuple $\mathrm{G} = < N, S, A, R, P, O, \gamma >$ characterizes proposed multi-agent RL setting, where N indicates the total number of agents, S is the state space, $A = \{A_1, \ldots, A_N\}$ represents the set of actions for every agent, P is the probability of a transition from one state to another, R is the reward function, and $O = \{O_1, \ldots, O_N\}$ is the set of observations for every single agent. and the gamma notation, γ, serves as a discount factor, as stated. In an environment with N agents, i^{th} agent observes the global state s_t at time-step t. It then applies the local stochastic policy π_i for carrying out an action of a_i^t and obtain a reward of r_i^t.

Here, a fully cooperative environment is proposed, so every agent observes a shared value for reward r_t at every time step; that is, $r_1^t = \cdots = r_N^t = r_t$. Cooperative MARL offers several advantages over zero-sum and general-sum environments. In a cooperative setting, agents work together to achieve a common goal, whereas in a zero-sum environment, one agent's gain comes at the expense of another's loss. Cooperative MARL promotes teamwork and collaboration, leading to potentially better overall outcomes. It also avoids the constant conflict and competitive dynamics present in zero-sum games, allowing for more stable and cooperative strategies to emerge. On the other hand, general-sum environments allow for more complex interactions and outcomes, they can be more challenging to navigate due to the varying nature of rewards. Cooperative MARL, on the other hand, focuses on agents working together towards a shared objective, simplifying the decision-making process and promoting joint learning.

Despite these advantages, the fully cooperative method in MARL has two significant limitations, one of them known as the "Tragedy of the Commons." This occurs when agents prioritize their individual interests over the collective goal, leading to suboptimal outcomes for the group as a whole. To address this limitation, a solution known as the "Nash Social Welfare" (NSW) has been proposed. The Nash Social Welfare is a modification of the fully cooperative method that aims to align individual and collective interests. It introduces a social welfare function that represents the overall welfare of all agents in the system. Mathematically, the NSW is defined as:

$$NSW = \sum_{i=1}^{N} V_i \tag{2.31}$$

In which the V_i is the value function for the ith agent. The NSW encourages agents to consider not only their individual rewards but also the rewards of other agents, leading to more cooperative behavior. By incorporating the NSW into the fully cooperative method, agents can learn to balance their individual interests with the collective goal, resulting in more efficient and effective cooperation in MARL.

Moreover, from the viewpoint of each given agent, the environment gets non-stationary over the training process, since each agent's policy π_i varies:

$$P(s' | s, a_i, \pi_1, \ldots, \pi_N) \neq P(s' | s, a_i, \pi'_1, \ldots, \pi'_N) \tag{2.32}$$

When $\pi_i \neq \pi'_i$, the fundamental presumption of MDP will be violated. This indicates every agent's perception of co-player policies is distinctive. The discrepancy in co-player policies is an acute issue as it makes restoring them intrinsically challenging and causes undesired fluctuations in the training phase. The matter is resolved here employing the Q-based Bellman equation. The Q-based Bellman equation helps resolve this issue by explicitly modeling the interactions and dependencies between agents. By defining the Q-function as the expected cumulative reward starting from a state, taking a joint action, and then following a joint policy thereafter, the Q-based Bellman equation captures the dynamics of the joint action-value function. This allows agents to evaluate the long-term consequences of their actions, considering the policies of all agents involved. By jointly optimizing the Q-values, agents can learn a joint policy that not only

maximizes their individual rewards but also takes into account the impact of their actions on other agents. This promotes coordination and cooperation among agents, leading to more stable and effective behavior in complex environments like WECS. Additionally, by using the Q-based Bellman equation, agents can learn to adapt to changes in the policies of co-players, ensuring robustness and flexibility in their decision-making processes. So, the proposed Q-function for MPPT purpose is considered to be:

$$Q_i^* (s, a_i \,|\pi_{-i}) = \sum_{a_{-i}} \pi_{-i} (a_{-i}, s) \, [r(s, a_i, a_{-i}) + \gamma \sum_{S'} P(s' \,|s, a_i, a_{-i}) \, max_{a_i'} \, Q_i^*(s, a_i') \tag{2.33}$$

The notation $Q_i^*(s, a_i \mid \pi_{-i})$ represents the optimal action-value function for agent i when the other agents are following policies denoted by π_{-i}. The equation describes how the optimal Q-value for an action a_i in state s is determined by the sum of the immediate reward $r(s, a_i, a_{-i})$ and the discounted future expected maximum Q-value, where a_{-i} represents the joint actions of all other agents except i. The optimal policy for agent i, which denoted as $\pi_i^*(s)$, is derived from the optimal Q-values. It is the policy that selects actions in each state to maximize the expected cumulative reward:

$$\pi_i^*(s) = \arg\max_a (Q_i(s, a_i)) \tag{2.34}$$

The joint optimal policy, π^*, is the combination of optimal policies for all agents:

$$\pi^* = (\pi_1^*, \pi_2^*, \dots, \pi_N^*) \tag{2.35}$$

In other words, the goal is to find the joint optimal policy that maximizes the expected cumulative reward stated in (2.36) by iteratively improving the Q-values over time t and updating the policies until convergence.

$$\max_{\pi_i} \mathbb{E}[\sum_t^\infty \gamma_t R(s, a_i, a_{-i})_t] \tag{2.36}$$

As the other challenges in the realm of MARL, the intricacies of inter-agent dependencies give rise noise and variance in rewards, leading to training instability. To address this issue, the Nash Equilibria is proposed. Nash Equilibria provide a theoretical framework for achieving stable states in strategic interactions among agents. In the context of MARL, a Nash Equilibrium is a set of strategies, one for each agent, where no agent has an incentive to unilaterally deviate from its current strategy given the strategies of the other agents. In other words, at a Nash Equilibrium, each agent's strategy is optimal given the strategies of the other agents, and no agent can improve its payoff by changing its strategy unilaterally. By aiming to reach a Nash Equilibrium, MARL agents can achieve stability in their interactions, reducing the noise and variance in rewards that can lead to training instability. Agents learn to anticipate and respond to the actions of other agents in a way that leads to a mutually beneficial outcome, rather than engaging in erratic or unpredictable behavior. This not only improves the efficiency of learning but also promotes cooperation and coordination among agents, leading to more robust and effective decision-making in complex environments. According to the concept:

$$\forall \pi_i \in \Pi_i, \forall s \in S, v_i^{(\pi_i^*, \pi_{-i}^*)}(s) \geq v_i^{(\pi_i, \pi_{-i}^*)}(s) \tag{2.37}$$

where Π_i is an array of all potential policies for agent i and $v_i(s)$ is the anticipated cumulative reward of agent i in state s. In the context of WECS applications, state-value functions of V have limitations when different actions in the same state can lead to significantly different outcomes. For example, in a wind farm, the same wind conditions at a turbine can result in different power outputs based on how the turbine is controlled. State-value functions would provide a single value for the expected return from that state, regardless of the control action taken, potentially leading to suboptimal control decisions. To address this limitation, the concept of the advantage function is proposed for WECS applications, in this work. The advantage function $A(s_t, a_t)$ provides a measure of how much better or worse taking action a in state s is compared to the average action in that state. In the context of this study, the advantage function is quantifying objectives from choosing a particular control action compared to the average control action. Mathematically, the advantage function is defined as the difference between the action-value function and the state-value function:

$$A_i(s_t, a_t) = Q_i(s_t, a_t) - V_i(s_t) \tag{2.38}$$

By incorporating the advantage function into RL algorithms for WECS, such as for wind turbine control, agents can make more informed decisions. This allows for more efficient learning and better performance, as the advantage function helps the agent understand the impact of its actions in different states, leading to more optimal control strategies that maximize power output or other objectives.

The structure of the proposed MARL strategy has been developed, with a focus on addressing challenges such as disparate perceptions of co-player policies and training instability arising from inter-agent dependencies. One crucial hyperparameter in MARL is the discount factor, which determines the importance of future rewards in the agents' decision-making process. In the next section, the proposed method for optimally tuning the discount factor will be presented. This method aims to find a balance between the short-term rewards and long-term goals of the agents, ensuring that the MARL system converges to stable and optimal strategies

2.6.1 Proposed Optimal Discount Factor Tunning Strategy

In the context of proposed MARL, the γ hyperparameter which can be seen in (2.33), plays a pivotal role in shaping the learning dynamics and long-term rewards within the environment, thereby significantly impacts the eventual policy convergence.

Using a fixed discount factor or grid-searched values for the discount factor, especially in WECS applications can lead to several limitations. Firstly, a fixed discount factor does not adapt to changes in the environment or system dynamics, which are common in WECS due to the variability of wind conditions. This can result in suboptimal control strategies, as the importance of future rewards relative to immediate rewards may vary over time. Additionally, using a fixed discount factor may not capture the complex interactions and dependencies between actions and rewards in WECS, where the impact of an action in one time step can influence future rewards in non-linear ways.

There is another way to tune the gamma hyperparameters, known as grid search strategy. Grid-searched values for the discount factor suffer from similar limitations. While grid search can potentially find a better discount factor than a fixed value, it is computationally expensive and may not scale well to large state and action spaces in WECS. Grid search also requires prior knowledge or assumptions about the range of possible discount factors, which may not always be accurate or realistic. Moreover, grid search does not adapt to changes in the environment, and the optimal discount factor found through grid search may not generalize well to different wind conditions or turbine configurations. Overall, using fixed or grid-searched values for the discount factor in WECS can limit the ability of RL algorithms to learn optimal control policies and adapt to changing environmental conditions.

Meta-learning strategy is proposed to address the DF tuning, optimally, where it is learned dynamically during the training process. Specifically, the meta-learned DF is minimized through stochastic gradient descent over multiple episodes or iterations. SGD is employed in proposed MARL, due to the ability in handling noisy gradients effectively, navigating complex landscapes efficiently, and facilitating stable convergence in high-dimensional parameter spaces like WECS. This adaptive mechanism allows the agents to autonomously adjust the discount factor, enhancing adaptability to changing environments and potentially accelerating convergence towards better policies. To do so, the loss function can be defined as the negative Q-value for state-action pairs, and then update gamma using gradient descent to minimize the loss function of L. In this regard:

$$L = -Q_i^* (s, a_i \,|\pi_{-i}) \tag{2.39}$$

$$L = -\left(\sum_{a_{-i}} \pi_{-i} (a_{-i}, s) \left[r(s, a_i, a_{-i}) + \gamma \sum_{S'} P(s' \,|s, a_i, a_{-i}) \, max_{a_i'} \, Q_i^*(s, a_i')\right)\right. \tag{2.40}$$

To update the γ by SGD:

$$\gamma_t^{updated} = \gamma_t - \beta \frac{\partial L(\gamma)}{\partial \gamma} \tag{2.41}$$

The parameter β in (2.41) demonstrates the step size in gradient descent. This process can be repeated iteratively until convergence, resulting in an estimate of the optimal gamma that maximizes the Q-value.

2.7 conclusion

In this chapter, the application of machine learning ML in WECSs was explored, beginning with the classification techniques and solution development. While significant potential was identified with ML methods, limitations affecting their effectiveness in WECS applications were also noted. To address these challenges, RL and its advantages were introduced, with the mathematical foundations and discussions provided. Despite the benefits, it was observed that general RL models may not always achieve optimal performance without further refinement. Consequently, a MARL model was proposed and developed step by step to enhance the RL approach. Recognizing the necessity for more precise and adaptive tuning, a meta-learning strategy was introduced to dynamically and accurately adjust the discount factor within the proposed MARL model. This innovative approach aimed to optimize the performance of the MARL method, ensuring a more robust and efficient solution for WECS applications. Through this comprehensive development and refinement, contributions were made to advancing the state of the art in power management and optimization for WECS, paving the way for more intelligent and adaptive energy solutions.

Chapter 3
RL-Based Control Strategy of a Three-Phase NPC Power Converter for WECS

3.1 Introduction

This section provides a comprehensive exploration of the cutting-edge wind energy power converter control method utilizing proposed MARL in previous chapter. The aim is to evaluate the performance of the suggested learning-based approach for a promising and practical wind energy conversion system configuration, which is direct-drive permanent magnet synchronous generator-based one. The effort concentrates on both satisfying grid code requirements in terms of THD, and increasing the power quality, while reducing system modeling sophistication. The full-scale back-to-back neutral-point clamped power converter is determined to be the case of control, employed in direct-drive PMSG-based WECS. Grid-side converter is controlled with high precision using proposed multi-agent reinforcement learning algorithm, which necessitates no offline training, and system modeling, unlike machine learning and neural network-based techniques. The simulations and subsequent discussion commence by implementing the polynomial locally weighted regularized regression machine learning strategy. This approach is initially evaluated to identify any inherent shortcomings. Upon observing these limitations, a proposed method is introduced to effectively mitigate these issues. The proposed method is meticulously applied to refine the ML-based strategy and voltage-oriented control method, ensuring a more robust and efficient approach in addressing the identified shortcomings. To do so, single-phase and double-phase voltage sag are considered to assess the functionality of the suggested MARL in unbalance scenarios. To empower the proposed strategy, as mentioned, meta-learning is employed within to optimize the discount factor value. Compared to fixed DF and conventional DF tuning methods, meta-learnt DF provide superior adaptability, and convergence rate. Literature review, MARL application, comparative analysis, findings and related discussion are provided in following sections.

3.2 Literature Review

Numerous converter technologies and control methodologies have been investigated to surpass simply generating wind energy and to assist in bolstering the power system. Presently, wind turbines employ direct-drive permanent magnet synchronous generators or doubly-fed induction generators, are the most widespread wind energy conversion technologies in the field of electrical engineering, as academic search for viable solutions [80], [81]. In comparison to DFIG, PMSG grants several advantages, as mentioned in the first chapter. For instance, it maintains grid voltage by supplying additional reactive power, since it is connected to the power system through a full-scale converter. It also eradicates the need for a gearbox, which diminishes the weight and size of nacelle equipment, beside the mechanical losses and preservation requirements [82].

With respect to power converter, on the other hand, a three-level neutral-point clamped back-to-back (NPC-BTB) converter offers more than two voltage levels, yet requires fewer components than, for instance, five-level topologies [83]. Furthermore, this system meets all of the necessary grid requirements and demonstrates adequate tolerance for faulty scenarios [84]. Therefore, for high-power wind energy deployments, a direct-drive PMSG wind turbine system that employs a 3-level NPC-BTB power converter, as depicted in Fig. 3.1, is a desirable configuration of choice.

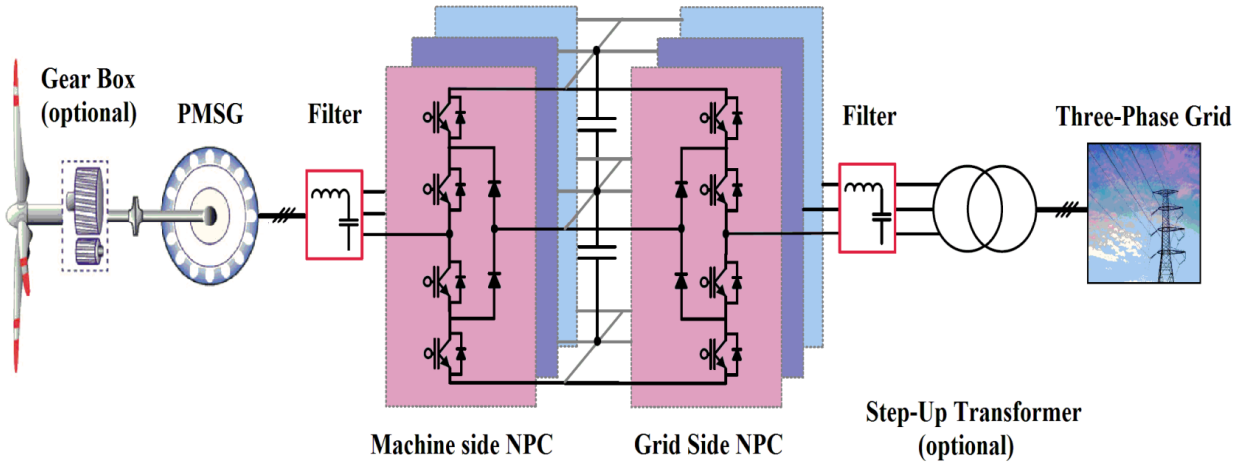


Fig. 3.1: The direct-drive PMSG-based WECS employed a 3 levels full-scale back-to-back neutral point clamped power converter.

When it comes to design control strategies for above-mentioned WECS structure, traditional PI-based controllers heavily rely on the exact calibration of the controller's parameters and

bandwidth [85]. Additionally, the parameters achieved in a particular system cannot be applied to other setups and must be re-tuned [86]. This type of linear control methods, primarily rely on linear approximations of converters, achieved through techniques like averaging and state space modeling. While these methods are straightforward to study and implement and work well for simpler, linear, or weakly non-linear converters, they falter when confronted with highly non-linear operating conditions, uncertainties, and disturbances [87]. Some solutions exist in the literature to mitigate these limitations, often involving the use of multiple controllers tailored to different operating modes. Conversely, non-linear control methods address converter non-linearity using techniques like feedback linearization, sliding mode control, and model predictive control. Feedback linearization transforms the non-linearity into a linear representation, though it demands an accurate model of the converter and its inverse, which can be challenging to obtain for complex converters [88]. As a result, model predictive control has gained more attention to overcome the shortcomings. In [89], a universal and moderately complex MPC approach was developed for DFIG-based WECSs, which could operate effectively in both balanced and unbalanced conditions. Or, to manage the three-level boost converter and grid-side NPC inverter, a model predictive technique was outlined in [90]. Two independent cost functions were employed to evaluate the predictions and reduce the switching states. However, the MPC controller required extensive exploration of all possible switching signals in each controller interval to identify the appropriate one, demanded a high computational cost. Therefore, in [91], a proficient MPC strategy was introduced for regulating torque and power in BTB-NPC converters for WECSs with PMSG. This method employed hexagonal and triangular designated domain to significantly reduce the count of possible switching states, resulting in reduced computation burden. Although, High power oscillations and substantial processing work have hindered MPC solutions, yet [92].

On the other hand, the sliding mode control is a promising approach for managing the NPC converter due to its features, such as finite-time convergence and robustness against system uncertainties [93]. However, despite these benefits, conventional SMCs are prone to the chattering problem, negatively impacting power system operations. The super-twisting algorithm, a type of second order SM control, offers an attractive solution for reducing chattering [94]. This algorithm hides the discontinuous function used in common SMC within the integral, significantly reducing the chattering effect. For instance, a second order SM controller has been proposed for the LC-coupling hybrid active power filter as a current controller, demonstrating excellent dynamic and

steady-state performance in terms of THD. Additionally, the controller has been applied to both the voltage and current loops of a three-level NPC converter, and its use has been extended to grid-connected WECS converters, dc-dc converters, and motor drives. However, this type of controller requires knowledge of the disturbance derivative's boundary, and its gains must be carefully managed to avoid excessive chattering amplification, limiting its performance [96].

Artificial intelligence is an expanding area of study that has gained significant attention in recent years due to the potential to provide assistance in various fields, as explained in chapter 2. Within the AI concept, fuzzy logic (FL), machine learning and reinforcement learning are three significant subfields. In [97], an FL control technique was proposed for three-phase, 3L-NPC power converters to enhance disturbance declination, power tracing, dc-link voltage adjustment, and suppress chattering. However, the FL mainly suffers from high computational cost due to the requirement of multiple control loops, and it relies heavily on empirical knowledge and is not suitable for complex systems [98]. The adoption of machine learning in power electronics has the potential to mitigate the drawbacks of traditional and FL-based solutions. Nevertheless, prior research on ML dominantly concentrated on employing ML as a mechanism for condition monitoring purposes, or converter defect detection, [99], [100]. In this context, using MPC for the training phase, [101] offered an ML-based technique for modular multilevel converters. The developed framework yielded results comparable to MPC but requiring substantially less computation. In [102] the voltage-oriented control (VOC) procedure was employed to train a ML model instead of MPC, which was subsequently utilized to effectively and precisely govern the grid-side NPC power converter. As a result, computational load was significantly reduced and moved from online to offline execution. Also, since the PI controllers in VOC were removed, bandwidth and PI controller parameters were no longer necessary. Moreover, there was a considerable drop in calculations. Several optimizers were tested on developed ML model and the findings indicated that in terms of training time, total harmonic distortion capability, and precision, Broyden-Fletcher-Goldfarb-Shanno optimizer outscored other cost function optimizers. It should be highlighted, although if the ML model used to regulate the three-phase grid-side NPC converter was adept at addressing balanced circumstances, the system was not tested in unbalanced instances. However, there are some disadvantages linked to the shift in computing load from online to offline retrieval. In an offline learning, the system cannot adjust to environmental alterations instantly. In the case of power converter control strategy, where unplanned changes can

happen and needs for prompt adaptations, this can be very troublesome. Online learning can therefore provide more adaptability and responsiveness in unbalance situation.

Hence, as stated, reinforcement learning is a subfield of AI that is particularly suited to online learning and can address the limitations of offline learning for power converter control strategy, mentioned earlier. RL algorithms enable the system to adapt to changes in real-time by learning from the feedback (reward or penalty) it receives through interactions with the environment. This allows the system to continually improve its performance and optimize the control policy. While some references can be found which addressed the reinforcement learning application in controlling DC-DC converters, there has been relatively little or even no attention paid to RL-based control of NPC power converters [103], [104].

In the following sections, the simulations and ensuing discussion will begin with the implementation of a strategy based on polynomial locally weighted regularized regression machine learning. This approach will be carefully assessed to identify any shortcomings it may have. Following the identification of these limitations, a proposed method will be introduced to address and overcome these issues. This proposed method will be applied in a thorough manner to enhance the ML-based strategy, aiming to create a more reliable and effective approach in dealing with the identified shortcomings. The proposed multi-agent reinforcement learning approach detailed in previous chapter is used to control the grid-side neutral point clamped power converter employed in direct drive permanent magnet synchronous generator-based WECS. The MARL in NPC control offers several advantages over earlier investigation, including improved adaptability and flexibility, as well as the ability to optimize the system in real-time. Additionally, the online learning feature of the proposed strategy allows the system to continuously improve its performance over time, as it adapts to changing environmental conditions and system dynamics. Furthermore, the MARL is capable of handling the single-phase and double-phase voltage sag unbalance conditions, in contrast to the ML-based approaches. By dividing the task into smaller sub-tasks and assigning them to different agents, the overall problem can be decomposed and solved in a more scalable and efficient manner, compared to the single agent RL. Besides, MARL allows for the exploration of diverse strategies and policies, which can lead to better overall performance and increased robustness in face of changing environmental conditions or system dynamics. To further improve the performance of the proposed control strategy, the discount factor

value, which explained that it is a vital hyperparameter of the RL algorithm, is learned optimally through a proposed meta-learning approach. This is found to be superior to fixed discount factor values or other traditional methods of tuning the discount factor, in terms of adaptability, and convergence rate. Outcomes of the comparison evaluation and simulation indicate the proposed method outperforms the alternatives and may serve as a solid contender for traditional and even the most lately researched control strategies. The steps of the method are summarized as follows:

1. The initial step involves the definition of NPC and grid parameters. These parameters serve as the foundational building blocks for the control strategy, providing essential information about the system's characteristics and specifications.
2. Following the parameter definition, the next crucial phase involves specifying hyperparameters. In this context, the hyperparameters include the number of episodes, the number of agents, the state size, the action size, and learning and exploration rates. These hyperparameters play a pivotal role in shaping the behavior and learning process of the proposed MARL algorithm, ensuring it aligns with the unique requirements of the grid side NPC control in WECS.
3. Subsequently, the definition of states and actions is imperative for the successful implementation of the control strategy. In this specific problem, the states are designated as grid active power, grid side line-to-line voltage, DC link voltage, and THD. These states collectively provide a comprehensive snapshot of the system's condition, enabling the algorithm to make informed decisions based on real-time data. For the actions, the modulation index has been chosen as the variable to be controlled. This is a key parameter that directly influences the converter's operation and, therefore, holds great significance in optimizing grid integration.
4. Now, in accordance with the explanation of the proposed MARL algorithm equipped with the meta-learned discount factor, it is essential to initialize the Q table, episode reward, and discount factor. The initialization process is a critical precursor to the algorithm's learning and decision-making phases. Methods for initializing these parameters must be carefully chosen to ensure effective and efficient learning.

   Zero and random initialization are two commonly used methods in reinforcement learning. Zero initialization involves setting all Q-table values to zero, while random initialization assigns random values to the Q-table. However, zero initialization often suffers

from convergence issues due to the agents initially having identical Q-values and making similar decisions. This can lead to slow learning and a lack of exploration. In this particular work, the random initialization method has been selected for the purpose of initializing the Q-table, episode reward, and discount factor. Random initialization introduces diversity in the initial Q-values, enabling agents to explore a wider range of actions and states from the outset. This diversity can help overcome convergence issues and promote faster learning, which is crucial for effective control of the grid side NPC in WECS.

5. As mentioned, an essential component of the proposed methodology involves the implementation of a meta-learning process. This meta learning is conducted to continuously update the discount factor used in the proposed MARL algorithm. The objective is to refine and optimize the DF through iterative updates until convergence is achieved. This iterative approach ensures that the DF is finely tuned and aligned with the specific dynamics and requirements of the grid side NPC.
6. Once the optimized discount factor is determined, it serves as a critical parameter within the proposed MARL algorithm. The DF is used to update the Q-table parameters within MARL, aiming to maximize the cumulative reward over time. This iterative updating process ensures that the Q-table is continually refined to make more informed decisions, ultimately enhancing the control strategy's effectiveness. The updating process perseveres until convergence is reached, signifying that the MARL algorithm has learned to make optimal decisions.
7. Following the successful control of the modulation index, it becomes instrumental in the next step of the process. The modulation index is employed in the Pulse Width Modulation (PWM) technique, which is utilized to generate the switching signals for the NPC. PWM allows for precise control over the output waveform, ensuring that the power converter operates in a manner that maximizes efficiency and meets the desired grid integration requirements.

This seamless integration of meta-learning, MARL, modulation index control, and PWM signal generation forms a comprehensive and sophisticated control strategy for the grid side NPC in WECS, contributing to enhanced performance and grid stability. Certainly, to provide a visual representation of the entire process, a detailed flowchart has been presented in Fig. 3.2.

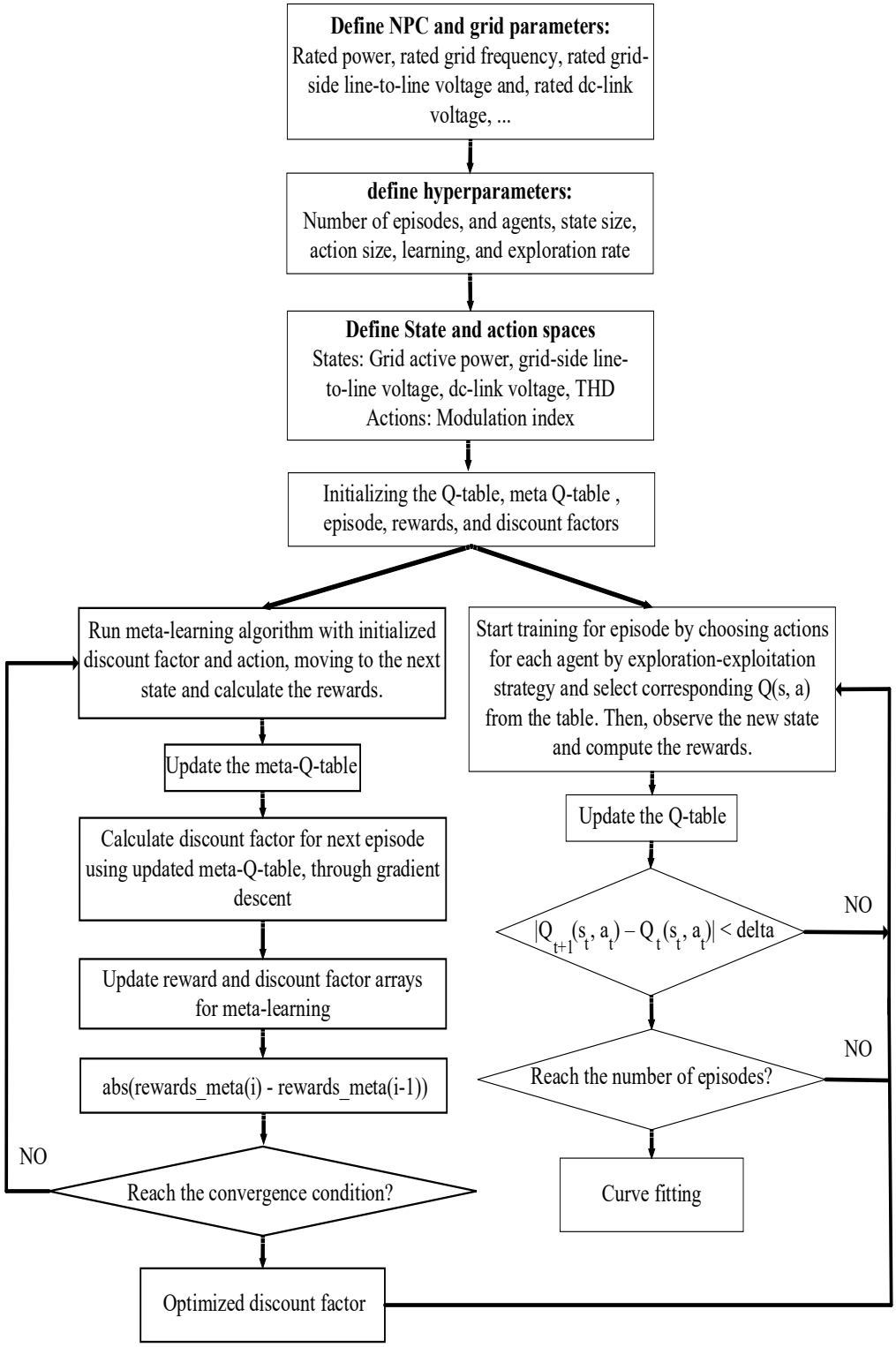


Fig. 3.2: The flowchart of the proposed grid-side NPC control strategy

3.3 ML-based and Proposed MARL Diagrams

The system parameters outlined in Fig. 3.1 are detailed in Table 3.1. The Turbine (machine) side NPC converter is controlled using a field-oriented control strategy, aiming to produce nearly 1200 V of pure DC voltage in the dc-link. Subsequently, the capacitors filter and equalize the DC voltage to 600 V each. Following this, the grid-side NPC inverter is controlled to supply 1.6 MW of active power to the grid, maintaining a line-to-line voltage of 800 V based on the VOC strategy.

Table 3.1: System Parameters

Parameters	Values
Rated BTB converter apparent power	2 MVA
Switching device	IGBT
Generator output voltage	~1200 V
Reference DC voltage for C_1	600 V
dc-link capacitor (each)	25 mF
dc-link resistor	50 mΩ
Switching frequency	2 KHz
Grid frequency	60 Hz
Grid side line-line voltage	800 V
Active power delivered to the grid	1.6 MW

VOC approach is based on the transformation between the abc stationary reference frame and the dq synchronous frame. The control technique is applied in a grid-voltage synchronous reference frame, with all variables in steady state components. This simplifies the inverter's design and control. The grid voltage is monitored and its angle θ_g is determined for the voltage orientation to achieve the VOC, as follows:

$$V_\alpha = \frac{2}{3}\left(V_{ag} - \frac{1}{2}V_{bg} - \frac{1}{2}V_{cg}\right) = V_{ag} \tag{3.1}$$

$$V_\beta = \frac{2}{3}\left(\frac{\sqrt{3}}{2}V_{bg} - \frac{\sqrt{3}}{2}V_{cg}\right) = \frac{\sqrt{3}}{3}(V_{ag} + 2V_{bg}) \tag{3.2}$$

$$\theta_g = tan^{-1}\frac{V_\beta}{V_\alpha} \tag{3.3}$$

where, θ_g, is the grid voltage angle, $v_{ag}, v_{bg},$ and v_{cg} are phase voltages. The VOC strategy is achieved by aligning the frame's d-axis with the grid voltage, resulting in $v_{dg} = v_g$, out of which:

$$v_{qg} = \sqrt{v_g^2 - v_{dg}^2} = 0 \tag{3.4}$$

This may be used to compute the active and reactive power of the system, as follows:

$$P_g = \frac{3}{2}\left(v_{dg}i_{dg} + v_{qg}i_{qg}\right) = \frac{3}{2}V_{dg}i_{dg} \tag{3.5}$$

$$Q_g = \frac{3}{2}\left(v_{qg}i_{dg} - v_{dg}i_{qg}\right) = -\frac{3}{2}V_{dg}i_{qg} \tag{3.6}$$

According to the (3.6) the q-axis reference current would be obtained:

$$i_{qg^*} = \frac{Q^*}{-1.5v_{dg}} \tag{3.7}$$

The overall block diagram of the VOC is illustrated in Fig. 3.3.

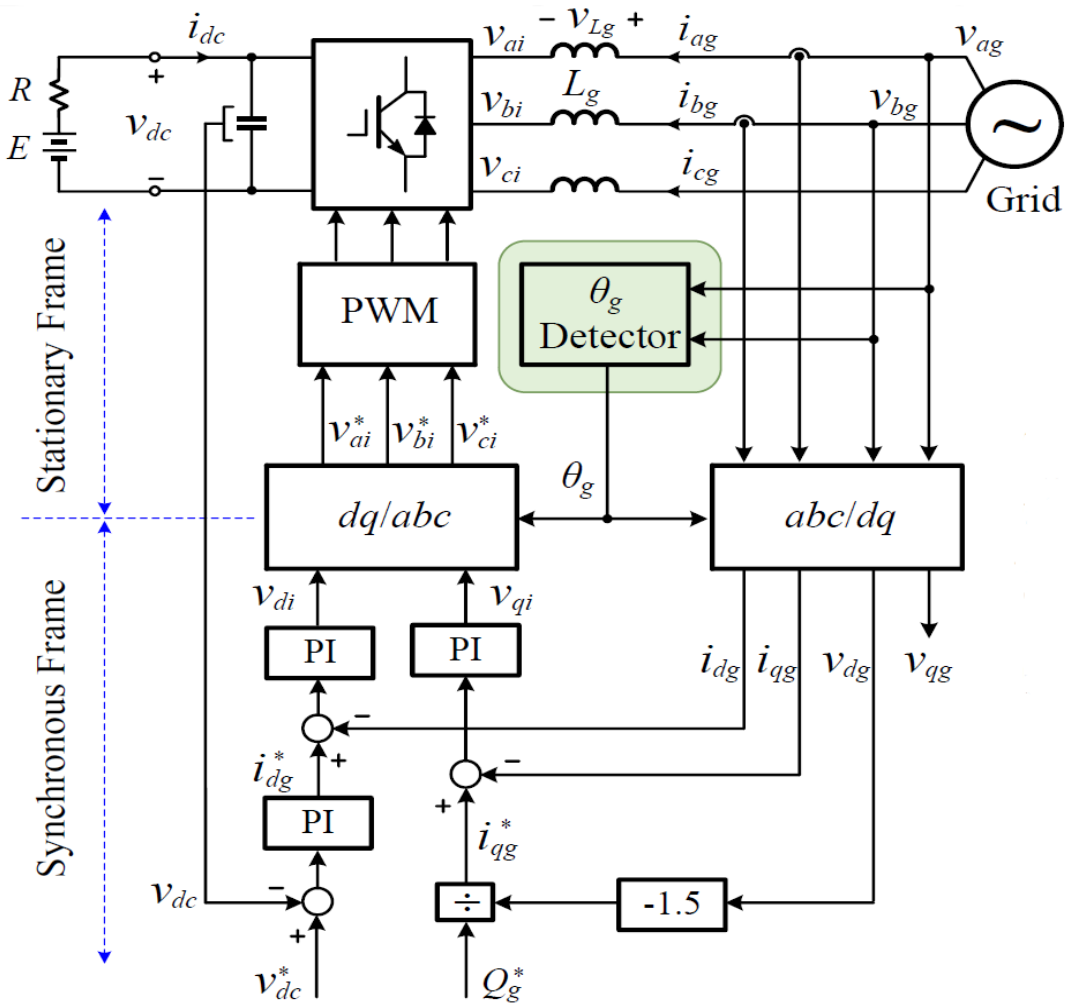


Fig. 3.3: Block diagram of the conventional PI-based voltage-oriented control.

This is used to collect the required data for training purpose of the ML model parameters. Three sensors are needed to measure the DC-link voltage, output voltage, and current. The input feature matrix of the algorithm is created after running successful simulation of the system for $m = 50410$ sample data, each contains $n = 5$ features. In this matrix $V_{dc1}$is the upper capacitor voltage of the DC-link, V_{dg} and V_{qg}, along with I_{dg} and I_{qg} are the phase voltages and currents in the dq-reference frame, respectively.

The data obtained from the results was gathered and split into two subsets through cross-validation: 70% for training data and 30% for blind-testing data, which verifies the ML training's accuracy. The training dataset is utilized to establish the fitting parameters ω based on the

optimizers used to minimize the cost function. Following the training phase, the blind, unseen data is tested to confirm the algorithm's reliable operation. The overall control strategy for the mentioned ML-based grid-side NPC converter is depicted in Fig. 3.4. The outputs of the ML-based control scheme are the d-q axis control voltages, which generate the in-phase disposition pulse width modulation (IPD-PWM) signals for the converter switches.

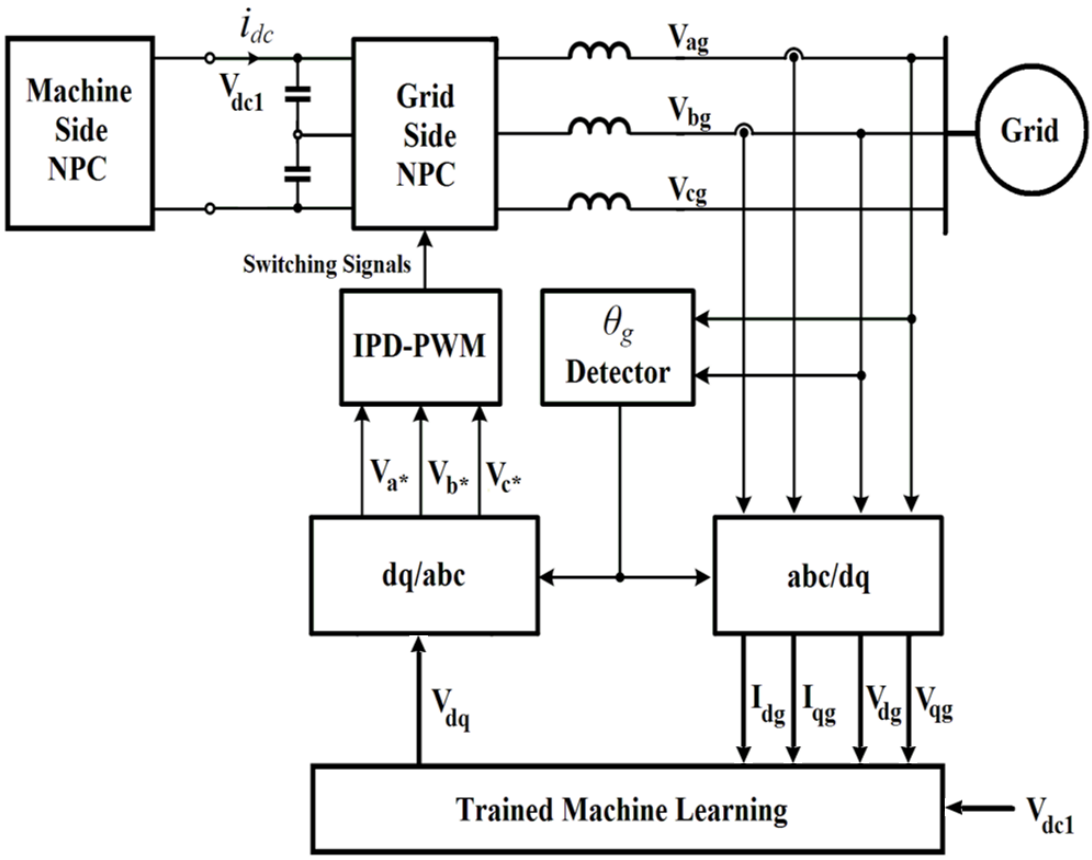


Fig. 3.4: System configuration for GSC control

Moreover, the proposed MARL diagram is depicted here, providing a clear representation of the overall procedure. Initially, the observations, which consist of sensor-measured signals, are carefully selected to ensure accurate input data for the agents. These are the problem state space and required information for the agents to make decision. Following this, the reward function is designed to reflect the objectives of the system, and the action space is defined to encompass the possible control actions for the power converters. Lastly, the implementation of the MARL is applied to the grid-side three-level NPC converter, where it functions as both the switching and

control method. The diagram in Fig. 3.5 visually illustrates the interactions between these components, showing the step-by-step flow of the proposed methodology.

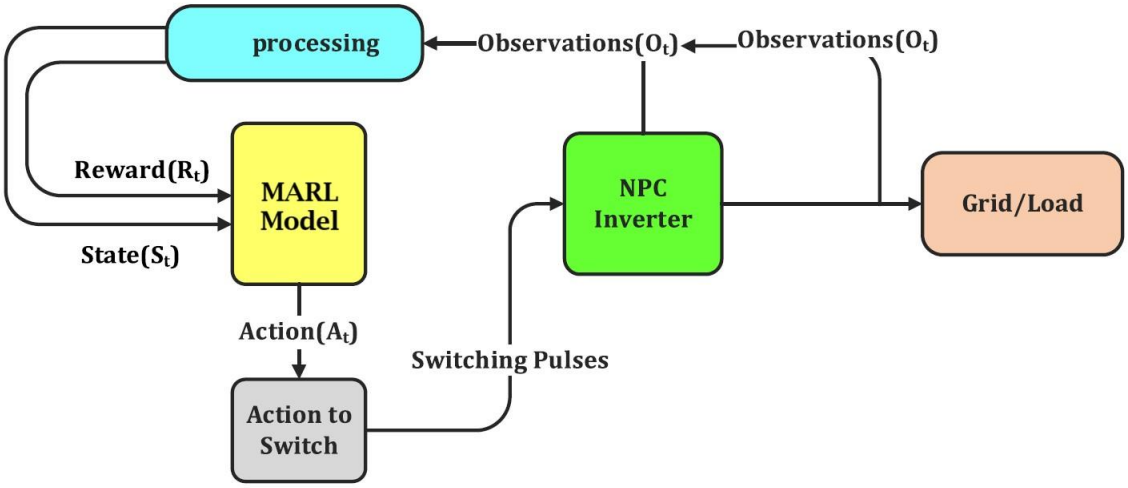


Fig. 3.5: Block diagram of the proposed MARL controller.

3.4 Simulation Results

In this section, a comprehensive simulation study is conducted to evaluate the effectiveness of the discussed ML-based control scheme for the BTB-NPC grid-connected inverter. Initially, the dc-link voltage and the voltage across capacitor C_1, which are controlled by a Proportional-Integral (PI) block in the field-oriented control of the machine-side converter, are depicted in Fig. 3.5. This figure also includes the 600 V reference for comparison purposes.

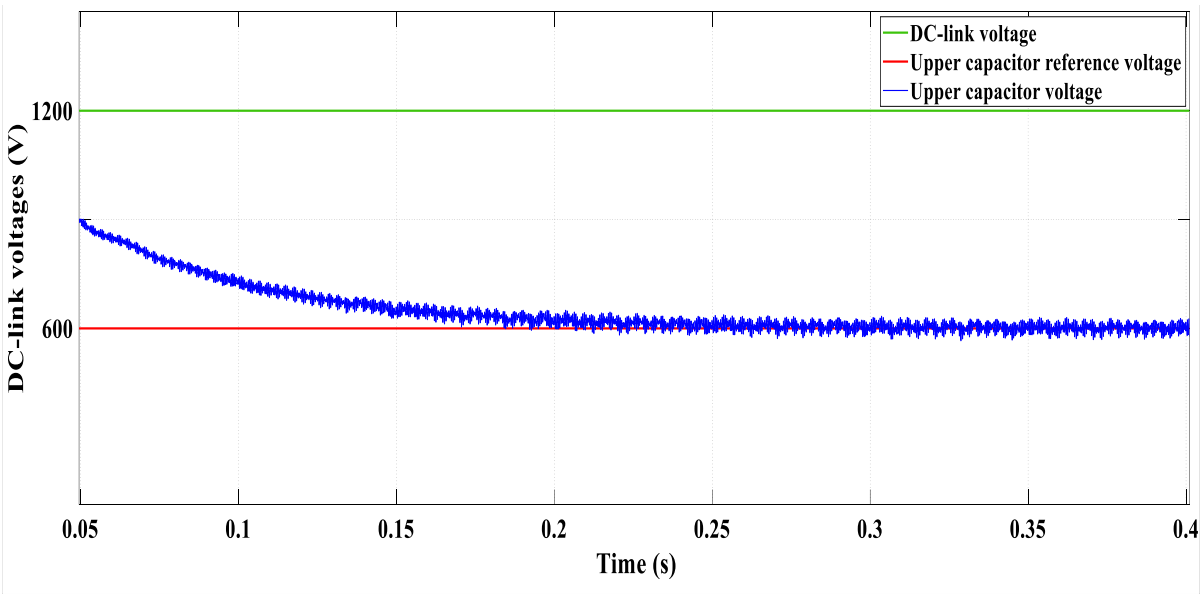


Fig. 3.5: The dc-link voltage, C_1 capacitor voltage, and the reference.

Then, the three-phase current of the grid side converter controlled by VOC is shown in Fig. 3.6. As noted earlier, the VOC is employed not only as a means of data collecting, also used for a better understanding of the functionality of the suggested solution, by preparing a comparative condition. In addition to the three-phase current, the THD analysis is performed to investigate the controller harmonic performance, as well. The THD in this instance was 3.79 %. These results are considered as the base study case of comparison and a tool for assessment of the efficacy of the four optimizers, including, SGD, MGD, BFGS, and L-BFGS. If the total harmonic distortion resulting from the ML model is found to be less than the THD acquired by the Voltage-oriented control, it will indicate an improved performance and thereby enhance the overall efficiency and power quality of the grid-connected WECS.

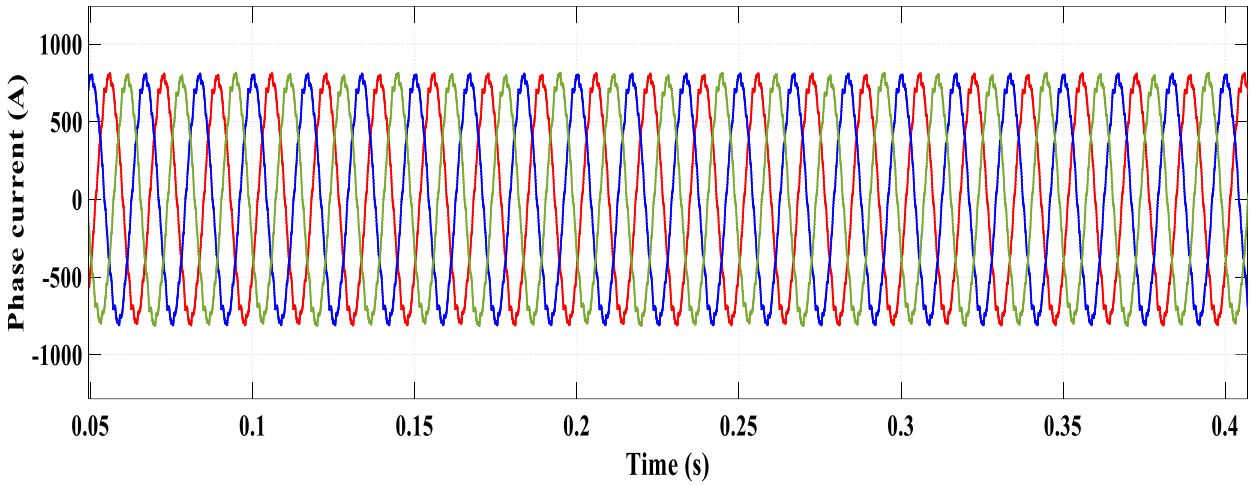


Fig. 3.6: Three-phase current of the grid side NPC converter controlled by VOC.

In this analysis, the three-phase balance condition is considered. The stochastic gradient descent optimizer is the first ML optimizer used to minimize the locally weighted polynomial regression cost function, trained by the VOC. Despite being one of the more traditional optimization algorithms, SGD demonstrates its effectiveness in this context. With a fixed step size of $\alpha = 0.001$, the training process took approximately 6 minutes. The accuracy of mapping the blind-testing data to the real output was calculated to be 96.12%, while the accuracy on the training dataset was 97.84%. These high percentages indicate that there is no overfitting, as the accuracy on the testing dataset is not significantly lower than that on the training dataset. The simulation results for the three-phase current of the grid-side converter are presented in Fig. 3.7(a), showcasing the system's successful performance. However, the system did experience a slightly higher THD of 3.93% compared to the one controlled by VOC. To further evaluate the SGD performance in this specific problem, the loss function value versus iterations has been plotted in Fig. 3.7(b). It is evident that after approximately 280 iterations, the response converges, demonstrating the effectiveness and convergence of the SGD optimizer in this scenario.

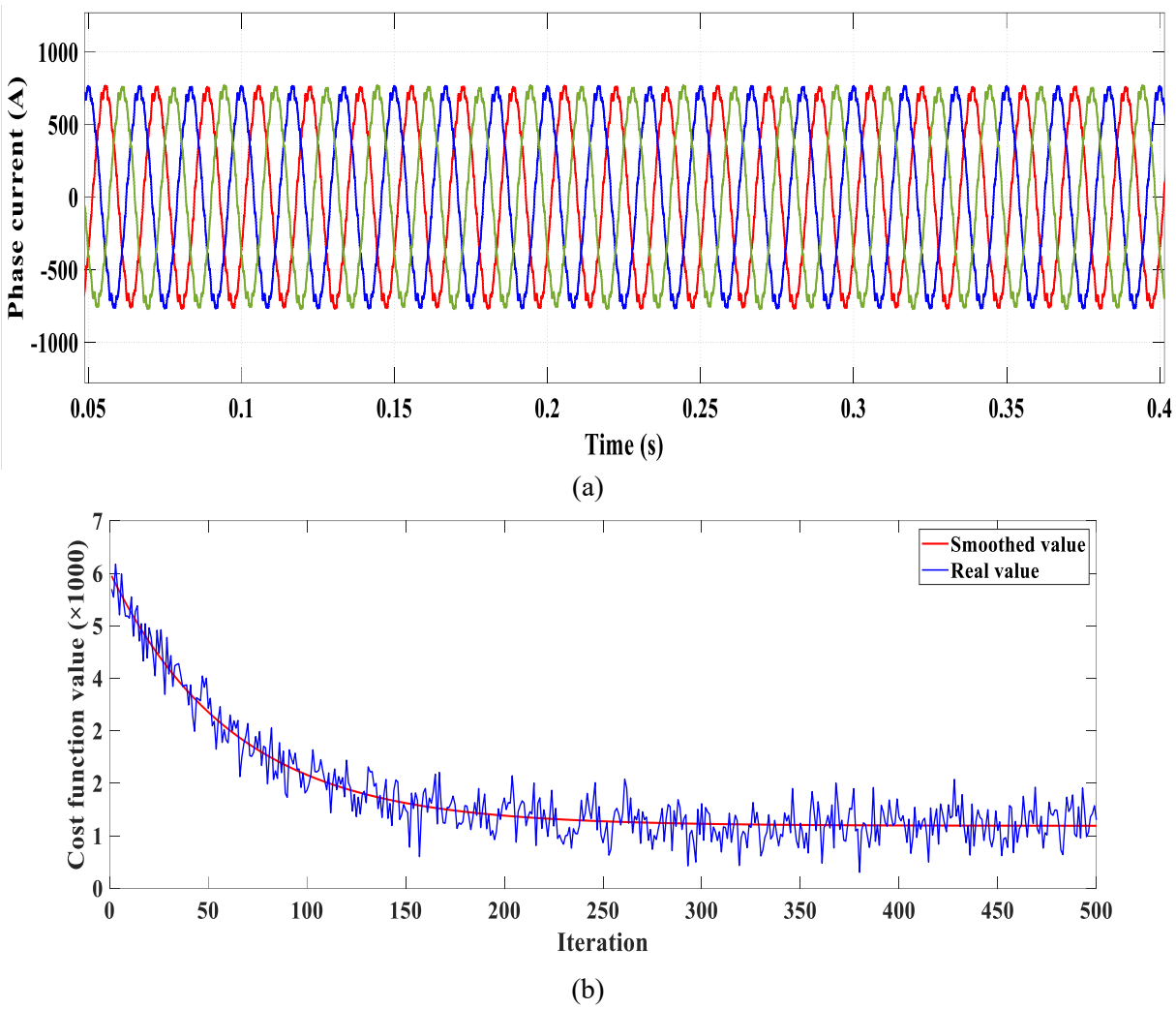
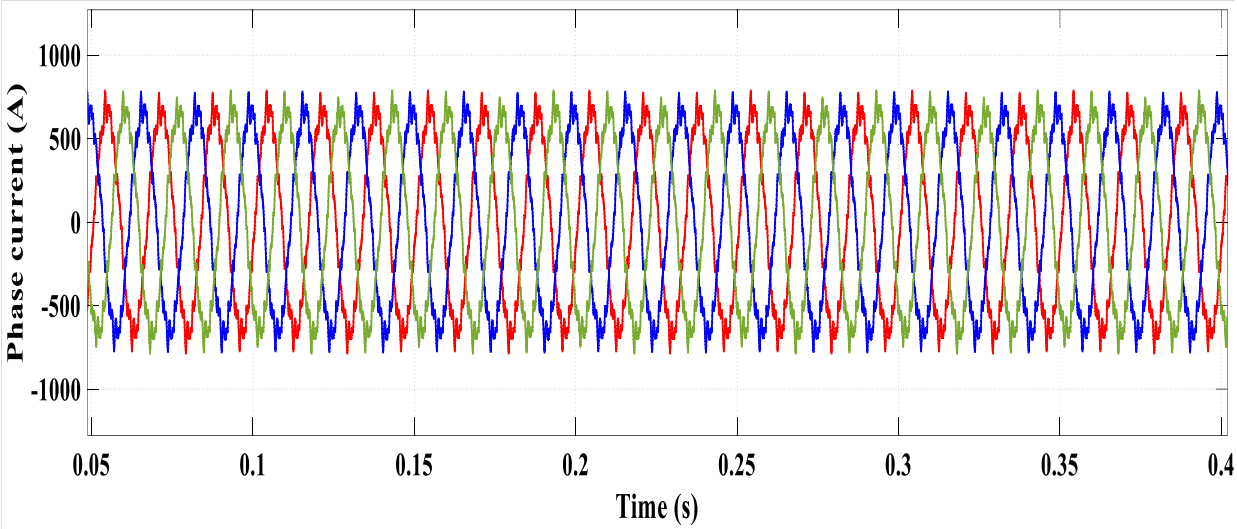


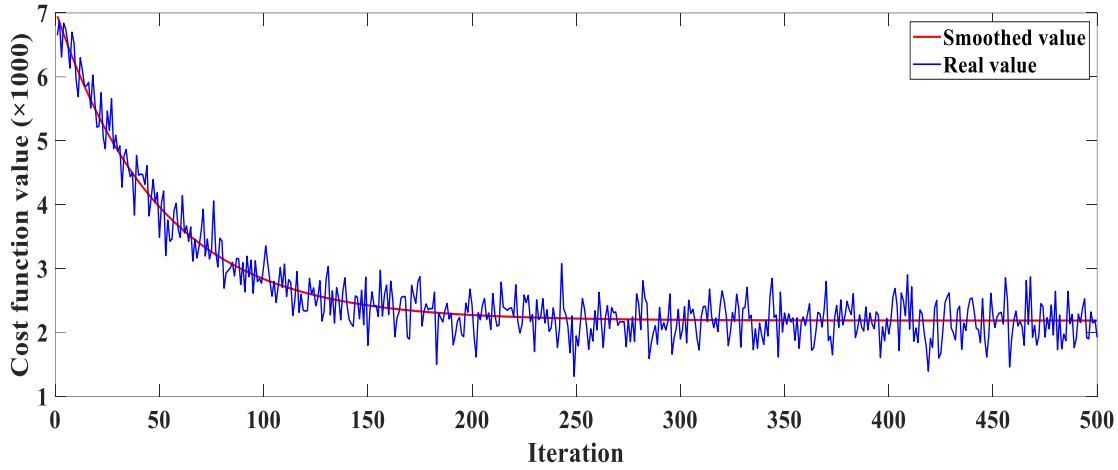
Fig. 3.7: Stochastic gradient descent results; (a) three-phase current, (b) convergence behavior.

For the analysis of the second optimizer, the MGD was chosen. Fig. 3.8(a) shows the three-phase output currents of the permanent magnet synchronous generator, while Fig. 3-8(b) illustrates the convergence behavior of the algorithm. Notably, a step size of 0.003 was selected for this analysis, using trial and error. Unlike SGD, MGD considers a batch of 64 data points in each iteration to update the parameters, instead of a single data, leading to a less training time of almost 4 minutes. However, this optimization method resulted in lower accuracies compared to SGD, with 89.89% accuracy on unseen data and 93.18% on training data. The THD also increased to 9.06%, indicating that BGD may not be a suitable candidate for this research problem as it fails to meet the THD criteria set for the grid, even in normal operation condition. This is further evidenced by the phase voltages shifting away from a sinusoidal pattern to a significant extent. Despite these

drawbacks, the convergence behavior of the BGD optimizer, as shown in Fig. 3.8(b), demonstrates convergence after approximately 200 iterations.



(a)

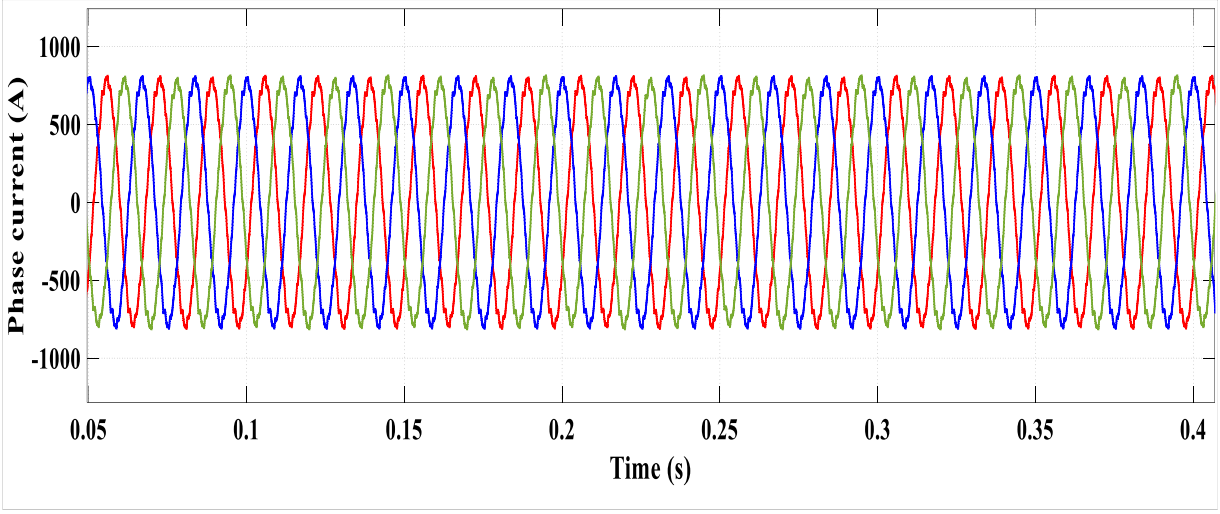


(b)

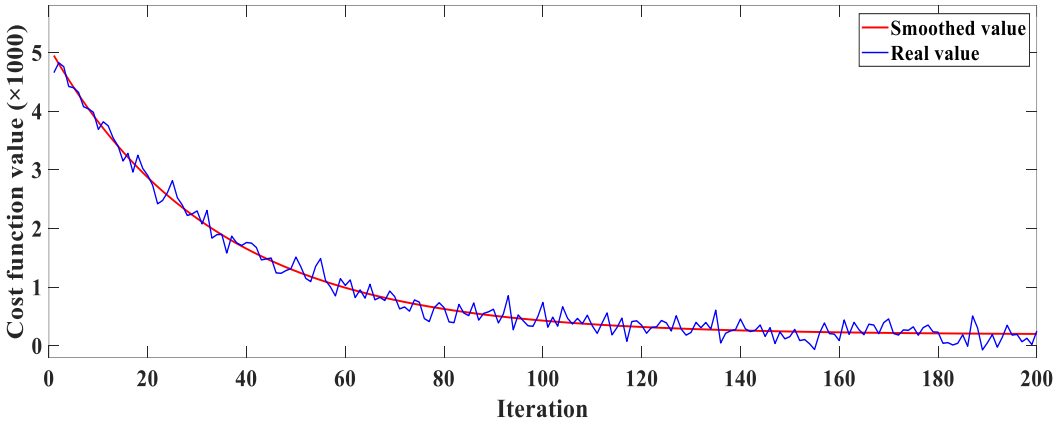
Fig. 3. 8: Batch gradient descent (BGD) results; (a) three-phase current, (b) convergence behavior.

The third optimizer considered is BFGS, a recent quasi-Newton based method discussed in the previous chapter. Unlike previous optimizers, BFGS does not require manual step size tuning. Instead, it performs a line search in the desired direction to determine the optimal step size for minimizing the loss function. The grid current waveform resulting from this optimization is depicted in Fig. 3.9(a). Compared to the SGD-optimized case study, the training time was significantly reduced to 2 minutes and 34 seconds, while simultaneously improving the accuracy

on both the testing and training datasets to 98.70% and 98.33%, respectively. These high accuracies indicate that the training process was conducted effectively, avoiding overfitting and ensuring that the algorithm can be applied to new, unseen data with almost identical outcomes. The THD analysis revealed a THD of 3.24%, which demonstrates the effectiveness of BFGS in minimizing harmonic distortions, and the performance surpasses the one controlled by VOC. This improvement in THD performance signifies the effectiveness of the BFGS optimization method in reducing harmonic distortions, leading to a more efficient and reliable grid-connected inverter system. The loss function values plotted against iterations in Fig. 3.9(b) provide a visual assessment of BFGS's performance in this specific scenario, showing convergence in approximately 140 iterations. The convergence for the BFGS algorithm is observed to be faster than that of SGD. As mentioned BFGS is a quasi-Newton method that approximates the Hessian matrix of second derivatives, allowing it to use curvature information to make more informed and efficient updates to the parameters. This leads to a faster convergence.



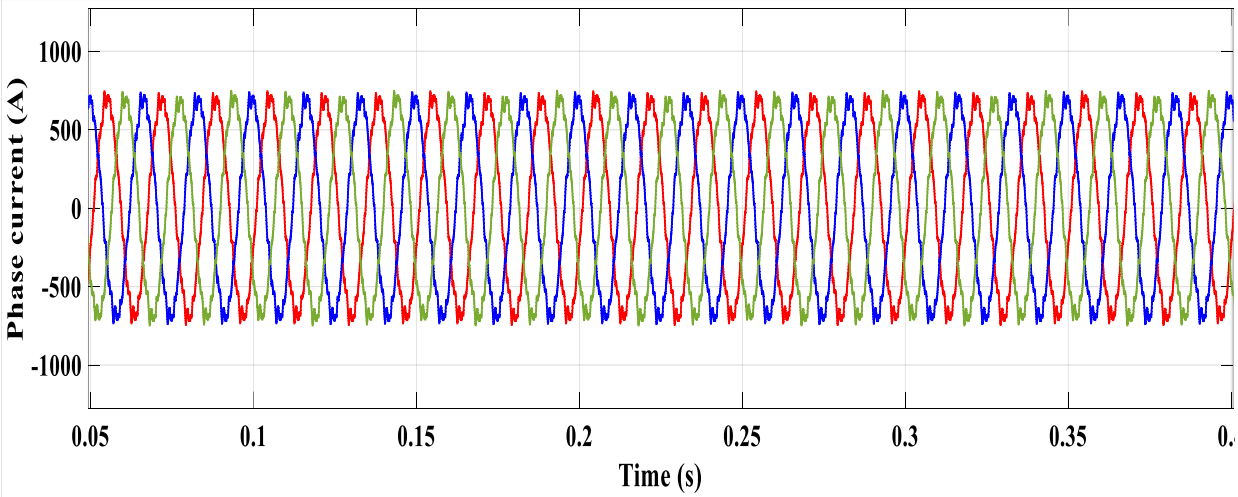
(a)



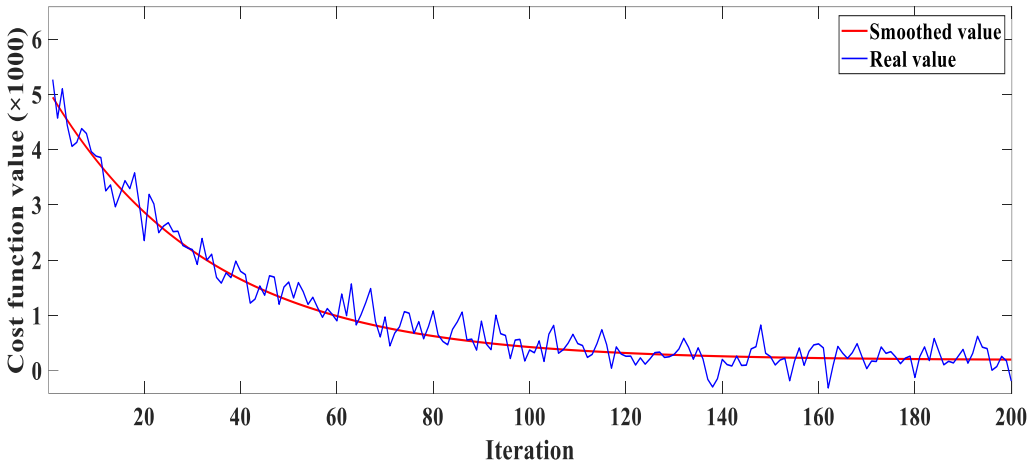
(b)

Fig. 3.9: Broyden–Fletcher–Goldfarb–Shanno results; (a) three-phase current, (b) convergence behavior.

The L-BFGS method, a variant of the BFGS algorithm, is known for its efficient memory usage, making it a popular choice in machine learning applications. L-BFGS navigates the variable space by using an estimate of the inverse Hessian matrix, similar to BFGS. However, unlike BFGS, which stores the entire estimate of the inverse Hessian, L-BFGS only stores a few vectors that implicitly represent the estimate. Fig. 3.10(a) displays the simulation current waveforms relevant to this optimization method. Despite its efficiency, L-BFGS required the least amount of training computation time at almost 2 minutes. However, compared to the simple BFGS, the accuracy and THD were slightly compromised, with a THD of 4.26% and accuracies of 97.00% and 94.61% on the training and testing datasets, respectively. The convergence behavior of L-BFGS, as shown in Fig. 3.10(b), is similar to that of the simple BFGS algorithm.



(a)



(b)

Fig. 3.10: Limited-Memory Broyden–Fletcher–Goldfarb–Shanno results; (a) three-phase current, (b) convergence behavior

To facilitate a clearer comparison, Table 3.2 is included. The table highlights that the L-BFGS optimization algorithm was the fastest, with a reduction of 2 minutes in training time compared to the other methods. However, the BFGS optimizer achieved the best performance in terms of THD and accuracy, performing better than the base case study conducted under VOC method. Also, compared to the base case study, the ML solution optimized by BFGS significantly reduced computational complexity. Furthermore, there is no longer a need to tune the PI controller parameters. Overall, the simple BFGS algorithm demonstrated effective functionality across all metrics, in normal operation condition of the PMSG-based WECS.

Table 3.2: Comparison Study

Optimizer	Complexity	Memory requirement (Mb)	Training time (s)	Accuracy on training dataset (%)	Accuracy on testing dataset (%)	THD (%)	Conversion iterations
SGD	low	1.18	366	97.84	96.12	3.93	280
MGD	Moderate	0.93	238	93.18	89.89	9.06	200
BFGS	Moderate	2.41	154	98.70	98.33	3.24	140
L-BFGS	high	2.07	117	97.00	94.61	4.26	140

To better understand the impact of pre-processing and the addition of Lasso regularization to the cost function, the problem was also solved using four different optimizers without these features. The results, presented in Table 3.3, show a significant reduction in training time and memory usage, which is expected since the cost function is simpler and no data preprocessing is required. However, this simplification led to decreased accuracy on both training and testing datasets. Additionally, the increased discrepancy between training and testing accuracies indicates overfitting, meaning the model performs well on training data but poorly on new test data. Furthermore, the system's performance in terms of THD was not acceptable across all optimizers. Therefore, it can be concluded that the inclusion of regularization and data processing enhances the machine learning model's accuracy and THD performance.

Table 3.3: Comparison study (non-regularized and non-pre-processed).

Optimizer	Memory requirement (Mb)	Training time (s)	Accuracy on training dataset (%)	Accuracy on testing dataset (%)	THD (%)	Conversion iterations
SGD	0.88	147	83.79	75.51	13.94	320
MGD	0.60	109	86.64	79.28	10.92	300
BFGS	1.12	77	92.48	85.33	6.78	190
L-BFGS	0.97	48	88.59	81.48	7.51	190

3.4.1 Simulation Results for the Proposed MARL Method

The ML models, although successful in normal operating conditions, was also tested under single-phase voltage drop scenarios. For this matter, the model resulted the best (optimized by BFGS) is considered to be analyzed under unbalance condition. However, the investigation revealed that the ML model was ineffective in handling unbalanced scenarios, as evidenced by the results shown in Fig. 3.11. Following the clearance of the single-phase voltage droop unbalance situation started at 0.1 sand cleared at approximately 0.15 seconds, the system was unable to return to its previous operating conditions under normal circumstances. This outcome suggests that the

ML model struggled to adapt to the unpredictable nature of unbalanced situations. Controlling wind energy conversion systems and power systems in unbalanced conditions is also a crucial point. Because these situations can occur frequently in real-world scenarios. Therefore, the proposed MARL-based approach optimally tuned by meta-learning method may offer a solution to this challenge by providing more adaptive and robust control in such scenarios.

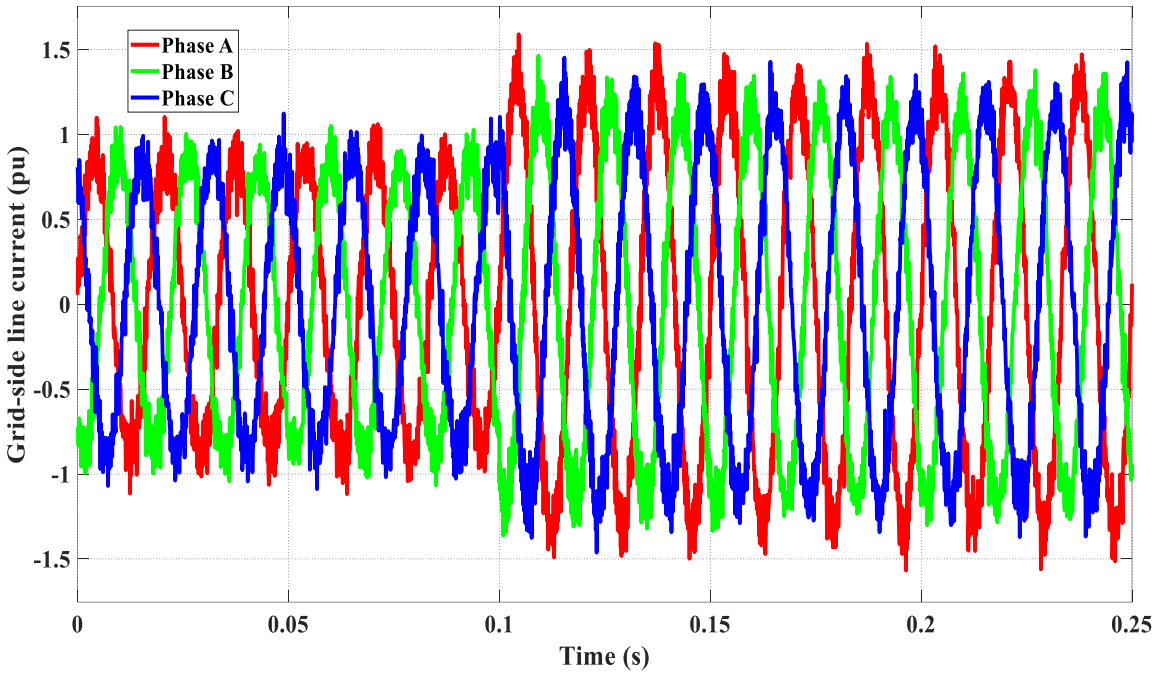


Fig. 3.11: Three-phase grid-side current under ML-based model controller for single-phase voltage drop unbalance condition.

Herewith, two MARL-based approaches are tested on a predefined WECS and their respective results have been graphically depicted below. The methods include MARL with constant discount factor DF within the learning and exploration process, and proposed MARL incorporates meta-learnt DF. Each method was carefully evaluated to determine the efficacy in optimizing the performance of the WECS. It is worth mentioning that the α and β are design to be 0.03, and 0.5, respectively, acquired by trial and error. To do so, two unbalance scenarios; single phase 90% voltage drop, and two phases voltage drop, have been decided to be examined. Afterwards, the functionality of the proposed method to control the grid-side NPC employed in PMSG-based WECS will be defined.

First Case Study; Single-Phase Voltage Drop

In order to approximate the real-world problem response behavior of the system, the voltage for phase-A declines to almost 10% of rated value throughout 0.06 s starting at t = 0.093 s, depicted in Fig. 3.12. Both MARL with constant DF and proposed MARL equipped with meta-learnt DF have been employed to observe the functionality of each, in this case study. The Fig. 3.13(a) and Fig. 3.13(b) illustrate the grid-side three-phase current for mentioned strategies, respectively. The maximum peak current alters to almost 1.35 pu since the grid-side NPC should convey the rated current as a reactive component to the grid in accordance with the low voltage ride through (LVRT) standards [105], within the unbalance interval. At any point, the current flowing towards the grid remained balanced for both methods, as illustrated. However, the constant DF have led to higher THD (4.65 % THD for fixed DF and 3.97 % for the meta-learnt DF) and higher current overshoot, as can be observed clearly from Fig. 7. This exhibit more rapid and precise transient responsiveness without notable overshoot of the proposed MARL strategy to control the grid-side NPC in a PMSG-based WECS.

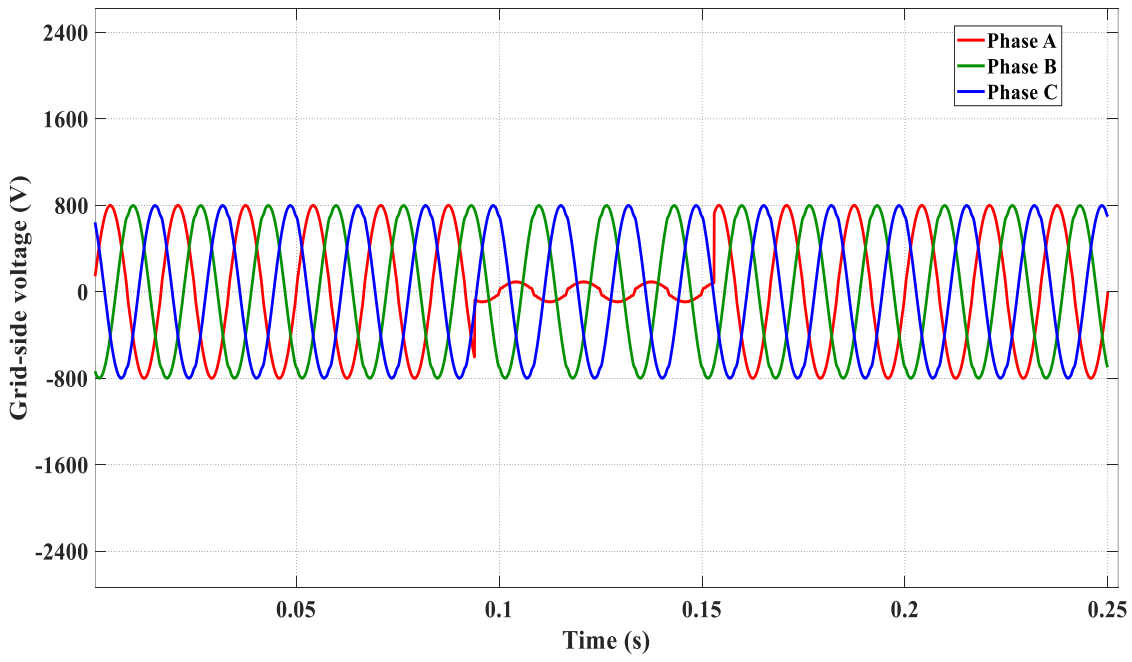
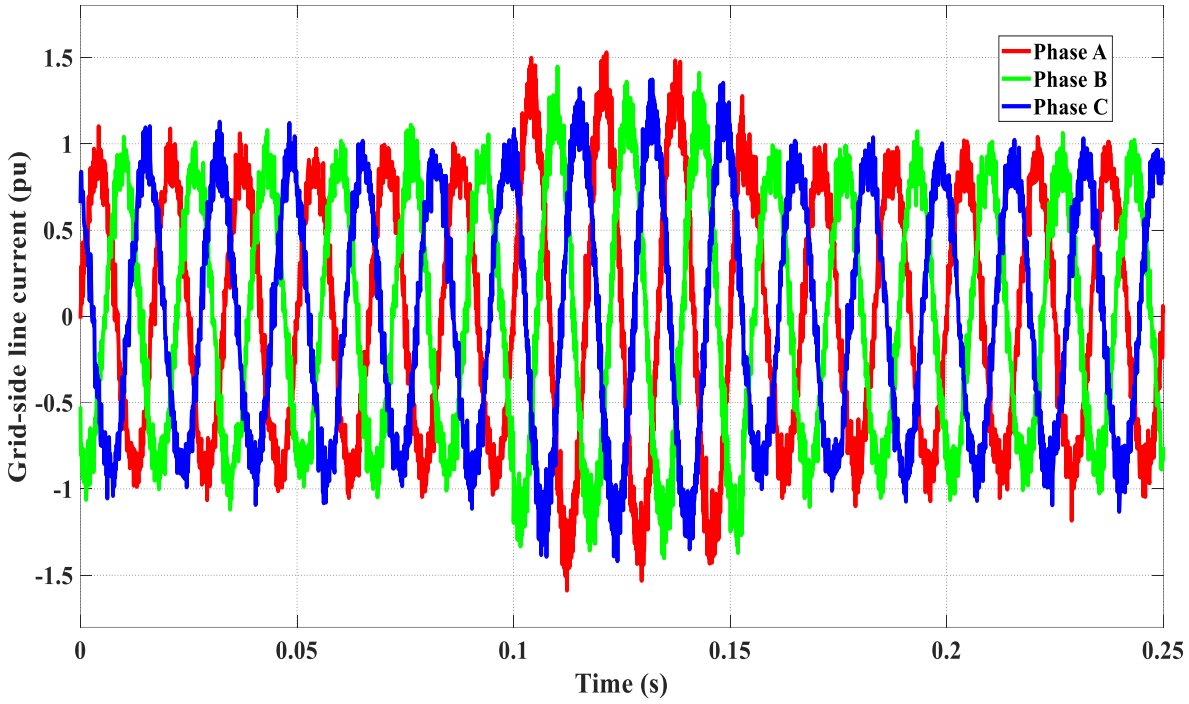
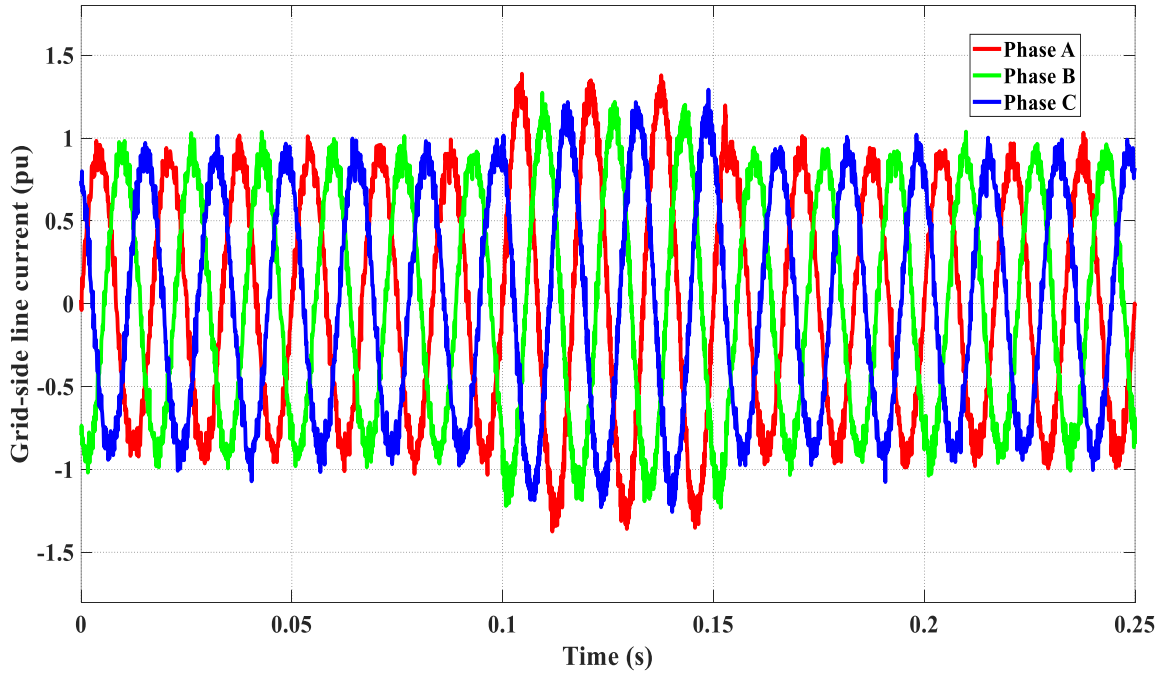


Fig. 3.12: Unbalance three-phase grid-side



(a)



(b)

Fig. 3.13: Three-phase grid-side current in single-phase voltage drop scenario; (a) MARL with constant DF, (b) proposed MARL employed meta-learnt DF.

As mentioned earlier, the WECS is designed to deliver 1.6 MW to the grid. To illustrate the performance of the proposed MARL approach with a meta-learned discount factor, Fig. 3.14 displays the average active and reactive power values. According to this figure, during steady-state operation, from 0 to approximately 0.1 seconds, the WECS operates at unity power factor. In this condition, the system achieves an average active power output of roughly 1.6 MW and an average reactive power output of 0 VAR supplied to the grid, indicating that all the power delivered is active power with no reactive component. However, when a single-phase voltage drop occurs, the scenario changes significantly. The active power transmitted to the grid drops to 0 W on average, while the reactive power increases substantially to nearly 1.6 MVAR. This change indicates the system's compliance with low voltage ride through (LVRT) standards, which require the system to support the grid by supplying reactive power during voltage sags. The specific voltage sag in this case is nearly 90%. After the unbalanced condition is resolved and the system returns to normal operation, the active power output increases back to 1.6 MW, and the reactive power output decreases to 0 VAR, resuming the initial steady-state performance. These results demonstrate the system's ability to meet LVRT requirements and its effective response to voltage disturbances, ensuring grid stability and reliability.

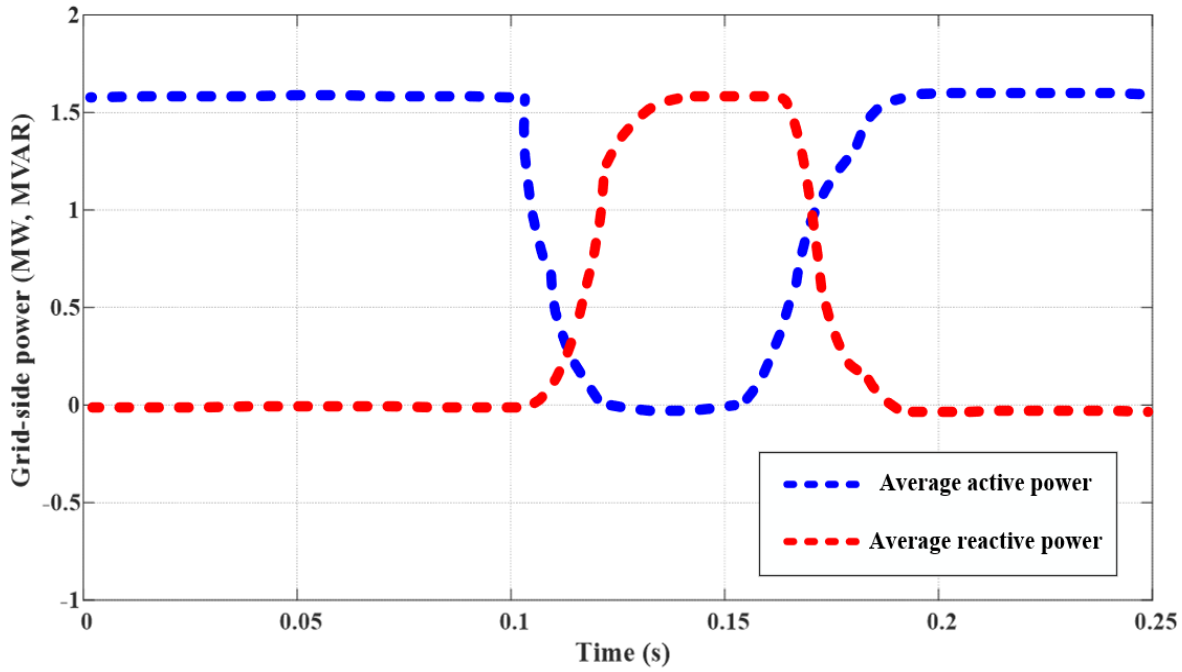


Fig. 3.14: Average active and reactive power delivered to the grid by proposed MARL controller model, in single-phase voltage drop scenario.

Second Case Study; double-Phase Voltage Drop

Figure 3.15 presents the simulation findings of the grid-side three-phase voltages for the second scenario. In this scenario, starting at 0.093 seconds, the voltages of phase A and phase B decrease to almost 60% of their rated voltage value over an interval of approximately 0.06 seconds. This voltage drop necessitates a response from the system to maintain grid stability.

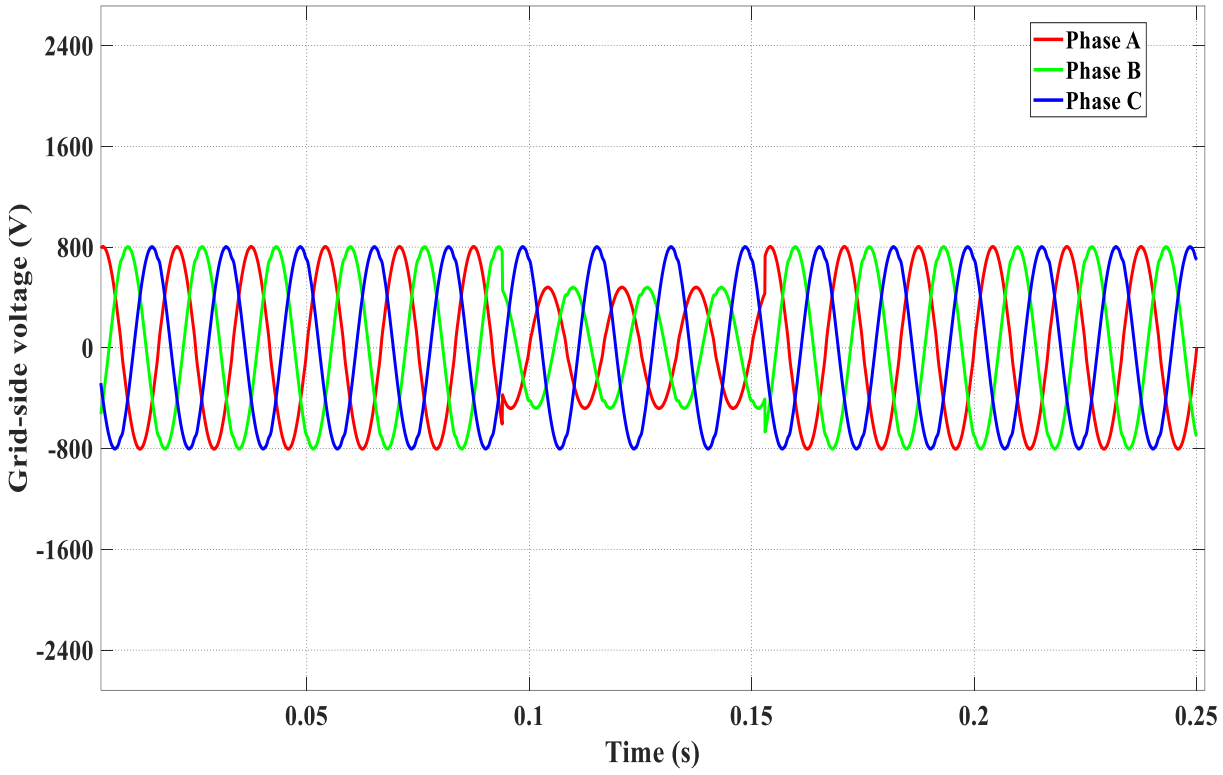
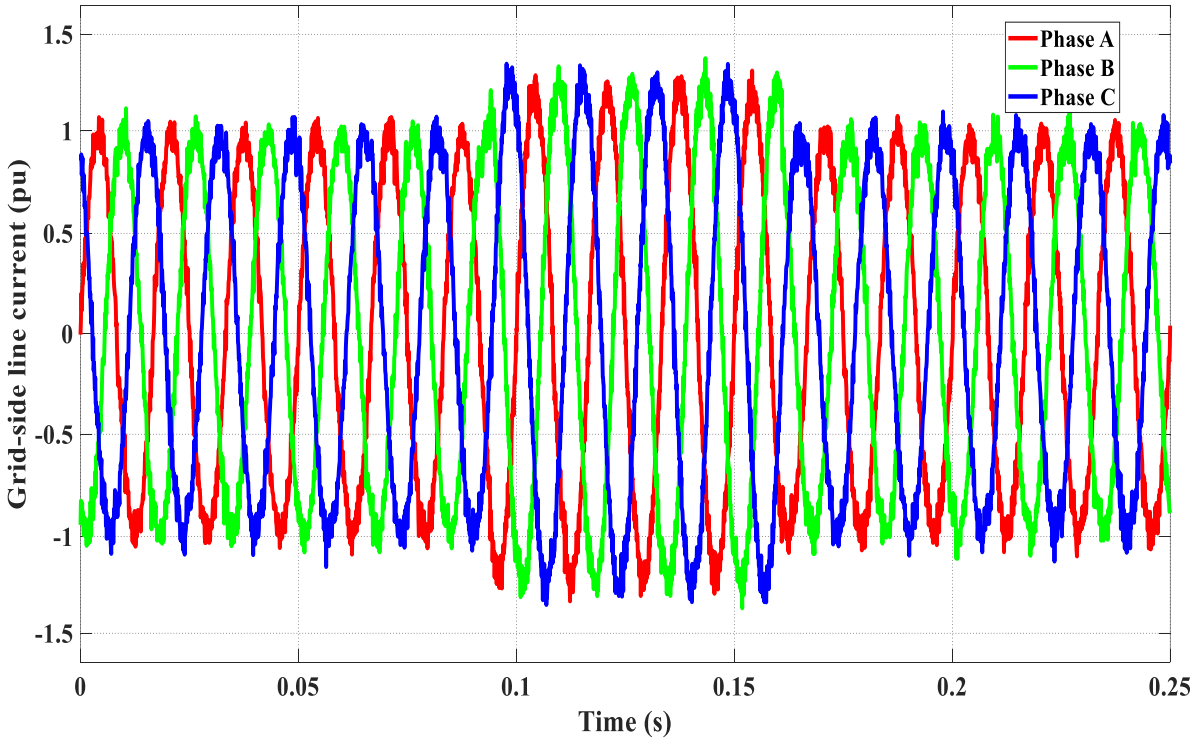


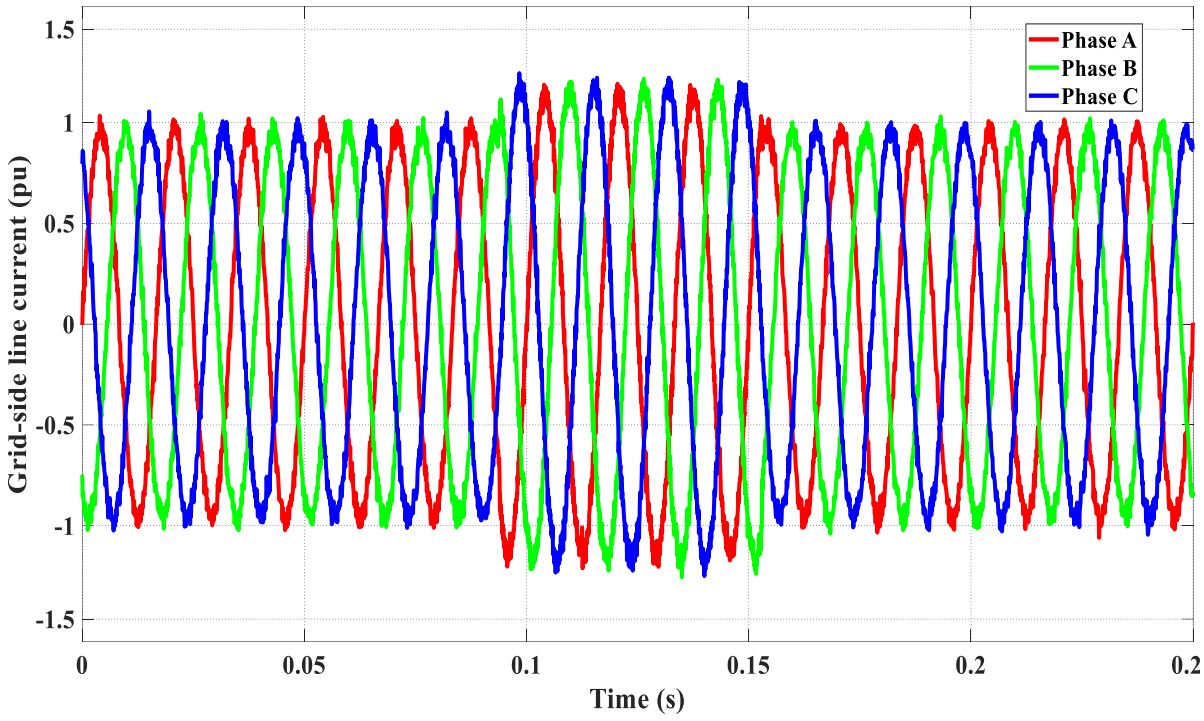
Fig. 3.15: Unbalance three-phase grid-side voltage in double-phase voltage drop scenario

During this period, the current adjusts to almost 1.3 per unit (pu) because the NPC power converter is required to provide at least 75% of the nominal current as a reactive component to meet the LVRT specifications. The remaining portion of the nominal current is then imposed as active current, ensuring that the system adheres to the necessary reactive power support during the voltage sag. In this instance, results similar to those for a single-phase grid voltage drop are demonstrated. According to the outcomes, the grid-side current remained balanced throughout the intervals for both the MARL approach with a fixed discount factor and the proposed MARL

approach with a meta-learned DF, as illustrated in Fig. 3.16 The proposed strategy exhibits almost no overshoot during the transition from normal operation to a two-phase voltage sag. Additionally, the THD performance is superior, with 4.11% THD for the fixed DF approach and 3.36% for the proposed model, as shown in Fig. 3-16(b). Moreover, the response time is quicker during the transitions compared to the fixed DF-based MARL, which can be seen in Figure 16(a). After the clearance of the unbalanced condition, the proposed strategy achieved normal operation almost instantly, whereas the fixed DF approach exhibited a delay of approximately 0.02 seconds. This rapid recovery and superior performance of the proposed MARL approach underline its potential as a promising alternative in managing grid stability during voltage disturbances. These results highlight the effectiveness and efficiency of using a meta-learned DF in improving the system's response to unbalanced grid conditions.



(a)



(b)

Fig. 3.16: Three-phase grid-side current in double-phase voltage drop scenario; (a) MARL with constant DF, (b) proposed MARL employed meta-learnt DF

In terms of the power delivered to the grid by the WECS under the proposed grid-side NPC power converter control method, both active and reactive power are supplied to the grid during the unbalanced interval of a double-phase voltage sag. This differs from the previously-studied scenario where only reactive power was fed to the grid during similar disturbances. Figure 3.17 illustrates the average active and reactive power delivered to the grid. In the steady-state period, the behavior of the power is identical to the one observed in the previous section following a single-phase voltage drop. Specifically, the system maintains stable power delivery with an average active power of approximately 1.6 MW and an average reactive power of 0 VAR. During the interval of the two-phase voltage drop, however, the average active and reactive power values change significantly. The average active power decreases to about 0.95 MW, while the average reactive power increases to approximately 1.1 MVAR. This adjustment in power levels indicates that the system is effectively providing the necessary reactive power support to meet the LVRT requirements during the voltage sag.

Once the unbalanced condition is cleared, the system quickly restores its former operating conditions. The active power output returns to 1.6 MW, and the reactive power output reverts to 0 VAR, indicating a swift recovery to normal operation without any voltage drop. This capability to regain normal operation promptly after clearing the unbalanced condition underscores the effectiveness of the proposed NPC power converter control method in maintaining grid stability and reliability under various voltage sag conditions.

Overall, these results highlight the robustness and efficiency of the proposed control strategy in ensuring higher power quality with less THD, and continuous and stable power delivery to the grid, even during significant voltage disturbances. The ability to supply both active and reactive power during two-phase voltage sags demonstrates the system's enhanced performance compared to scenarios where only reactive power was provided.

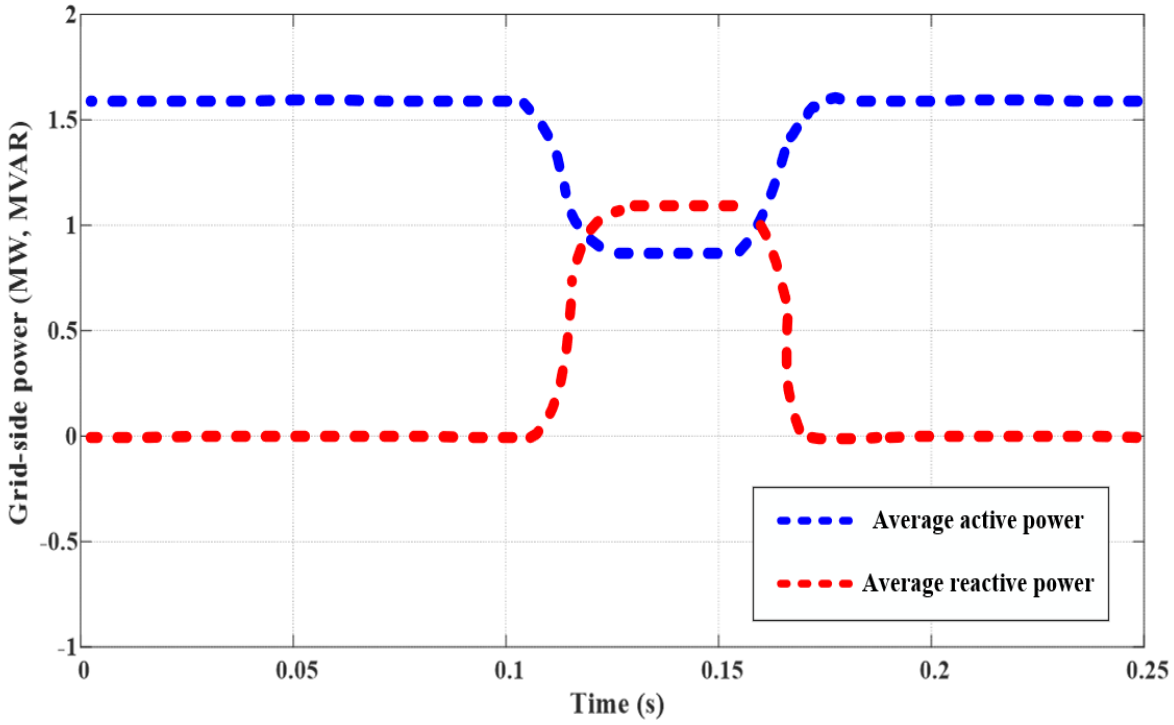


Fig. 3.17: Average active and reactive power delivered to the grid by proposed MARL controller model, in double-phase voltage drop scenario

3.4.2 Comparative study

Figure 3.18 is provided to highlight the recommended MARL approach in the context of learning and exploration behavior compared to the commonly-used MARL method with fix DF. The cumulative rewards of the prescribed MARL with a meta-learned discount factor are shown in green. Initially, the cumulative rewards for this technique are lower than those for the MARL instance where the constant DF is set to a value of 0.7, depicted in blue. This initial underperformance is attributed to the nature of MARL, which involves multiple agents learning to cooperate collaboratively to achieve a shared goal. This collaborative learning process can be problematic in the early episodes as the agents explore various strategies to find the optimal DF. The complexity of coordinating multiple agents results in slower initial progress compared to using a fixed DF. However, after approximately 120 episodes, the recommended MARL approach begins to outperform the conventional one. This improvement indicates that the agents have effectively learned to cooperate and optimize their strategies, leveraging the advantages of the meta-learned DF. The performance gains become evident as the cumulative rewards for the recommended approach surpass those of the fixed DF method. Additionally, the moving average of the cumulative rewards over the last 10 episodes is shown in solid red. This moving average provides a smoothed representation of the performance trends, highlighting the stability and consistency of the recommended MARL approach as it progresses through the episodes.

Overall, it the long-term benefits of the proposed MARL with a meta-learned DF. While it may initially lag behind due to the complexities of multi-agent learning, it ultimately achieves superior performance as the agents effectively adapt and optimize their behaviors. This highlights the potential of the recommended MARL approach to deliver better results in dynamic and complex environments compared to traditional MARL method with fixed parameters.

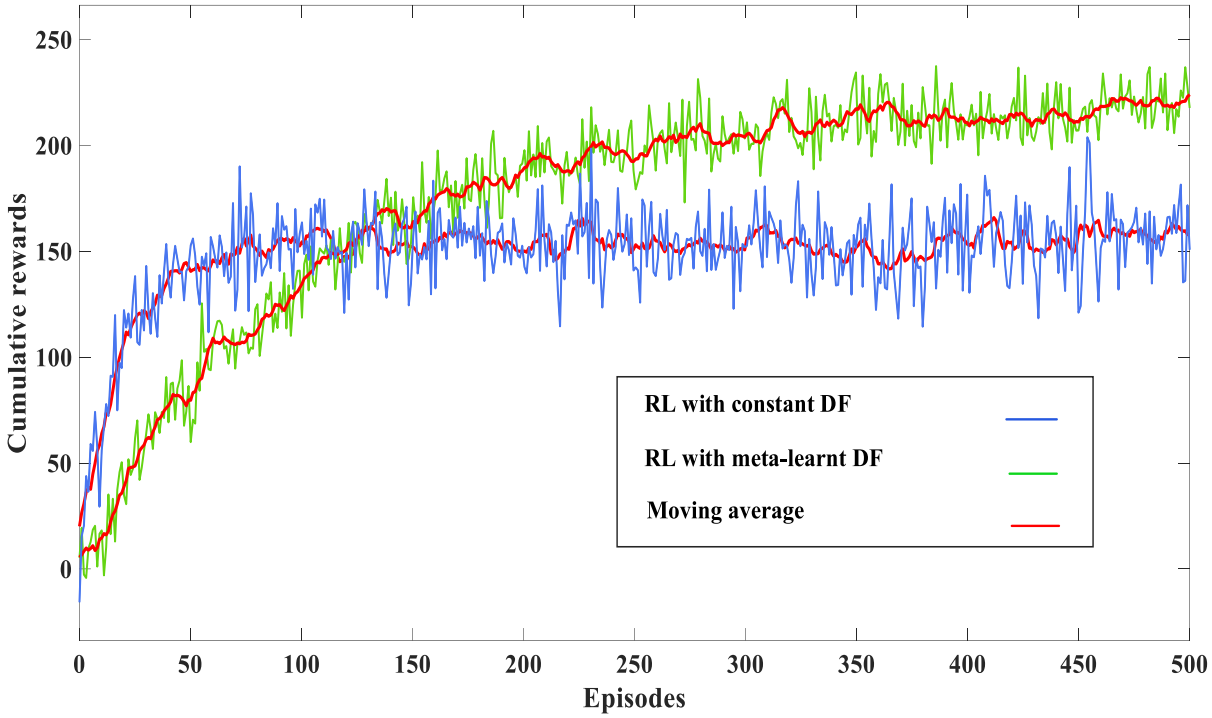
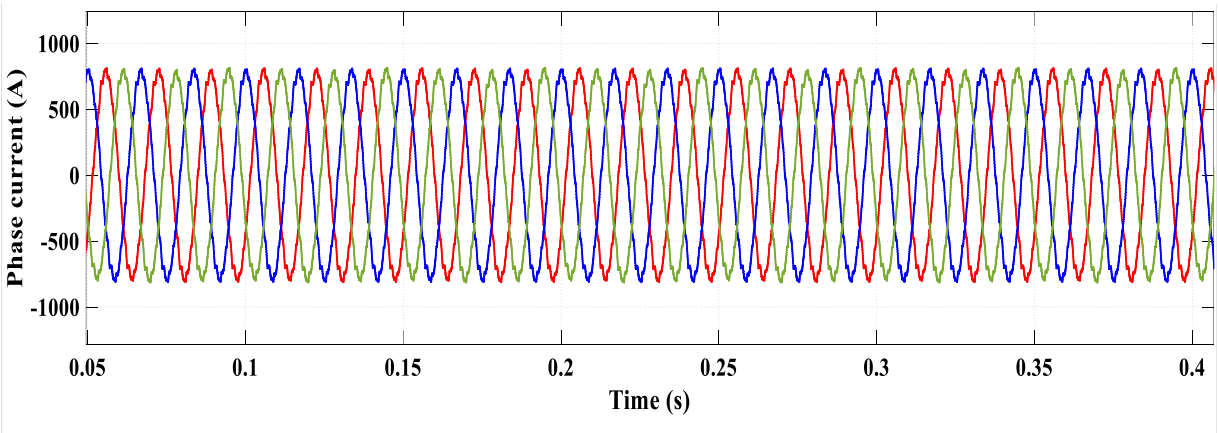
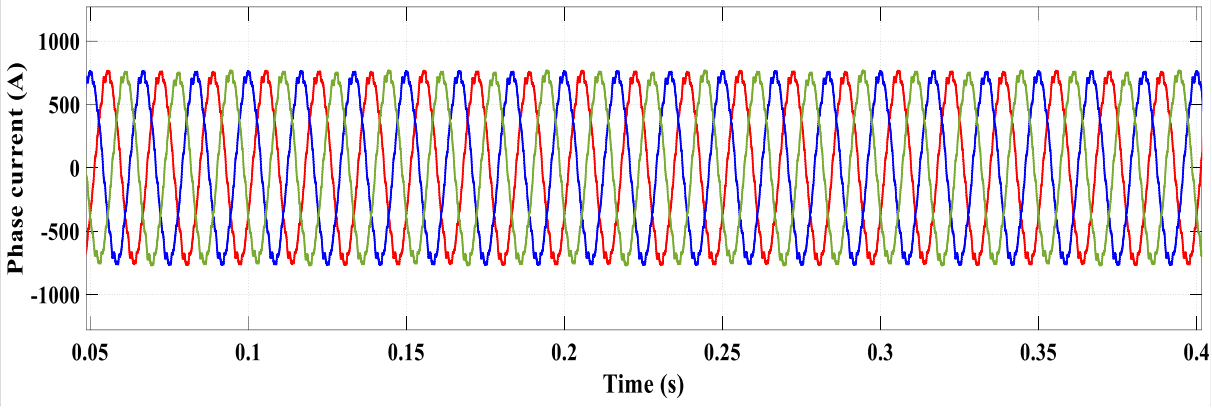


Fig. 3.18: Cumulative rewards in episodes for RL with constant and also meta-learnt DF.

The proposed method is also comprehensively compared to the VOC in many aspects. Hence, a normal operation condition is employed to analyze the performance of the proposed method and the VOC-based one, in the same WECS configuration. The three phase grid-side currents acquired by these two methods are demonstrated in Fig. 3.19. According to this figure, MRAL solution provide almost 0.4% improvement in THD performance, compared to the VOC.



(a)



(b)

Fig. 3.119: Three-phase current acquired by, (a) proposed MARL, (b) VOC.

Moreover, Table 3.4 highlights several key performance metrics for comparison purposes, including control strategy, complexity, adaptability, performance, response time, robustness, scalability, energy efficiency, hardware requirements, sensor requirements, THD, and so forth. In terms of efficiency, the 4% improvement by using MARL over VOC in the NPC power converter comes down to MARL's ability to dynamically adapt and optimize control under varying conditions. VOC is based on fixed parameters, which lead to inefficiencies such as increased harmonics, as mentioned, and increased reactive power. Augmented reactive power reduces the efficiency of a power system by elevating the total apparent power required without contributing to productive output. While reactive power supports the establishment of electric and magnetic fields in inductive and capacitive components, it does not generate useful work. As reactive power rises, the overall apparent power increases, yet the active power, which performs actual work, remains unchanged. This results in a lower power factor and efficiency, as more energy is consumed to sustain the system's operation without increasing useful power delivery. Furthermore, the increase in reactive power leads to higher losses due to resistive heating in conductors and transformers, compounding the reduction in overall system efficiency. MARL, on the other hand, learns and adjusts in real-time, reducing these inefficiencies by optimizing the converter's switching and minimizing harmonics, improving both power quality and overall efficiency. Moreover, VOC assumes linearity and ideal conditions, but real-world nonlinearities, such as switching variations and voltage imbalances, degrade its performance. MARL is better at

managing these nonlinearities by continuously adapting to system changes, ensuring closer operation to the converter's theoretical maximum efficiency.

Overall, the MARL method shows advantages over VOC, particularly in terms of adaptability, robustness, scalability, energy efficiency, and THD. However, MARL is more complex, requires higher memory usage, longer response times, compared to VOC. However, in a 60 Hz system, each cycle lasts approximately 16.67 ms. While VOC has a faster response time of 1.7 ms compared to MARL's 4.2 ms, MARL's response time is still sufficient to adapt effectively within a single cycle. With such a response time, MARL can respond to changes four times per cycle, allowing it to dynamically adjust control strategies in real time and maintain efficient performance.

Table 3.4: Comparison Study

Aspect	VOC	Proposed MARL
Control Strategy	Traditional control method using PI controllers for voltage and current regulation.	Advanced control method using multiple reinforcement learning agents for decision making.
Complexity	Moderate complexity. Requires tuning of PI controllers.	High complexity. Requires training of multiple agents and fine-tuning of learning parameters.
Adaptability	Limited adaptability. Requires retuning for different operating conditions.	High adaptability. Agents can learn and adapt to changing conditions.
Response Time	Fast response due to predefined control laws. (1.7 ms)	Require more time for agents to learn optimal actions. (4.2 ms)
Scalability	Limited scalability. Adding new components requires retuning.	High scalability. New agents can be added for new components.
Energy Efficiency	Efficient under well-tuned conditions (93%)	Potential for higher efficiency through continuous optimization (97%)
Sensor Requirements	Yes, requires accurate voltage and current sensors.	Yes, requires accurate voltage and current sensors
Simulation Time	Shorter simulation time. Well-defined control laws. (14s)	Longer simulation time. Involves extensive training of agents. (33s)
Reliability	High reliability in stable conditions.	High reliability in dynamic conditions due to continuous learning.
Ease of Use	Easier to use with well-documented methods and guidelines.	Requires expertise in reinforcement learning and control systems.
THD	Moderate THD. (3.79%)	Lower THD due to optimized control. (3.36%)
Memory Use	Lower memory usage. (1.6 MB)	Higher memory usage due to agent storage and training data. (3.4 MB)
Computational Requirements	Moderate computational requirements. (Standard CPU)	High computational requirements. (High-performance GPUs or CPUs)

3.5 Conclusion

This chapter utilized a novel reinforcement learning based approach for controlling the grid-side neutral point clamped power converter used in direct-drive permanent magnet synchronous generator-based wind energy conversion systems. The innovative technique utilized a multi-agent reinforcement learning strategy, detailed in previous chapter, which provided several advantages over previous machine learning and RL-based methods. These advantages included enhanced adaptability and flexibility, as well as the ability to optimize the system online continuously. One of the key benefits of the MARL approach was its online learning capability, which allowed for the continual improvement of system performance as it adapted to changes in environmental conditions and system dynamics. The proposed MARL method could handle both single-phase and double-phase voltage sag unbalance conditions, marking a significant improvement over traditional ML-based method. By dividing the problem into smaller sub-tasks and assigning them to different agents, the proposed MARL approach enabled more scalable and efficient problem decomposition and resolution compared to single-agent reinforcement learning. This division allowed for the exploration of various strategies and policies, leading to better overall performance and increased robustness in the face of changing environmental conditions or system dynamics.

To further enhance the performance of the proposed control approach, a meta-learning technique was employed to optimally learn the discount factor value, a crucial hyperparameter of the reinforcement learning algorithm. This meta-learning method was found to be superior to fixed DF values or other conventional methods of tuning the discount factor in terms of adaptability, cumulative rewards, total harmonic distortion, and convergence rate.

Evaluation and simulation results demonstrated that the proposed method outperformed both traditional and recently researched control strategies, establishing it as a strong competitor in the field. The combination of MARL and meta-learning techniques resulted in a control approach that was not only more adaptable and flexible but also more efficient and robust, making it a significant advancement in the control of grid-side NPC power converters in PMSG-based WECS.

Chapter 4
Reinforcement Learning-based Maximum Power Point Tracking Approach for PMSG-based WECSs

4.1 Introduction

Wind energy conversion systems have been instrumental in harnessing renewable energy, yet they often face challenges with conventional maximum power point tracking methods. These traditional approaches can suffer from slow tracking speeds, poor precision, and a high sensitivity to fluctuations in wind speed. To address these issues, this chapter employs the cutting-edge multi-agent reinforcement learning strategy explained in chapter 2, to optimize MPPT in variable WECS. The proposed MARL approach distinguishes itself by significantly improving energy output and enhancing the system's ability to swiftly respond to wind speed variations. Unlike traditional single-agent methods, the MARL approach employs multiple agents working in tandem. This decentralized cooperation allows the system to achieve superior precision and interaction capabilities, thus maximizing power generation more effectively. One of the key innovations of this approach is the incorporation of a meta-learned discount factor. This enhancement optimizes the MARL algorithm, reducing the learning phase duration and improving the convergence rate. The result is a more robust and efficient solution that adapts quickly to changing wind conditions, ensuring consistent and reliable performance.

Extensive simulations provide a comprehensive performance assessment of the proposed model-free MARL approach. The results highlight its potential applicability and underscore its advantages over conventional methods. These simulations demonstrate the MARL strategy's ability to maintain high energy output and quick adaptation to wind speed changes, showcasing its practical benefits. To further validate the effectiveness of the proposed algorithm, a 1000W prototype was implemented. This real-world application confirms the functionality of the MARL strategy in wind MPPT scenarios, providing tangible evidence of its superiority. The prototype testing reveals that the MARL approach not only meets but exceeds the performance of traditional MPPT methods, offering a reliable and efficient solution for wind energy systems.

4.2 Literature Review

Given the growing prices and detrimental effects of fossil fuels on the environment, there is an urge more than ever for viable renewable energy alternatives. In this regard, wind energy has experienced growth and is well-positioned to emerge as a significant sustainable and economical source of power [106]. Although, in order to boost the efficiency, a maximum power point tracking approach is required to extract maximum power from the wind energy conversion system. Three of the main varieties of MPPT methodology for WECS include indirect power control (IPC), direct power control (DPC), and smart or artificial intelligence (AI) based schemes. The most popular IPC-based conventional MPPT approach are tip speed ratio (TSR) [107], optimal torque (OT) [108], [109], and power signal feedback (PSF). These methods may not always be as efficient as direct methods in extracting maximum power. furthermore, they can be sensitive to inaccuracies in the wind turbine model parameters, which affect the performance and reliability. Also, major performance downsides might arise from any mismatch between the predicted and actual characteristics for WECS [110].

DPC is another widely recognized primary categorization of MPPT scheme. Among the most commonly-used DPC-based sub-approaches, incremental conductance (INC) [111], and optimum relation based (ORB) [112], can be named. These approaches suffer from fluctuations around maximum power points, lower convergence rate, and also requirement of additional sensors and equipment, which raises the initial, operation, and maintenance costs [113]. The Perturb and Observe (P&O) algorithm is another widely used DPC-based method for MPPT purposes [114]. Its simplicity and ease of implementation are notable advantages, making it a popular choice. However, P&O has shortcomings, such as struggling to accurately track the MPP under rapidly changing wind speeds. While P&O is a commonly used method, there is ongoing research to improve the performance.

There are also methods such as; whale optimization, and droop-based control, that are not popular because of global optimum and power fluctuation issues [115], [116].

Obstacles encountered in above-mentioned strategies for nonlinear and unpredictable WECS have given rise to the popularity of AI or learning control-based methodologies including fuzzy logic, artificial neural networks, model predictive control, and reinforcement learning. In an attempt to address the shortcomings, in [117] and [118] FL-based strategies were considered to

control the current loop for MPPT purposes. Or a direct drive permanent magnet synchronous generator based WECS was implemented in [119] to deal with MPPT challenge by FL for wide range of wind speeds. However, primary drawback of FL as a state feedback-based controller, is need of comprehensive data measurement in addition to thorough dynamic model of the system. Moreover, literature survey on FL-based MPPT predominantly concentrates on refining dynamic performance of WECS, while less attention has been paid to investigation of optimal performance attributes [120].

Therefore, it turns out, more intelligent optimization methods may offer more reliable and consistent results for MPPT, while reduce sophistication and provide a model-free solution [121], [122] Reference [113] designed a MPPT technique to identify the optimal relationship between the rotor speed and permanent magnet generator power using ANNs. The procedure switched from the ANN to the ORB since MPP was attained. References [123] and [124] employed an ANN and deep neural network approaches which allowed for adjusted electromagnetic torque so that the system could enforce the generator speed to follow the reference speed and reference power, respectively. But, ANN-based solutions carry some problematic concerns. Large-scale training datasets are necessary for ANNs, and this might not always adequately capture the intricate and dynamic wind patterns.

On the other hand, an enhanced MPC strategy was introduced for a back-to-back (BTB) three-level neutral point clamped (NPC) converter in [125], designed for high-power PMSG-based WECSs. This approach incorporated virtual space vectors and utilizes sensors to measure machine side electrical parameters along with rotor position. However, two major disadvantages of MPC-based methods are the computational complexity and sensitivity to model inaccuracies [126].

The shortcomings for neural network and MPC based MPPT solutions motivated researchers to move towards RL-based MPPT. RL is a machine learning paradigm where an agent learns to make decisions by interacting with an environment. RL can adapt to changing environmental conditions and optimize the turbine's performance without requiring detailed models of the system. In this regard, [127] proposed a variable-speed WECS intelligent MPPT method based on RL. By updating the action values in line with the rewards earned, the controller applied a model-free Q-learning approach to continually gain a mapping policy from states to optimal control actions. Even though the convergence rate and exploration were not optimal yet, [128] suggested a

maximum detection method, a component of Q-learning, that accelerates real-time tracking by pushing the WECS towards MPPT performance. Reference [129] is concerned with the problem of MPPT for a permanent magnet vernier generator based WECS using the Q-learning with the finite-time control. The major disadvantages of such single-agent RL for MPPT problems are the limited ability to explore and exploit the vast and complex search space effectively, and slow learning rates as it relies on a single agent to adapt to changing environmental conditions.

In this research endeavor, a novel approach employing multi-agent reinforcement learning is presented to address the MPPT limitations encountered in PMSG-based WECSs employed with BTB-NPC power converter, illustrated in previous chapter. In comparison to traditional methods, the proposed approach enhances energy output and the ability to swiftly respond to changes in wind speed. Furthermore, due to its decentralized nature, multiple agents collaborate to maximize power generation, ensuring superior precision and enhanced interaction capacity compared to single-agent RL methods. Additionally, by incorporating a meta-learnt discount factor (DF), a pivotal hyperparameter in the RL process, the MARL algorithm improves learning phase time and convergence rate, resulting in a more robust solution. Once the optimal rotor speed-power (ω_r, P_e) curve is determined throughout the proposed strategy, the method utilizes P&O to accurately pinpoint the MPP. Extensive simulation results demonstrate the performance of the model-free MARL approach, indicating its potential applicability. Moreover, a 1000W prototype validates the functionality of the algorithm in real-world wind MPPT problems.

4.3 Proposed MARL Application in MPPT

Illustrated in Fig. 4.1, proposed MARL approach involves two distinct phases within MPPT; online application phase and online learning phase. In online learning phase, controller serves as agents, engaging with the environment to acquire knowledge of maximum power points through Q-value. These MPPs are then utilized to derive optimal rotor speed and electrical power profiles, which are subsequently employed for P&O maximum power point tracking control over online application phase.

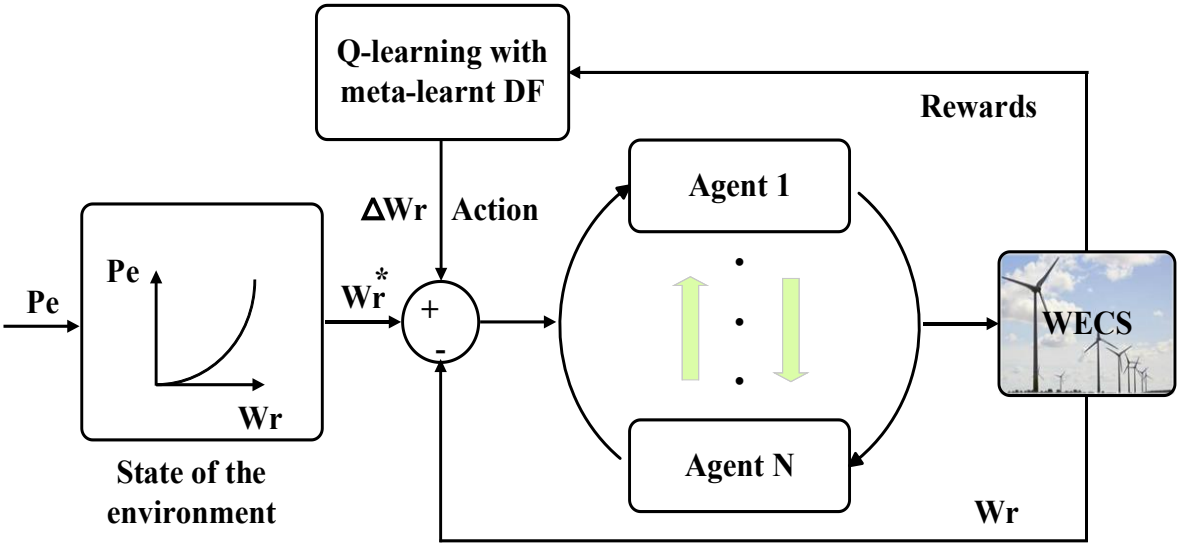


Fig. 4.1: The concept of the proposed MARL for MPPT scheme of WECS

During the initial online application phase, it is crucial to clearly define the concepts of state and action spaces, along with the reward, when employing proposed algorithm for determining the optimal $\omega_r - P_e$ curve. The controllers acting as agents in the WECS identify the current state and choose a discrete action based on Q-values associated with all viable actions using Q-function defined in chapter 2. Subsequently, the agent receives a reward, to adjust the values of the executed actions in previous state.

To determine the problem control strategy and the state space for the proposed method, PMSG modeling, as well as the overall mechanical power of the WECS should be pre-defined. For the PMSG, the following equations would describe the dynamics:

$$v_d = Ri_d + \frac{L_d di_d}{dt} - L_q i_q \omega_s \tag{4.1}$$

$$v_q = Ri_q + \frac{L_q di_q}{dt} + (L_d i_d + \varphi_m)\omega_s \tag{4.2}$$

In which, R is the stator resistor, v_d and v_q are the stator voltages after Park's transformation, L_d and L_q denote the inductances, stator pulsation and the flux are shown by ω_s, and φ_m.

As the WECS is considered to be grid-connected, the nonlinear current equations for PMSG would be:

$$i_d = -\frac{1}{L_d+L_l}\left[(R+R_l)i_d + P\left(L_q+L_l\right)i_q\omega\right] \tag{4.3}$$

$$i_q = -\frac{1}{L_q+L_l}\left[(R+R_l)i_q - P(L_d+L_l)i_d\omega + P\varphi_m\omega\right] \tag{4.4}$$

The R_l and L_l are the load resistor and inductance. And the ω is considered to be the angular speed. Considering $L_d = L_q$, for the PMSG with p poles, the electromagnetic torque, and subsequently the power output can be presented as:

$$T_{em} = p\varphi_m i_q \tag{4.5}$$

$$P = T_{em}\omega_r \tag{4.6}$$

In terms of wind turbine modeling, the turbine rotor radius size (R), air density (ρ), speed of the wind (V_w), and the conversion efficient of the generator (C_p) are the characteristics that impact the power output shown by:

$$P_{WT} = \frac{1}{2}\rho\pi R^2 V_w^3 C_p \tag{4.7}$$

The C_p is correlated to TSR of λ and pitch angle of θ:

$$C_p = 0.5 - 0.0167(\theta - 2)\, sin\left[\frac{\pi(\lambda+0.1)}{18-0.3(\theta-2)}\right] - 0.00184(\lambda - 3)(\theta - 2) \tag{4.8}$$

Hence, the maximum power can be extracted from the wind, when the optimum tip speed ratio of λ_{opt} is reflected upon:

$$P_{max} = \frac{1}{2}\frac{\rho A R^3 C_{pmax}}{\lambda_{opt}^3}\omega_t^3 = k\omega_t^3 \tag{4.9}$$

In (4.7), the A is the area of the swept area of the turbine blades, and ω_t considered as the WT rotation speed. Then, for each wind speed, there exists a corresponding $\omega_r - P_e$ curve. Within this curve, there is a singular rotor speed that is optimal for maximizing wind power extraction. Therefore, the state space can be defined effectively as a pair representing the operating point of speed and power, as follows:

$$S = \{s | s_{k,j} = (\omega_{r,k}, P_{e,j})\} \tag{4.10}$$

It is worth mentioning that $k \in [1, 2, \dots, K]$ and $j \in [1, 2, \dots, J]$, denote the total number of uniformly divided segments across the entire range of ω_r and P_e, respectively.

Under the condition that the constraints on rotor speed ω_r are observed and not surpassed, the agent has at its disposal three distinct action possibilities for each state: augmentation ($\triangle \omega_r$), reduction ($-\triangle \omega_r$), and maintaining an idle state (0). Driven by the evaluation of the current state represented by (ω_r, P_e) and the prescribed action policy, the agent strategically chooses an action from the designated action space. This decision, is geared towards optimizing the adjustment of the rotor speed. In this regard, the action space is demonstrated as follows:

$$A = \{a| +\triangle \omega_r, -\triangle \omega_r, 0\} \tag{4.11}$$

After accomplishing a given task, the agents receive a reward to evaluate the effectiveness of the chosen action. The reward function is crafted to incentivize the agent in actively pursuing MPPT while steering clear of operating the turbine at points where power output falls below the optimal threshold. In this context, the reward function may be conceptualized and formulated as follows:

$$R_{t+1} = \begin{cases} +1 & when\ P_{e,t+1} > P_{e,t} \\ 0 & when\ P_{e,t+1} = P_{e,t} \\ -1 & when\ P_{e,t+1} < P_{e,t} \end{cases} \tag{4.12}$$

Simply, if the specified action makes P_e to climb, the agent would attain a positive reward; however, if it causes the P_e to decrease, compared to the previous stage, the agent will get a negative reward which can be named as penalty. Otherwise, if there is no change in the power, the cumulative reward would be unchanged

The flow chart of the proposed method employed with meta-learnt DF is illustrated in Fig. 4.2, where the updating criteria is defined by (4.13). The δ stands for the optimizing convergence rate and is considered to be 0.00001 for this study.

$$|Q_{t+1}(s_t, a_t) - Q_t(s_t, a_t)| < \delta \tag{4.13}$$

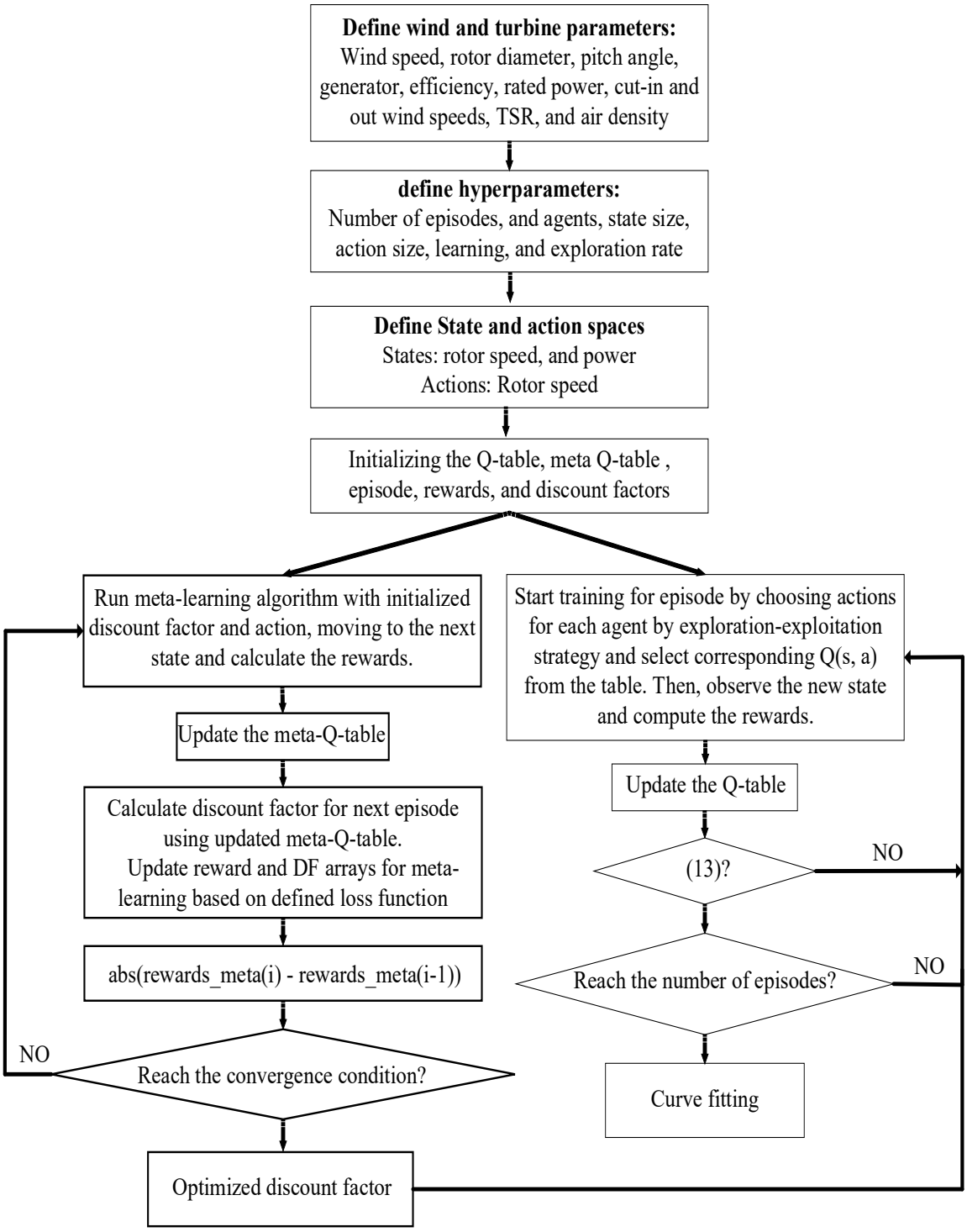


Fig. 4.2: Flow chart for proposed MARL employed with meta-learnt DF [130].

4.4 System Configuration

In this study, the same system of the previous chapter, a wind turbine system featuring a direct-drive permanent magnet synchronous generator, is employed for high-power wind energy applications to evaluate the effectiveness of the proposed MARL method in MPPT. The system incorporates a 3-level neutral point clamped back-to-back power converter. The voltage-oriented control method is employed to regulate the grid-side NPC converter, while the machine-side NPC converter is slated to be controlled through the suggested MARL online strategy, detailed in previous section. The primary objective of the VOC is to generate a pure DC voltage of around 1200 V in the dc-link. This DC voltage is subsequently evenly filtered and distributed across two capacitors, each maintaining a voltage level of 600 V. The grid-side NPC inverter, under the governance of the VOC, transmits 1.6 MW of active power to the grid, featuring a line-to-line voltage of 800 V. For detailed system parameters, please refer to Table 3-1.

4.5 Simulation Results

In the simulation study, the evaluation of the proposed MARL method for MPPT in PMSG-based WECS has been conducted across three distinct scenarios, leveraging Simulink MATLAB software for comprehensive analysis. These scenarios encompass varying wind speed conditions, allowing for a thorough investigation into the MARL method's efficacy under different environmental settings. A 5-minute period is considered for online learning simulation, which strikes a balance between computational efficiency and exploring the state space, adequately. It should be mentioned that the PMSG and proposed method parameters are provided in Table 4.1 and Table 4.2.

Table 4.1: Simulated PMSG Parameters

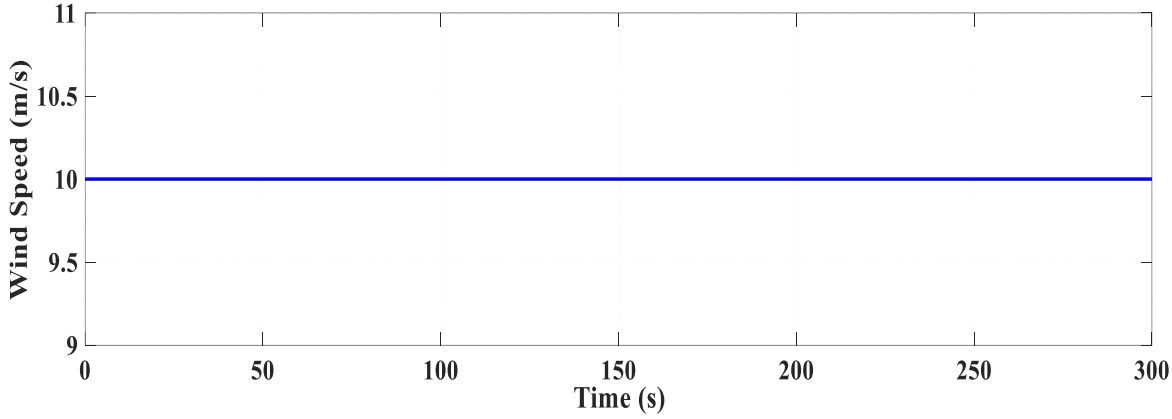
Parameters	Values
Rated wind power	2 MW
Mechanical torque	848.826 kN m
Rotor flux linkage	5.826 Wb
d-axis inductance	1.573 mH
q-axis inductance	1.573 mH
Stator resistance	0.821 ohm
Air gap flux density	0.00077 Tesla
Air gap	4 mm
Number of pair poles	26

Table 4.2: Proposed MARL Method Parameters

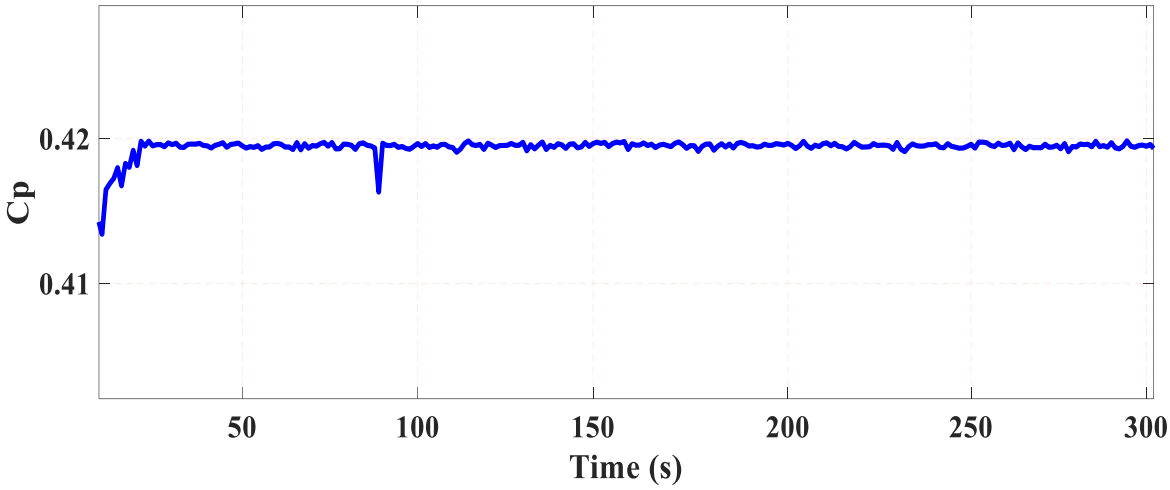
Parameters	Values
Number of agents N	4
Number of episodes	1000
Number of iterations in each episode	20
K	100
J	80
Convergence rate δ	0.000001
Updating step β	0.002
Exploration rate	0.3
Exploitation rate	0.7

4.5.1 Constant Wind Speed Scenario

To evaluate the system ability in optimizing the C_p in a scenario where a fixed wind speed of 10 m/s is considered, the Fig. 4.3 is presented. According to the figure, the initial pursuit of the maximizing C_p commenced at 0 seconds by the agents. In an efficient learning session, the agent swiftly navigated to the MPP within a concise timeframe of approximately 15 seconds. A subsequent exploration phase was noticeable around the 90-second mark, during which the agent diligently sought actions aimed at augmenting power production. Afterwards, the proposed controller fixes the C_p at the maximum possible value, around 0.42. Additionally, a comprehensive analysis has been provided in Table 4.3, showcasing some of the action values derived from Q-table after a 300-second learning period, focusing on specific state pairs (ω_r, P_e) in per unit. Notably, states (0.82, 0.735) and (0.83, 0.735) emerged as the most frequently visited. The MPP, crucial for optimized performance, was identified within the state (0.82, 0.735), where the action "stay" holds the highest value. Conversely, in other states values, either increment (Inc), or decrement (Dec) commands possess highest actions.



(a)



(b)

Fig. 4.3: Fixed wind speed scenario, (a) the wind speed, (b) the C_p in a 5-minute online learning.

Table 4.3: States and Actions for Some Values from Q-Table in Fixed Wind Speed Scenario

States		Action
ω_r (pu)	P_e (pu)	Inc/Dec/Stay
0.80	0.735	Inc
0.81	0.735	Inc
0.82	0.735	Stay
0.84	0.735	Dec
0.85	0.735	Dec
0.87	0.735	Dec

4.5.2 Step-Changed Wind Speed Scenario

The second scenario is examined to evaluate the proposed algorithm's ability to manage the step change wind speed condition. This is accomplished by implementing a repetitive wind speed profile that underwent step changes alternately from 12 to 8, and then from 8 to 10 m/s and again from 10 to 12 m/s, every 50 s (Fig. 4.4). The resulting 5-minute online learning process is also recorded, and the findings are depicted in Fig. 4.4(b). At the outset, the agent was "naive" and continued to explore at each wind step change, instantly. However, the agent's learning process

allowed it to adapt to the condition in approximately 150 seconds and fixed the C_p at almost 0.42, right after the first wind pattern cycle. Moreover, based on the figure, in each wind step changes, the C_p drop rate has decreased less, which shows improved exploration of the agents within system.

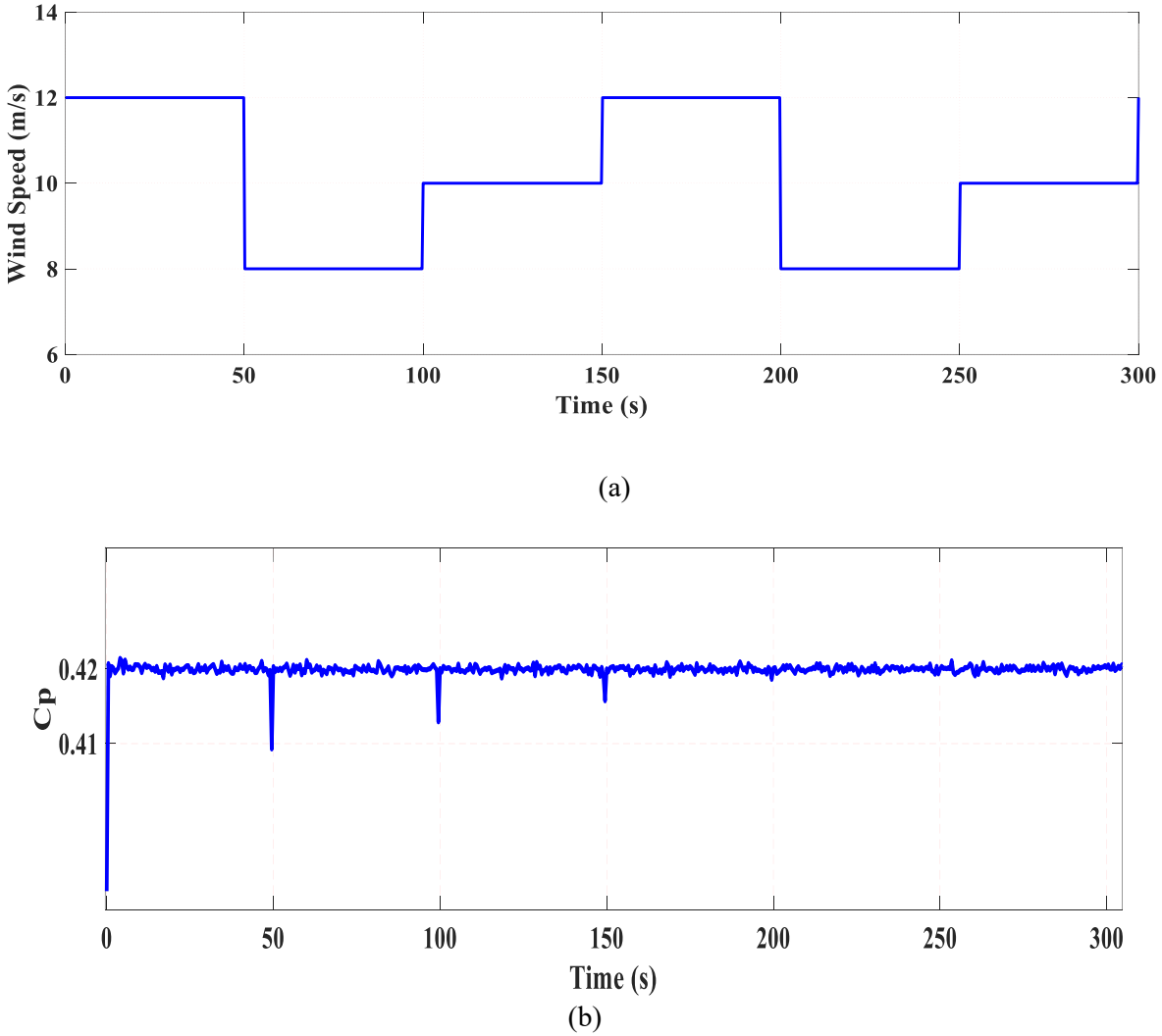


Fig. 4.4: Step-change wind speed scenario. (a) wind speed, (b) the C_p in a 5-minute online

The Fig. 4.5 also illustrates the dynamics of the interplay between step-changed wind speed and rotor speed under the proposed strategy. At the outset of the plot, when the wind speed is at the rated value of 12 m/s, the rotor speed initially starts to pick up and ultimately reaches a maximum rated value of 1 pu. This behavior reflects the rotor's response to the input of the wind speed, and it showcases how the rotor operates at its maximum capacity to harness the wind energy.

However, as the wind speed declines to 8 m/s, the rotor speed adjusts accordingly to approximately 0.6 pu. Finally, when the wind speed increases from 8 to 10 m/s, the rotor speed also increases and reaches almost 0.82 pu. The plot highlights the actual and the theoretical maximum power values of the rotor speed in blue and dotted red lines, respectively.

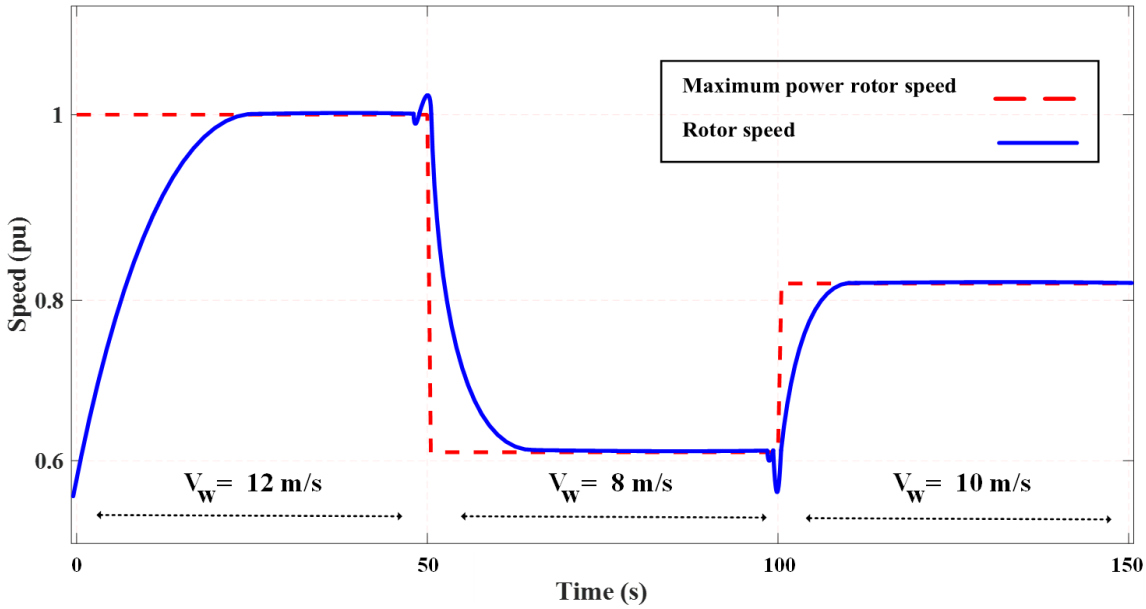


Fig. 4.5: Theoretical and actual rotor speed in step-changed wind speed.

Moreover, the provided information in Fig. 4.6, portrays a comparison between the actual power output of a wind energy conversion system and its theoretical maximum power output. Evidently, after almost one complete wind pattern cycle, the controller effectively tracks the maximum power point and operates in close proximity to the optimal operating point. For example, initially when the agents are completely naive, in 12 m/s rated wind speed from 0 to 50 s, the optimal power is 1 pu, while the actual power is averagely around 0.77 pu. However, with the same wind speed from 300 to 350 s, the actual power is matched with the optimal value at 1 pu, which demonstrates the trained agents' ability to track the MPP, accurately.

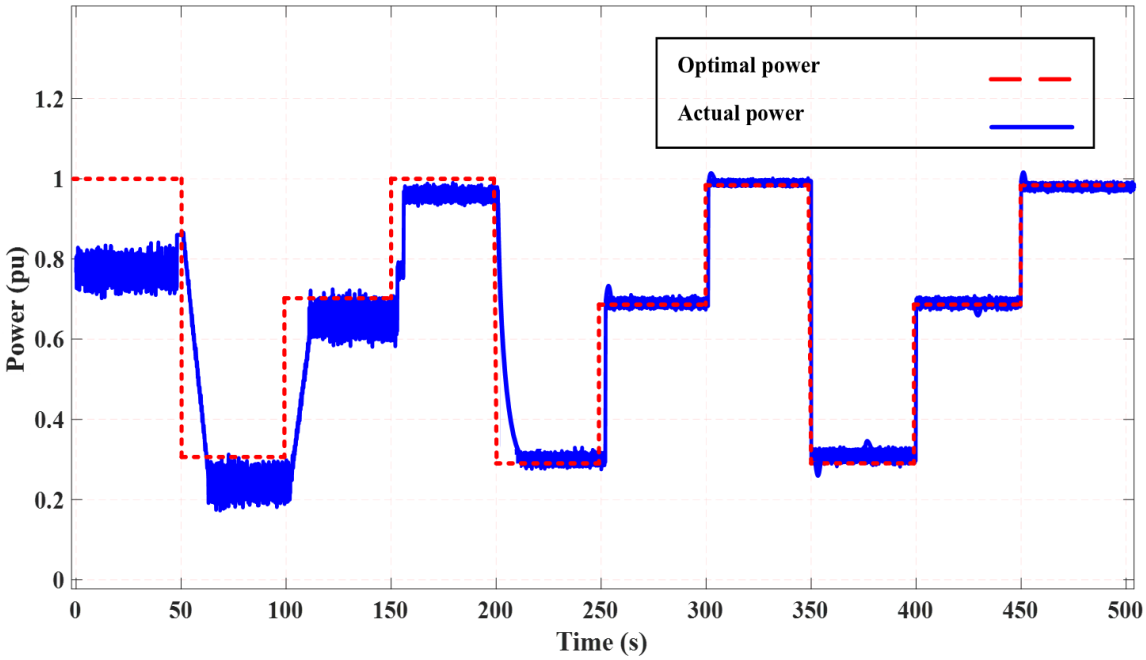


Fig. 4.6: Theoretical and actual power for step-change wind speed scenario.

Table 4.4 also presents subset of action values, as well as the related actions, revealing the presence of three distinct MPPs situated within states (0.6, 0.325), (0.82, 0.735), and (1.0, 0.985). Apart from these values, proposed method maximizes the extracted power by increasing or decreasing the rotor speed according to the acquired optimal policy by the agents. The successive actions then make the rotor speed reach one of the above-mentioned values. These findings serve as robust confirmation that the agent effectively assimilates knowledge and learns from its experiential interactions within the system.

Table 4.4: States and Actions for Some Values from Q-Table in Step-Change Wind Speed

States		Action
ω_r (pu)	P_e (pu)	Inc/Dec/Stay
0.58	0.325	Inc
0.60	0.325	Stay
0.65	0.325	Dec
0.81	0.735	Inc
0.82	0.735	Stay
0.98	0.985	Inc
1.00	0.985	Stay
1.01	0.985	Dec

4.5.3 Variable Wind Speed Scenario

The third scenario within the simulation study delves into variable wind speed conditions, representing a scenario that almost mirrors realistic environmental situations. In Fig. 4.7, a representation of the randomly generated variable wind pattern spanning between 7 and 12 m/s is depicted which is used for comprehensive assessment of the proposed MARL for C_p control.

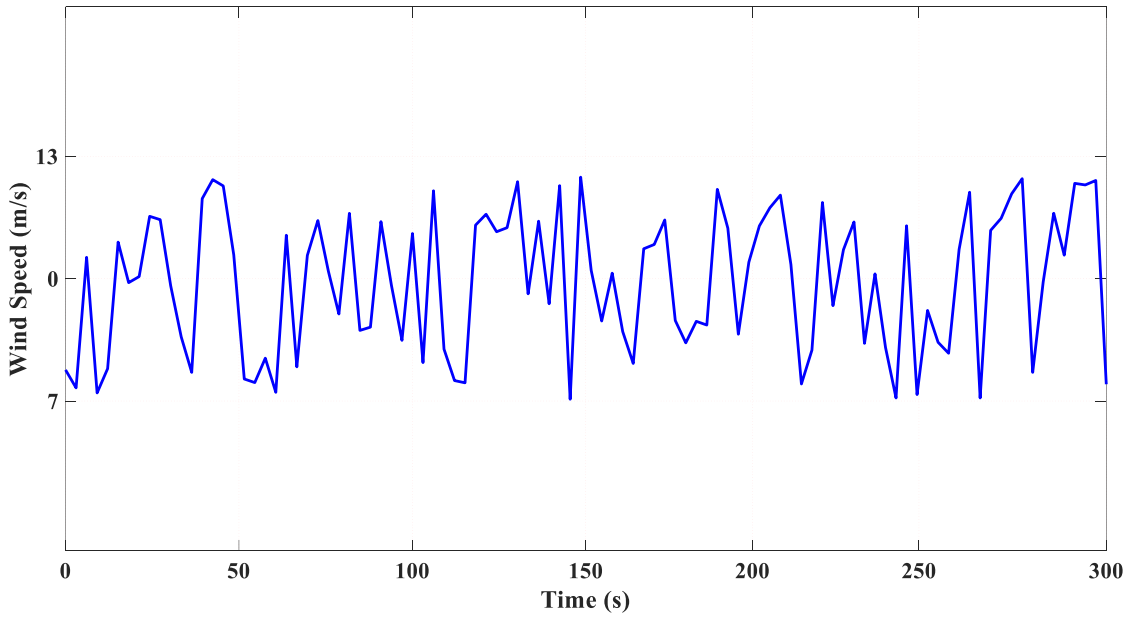


Fig. 4.7: The variable wind speed pattern.

The outcome is demonstrated in Fig. 4.8. However, despite the MARL system's attempt to regulate C_p, based on learned experiences, the C_p curve showcased in this figure exhibits oscillation. These changes coincide with rapid, almost large-magnitude changes in wind speed. But, observed oscillation in C_p, reveals a noteworthy trend of decreasing fluctuation amplitude, converged to 0.42. This intriguing pattern serves as a compelling indication of the advanced learning capabilities embedded within the proposed MARL framework. The diminishing oscillation underscores the system's adeptness in adapting and refining its strategies over time, reflecting a profound capacity to learn from experience. The Fig. 4.9 depicts the corresponding power output and as can be seen the proposed MPPT strategy is able to follow the wind pattern and adjust the rotor speed to extract the maximum operational power point. It is operating under variable wind speeds ranging from 7 to 12 m/s, in which 12 is considered as the rated wind speed value. In this case, the power fluctuates between 1 pu and almost 0.2 pu.

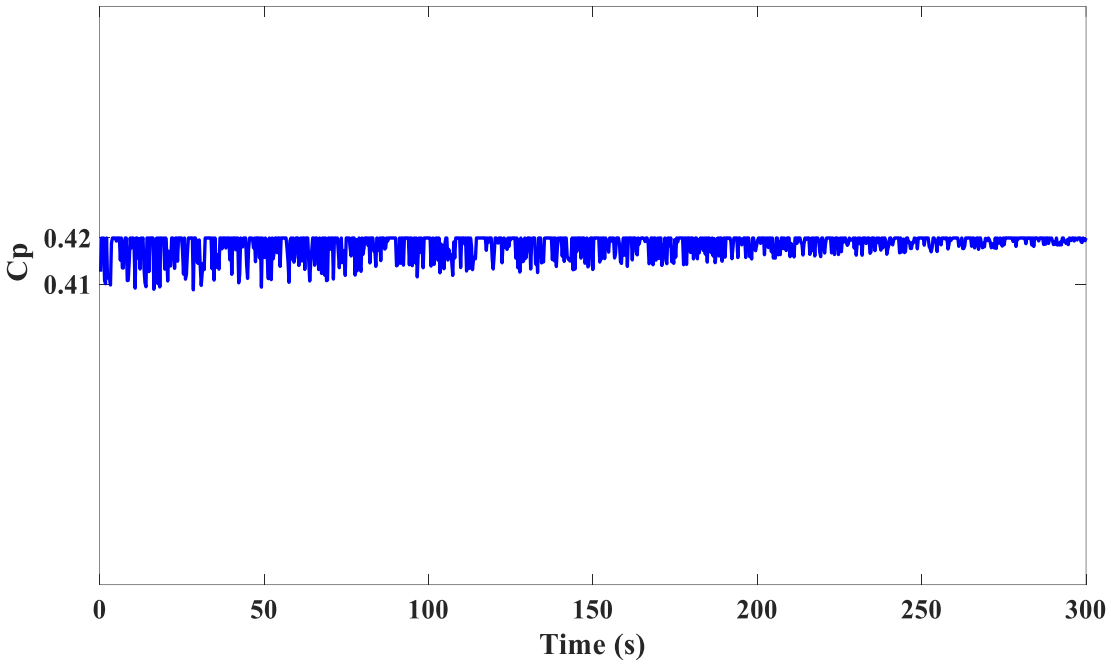


Fig. 4.8: The C_p in a 5-minute online learning in variable wind speed scenario.

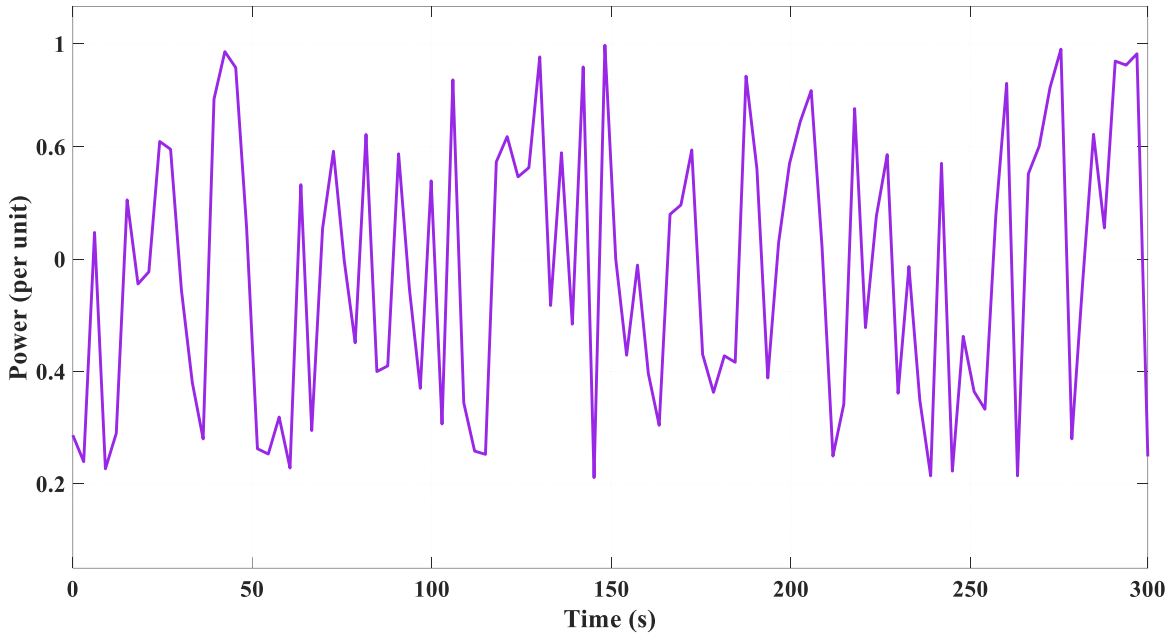


Fig. 4.9: The extracted power from the variable wind speed.

4.5.4 Comparative Study

The quantitative and qualitative comparison among well-known MPPT algorithms is also provided in Table 4.5, and it should be noted that the same wind pattern generated has been used for all the methods.

The tip speed ratio method requires prior knowledge of the wind turbine's characteristics to maintain the optimal tip speed ratio. This method does not involve online learning, which simplifies its implementation but limits its adaptability to changing wind conditions. TSR also requires additional sensors to measure wind speed accurately, increasing the system's complexity and cost. Despite these requirements, TSR achieves an average output power of 0.481 pu, a computational time of 3.10 ms, and a memory requirement of 1.034 kB. The need for precise wind speed measurements and additional sensors can be seen as a drawback.

The optimal rotor blade method also necessitates prior knowledge of the wind turbine's characteristics but does not require online learning or additional sensors. This simplicity makes ORB easy to implement and cost-effective. However, its average output power is the least among all the methods at 0.459 pu, with a computational time of 2.21 ms and a memory requirement of

1.279 kB. While ORB is less efficient, its lower complexity and sensor requirements make it a practical choice for simpler applications.

The P&O method stands out for not requiring prior knowledge or additional sensors, making it highly adaptable and easy to implement. P&O does not involve online learning, which reduces its complexity while maintaining a high average output power of 0.496 pu. The computational time for P&O is 2.84 ms, with a memory requirement of 1.488 kB. P&O's balance of high efficiency and moderate computational demands makes it a popular choice in various applications, although it can experience oscillations around the maximum power point.

The proposed MARL method combines advanced learning algorithms and control strategies to optimize MPPT performance. It does not require prior knowledge or additional sensors but leverages online learning to enhance adaptability. The proposed MARL method achieves the highest average output power of 0.511 pu, demonstrating superior efficiency and energy yield. Its computational time is 2.67 ms, and the memory requirement is 1.640 kB. Despite its higher complexity and memory requirements, the proposed MARL method offers significant benefits in terms of efficiency and adaptability, making it the most effective choice for maximizing energy capture in modern WECS.

According to the table, the proposed method includes the complexity introduced by inter-agent dependencies in MARL, as well as the parallelly-run meta-learning to tune the DF, leading to increased computational time and computational memory. Despite these limitations, the meta-learned DF and the use of multi-agent structure increase the MPPT efficiency, as reflected in the average output power, which surpasses that of the TSR, ORB, P&O methods.

Table 4.5: Comparative Study for Several MPPT Approaches

Strategies	Prior knowledge requirement	Online learning	Additional sensors	Average output power (pu)	Computational time (ms)	Memory requirement (kB)
TSR	Yes	No	Yes	0.481	3.10	1.034
ORB	Yes	No	No	0.459	2.21	1.279
P&O	No	No	No	0.496	2.84	1.488
Proposed MARL	**No**	**Yes**	**No**	**0.511**	**2.67**	**1.640**

The power extracted from the wind by the aforementioned MPPT methods is also depicted in Fig. 4.10, which serves to validate the results and discussion presented in the table. The figure illustrates the performance of each method, showing the varying efficiency and energy yield corresponding to each strategy. These visual representations in this figure corroborate the quantitative and qualitative analyses in the table, underscoring the superior performance of the proposed MARL method in maximizing energy capture in WECS.

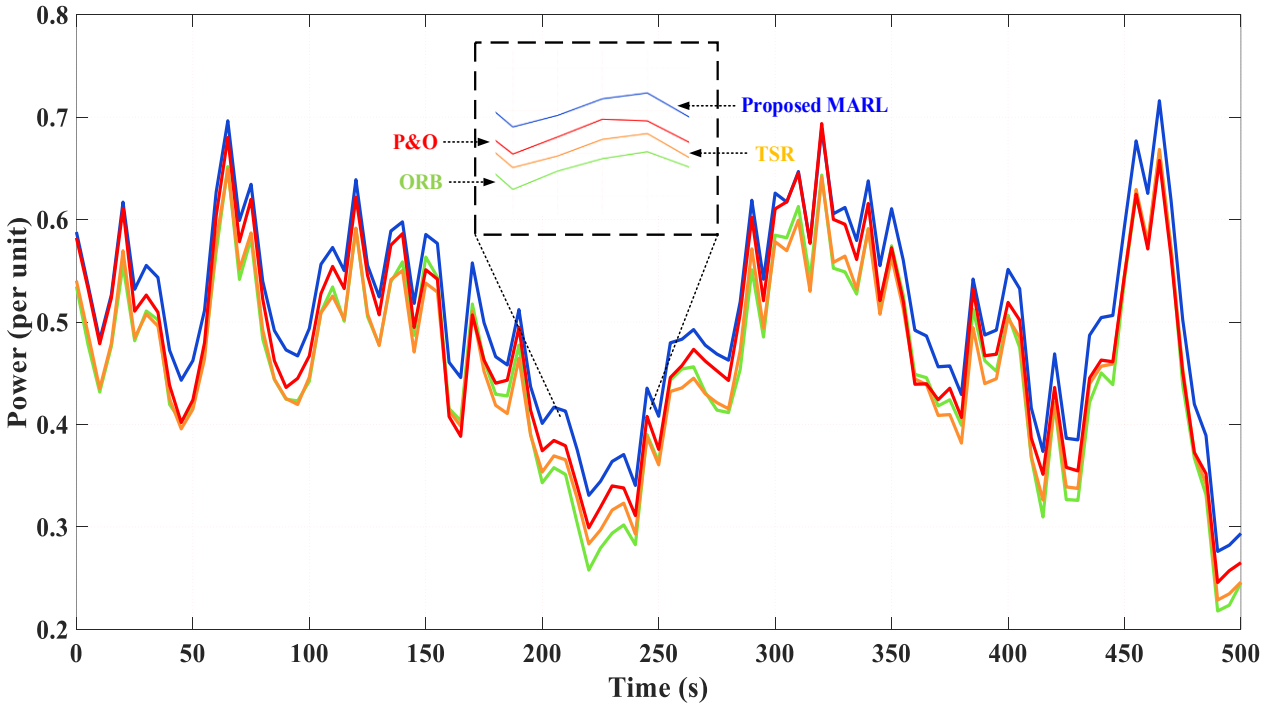


Fig. 4.10: The extracted power from the wind by four different MPPT methods.

To demonstrate the superior performance of the proposed MARL with meta-learnt DF in terms of cumulative rewards, three other scenarios are compared in the same variable wind speed condition which was considered in previous scenario; RL with a fixed discount factor, RL with a meta-learned DF, and a proposed MARL with a meta-learned DF. The results, presented in Fig. 4.11, reveal that RL with a fixed DF initially performs similarly to single-agent RL with a meta-learned DF but declines after 150 episodes. In contrast, MARL with a meta-learned DF, despite starting with lower cumulative rewards due to initial collaboration challenges, outperforms the others after around 400 episodes. At the end of the 1000th episode, the cumulative rewards for the mentioned scenarios converged to almost 440, 515, and 580, respectively. This improvement is aligned with tracking speed and precision and also learning capabilities of the MARL strategy, employed with meta-learnt DF, suggesting its potential for long-term performance.

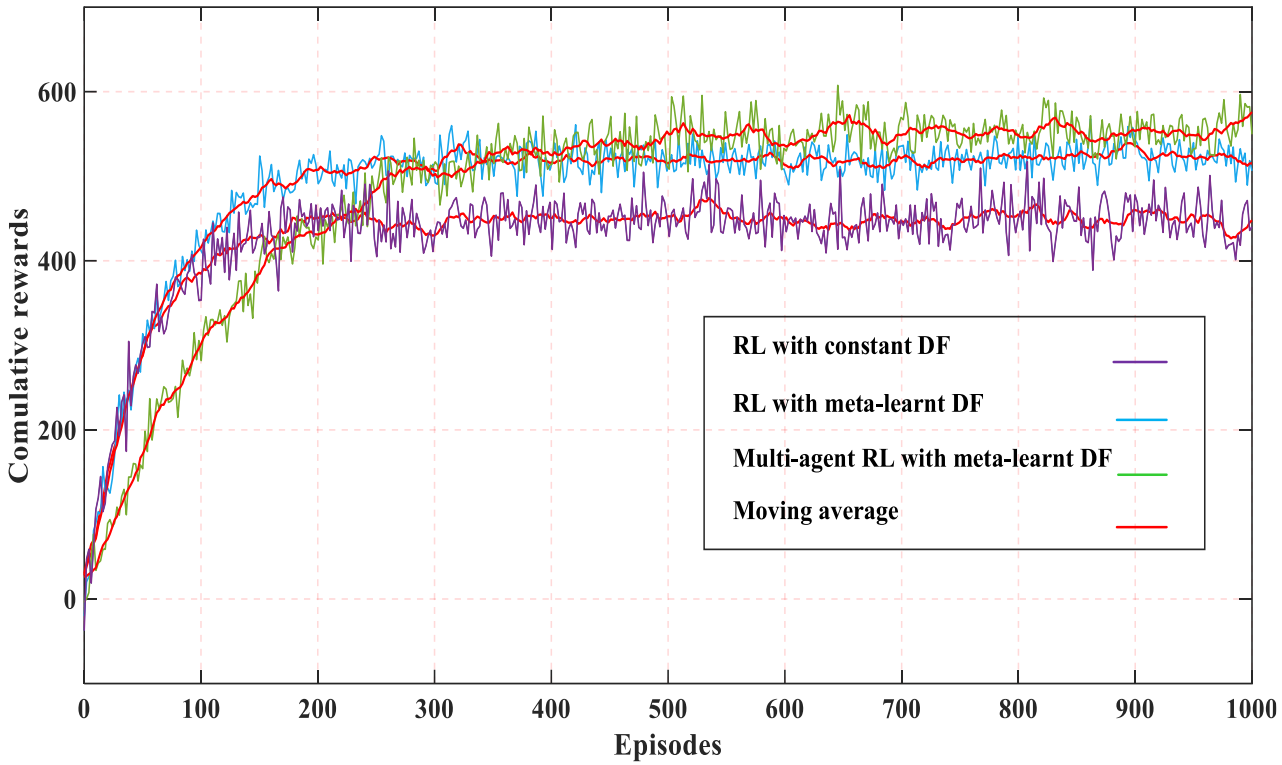


Fig. 4.11: cumulative rewards in 1000 episodes for different RL approaches.

4.6 Experimental Results

In this section, the implementation of the proposed controller is detailed, demonstrating its application in a direct-drive PMSG-based WECS. The schematic representation of the experimental setup is illustrated in Fig. 4.12. To emulate the dynamics of a wind turbine directly powering the PMSG, a DC motor, facilitated by a DC drive, is employed. The power generated by the PMSG is then channeled back to the DC source via a three-phase 2-Level voltage source converter. It should be mentioned that due to the lack of NPC power converter module in the lab, the 2-level voltage source converter in a BTB configuration has been used. The implementation of the proposed MARL control algorithms equipped with meta-learnt discount factor takes place on a DSP DSPTMS320F28335 real-time control board. In the experimental setup, the training process of the proposed multi-agent reinforcement learning algorithm is conducted on a computer, specifically utilizing MATLAB installed on a laptop. This process involves iteratively learning the optimal control policy through interactions of agents with the environment. Tasks such as defining the reinforcement learning problem, selecting appropriate algorithms, executing the training loop, and so forth are all performed on the computer. Once the training process is completed, the trained model, which encapsulates the learned control policy, is stored on the computer. During operation, the trained model is used to make real-time control decisions based on the current state of the system. The control processes, including the application of the learned control policy to the system, are implemented on a DSP board. This board receives inputs from sensors, executes the control algorithm based on the trained model for MPPT, and generates outputs to adjust the operation of the wind turbine, optimally. It should be mentioned that the laptop configuration is; 13th Generation Intel® Core™ i7-13700H Processor (E-cores up to 3.70 GHz P-cores up to 5.00 GHz), with the GPU of 8GB GDDR6.

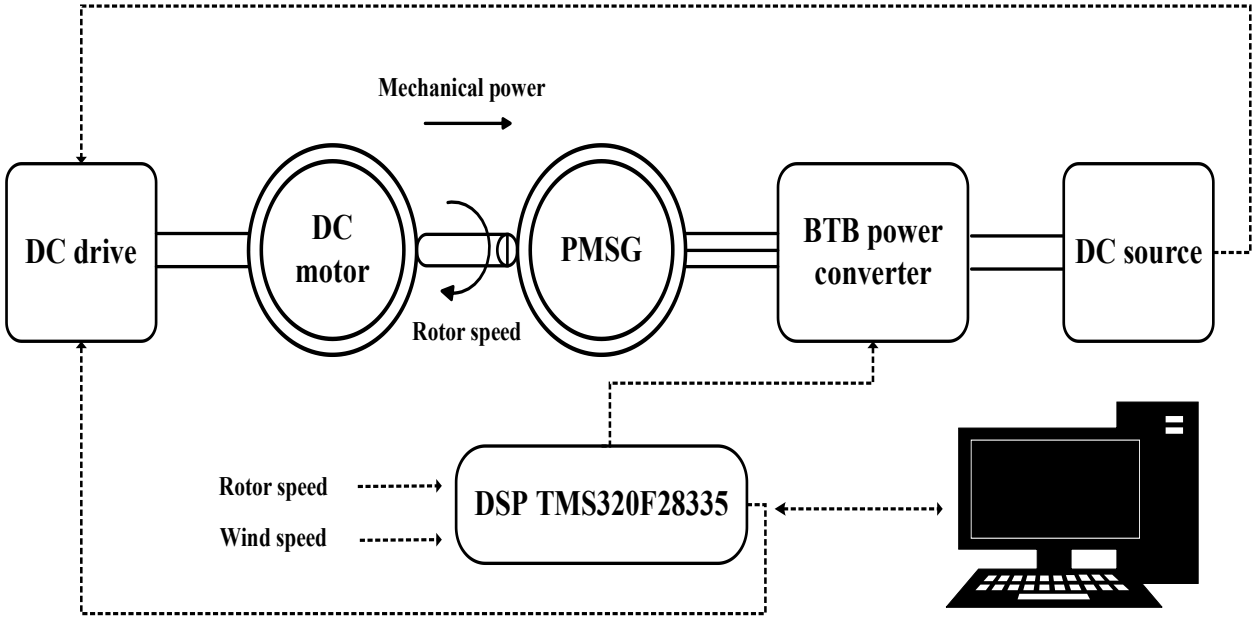


Fig. 4.12: Experimental schematic of the emulated WECS.

The arrangement of the experimental setup is depicted in Fig. 4.13, containing labeled equipment. For detailed specifications regarding the PMSG parameters, and the WECS model the Table 4.6 and Table 4.7 have been provided.

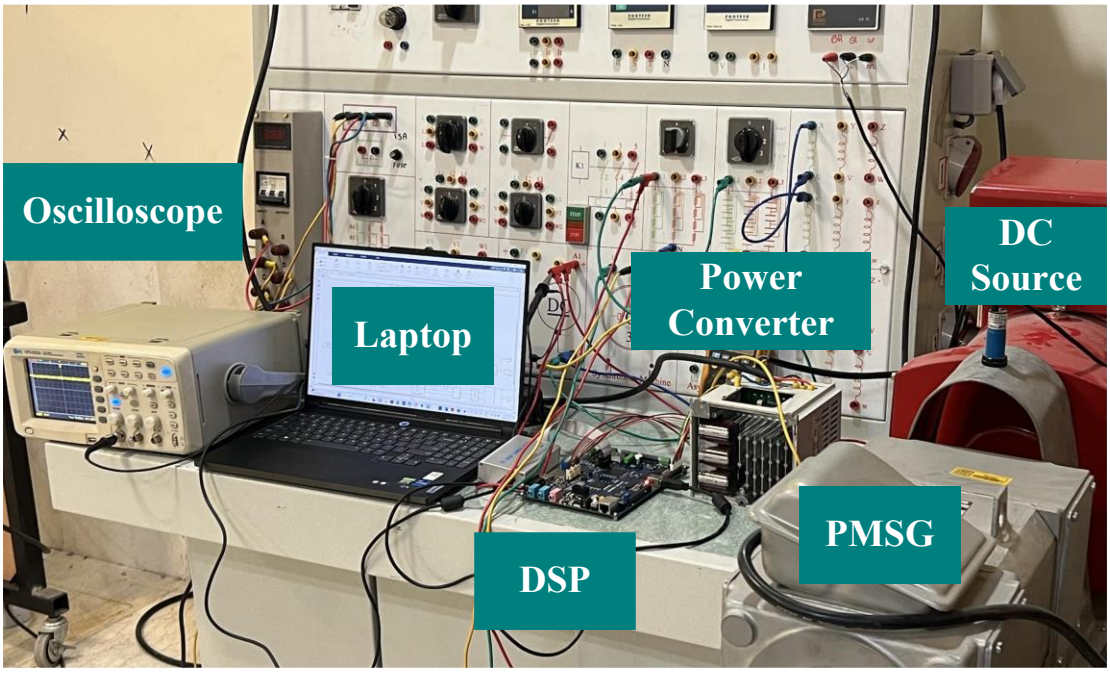


Fig. 4.13: Experimental set-up.

Table 4.6: Implemented WECS Parameters

Parameters	Values
Rated BTB converter apparent power	1.2 kVA
Machine side output DC voltage	120 V
Reference dc-link voltage	60 V
dc-link capacitor (each)	200 μF
dc-link resistor	1 Ω
Switching frequency	1 KHz
Grid frequency	60 Hz
Grid side line-line voltage	180 V
Active power delivered to the grid	1 kW

Table 4.7: Implemented PMSG Parameters

Parameters	Values
Rated wind power	1000 W
Moment of inertia	0.0006J Kg.m^2
EMF constant	11.5 V/krpm
d-axis inductance	0.375 mH
q-axis inductance	0.434 mH
Stator resistance	0.706 ohm
Air gap flux density	0.00041 Tesla
Air gap	2 mm
Number of pair poles	4

In the MPPT simulation, a switching frequency of 2000 Hz was used, while a switching frequency of 1000 Hz is chosen for the actual implementation. This discrepancy is primarily due to practical considerations related to IGBT operation, thermal management, and electromagnetic interference (EMI). Higher switching frequencies, such as 2000 Hz, can enhance control precision and system responsiveness in simulations. However, in practical implementations, particularly with IGBTs, higher frequencies can lead to increased switching losses and elevated heat generation. This requires more effective cooling solutions and can impact overall system efficiency. Furthermore, higher switching frequencies tend to generate more high-frequency noise, which can exacerbate EMI issues. Excessive EMI can interfere with nearby electronic devices and communication systems, posing compliance challenges with electromagnetic compatibility (EMC) standards. Using a lower switching frequency of 1000 Hz in the implementation helps mitigate these issues. This frequency strikes a practical balance between performance and thermal management, reducing switching losses and heat generation. Additionally, it helps minimize EMI, ensuring that the system operates within safe thermal limits and complies with EMC regulations. As 1000 Hz is a commonly used value in IGBT applications, it aligns with industry standards, optimizing both system reliability and performance.

Three distinct scenarios are systematically employed to assess the effectiveness of the proposed MARL method for MPPT applications. In the first scenario, the system's response to a step change in wind speed is thoroughly examined. The second scenario delve into the performance of the MARL approach when faced with variable wind speeds, and finally, the third scenario involve a comparative analysis between the proposed MARL-based MPPT method and the conventional P&O method.

4.6.1 Step-Changed Wind Speed Scenario

To validate the learning capability of the agents, a wind speed profile featuring step changes alternating between 8 and 10 m/s every 50 seconds is employed (Fig. 4.14(a)). The outcome for the implemented power coefficient is depicted in Fig. 4.14(b). Initially, the agent exhibits a short term "naive" state, resulting in a decrease in C_p due to incorrect exploration. This motif as well as the fluctuation in C_p amplitude, up to 3% is attributed to a misguided exploration direction,

specially at the time of wind speed step change like 50 s, 100 s, 150 s, and so on. Nevertheless, the agent demonstrates its capacity to adjust its actions, ultimately moving in the correct direction illustrated by decreasing the drop value at almost each wind speed step-change. This underscores the agent's ability to learn from erroneous experiences, showcasing its adaptive learning capabilities. It can be proven by almost omitting the C_p drop from 250 s.

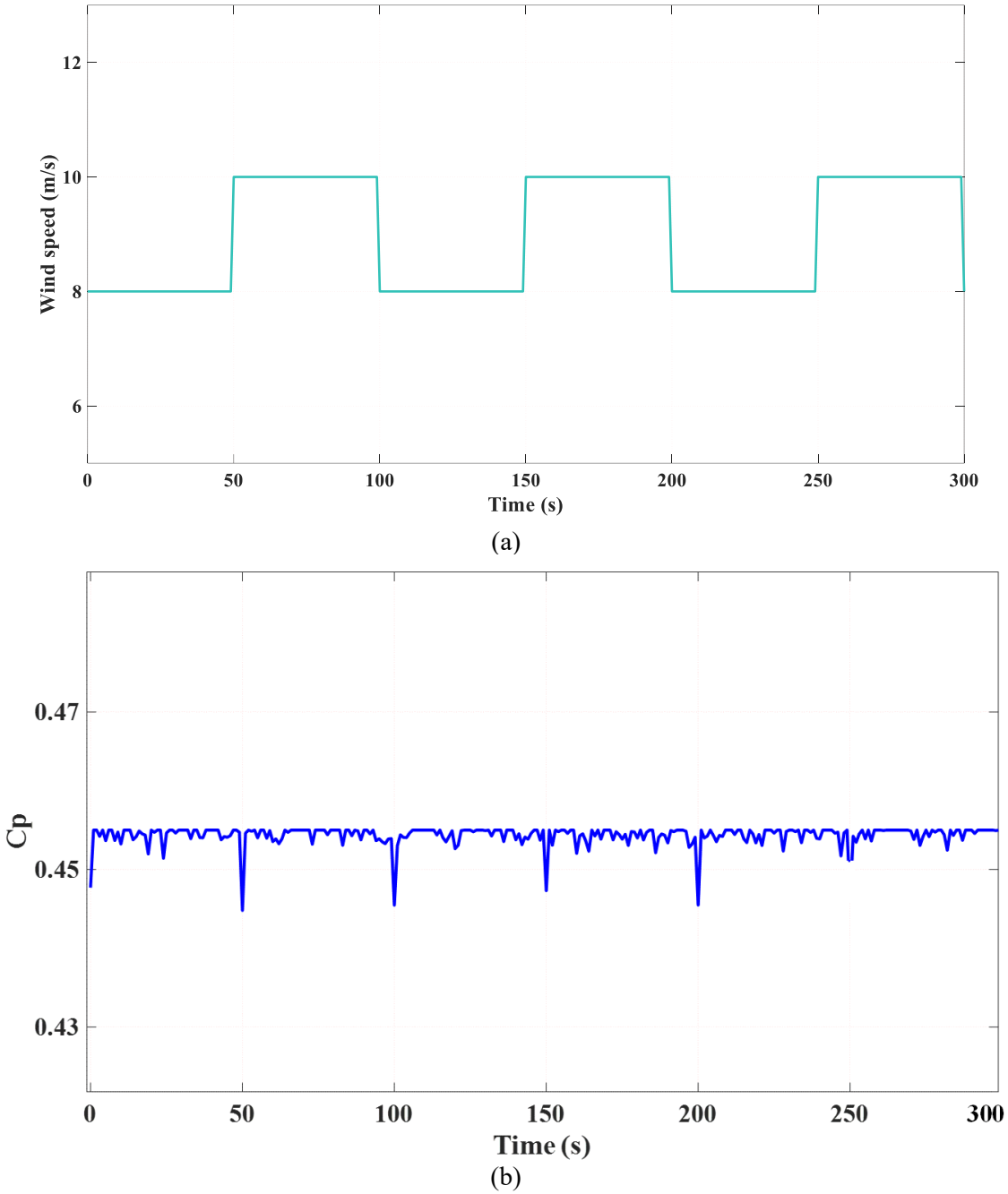
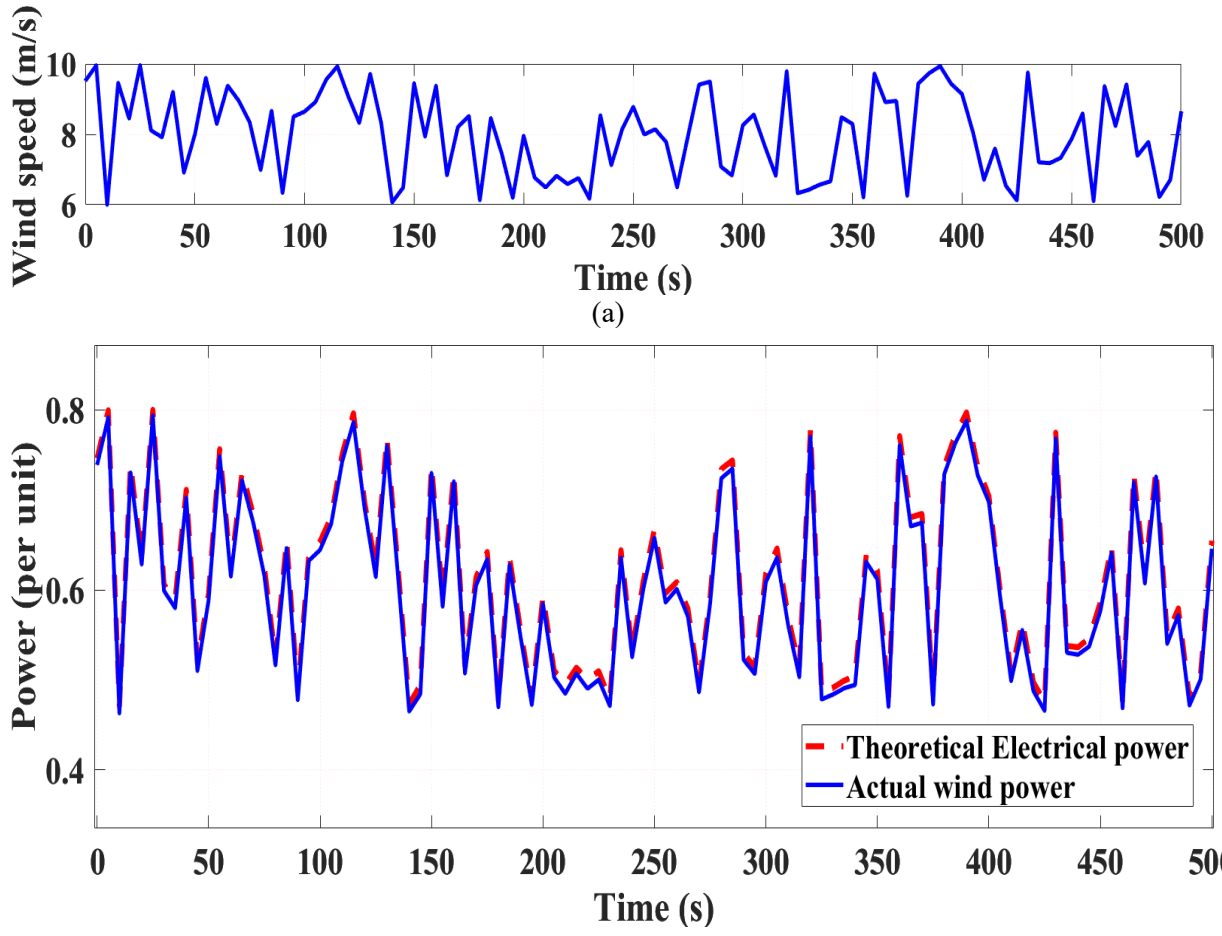


Fig. 4.14: Step-changed implemented scenario. (a) step-changed wind pattern, (b) C_p in a 5-minute online learning in step-changed wind speed scenario.

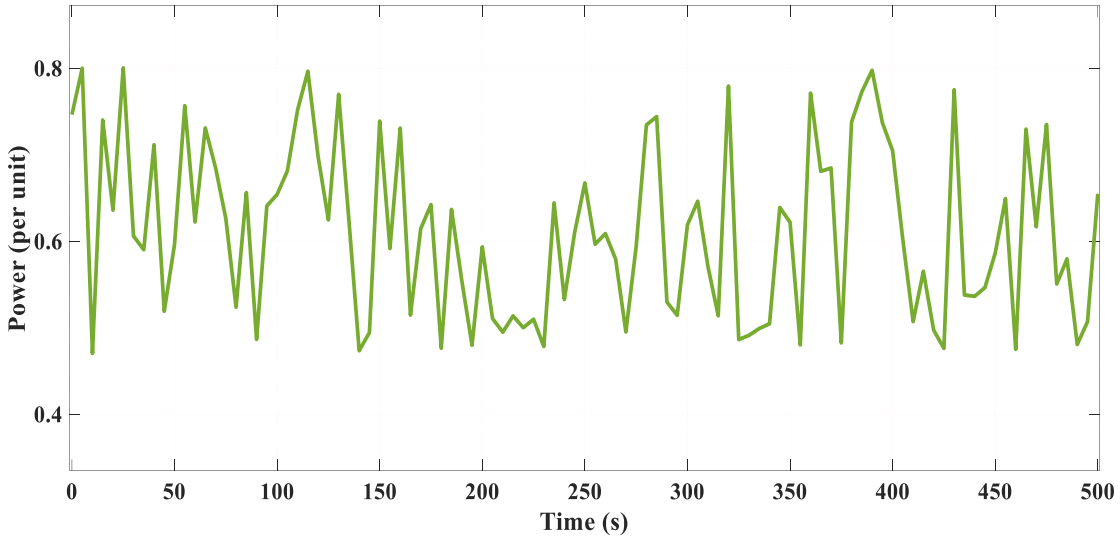
4.6.2 Variable Wind speed Scenario

The wind pattern, and the corresponding power output are graphically illustrated in the subsequent Fig. 4.15(a) and Fig. 4.15(b), providing a comprehensive analysis of the system's response to variable wind speeds. The theoretical power as well as the actual power gained by the proposed method are depicted at the same graph to show how successfully the algorithm can follow MPP. Moreover, Fig. 4.15(c) is provided to validate the performance of the proposed method by comparing with the same system simulation output. The alignment between the implemented and simulated output power further attests to effectiveness of the proposed approach in accurately tracking maximum power point. According to the findings, the average powers for the implemented and simulated system are 0.58 pu, and 0.61 pu, respectively. Although, the slight discrepancy can be justified by different initialized condition measurement accuracy, losses and environment condition.



(a)

(b)



(c)

Fig. 4.15: The variable wind speed scenario. (a) wind speed profile, (b) electrical power and actual wind power captured by implemented proposed MARL method, (c) wind power captured by simulated system.

The Fig. 4.16 also depicts the variation in implemented C_p throughout the learning process. Notably, fluctuations in C_p are observed, attributed to the variable wind patterns. However, as the learning process approaches completion, a discernible trend emerges— the variability in C_p diminishes. This reduction in fluctuation suggests that agents, adapting to the complexities of varying wind patterns, achieve a more stabilized and optimized performance over time.

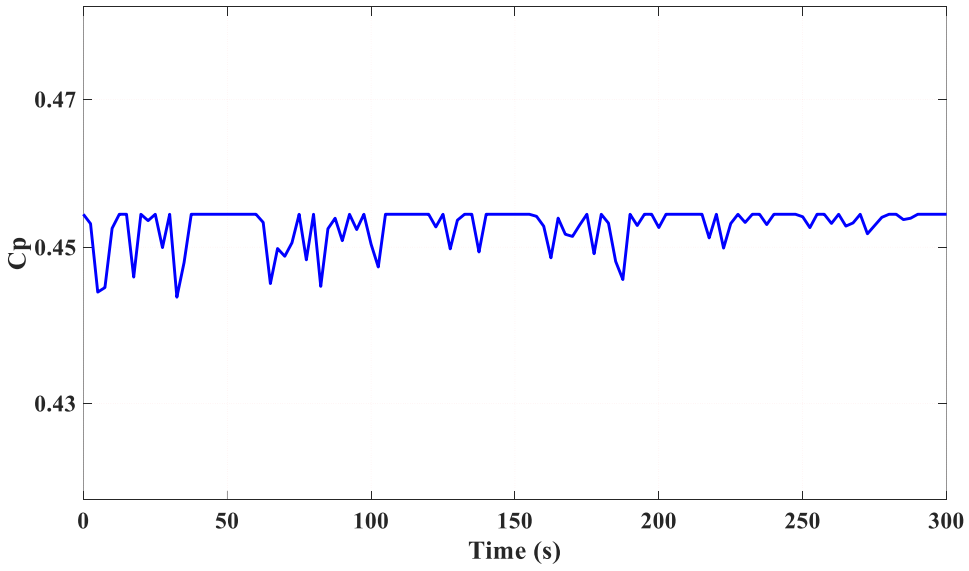
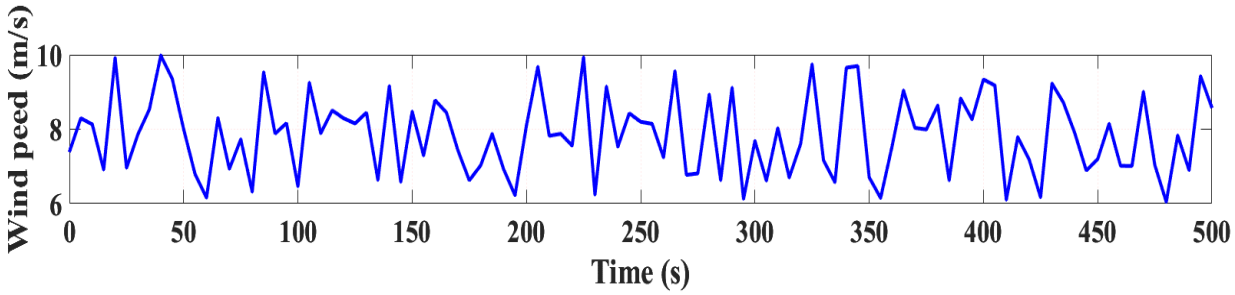


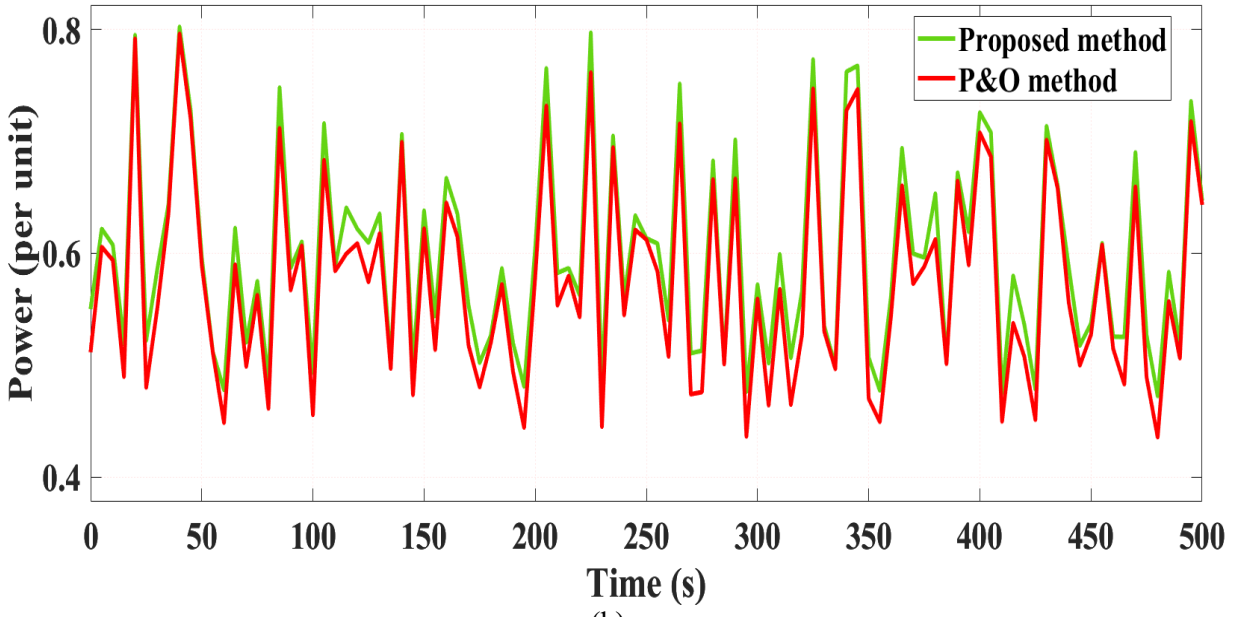
Fig. 4.16: The C_p in a 5-minute online learning in variable wind speed scenario.

4.6.3 Comparative Study

In the final scenario, an additional variable wind pattern (Fig. 4.17(a)) is taken into account, and the resulting power outputs from both the proposed method and the conventional P&O based MPPT approach are presented (Fig. 4.17(b)). The comparative analysis depicted in the figure indicates that the proposed method excels in extracting more power (averagely 0.629 pu) and operates more optimally in comparison to the general P&O approach (averagely 0.616 pu) for the entire 500 s time period. Moreover, the proposed method omits the wind speed sensor required in conventional P&O method, as well as the complicated offline system modeling.



(a)



(b)

Fig. 4.17: Comparative study between the proposed MARL and general P&O methods. (a) wind speed profile, (b) actual wind power by both methods.

In another scenario, a step change in wind speed from 6 to 8 m/s is considered to analyze the response rate to the rotor speed change for both the proposed method and the traditional P&O method to follow the maximum power point. This analysis focuses on the rate of change in rotor speed following the wind speed step change, with the results plotted for both methods. According to the plotted Fi. 4.18, both methods eventually converged to a rate of change of approximately 0.2. However, the proposed method demonstrated a faster response, reaching 5% of the convergence value within 34 ms. In contrast, the traditional P&O method took slightly longer, achieving the same 5% convergence threshold at 46 ms. This marginal yet crucial difference highlights the superior responsiveness and efficiency of the proposed method in adapting to changes in wind speed, ensuring quicker stabilization and optimal performance of the wind turbine system.

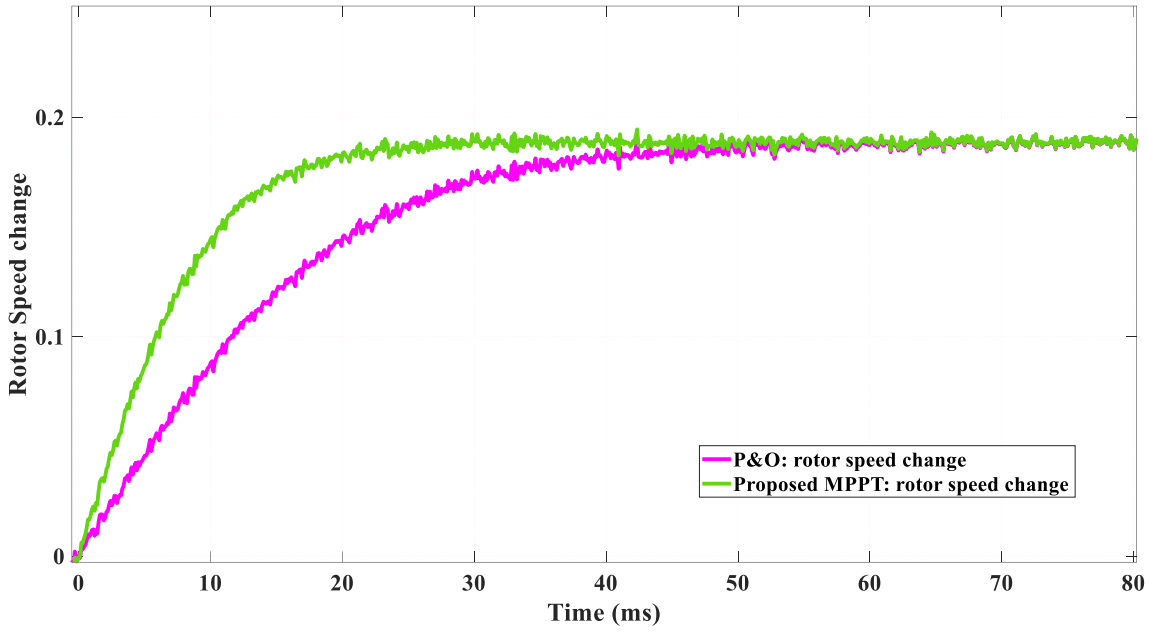


Fig. 4.18: Comparative study between the proposed MARL and general P&O methods in terms of rotor speed reaction to the wind speed change.

4.7 Conclusion

Wind energy conversion systems employing conventional MPPT methods often encounter challenges such as sluggish tracking speeds, imprecise adjustments, and sensitivity to wind speed

fluctuations. This study introduced a novel approach using multi-agent reinforcement learning MARL to overcome these limitations in variable wind speed PMSG-based WECS. In comparison to traditional methods, the proposed MARL strategy enhanced energy output and the system's ability to swiftly adapt to changes in wind speed. Furthermore, due to its decentralized nature, the approach involved multiple agents collaborating to maximize power generation, ensuring superior precision and enhanced interaction capabilities compared to single-agent RL methods. Additionally, by incorporating a meta-learned discount factor, the MARL algorithm was further refined in terms of learning efficiency and convergence speed, resulting in a more robust solution. Comprehensive simulation results were presented to evaluate the performance of the proposed MARL approach, demonstrating its potential applicability. Moreover, a prototype was also implemented to validate the effectiveness of the proposed algorithm in real-world wind MPPT scenarios.

Chapter 5
Robust, Optimal, and Environment-Friendly Power Management Scheme for Renewable Energy systems

5.1 Introduction

This chapter recommends utilizing recurrent neural networks and reinforcement learning-based approaches to address the challenges in power prediction for renewable energy resources, micro-grid (MG) power management, and energy scheduling. The proposed MG includes combined cooling, heating, and power (CCHP), wind turbines, photovoltaic systems, and battery energy storage systems (BESS). To accurately estimate the 24-hour generation of renewable resources, a multi-layer recurrent neural network (MLRNN) model is developed and trained using historical wind and solar data. The model's hyperparameters are optimized through a grid-search algorithm. This approach offers several advantages over traditional methods. MLRNNs provide higher accuracy in forecasting due to their ability to capture complex temporal dependencies and patterns in the data. They can effectively extract non-linear mapping functions from labeled data, which is essential for modeling the intricate relationships between different variables in renewable energy systems. Additionally, the MLRNN model can be easily scaled and adapted to incorporate additional data sources or extended time horizons, enhancing its versatility and applicability. MLRNNs are also more robust to noisy data compared to conventional linear models, ensuring more reliable predictions under varying conditions.

The forecasted values are then employed in the subsequent study phase to achieve optimal power management within the MG's generation units. A model-free, multi-agent reinforcement learning approach mainly proposed in chapter 2 is employed to minimize the multi-objective fuel and CO_2 emission cost function, ensuring a robust and environmentally friendly MG. This solution offers distinct advantages. MARL agents can adapt to changing environments and operational conditions, providing more flexible and responsive power management. Unlike centralized approaches, proposed MARL allows for decentralized control, reducing the computational burden and improving response quality. By employing multiple agents, the MARL method enhances coordination among various MG components, leading to more efficient and balanced power

distribution amongst the renewable energy resources, and the micro-turbine power generators. MARL outperforms single-agent methods by effectively finding optimal operating power points in complex and dynamic environments, ensuring better performance and resource utilization. Moreover, the optimal number of agents are also defined to improve the strategy, as well. According to the findings, the proposed MARL method effectively dispatches power generation among renewable resources and the MT in the CCHP system, while reducing fuel and CO_2 emission costs. It also manages the BESS efficiently to support MG operations, demonstrating its capability to maintain optimal performance while supporting the MG.

5.2 Literature Review

Several countries around the world are bringing carbon (CO2) emission and energy yield regulations into place to diminish the greenhouse gas emissions and augment the energy efficiency [131]. As a result, deployment of distributed generation (DG) systems, which primarily use renewable energy-based sources, is rapidly rising to meet the prescribed prospect. Infrastructures and system enlargement will be then required, alongside investments in storage technology, and power generation management. Integrating the DGs into the insightful micro-grids (MGs) concept plays a vital role in supplying safe, balanced and eco-friendly operations of the power system [132], [133]. MGs embrace a range of DG blocks, including wind turbine (WT), photovoltaic (PV), battery energy storage system (BESS), and so forth. Although, an energy management strategy (EMS) is necessary for a MG to operate in grid-connected and stand-alone scenarios to properly adapt the energy, supplied and exported from and to the main power network [1314]. EMS is frequently described as a nonlinear optimization issue. Numerous approaches, such as quadratic and mixed integer programming, particle swarm optimization, and genetic algorithm, have been suggested in studies to tackle the EMS problems with different objective functions [135]-[137]. Nevertheless, since the renewable energy sources and loads are unpredictable in real world problems, the mentioned methods rely on the day-ahead forecast's accuracy, and this has negative impact on how the system operates with these algorithms, in real time [138]. To solve the shortcomings, an alternative EMS solutions are available, depending on the Markov decision process (MDP). Reference [139], derived a near-optimal EMS for an MG, using a finite MDP-

based supplemented by deep neural network learning. The approach, meanwhile, required high calculation volume and complex system modeling.

Stochastic optimization (SO) based EMS is an option, as well. A finite collection of alternative situations is used by SO to explore the search area and makes decisions derived from empirical average of such possibilities. To reduce the real-time electricity cost, a SO-based EMS strategy for stand-alone grids was reported in [140]. With the same objective function, a real-time SO for EMS in domestic system contains WT, PV, and energy storage systems was suggested in [141]. But inherently, the SO is an iterative problem-solving approach, and in some cases the convergence cannot be assured Also, implementation in real-world situations may not always be feasible [142].

Model predictive control is another comprehensive approach that makes it through an explicit concept to forecast how the structure will respond in the upcoming and adapt the control plan over time. MPC has a proven track record of offering real-time solutions for multi-variable operations that are delayed and chaotic [143]. To reduce operating costs and carbon dioxide emission, an MPC-based EMS in an MG was provided in [144]. Reference [145], on the other hand, employed MPC-based solution to lessen the negative consequences of unforeseen renewable energy units' outputs, in an islanded MG. Conversely, MPC suffers from sophisticated calculation, high maintenance cost and lack of flexibility.

Power management improvement is also being treated by gravitational search algorithm [146], and artificial bee colony search algorithm [147]. Although, for the highest fitness fulfillment, these algorithms have complicated parameter calculations, restrictions, coding challenges, and formulations. Furthermore, the adaptive neuro-fuzzy inference system and fuzzy-logic controller are the other practical choices that of course still have drawbacks in power management problems, according to the literature [148]. In order to achieve consistent and refinement resolution, artificial intelligence (AI) based approaches like artificial neural network (ANN) and reinforcement learning (RL) based optimization approaches are getting attractive choices in recent years. RL, on the other hand, has achieved superior performance in optimizing agent in power systems' problems, like deriving a daily consumption plan to charge the electric vehicle fleet [149], or optimizing the profit function for electricity suppliers. RL can also be leveraged in strategic bidding for load-shedding in an MG [150]. As the other application, RL was used to strategically schedule batteries in order to reduce energy costs for customers in [151], or

reduce the estimated costs for both customer and a service provider in an MG, in [152]. Besides, an energy management problem of a MG was framed in [153] to minimize operational costs. This was achieved using RL to develop a real-time scheduling plan, taking into account the uncertainty of the load. Also, the RL was utilized to capture uncertainty in [154] without external prediction mechanism or probability distribution for controlling a residential multi-energy system.

However, RL-based power management strategies are still in their infancy and to the best of authors' knowledge, less research has been conducted in this field, and require further efforts to address the gaps. Hence, this research addresses the challenges in renewable energy resources output power prediction, and energy management in MGs, in two phases. The proposed MG incorporates WT, PV, and a BESS, adding complexity with numerous variables and constraints. Additionally, a combined cooling, heating, and power system is included (Fig. 5.1). Proposed MLRNN model is trained on historical wind and solar data to accurately predict day-ahead renewable energy outputs, in the first phase of study. This approach offers advantages, including improved accuracy in capturing non-linear relationships between variables and output power labeled datasets. The model's hyperparameters are optimized using a grid-search algorithm, resulting in superior prediction accuracy, compared to the trial-and-error strategy. The micro-grid then employs proposed model-free, MARL approach in the second phase to optimize power flow and energy scheduling, considering the predicted renewable energy output. The objective is to minimize fuel costs and CO2 emission cost, making the MG robust and environmentally friendly. The recommended MARL demonstrates significant exploration capabilities, achieving near-optimal cost function values. By distributing tasks among multiple agents, the MARL approach outperforms single-agent reinforcement learning in finding optimal operating points in a complex environment. Furthermore, by determining the optimal number of agents, this method effectively balances computational load and memory usage, ensuring efficient operation without compromising performance. Additionally, the MARL strategy provides a comprehensive evaluation through cost function values, cumulative rewards, and convergence speed, demonstrating superior performance in these metrics. The outcomes in a more robust and scalable solution, capable of handling complex power management scenarios with enhanced precision and reliability compared to single-agent RL and conventional approaches. Comprehensive MATLAB simulations validate the method's potential for real-world application, by conducting comparative analysis, as well as the BESS scheduling within the MG.

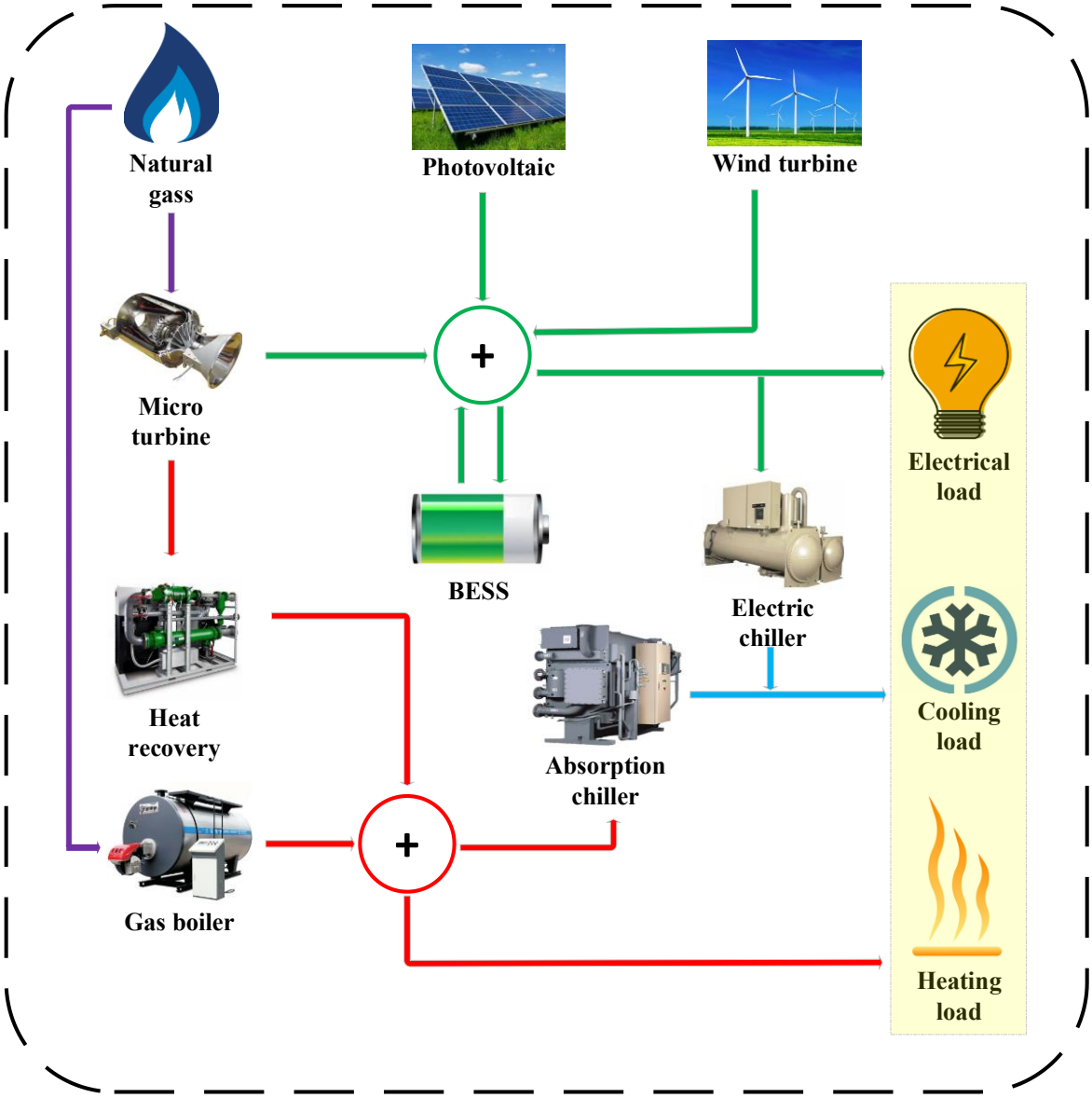


Fig. 5.1: Proposed MG configuration [155].

5.3 Proposed Solution

The proposed solution comprises two distinct phases. In power prediction phase, historical weather patterns data, acquired by the software supported by NREL called system advisor model (SAM), is used to predict the output power of renewable resources within the MG, accurately. Building upon the predictions, subsequent power management phase involves dynamic load

balancing and power sharing among the generation units aimed at minimizing fuel and CO_2 emission costs. This two-phase approach ensures efficient utilization of renewable resources while maintaining grid stability and reliability. As the world move toward a sustainable energy future, robust power management strategies play a pivotal role in achieving the goals. The phases are detailed, as follows.

5.3.1 Phase One-MLRNN for WT and PV Power Prediction

A recurrent neural network (RNN) is designed for sequential data processing, making it particularly effective for time series prediction problem. RNNs utilize their internal state to process sequences of inputs, allowing them to capture temporal dependencies. However, traditional RNNs suffer from significant shortcomings, including the vanishing gradient problem, which hinders the network's ability to learn long-term dependencies, and difficulties in training due to their inherently sequential nature. To address these issues, here, a multi-layer RNNs, also known as deep RNNs, is introduced to foresee short-term WT, PV output power generations, in order to accurately adapt to the changing weather patterns. These networks stack multiple RNN layers, enabling them to capture more complex patterns and hierarchies in the data. The strength of MLRNNs lies in their enhanced capability to model intricate temporal dynamics and long-range dependencies, leading to more accurate and robust predictions in various applications.

The configuration of the model for one layer is demonstrated in Fig. 5.2. The backpropagation procedure through the time is also shown in this figure. It involves computing the gradient of the loss function with respect to each weight by the chain rule, iteratively adjusting the weights to minimize the loss. According to the figure, the hidden state of h_t can be attained as follows. Where, the input at the time of t is shown by x_t, the recurrent hidden state at the time of $t-1$ is represented by h_{t-1}. The σ is the activation function, and the V and W are the weight matrices which should be initialized. It is worth mentioning that the bias vector b_h is concatenated to the mentioned matrices.

$$h_t = \sigma(Vx_t + Wh_{t-1}) \tag{5.1}$$

Then the output can be achieved by:

$$y_t = Vh_t \tag{5.2}$$

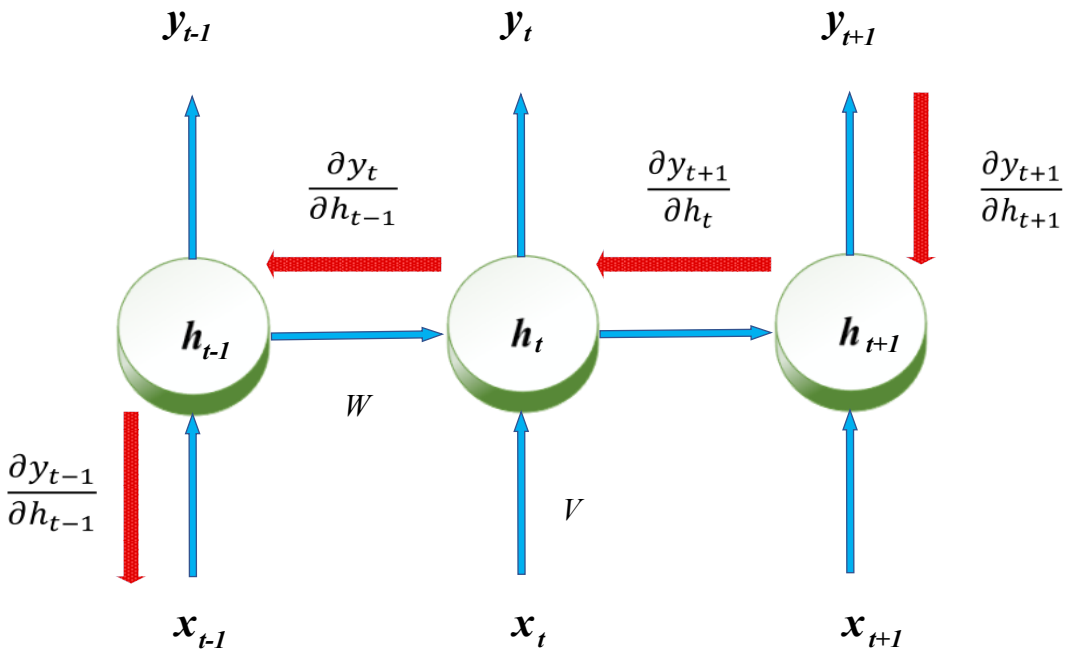


Fig. 5.2: The configuration of the MLRNN for one layer.

In a MLRNN, the hidden states are updated across multiple layers L. For layer l at time step t, the hidden state $h_t^{(l)}$ and the y_t are computed as follows:

$$h_t^{(l)} = \sigma(V^{(l)}x_t^{(l-1)} + W^{(l)}h_{t-1}^{(l)}) \tag{5.3}$$

$$y_t = V^{(l)}h_t^{(l)} \tag{5.4}$$

Normalized root mean square error (nRMSE) is used regarding assess the effectiveness of the forecast, known as the loss function L. In (5.5), the predicted and actual values of the power are shown by P_{pred} and P_{act}, within 24-hour period.

$$L = \sqrt{\frac{1}{24}\sum_{t=1}^{24}\left(P_{pred} - P_{act}\right)^2} \tag{5.5}$$

The backpropagation involves computing the gradients of the loss function L with respect to the weights at each layer. For a given loss L at time step t:

$$\frac{\partial L}{\partial W^{(l)}} = \sum_{t=1}^{24} \frac{\partial L}{\partial h_t^{(l)}} \frac{\partial h_t^{(l)}}{\partial W^{(l)}} \tag{5.6}$$

Where $\frac{\partial h_t^{(l)}}{\partial W^{(l)}}$ includes the dependencies from previous time steps and layers. Afterwards, an optimizer is required to update the weights which results in reducing overall losses and enhances the model accuracy. Stochastic gradient descent is suggested as the optimizer for the purpose [156].

According to the proposed model, there are two hyper parameters; number of layers in MLRNN and step size (learning rate) of the SGD. The traditional methods like Bayesian or trial and error methods result in the non-optimal selection of these hyperparameters. Hence, in this study, a grid-search method is proposed for the hyperparameters tuning. So, 10% of the samples are randomly selected as the validation dataset, and the model is run through the parameters' grid. Afterwards, the combination of the hyperparameters which results in the best performance will be obtained, and then will be used in training and testing process.

After highly-accurate prediction of the uncertain renewable resources, optimal power management should be conducted to minimize the defined multi-objective cost function. The subject is covered in second phase of study.

5.3.2 Phase Two-MARL for Power Management Scheme

The proposed method has already been developed in Chapter 2. In this chapter, it is focused on customizing it for power management problem-solving, specifically by defining the state, actions, and reward function as the three fundamental factors in RL-based problem. This customization is crucial to tailor the general approach outlined in Chapter 2 to address the unique challenges and requirements of power management. By precisely defining the state variables, possible actions, and the reward function, it can effectively implement and optimize the method for this specific problem domain.

For the state space, the output powers of the BESS, MT, WT, and PV systems, as well as the time of day, h(t) are opted. These selections are made, because the output powers are critical indicators of the energy supply available from different sources at any given time, which directly influences power management decisions, and the BESS is a must for supporting the grid in times of power generation shortage. Including these as state variables ensures that the model can accurately capture the dynamic nature of the power supply and adjust strategies accordingly. Second, the h(t), is a vital component in the state space because energy generation and consumption patterns vary significantly throughout the day. So, the state space is defined as:

$$s_t = \{P_{PV}, P_{PV}, P_{MT}, P_{BESS}, h(t)\}, s_t \in S \tag{5.12}$$

Based on the state space, the agent can issue decision commands for the generation units and BESS at each dispatching interval. The action space is divided into three distinct sets: "idle", which means maintaining the current value; "increase", used to increase power generation or charge the BESS; and "decrease", used to decrease power generation or discharge the BESS.

The reward function serves as a pivotal feedback mechanism, providing the agent with insights into the effectiveness of its actions. This feedback loop is crucial for the agent to glean lessons from its prior experiences and refine its decision-making capabilities over time [157]. A meticulously crafted reward function is imperative, as it enables the agent to discern how best to optimize the actions toward the overarching objectives of cost minimization and efficient power management within the microgrid. The reward function is defined as follows:

$$R_{t+1} = \begin{cases} +1 & when\ CF_{total,t+1} > CF_{total,t} \\ 0 & when\ CF_{total,t+1} = CF_{total,t} \\ -1 & when\ CF_{total,t+1} < CF_{total,t} \end{cases} \tag{5.13}$$

The proposed strategy entails an initial focus on defining the output power of renewable energy resources, followed by the power management solution based on the proposed MARL framework. However, to effectively implement this strategy, it is necessary to model the microgrid, including its constraints and operational parameters, provided in next section.

5.4 MG Modeling and Problem Formulation

Power management in an MG can be accomplished by making sure that power supply and demand are always balanced. This can be quantitatively represented by the following equality equation for the system illustrated in Fig. 5.1. Here, the electric chiller (EC) power is shown by P_{EC}, D_{elec} is the electrical demand, Q_{boiler} and Q_{AC} represent the boiler and absorption chiller capacity, respectively. H_{MT} is the recovered heat from MT, and Q_{hu} considered as fed power to the heating unit.

$$P_{WT}(t) + P_{PV}(t) + P_{MT}(t) + P_{battery,dch}(t) - P_{battery,ch}(t) - P_{EC}(t) = D_{elec}(t) \tag{5.14}$$

$$Q_{boiler} - Q_{AC} + H_{MT} = Q_{hu} \tag{5.15}$$

$$Q_{AC}.COP_{AC} + P_{EC}.COP_{EC} = D_{cooling} \tag{5.16}$$

$$Q_{hu}.\eta_{hu} = D_{heating} \tag{5.17}$$

$$P_{MT}.\frac{\eta_{th}}{\eta_e} = H_{MT} \tag{5.18}$$

In (16), the cooling demand is illustrated by $D_{cooling}$ and, COP_{AC} and COP_{EC} stand for performance coefficients of absorption chiller (AC) and electric chiller. In (5. 17), and (5.18) the η_{hu} is the MT thermal efficiency, η_e is the MT electrical efficiency, and η_{hu} presents the heating unit efficiency.

Then, the overall cost of generating and supplying energy is expressed mathematically by the cost function (CF) in power management problems. By adopting suitable methodology, the optimal solution would be acquired aiming to reduce the cost function. Considering that CO2 emissions, which have enormous negative effects on the environment and the economy, is a

significant challenge. A comprehensive and robust approach to combat climate change and establish sustainable and cost-effective energy systems must include minimizing CO2 emission costs, as well. Hence, in this work, a cost function is established considering both costs of fuel, and CO2 emission. The first objective function demonstrated in (19) depends on natural gas consumption by MT and the boiler (F_{MT}, and F_{boiler}). And the C_{NG} demonstrates the natural gas price per m^3. It should be noted that the PV and WT are considered as free energy resources. Second cost function related to the CO2 emission from burning natural gas is provided in (5.22).

$$CF_1 = \sum_{t=1}^{24}(F_{MT}(t) + F_{boiler}(t)).C_{NG} \tag{5.19}$$

Where:

$$F_{MT}(t) = \frac{P_{MT}(t)}{\eta_e}\beta \tag{5.20}$$

$$F_{boiler}(t) = \frac{Q_{boiler}(t)}{\eta_b}\beta \tag{5.21}$$

$$CF_2 = \sum_{t=1}^{24}(F_{MT}(t) + F_{boiler}(t)).\mu_{CO_2} \times \rho_{CO_2} \tag{5.22}$$

In these equations, β is the conversion factor of the natural gas, and η_b is boiler efficiency. And μ_{CO_2} is the emission conversion factor, and ρ_{CO_2} is the penalty cost of the emission in \$/kg. Finally, the total cost function that are aimed to be minimized, would be:

$$CF_{total} = CF_1 + CF_2 \tag{5.23}$$

The penalty factor for MG power management in the presence of wind turbines and photovoltaic systems needs to closely mirror real-world conditions to ensure the developed model's practicality and effectiveness. This requirement arises from the inherent variability and intermittency of renewable energy sources like WT and PV, which significantly impact the

stability and reliability of power management in a real-world setting. By incorporating a penalty factor that accurately reflects these challenges, the model can better simulate the operational constraints and economic impacts faced by actual power systems. This factor will be integrated into the cost function, ensuring that the optimization process accounts for the additional costs and operational considerations associated with integrating renewable energy sources. The cost function will thus include terms that represent the penalty for deviations from optimal power management, capturing the costs associated with under- or over-utilization of WT and PV systems, and ensuring that the solution is robust, efficient, and reflective of real-world dynamics. This equation illustrates the definition of the penalties, expressed through the symbols ρ_{WT} and ρ_{PV}, which are the fees times the disparity among the calculated and expected generation.

$$CF_{total} = \sum_{t=1}^{24} \left\{ \left(F_{MT}(t) + F_{boiler}(t) \right). C_{NG} + \rho_{WT}. \left(P^{wind,actual}(t) - P^{wind,forecasted}(t) \right) + \rho_{PV}. \left(P^{PV,actual}(t) - P^{PV,forecasted}(t) \right) \right\} \quad (5.24)$$

To comprehensively understand the impact of the penalty factor on microgrid (MG) power management, both scenarios with and without the penalty factors will be considered. Analyzing these scenarios allows us to assess the extent to which the penalty factor influences the optimization process and the overall performance of the power management system.

Optimizing the cost function in a MG may be subject to a number of restrictions and constraints that must be taken into account. For the CCHP, which generates heat, power, and cooling all simultaneously using a single power source, with the purpose of enhancing reliability and stability in MGs, while cutting down the prices and greenhouse gas emissions, the AC, EC, and the boiler must not exceed their rated capacity. So, the CCHP constraints considered as follows:

$$0 \leq Q_{AC}(t) \leq Q_{AC_{max}}(t) \quad (5.25)$$

$$0 \leq P_{EC}(t) \leq P_{EC_{max}}(t) \quad (5.26)$$

$$0 \leq Q_{boiler}(t) \leq Q_{boiler_{max}}(t) \tag{5.27}$$

Micro turbine in a CCHP system produces energy (P_{MT}) by combustion of fuel, such as natural gas, or propane. The steam or hot water released by the turbine's exhaust can subsequently serve for heating purposes or to drive an absorption chiller for cooling. For the MT, when an MT's output power falls below a certain level, the efficiency drastically decreases. So:

$$P_{MT_{min}}(t) \leq P_{MT}(t) \leq P_{MT_{max}}(t) \tag{5.28}$$

On the renewable energy resources of PV and WT, the constraints are provided, as well. According to the (5.29) and (5.30), the PV and WT output power can be limited within zero and rated generation capacity of the units.

$$0 \leq P_{WT} \leq P_{WT_{max}} \tag{5.529}$$

$$0 \leq P_{PV} \leq P_{PV_{max}} \tag{5.30}$$

An energy storage system known as BESS in a microgrid serves to store excess energy ($P_{battery,ch}$) provided by the microgrid for future consumption ($P_{battery,dch}$). The microgrid's energy demand and supply are balanced in part by the BESS, which can also offer backup power during blackouts. The following are some restrictions to take into account when integrating BESS into power management in a microgrid. It should be noted X_{ch} is 1 when the BESS is being charged, and X_{dch} is 1 when the BESS is being discharged, otherwise they take the 0 value.

$$0 \leq P_{BESS,ch}(t) \leq \frac{P_{BESS,ch_{max}}}{\eta_{ch}} . X_{ch}(t) \tag{5.31}$$

$$0 \leq P_{BESS,dch}(t) \leq P_{BESS,dch_{max}} . \eta_{ch} . X_{dch}(t) \tag{5.32}$$

$$X_{ch}(t) + X_{dch}(t) \leq 1 \tag{5.33}$$

The state of charge (SoC) in a BESS is a crucial factor in determining the amount of energy possible to be extracted from and the capacity which is still available for future usage. Monitoring SoC is crucial to prevent overcharging or undercharging, which can impair performance or possibly permanently damage the BESS [150]:

$$q_{BESS}(t) = q_{BESS}(t-1) + P_{BESS,ch}(t).\eta_{ch} - \frac{P_{BESS,ch}}{\eta_{ch}} \tag{5.34}$$

In which q_{BESS} is the battery energy in kWh. Moreover, η_{ch} and η_{dch} are charging and discharging efficiency of the battery energy storage system.

$$q_{BESS_{min}} \leq q_{BESS}(t) \leq q_{BESS_{max}} \tag{5.35}$$

$$\sum_{t=1}^{24} \eta_{ch}.P_{BESS,ch}(t) = \sum_{t=1}^{24} \frac{P_{BESS,dch}(t)}{\eta_{dch}} \tag{5.36}$$

The MG, CCHP, BESS, and modeling parameters defined in the equations are detailed in Table 5.1.

Table 5.1: The Units' Parameters [158].

Parameters	Values (unit)	Parameters	Values (unit)
$P_{MT_{max}}$	100 (kW)	η_e	0.5
$P_{MT_{min}}$	10 (kW)	η_b	0.8
$Q_{boiler_{max}}$	80 (kW)	β	0.09043 (m^3/kWh)
$Q_{AC_{max}}$	100 (kW)	COP_{AC}	0.7
$P_{EC_{max}}$	100 (kW)	COP_{EC}	3
η_{hu}	0.8	C_{NG}	0.157 ($\$/m^3$)
η_{th}	0.3	$\mu_{CO_2} \times \rho_{CO_2}$	0.062 ($\$/m^3$)
ρ_{PV}	0.09 ($\$/kWh$)	ρ_{WT}	0.1 ($\$/kWh$)
$q_{BESS_{max}}$	20 (kWh)	$P_{BESS,dch_{max}}$	6 (kW)
$q_{BESS_{min}}$	6 (kWh)	η_{ch}	85 %
$P_{BESS,ch_{max}}$	6 (kW)	η_{dch}	98 %

So far, the proposed method, along with the MG modeling, cost function definition, and problem constraints, have been meticulously determined. These elements establish a robust foundation for evaluating the performance and effectiveness of the method in managing power within the microgrid framework. In the next section, the proposed method will be rigorously tested on the suggested MG to assess its applicability and validate its potential benefits.

5.5 Simulation Results

MLRNN has been applied to forecast WT and PV output power, and the findings are impressively accurate. The MLRNN prediction model is trained from historical data and therefore can take into consideration the dynamic and nonlinear features of renewable energy resources to provide precise predictions of the 24-hour power output. The Northern California weather data from NREL software called SAM has been utilized for the training process, contains wind speed, wind direction, air temperature, air pressure, and humidity for WT, and solar irradiance, ambient temperature, panel temperature, sun position, and humidity for PV. Grid-search has been employed for optimal auto-tuning of the hyperparameters. In this regard, 10% of the data has been assigned as the validation data set for the mentioned purpose. It should be mentioned that for the grid-search,

the domain for two hyperparameters of number of hidden layers in MLRNN and the step size of the SGD are considered to be between $[1, 8]$ and $[10^{-3}, 10]$, accordingly. The domains are selected to ensure comprehensive exploration during grid search, balancing theoretical considerations and empirical evidence. To have visual understanding of the grid-search, the 3-D plots, which show the optimal hyperparameter values to have minimized nRMSE cost function for both WT and PV power output predictions, based on that 10% randomly selected weather data, are presented in Fig. 5.3. According to the fig. 5.3(a), the optimum hyperparameters of the WT output power prediction is calculated to be 4 hidden layers with the SGD step size of 0.03 for updating the weights. Fig. 5.3(b) demonstrates 3 layers and step size of 0.48 as the optimized hyperparameter values for PV.

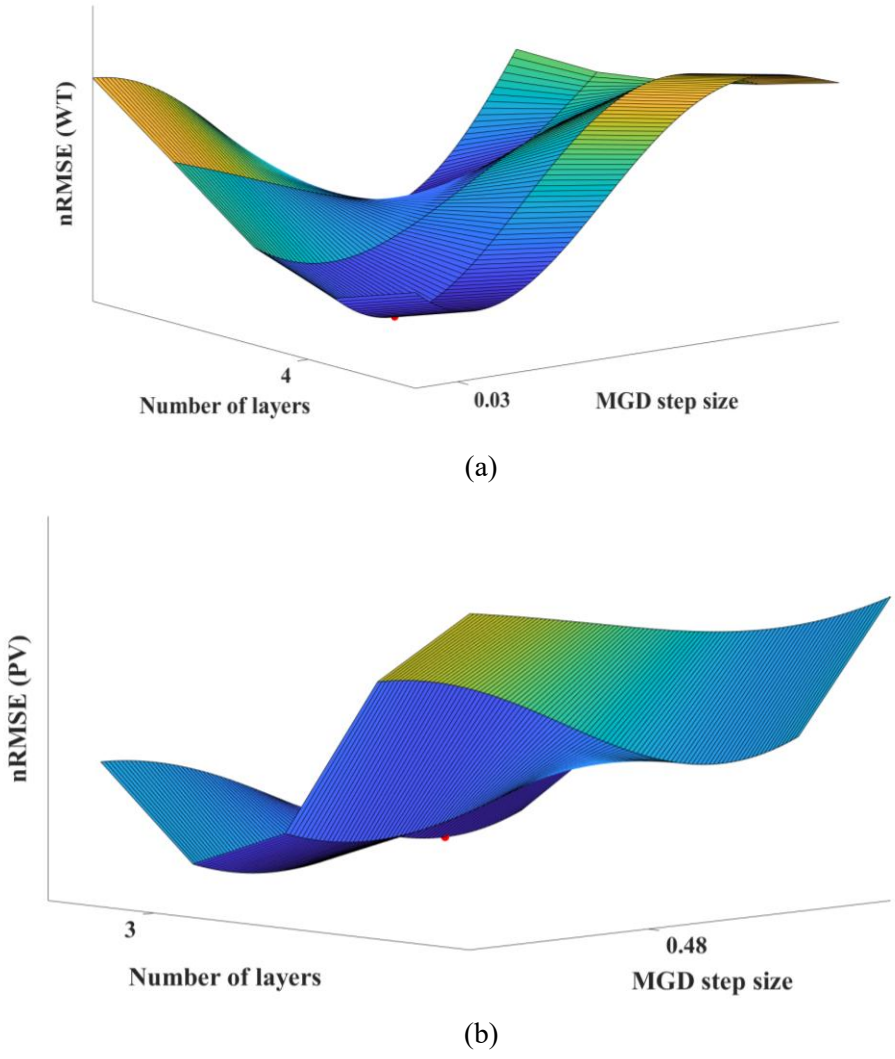


Fig. 5.3: Grid-search 3-D plots for optimum hyperparameters of; (a) WT, (b) PV power prediction.

Based on the defined parameters the proposed model has been trained on 70% of the dataset, using backpropagation optimized by SGD. And 30 epochs each of which contains 20 iterations have been assigned for MLRNN model training. To enhance the accuracy and convergence rate, the data has been preprocessed and shuffled, beforehand. The outcomes illustrate the accuracies of 96.45% and 98.70% for WT and PV output power prediction, respectively. To have a clear overview of accuracy and training behavior of the model within the training in 30 epochs, Fig. 5-4 is plotted for WT, Fig. 5.4(a) shows the accuracy and the Fig. 5.4(b) shows the nRMSE trend over the iterations. These characteristics are almost the same for the PV.

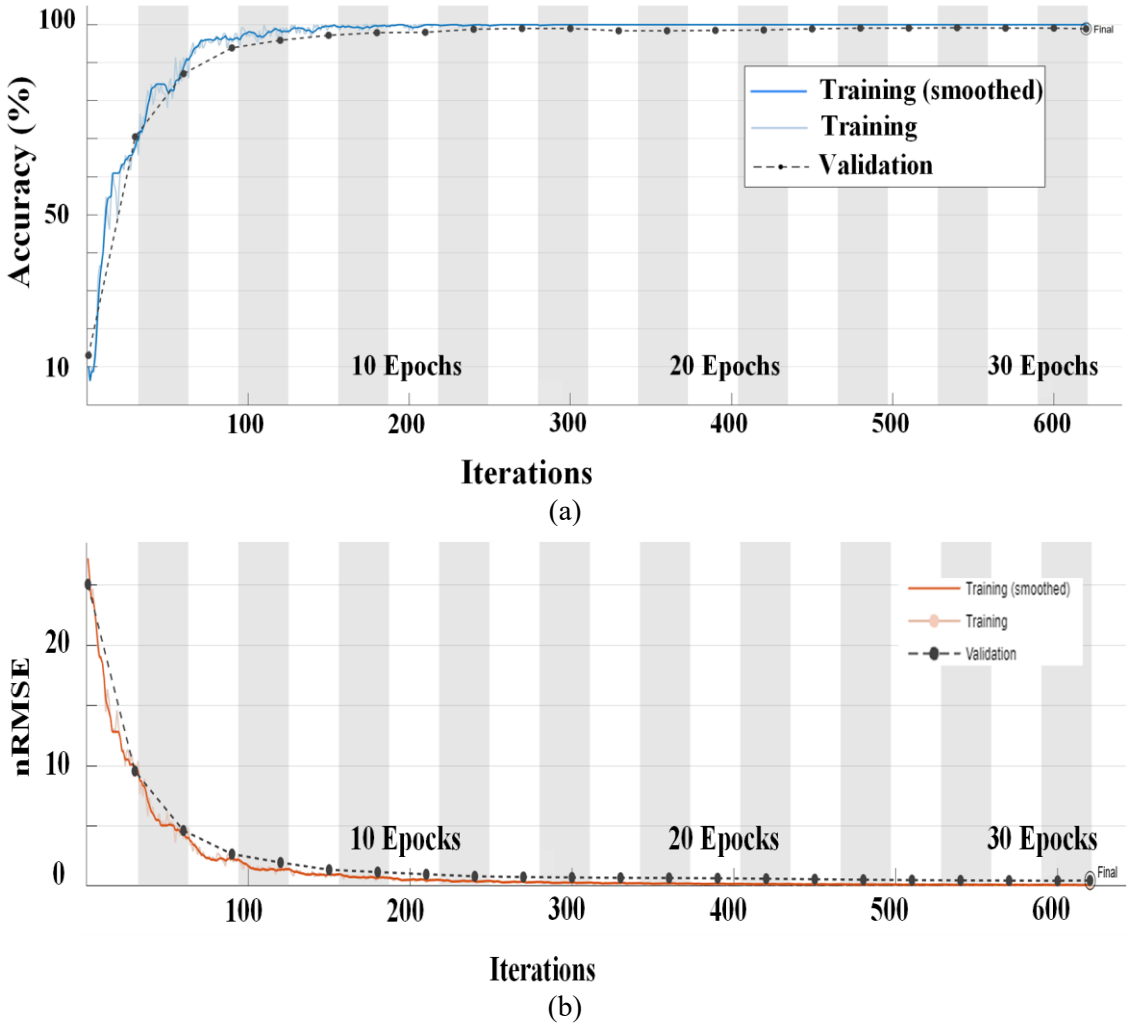


Fig. 5.4: Training process of MLRNN for WT output power prediction. (a) the accuracy trend, (b) the nRMSE error convergence trend.

After successful training, the rest of the 20% data which is determined as testing data set, has been applied to the trained model for predicting the WT and PV output power. The results are provided in Fig. 5.5 and Fig. 5.6, respectively. 24 actual and corresponding prediction samples are demonstrated to validate the accurate performance of the proposed model on unseen test data set. It should be mentioned that the accuracies on test data set for WT and PV output power prediction are measured to be 96.06% and 98.32%, accordingly.

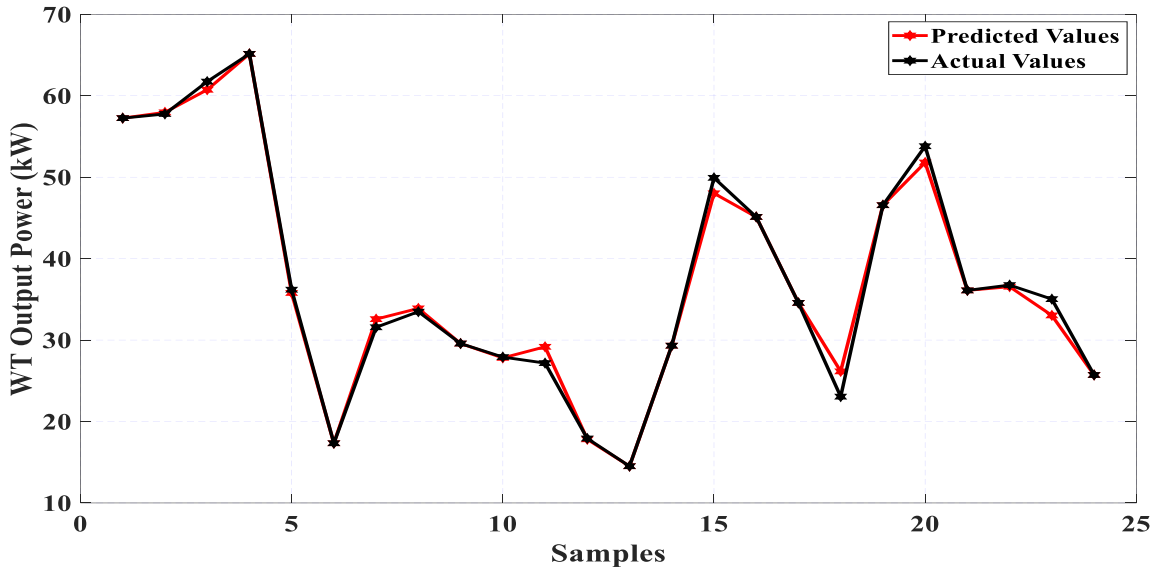


Fig. 5.5: Predicted and actual values of the generated power from WT in 24 hours.

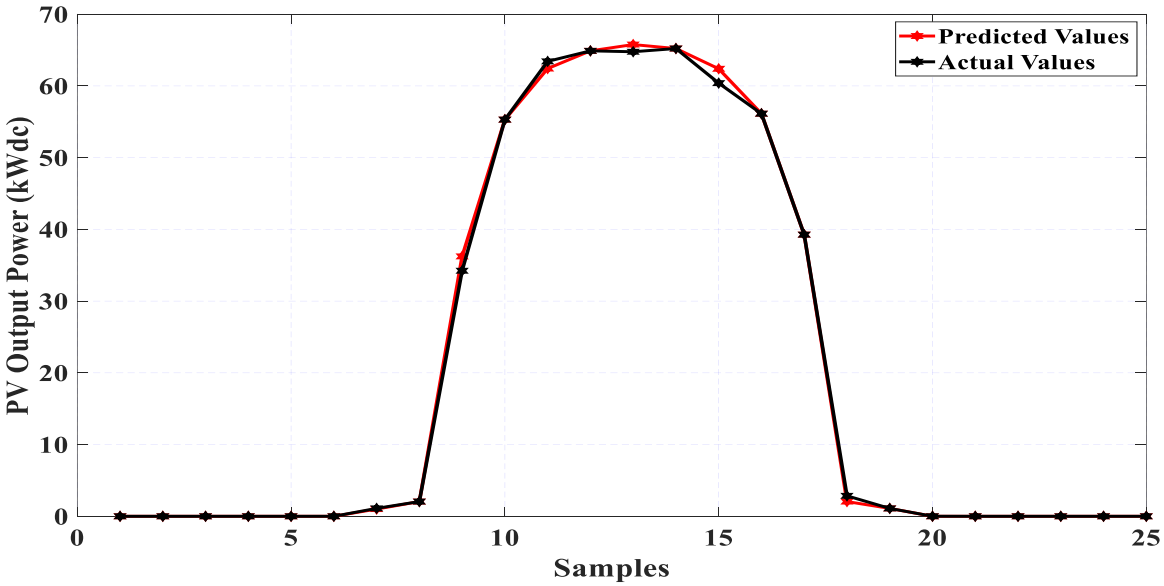


Fig. 5.6: Predicted and actual values of the generated power from PV in 24 hours.

The summary of proposed MLRNN parameters explained above, are gathered in Table 5.2. Moreover, the 24-hour WT and PV output power capabilities are shown in table 5.3, and Table 5.4, which would be used in next phase of study.

Table 5.2: Summary of The Proposed MLRNN Parameters.

WT	
RNN layers	4
Optimizer	MGD
MGD step size	0.03
Batch size	64
Epochs	30
Iteration	20×30
Test accuracy	96.06%

PV	
RNN layers	3
Optimizer	MGD
MGD step size	0.48
Batch size	128
Epochs	30
Iteration	20×30
Test accuracy	98.32%

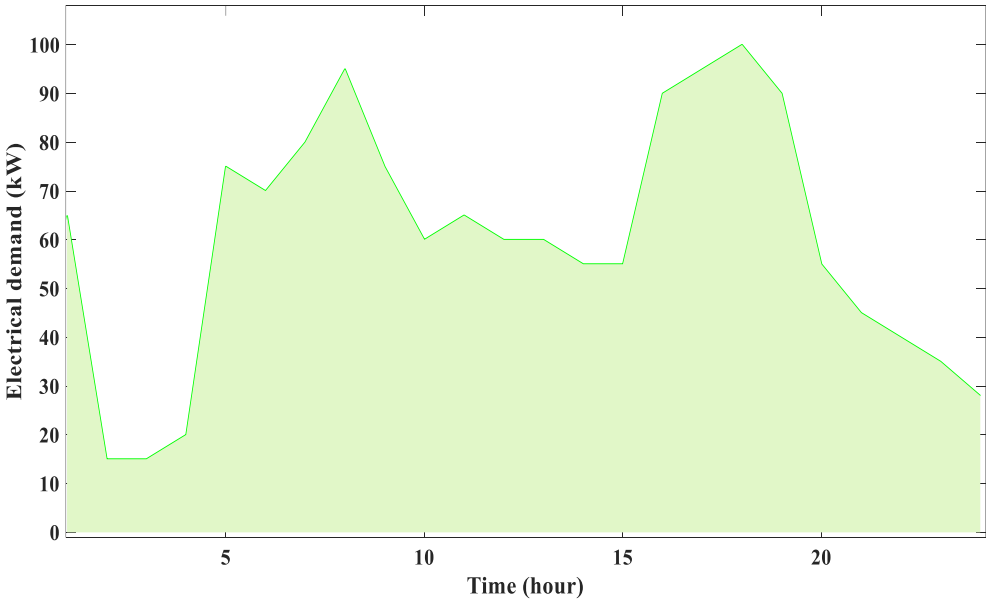
Table 5.3: WT Predicted Output Power Capability.

Hour	WT (kW)	Hour	WT (kW)	Hour	WT (kW)
1	57.2	9	29.5	17	34.4
2	58.5	10	28.6	18	26.8
3	61.3	11	29.1	19	46.5
4	65.0	12	18.4	20	51.1
5	36.7	13	16.2	21	36.3
6	17.5	14	36.9	22	36.7
7	32.1	15	48.9	23	33.0
8	33.2	16	45.3	24	25.2

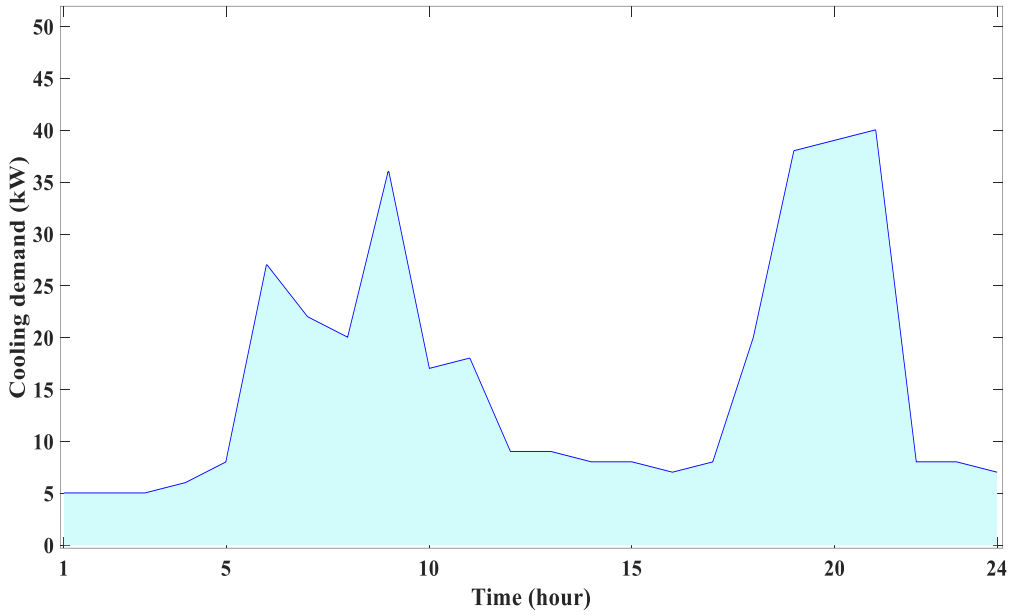
Table 5.4: PV Predicted Output Power Capability.

Hour	PV (kW)	Hour	PV (kW)	Hour	PV (kW)
1	0	9	36.6	17	391
2	0	10	55.4	18	2.2
3	0	11	62.7	19	1.6
4	0	12	65.3	20	0
5	0	13	65.7	21	0
6	0	14	65.9	22	0
7	1.0	15	62.4	23	0
8	2.1	16	56.0	24	0

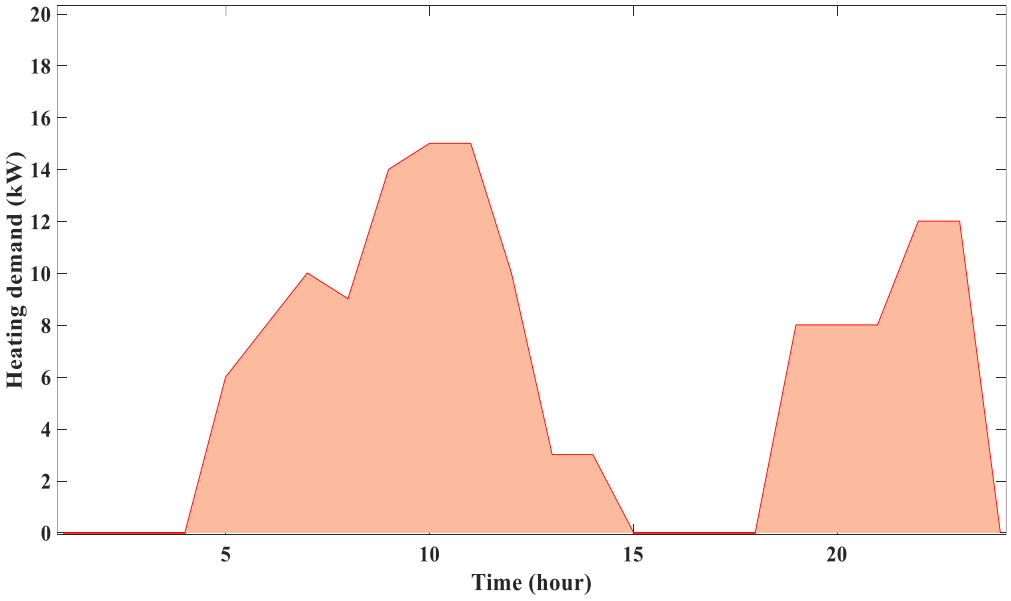
As the next step, proposed MARL algorithm is employed on the MG to validate the robust economic power flow performance of the method. Electrical, heating, and cooling demands are pictured in Fig. 5.7. According to the load profiles and the predicted power capabilities of the renewable energy resources, the proposed power management strategy would be executed.



(a)



(b)



(c)

Fig. 5.7: (a) Electrical, (b) cooling, (c) heating load profiles.

In this study, a dynamic time-sharing electricity price curve is employed in the microgrid model, replacing a fixed-price model. The time-sharing curve, depicted in Fig. 5.8, illustrates the fluctuation of electricity costs over the course of a day. This dynamic pricing strategy offers several advantages over a fixed-price model. The dynamic pricing reflects the real-world variability of electricity costs, providing a more accurate representation of the economic factors that influence energy management decisions. By incorporating this dynamic pricing approach, the model can better simulate and optimize the microgrid's operations under realistic conditions, leading to more effective and economically viable solutions.

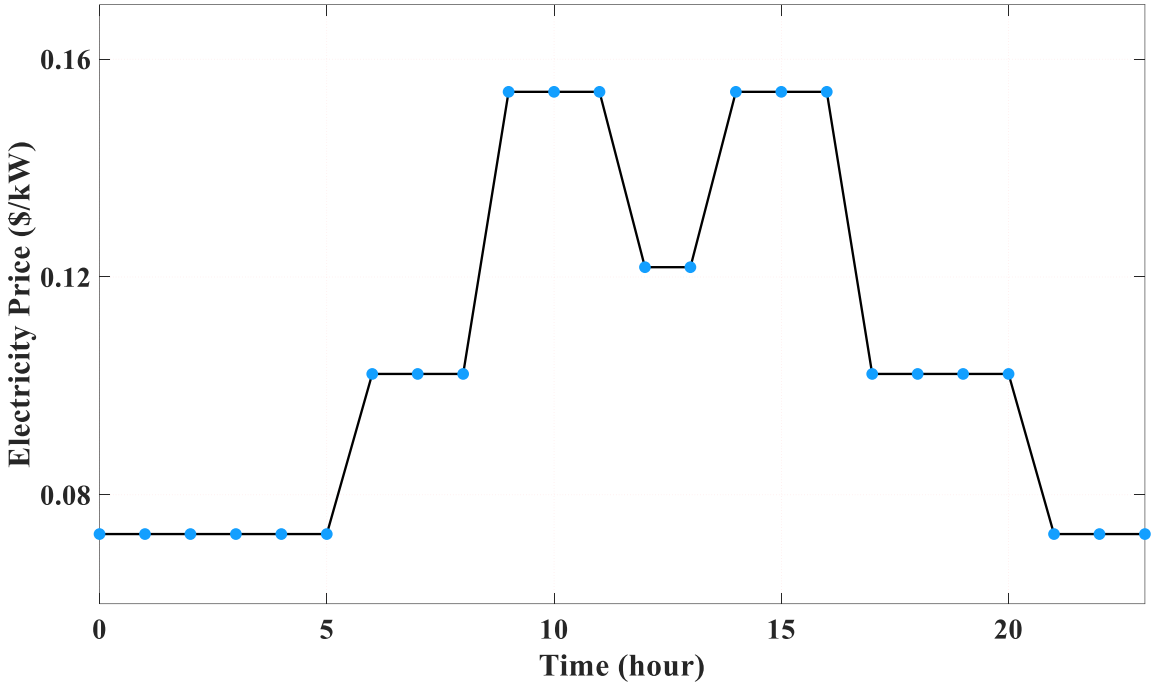


Fig. 5.8: Time-sharing electricity price curve.

Four agents have been assigned to find the optimal solution for the developed cost function without penalty factors for PV and WT. And the discount factor is chosen to be 0.91 by trial and error. The dispatch results are provided in Fig. 5.9, contains the generation units' output power within a 24-hour period. According to the figure and comparing it with the loads' profiles, it can be deduced that MT has been assigned mainly to feed heating and cooling demands, due to the natural gas price. And the WT and PV in high demand periods operate near their maximum power generation capability. Keeping a minimum generation level for MT in a microgrid helps ensure

stability, operational efficiency, power quality, economic viability, integration with renewable sources, and the provision of ancillary services. This minimum output power was defined to be 10% of the rated capacity. So, in low demand hours, MT has worked at its minimum operation point which is 10 kW. The mentioned observation can validate proper functionality of proposed MARL to have an economical and optimal power management, while considering the pre-defined constraints. And based on fuel price and CO2 emission cost which were provided in Table I, total cost for the acquired power management among the generation units attained to be almost 224$.

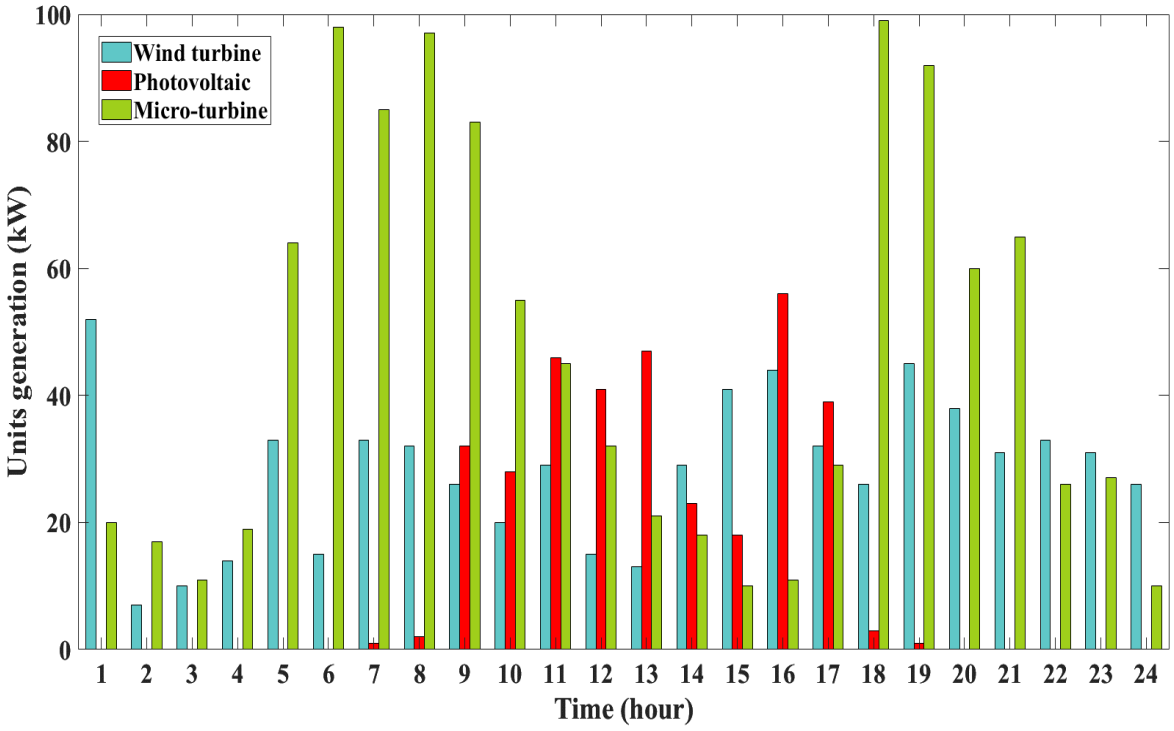


Fig. 5.9: The generation units' managed output power without penalty factor consideration.

The outcomes for the minimized cost function considering penalty factors for PV and WT are demonstrated here. The explanation mirrors the scenario without penalty factors, illustrating the effectiveness of the proposed method for this more realistic power management problem, as well. However, the cost calculated is $241, which represents an increase due to the inclusion of penalty factors that account for various operational constraints and inefficiencies. These penalty factors are essential for capturing the cost for differences between generated power from the renewable energy sources and the expected available power. By incorporating this element, the model provides a more accurate representation of the actual costs involved in managing PV and WT

systems. This inclusion inevitably results in a higher cost figure compared to the idealized scenario without penalties. The rest of the study is based on this scenario, and the result of energy dispatch amongst the generation units is shown in Fig. 5.10.

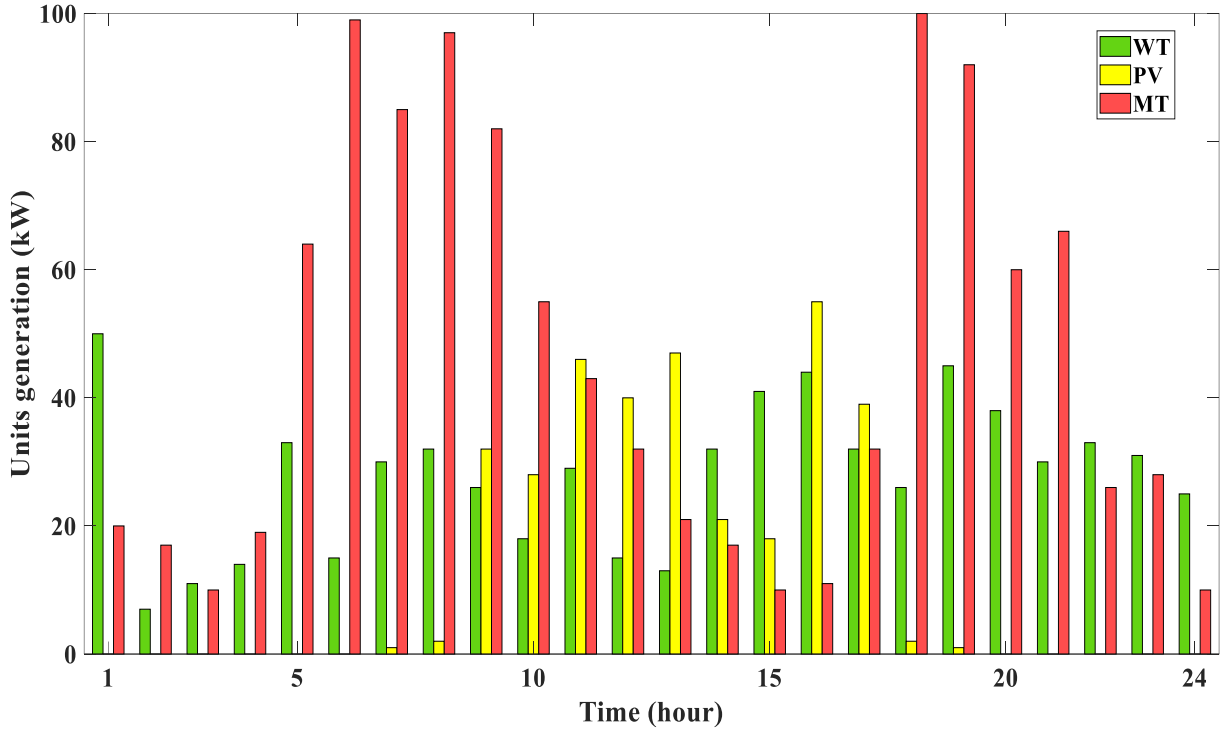


Fig. 5.10: The generation units' managed output power considering penalty factors.

The charging states of the BESS over 24 hours are illustrated in Fig. 5.11. The initial BESS charge is set at 10 kW at 1:00 am to ensure there is a baseline storage level to manage early morning demand when generation might be lower. During times when the generation units cannot meet the electrical demand, the battery discharges to supply the load, with a maximum discharge rate of 6 kW per step. Conversely, the battery is scheduled to recharge at the earliest possible moment when excess generation is available, with a maximum charge rate of 6 kW per step. This deliberate limitation on charging and discharging rates is strategically implemented to achieve several critical objectives. First, by capping the maximum step at 6 kW, we reduce the stress on the battery cells, which enhances their longevity. Rapid charging and discharging can lead to significant wear and tear on the cells, reducing the overall lifespan of the BESS. Second, this approach minimizes the risk of overheating, which is crucial for ensuring safe operation. Overheating can not only damage the battery but also pose safety risks. Third, operating within this optimal power range improves

the efficiency of the BESS. Batteries typically perform better and are more efficient when charged and discharged at moderate rates rather than at their maximum capacity. Moreover, the charging and discharging states of the BESS with positive and negative values, respectively, are illustrated in Fig. 5.12.

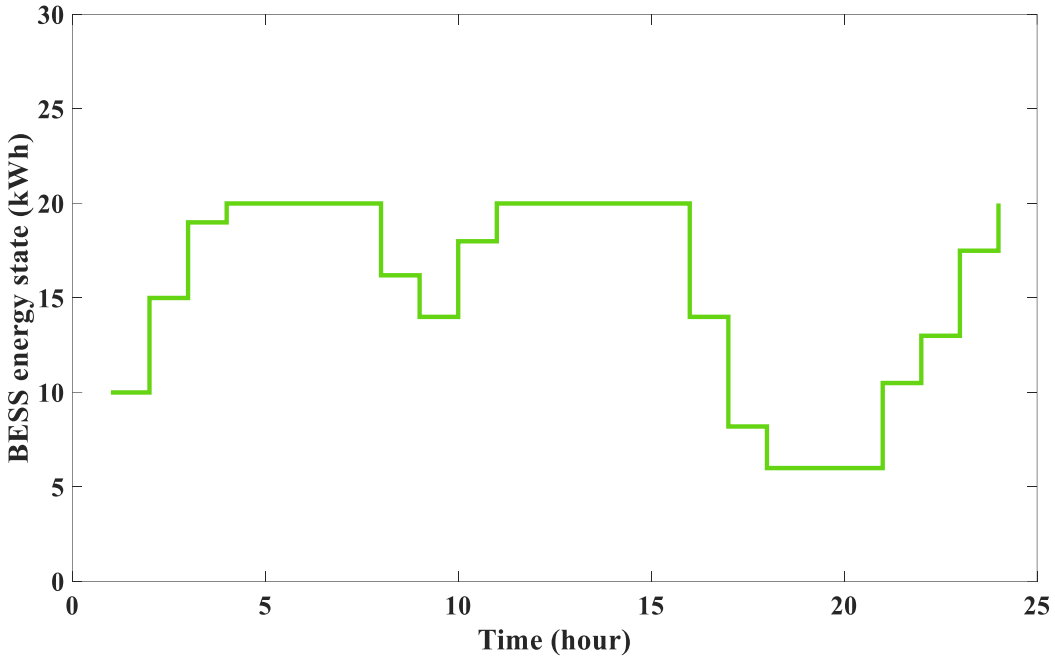


Fig. 5.11: The BESS energy.

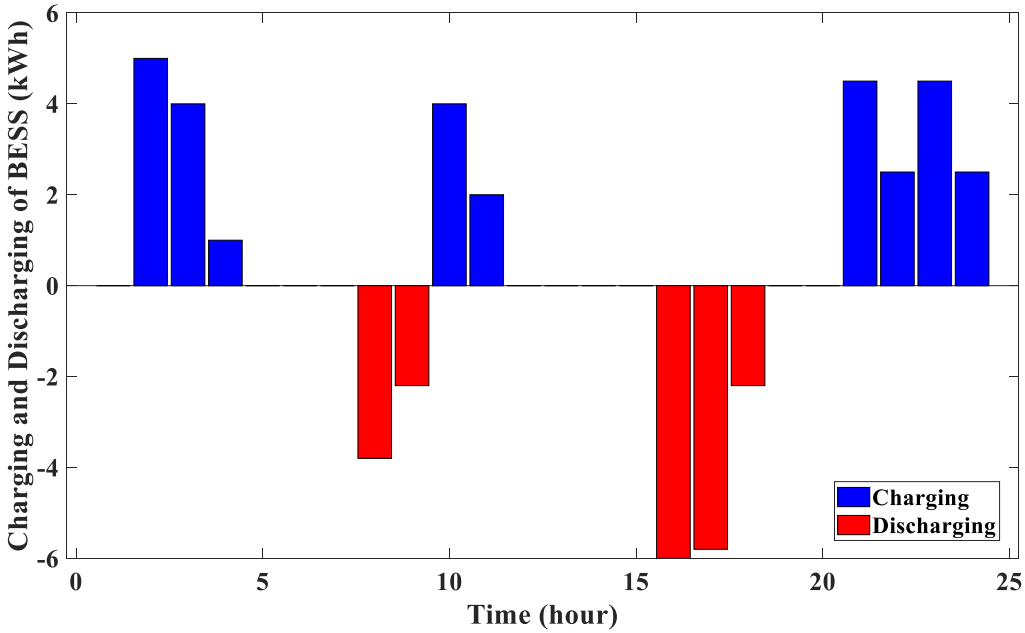


Fig. 5.12: The BESS charging and discharging.

5.5.1 Comparative Study

Comparative analysis is also provided to validate the superior performance of the proposed MARL solution over single agent RL. The provided Fig. 5.13 illustrates a comparative analysis focusing on their cumulative rewards. In the figure, the single-agent RL method starts with higher cumulative rewards and converges quickly within almost 90 iterations, reaching a final value around 50. On the other hand, the proposed MARL solution exhibits a slower start with noticeable fluctuations in cumulative rewards. The exploration phase leads to fluctuations in individual agent rewards as they try to learn the beneficial actions for different scenarios. Moreover, communication and coordination among agents introduce complexity, potentially causing fluctuations as agents adjust their strategies based on shared information. Despite this initial phase, the MARL approach shows a remarkable improvement over time and almost converges to a final value around 62, which is significantly higher than the single-agent performance. The superior performance of the MARL solution is attributed to its ability to leverage the collective intelligence and coordination among multiple agents. By allowing agents to learn and collaborate simultaneously, the MARL approach can explore a wider range, adapt to dynamic environments of the MG more effectively, and ultimately achieve higher cumulative rewards for power management purpose.

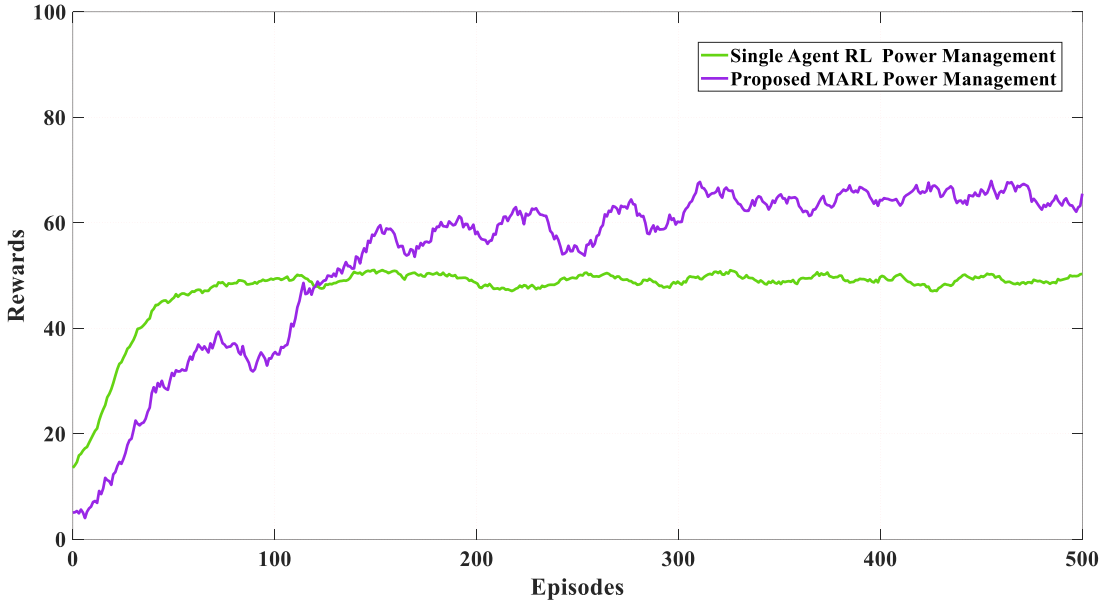
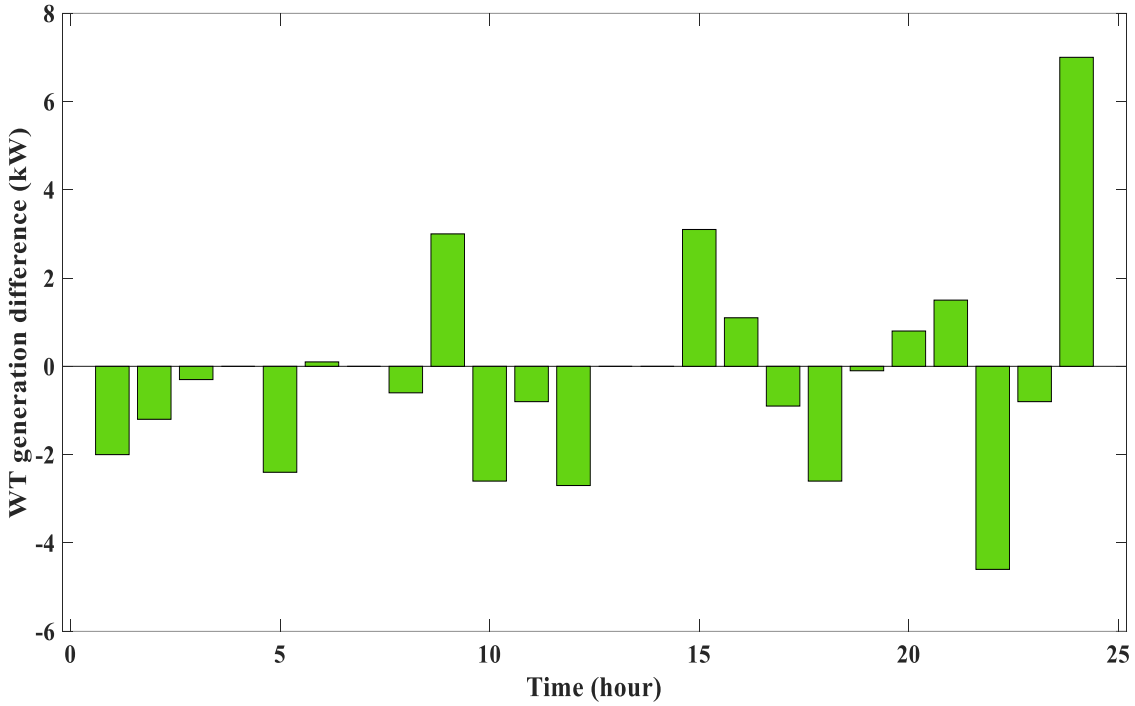
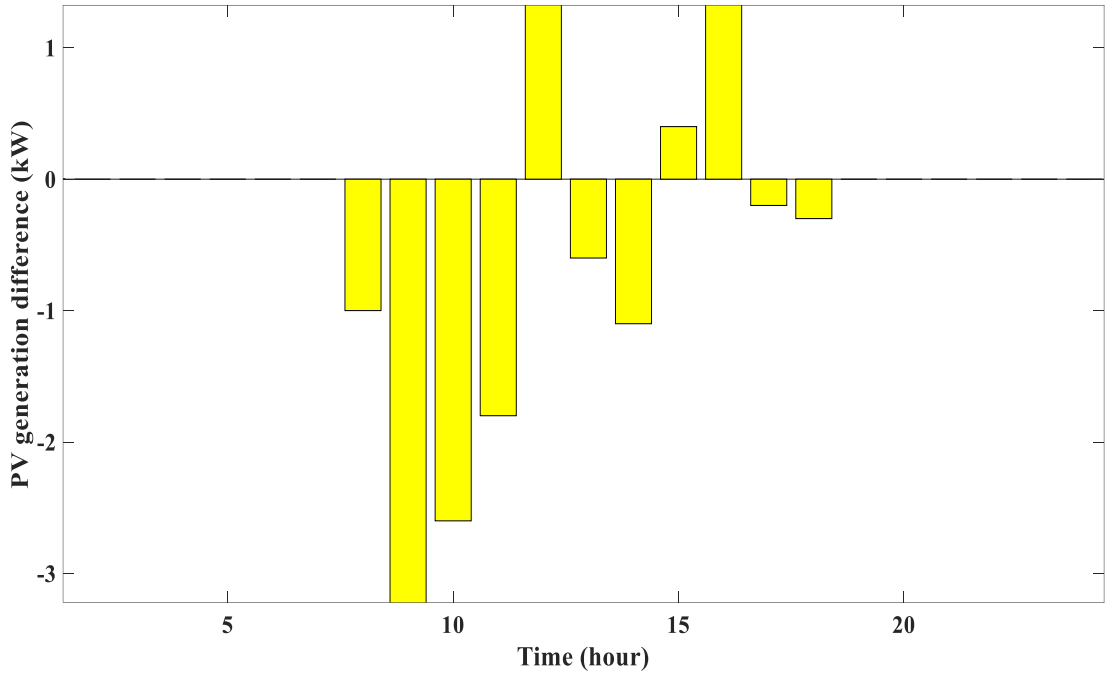


Fig. 5.13: The comparative study in terms of cumulative rewards.

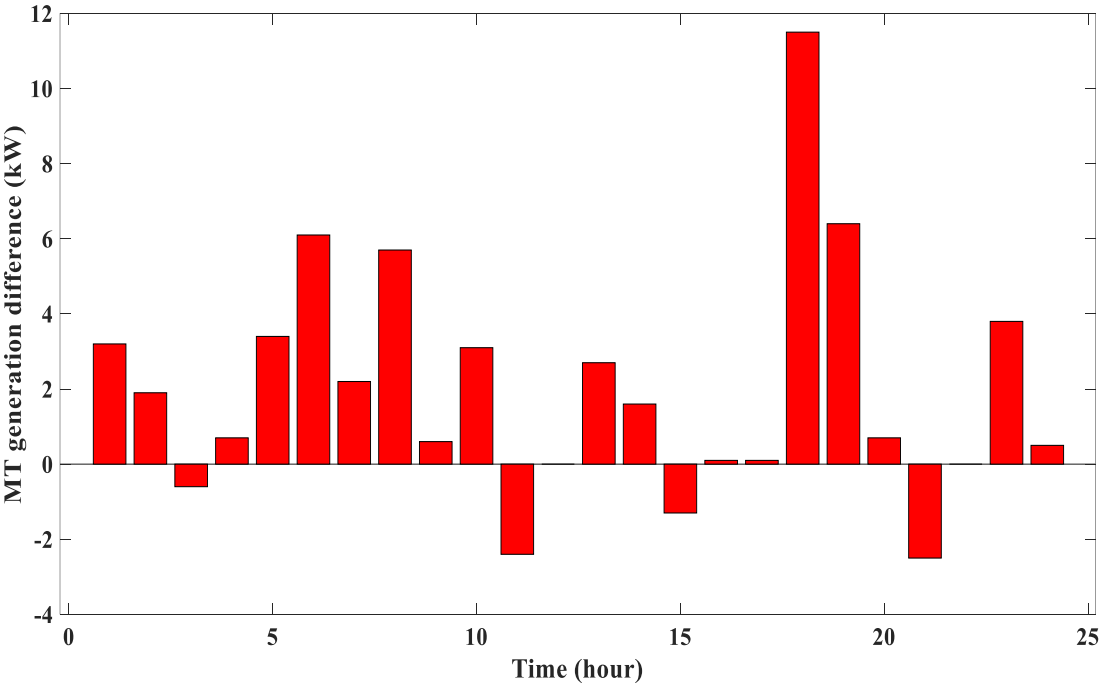
Furthermore, the single agent RL is employed to optimize the provided cost function in the same MG configuration. Through the analysis and calculations, the cost is determined to be 257 dollars. In comparison, the proposed MARL solution achieved a lower cost of 241 dollars, indicating an improvement in cost efficiency of the proposed power management solution. The Fig. 5.14 is also depicting the difference in units' generation between the single-agent RL and MARL approaches. The negative values show less generation and the positive values represent the higher generation compared to the proposed MARL power management strategy. According to the figure, single agent RL increases the power generation from micro-turbine (Fig. 5.14(c)) and drop the renewable energy free resources generation in some hours, which leads to higher operational cost, compared to the suggested method.



(a)



(b)



(c)

Fig. 5.14: The units' generations differences compared to the proposed MARL employed single agent RL by (a) WT, (b) PV, (c) MT.

Furthermore, the proposed method is tested with various number of agents and the results are demonstrated in Table 5.5. The provided table offers a detailed comparison of different numbers of agents in proposed reinforcement learning framework, considering their impact on cost function value, convergence speed, memory usage, calculation time, and cumulative rewards. As the number of agents increases, the cost function value generally decreases up to four agents, reaching its minimum at four agents by 241$. This suggests that increasing the number of agents improves the ability to optimize long-term planning and reduce overall costs. Beyond four agents, the cost function value starts to increase, indicating that adding more agents does not necessarily lead to better performance and might even degrade it.

The convergence speed, indicated by the number of iterations required, initially increases with the number of agents, as more agents mean more complex interactions and dependencies that need to be learned. Specifically, the convergence speed increases from 90 iterations for one agent to 320 iterations for four agents. However, after four agents, convergence speed begins to slow down significantly, highlighting the diminishing returns in performance with additional agents. For instance, with six agents, convergence speed is at 410 iterations, and beyond six agents, the system fails to converge.

Memory usage also increases with the number of agents. As more agents are introduced, the system needs to store more state information and interaction data, leading to higher memory consumption. The memory usage increases from 327 kB for one agent to 809 kB for four agents and spikes to 4614 kB for six agents, indicating substantial resource requirements. The sharp increase in memory usage beyond four agents suggests that the system becomes more resource-intensive without corresponding improvements in performance. Calculation time shows a similar trend, increasing with the number of agents due to the additional computational complexity. Calculation time rises from 0.81 seconds for one agent to 3.76 seconds for four agents and jumps significantly to 13.98 seconds for six agents. This indicates that while more agents can enhance decision-making, the computational cost increases substantially, especially beyond four agents.

Cumulative rewards increase with the number of agents up to four, suggesting that more agents help the system accumulate more rewards by making better long-term decisions. The cumulative rewards increase from 50 for one agent to 62 for four agents. However, beyond four agents, cumulative rewards start to decline, indicating that the system becomes less effective in

optimizing long-term rewards, possibly due to the increased complexity and resource contention among the agents.

Overall, the proposed configuration of four agents strikes an optimal balance among cost, convergence speed, memory usage, calculation time, and cumulative rewards, and achieving the lowest cost function value at $241 and the highest cumulative rewards at 62. This configuration minimizes the cost function value and maximizes cumulative rewards while maintaining reasonable convergence speed and resource usage. Adding more agents beyond this point results in diminishing returns, higher resource consumption, and potentially instability, making four agents the most effective choice for this power management problem.

Table 5.5: Comparative Study in Terms of Number of Agents.

Number of agents	Cost function value ($)	Convergence speed (approximate iterations)	Memory usage (kB)	Calculation time (s)	Cumulative rewards (Approximately)
1	257.16	90	327	0.81	50
2	255.54	130	411	1.44	54
3	246.01	150	798	2.25	57
4 (proposed)	**241.76**	**320**	**809**	**3.76**	**62**
5	246.22	270	1324	2.31	56
6	283.76	410	4614	13.98	39
7	inf	-	-	59.99	-inf
8	inf	-	-	48.31	-inf

As mentioned, a discount factor of 0.91 was selected for the four-agent RL power management solution. To evaluate the effect of different DF values, various alternatives were considered, and the results are provided in the table below. The DF of 0.91 was chosen as it offered the best balance across several key metrics, including response time, memory use, cumulative

rewards, and system stability. The advantages of a DF of 0.91 include its ability to adequately prioritize future rewards while maintaining efficient short-term decision-making, resulting the best cost function value among other DF values. This balance helps in achieving a more stable and effective power management strategy. The Table 5.6 is provided to comprehensively shed light on the discussed material.

The provided table offers a comprehensive comparison of different discount factor (DF) values in terms of cost function value, convergence speed, memory usage, calculation time, and cumulative rewards. As the DF increases, the cost function value decreases, reaching its minimum at a DF of 0.91. This indicates that higher DFs (up to a certain point) enhance long-term planning, thereby reducing overall costs. However, beyond a DF of 0.91, the cost function value begins to rise slightly, suggesting diminishing returns or potential over-optimization for future rewards at the expense of immediate costs.

The convergence speed, represented by the number of iterations required, increases with the DF from 0.3 to 0.91. This trend is expected because a higher DF means the agent considers more future rewards, making the value function more complex and harder to optimize quickly. At a DF of 1, the agent fails to converge, indicating instability and unrealistic value estimations due to the overvaluation of future rewards. Memory usage also increases steadily with the DF. Higher DFs necessitate storing more detailed future reward information, hence the higher memory consumption. The significant jump in memory usage from a DF of 0.91 to 0.95 highlights the steep increase in resource requirements for small increments in DF at higher values. Similarly, calculation time increases with DF due to the additional computational complexity associated with considering a longer horizon of future rewards. The calculation time at DF = 1 is notably low because the agent fails to converge, leading to premature termination.

Cumulative rewards increase with the DF up to 0.91, indicating that higher DFs help the agent accumulate more rewards by making better long-term decisions. The slight drop in cumulative rewards at DF = 0.95 and the unrealistic values at DF = 1 suggest that excessively high DFs may not effectively balance immediate and future rewards, leading to suboptimal performance.

The DF of 0.91 stands out as it provides the best balance between short-term and long-term rewards, resulting in the lowest cost function value and the highest cumulative rewards without excessive memory usage or convergence issues. The convergence speed and calculation time at

this DF are reasonable, making it a practical choice for real-world power management applications where computational resources and time are limited. Additionally, the agent remains stable and performs well, avoiding the pitfalls of very high DFs that can lead to unrealistic valuations of future rewards.

Table 5.6: Comparative Study in Terms of DF Value.

DF value	Cost function value ($)	Convergence speed (approximate iterations)	Memory usage (kB)	Calculation time (s)	Cumulative rewards (Approximately)
0.3	259.87	200	721	3.37	57
0.5	277.61	220	738	3.40	57
0.7	271.84	280	766	3.51	57
0.9	247.12	320	781	3.98	60
0.91 (Proposed)	**241.76**	**320**	**809**	**3.76**	**62**
0.95	248.33	330	1307	4.45	59
1	inf	-	-	0.99	-inf

5.6 Conclusion

In order to overcome the obstacles in micro-grid power management, a recurrent neural network along with reinforcement learning based method was proposed in this work. The projected MG included a BESS, WT, PV, and CCHP systems. A multi-layer recurrent neural network model was developed and trained using historical wind and solar dependent factors dataset to efficiently estimate the generation of day-ahead renewable WT and PV resources. The strategy featured advantages over previously investigated methods, including higher accuracy and the enhanced ability to extract non-linear mapping functions between the input and output. To improve the model, the hyperparameters of number of hidden layers and learning rate for wights updating

were auto-tunned via the grid-search. The anticipated values would subsequently be used in the following phase, which was the optimal power management throughout MG's generation units. In order to achieve a dependable and environmentally friendly MG, a model free reinforcement learning was used to maintain the multi-objective fuel and CO2 emission cost function minimized. Hence, a four-agent RL approach was suggested to cope with the task. By integrating the smaller tasks that each agent performed, multi-agent solutions aimed to solve the complicated issue of discovering the best operating points in the challenging environment, rather than four agents. The results demonstrated that the proposed MARL successfully distributed the power generation amongst micro turbines in CCHP system and other renewable resources, economically. The outcomes also confirmed that the proposed MG's BESS performed properly under both light and heavy load conditions within a day.

Chapter 6
Conclusion

6.1 Summary

This PhD thesis aimed to enhance wind energy conversion systems by leveraging advanced techniques unified by a proposed multi-agent reinforcement learning method to address challenges in fossil fuel consumption, carbon dioxide emissions, and global warming.

In Chapter-1, an overview of WECS was provided, including different configurations, power converters, and control strategies. Historical numerical information about WECS was also reviewed, laying the groundwork for understanding the current state and potential advancements in wind energy technology.

Chapter-2 detailed an AI-based controller and the step-by-step development of the proposed MARL method equipped with meta-learned discount factors. This approach aimed to enhanced system adaptability and performance, distinguishing itself from traditional control strategies that often struggled with parameter tuning and adaptability under varying conditions.

Chapter-3 delved into the control of a neutral point clamped power converter in a direct-drive permanent magnet synchronous generator (PMSG)-based WECS. The focus was on improving power quality and meeting grid code requirements for total harmonic distortion. The proposed MARL approach eliminated the need for extensive offline training and system modeling, offering superior performance in dynamic and unbalanced scenarios compared to traditional and AI-based methods.

In Chapter-4, the thesis addressed maximum power point tracking challenges in WECS. Traditional methods like Perturb and Observe and Incremental Conductance were compared with the customized MARL approach, which employed multiple agents for improved energy output and responsiveness. The MARL strategy, validated through simulations and a 1000W prototype implementation, demonstrated significant improvements over traditional MPPT methods.

Finally, Chapter-5 explored power prediction and management within a microgrid environment, integrating renewable energy sources such as WT, PV systems, BESS, and CCHP. A multi-layer recurrent neural network was developed for accurate 24-hour forecasting, and also optimizing power management using the proposed MARL method. This approach minimized fuel and CO_2 emissions costs, enhanced coordination among MG components, and ensured efficient resource utilization and BESS scheduling.

Overall, this thesis presented a robust and innovative framework for enhancing wind energy conversion, renewable energy forecasting, and microgrid power management through a unified MARL-based approach. Python and Simulink MATLAB used for coding and simulation purposes. The findings underscored the potential of MARL to significantly improve the efficiency, reliability, and environmental sustainability of renewable energy systems and WECS.

6.2 contributions

The contributions of this research on power converter power quality are summarized as follows:

- The research introduced MARL to control power converters, extending the capabilities of RL by allowing multiple agents to interact. This was an innovative application of MARL in wind energy systems, representing a pioneering contribution to the field of control strategies. Also, a novel approach was suggested to optimize the MARL hyperparameter known as the discount factor, significantly enhancing the performance and response of the system compared to traditional tuning methods that often relied on manual adjustments.
- In a fully cooperative environment like WECS, agents typically focus on maximizing their own rewards, often neglecting the rewards of other agents. This behavior tends to lead to suboptimal performance. To resolve this issue, the author proposed incorporating Nash equilibrium and Nash social welfare strategies, encouraging the agents to account for both their individual rewards and the cumulative rewards of the entire system.
- By eliminating the requirement for detailed system modeling, this study substantially reduced the computational workload, making the control system more practical and less resource-intensive. Moreover, the elimination of some PI controllers removed the need for parameter

tuning and bandwidth adjustments, simplifying the control process and reducing maintenance complexities.

- This study reduced power quality issues like single or double phase voltage drop, which are often challenging for other AI models. By leveraging the adaptability and dynamic response capabilities of the MARL, this work ensured a stable and reliable power supply, and preserving the integrity of the electrical grid. This contribution was vital for the sustainable integration of renewable energy sources into the power grid, enhancing overall system resilience and efficiency.

The impact of this research on highly efficient MPPT is summarized as follows:

- The research introduced a novel application of multi-agent reinforcement learning, extending its use beyond power converter control to include maximum power point tracking. This innovative approach expanded the horizon of research in this domain. Additionally, the highly efficient meta-learning method was proposed for optimizing the discount factor, significantly enhancing the system's performance, efficiency, and responsiveness compared to traditional methods that often relied on manual adjustments. The proposed algorithm improved the energy capture capabilities of wind turbines, ensuring that the system operated at peak efficiency under different wind condition scenarios and provided a fast response to wind speed changes, compared to the counterparts.
- At times, agents may deviate from the learned optimal policy, leading to suboptimal model performance and fluctuations in the MARL learning process. To overcome this challenge, the Q-Bellman equation was introduced by the author to guide agents within the MARL framework for the MPPT problem, ensuring they adhere to the optimal policy that maximizes cumulative rewards. This approach helps maintain consistent performance and improves overall system efficiency.
- The study emphasized the practicality of the proposed multi-agent reinforcement learning method for maximum power point tracking by implementing it in a laboratory environment. This hands-on validation bridged the gap between theoretical RL-based models and real-world applications in power systems. The implementation took place on a DSP TMS320F28335 real-

time control board, showcasing the robustness and effectiveness of the proposed control algorithms equipped with a meta-learned discount factor.

The impact of this research on environmentally friendly power management and BESS energy scheduling is summarized as follows:

- The research proposed a customized MARL method for power management in a comprehensive microgrid integrating PV, WT, CCHP units, etc. method aimed to optimize a multi-objective function to reduce both fuel consumption and CO_2 emissions while considering the penalty factors for PV and WT generations. Additionally, the study considered the BESS scheduling to support the microgrid in stability and reliability.
- The study focused on determining the optimal number of agents and optimal discount value in the MARL, an aspect that had not been explored previously. This optimization ensured a balanced approach between complexity and effectiveness in managing the microgrid, leading to improved performance and efficiency.
- MLRNN was proposed to accurately predict the WT and PV output power which would be used as the maximum capability of the generation resources for the power management purpose. The research employed grid-search optimization to fine-tune the MLRNN hyperparameters; number of layers and SGD step size. These critical hyperparameters significantly influenced the learning process and performance, resulted in better convergence and overall system performance compared to traditional manual tuning methods.

6.3 Future Work

As future work of the doctoral study the following projects are proposed to achieve more robust and efficient direct-drive PMSG-based WECSs.

- In exploring optimal reinforcement learning models for WECS applications, there are other recently-studied models such as; Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), Soft Actor-Critic (SAC), etc., that are valuable to be investigated. Comparing these approaches will provide insights into

their effectiveness in optimizing the responses, as well as ensuring computational efficiency.

- Integrating MARL with a meta-learned discount factor substantially enhances performance in WECS by improving adaptability and optimizing power output, according to the outcomes. However, this integration leads to increased computational burden and memory usage. To mitigate these challenges and achieve a more optimal solution, the code can be refined through the implementation of more efficient algorithms that streamline processes and reduce complexity, by employing techniques such as model compression and parallel processing which can significantly decrease computational load
- In this study, an islanded microgrid was analyzed. Future research could focus on grid-connected MGs, as well as the transition between islanded and grid-connected modes. Exploring these dynamics is crucial for understanding the resilience and flexibility of MGs, ensuring stable power supply during grid fluctuations or outages. Studying the seamless transition between modes can enhance the reliability of renewable energy integration and improve the overall stability of the power system.

6.4 Conclusion

To mitigate fossil fuel consumption, which leads to carbon dioxide emissions and global warming, renewable energy gained prominence. Among the various renewable sources, wind energy emerged as one of the most cost-effective solutions for electricity generation. Numerous studies focused on enhancing the performance of wind energy conversion systems (WECS) in different areas. However, traditional control strategies employed in WECS often resulted in lower efficiency, complicated implementation, complex system modeling, intricate drive circuit design, and suboptimal responses. This PhD thesis presented a thorough investigation of advanced techniques to optimize WECS, centered around a proposed multi-agent reinforcement learning method. The research was organized into three primary objectives, each contributing to the advancement of renewable energy technologies through the innovative application of MARL.

Firstly, the thesis examined the control of a neutral point clamped power converter used in a direct-drive permanent magnet synchronous generator based WECS. The aim was to improve

power quality and comply with grid code requirements for total harmonic distortion. Traditional controllers, such as PI, often struggled with parameter tuning and adapting to varying operating conditions, leading to suboptimal performance in dynamic and unbalanced scenarios. The proposed approach eliminated the need for offline training and extensive system modeling, distinguishing it from traditional machine learning based techniques, and PI-based methods. Simulations and comparative analyses validated the MARL strategy, particularly in addressing unbalanced voltage sag conditions. The integration of meta-learning to optimize the discount factor (DF), a critical hyperparameter in RL-based methods, further improved the control system's adaptability and convergence rate, ensuring power quality, and higher power converter efficiency.

Subsequently, the research tackled challenges in maximum power point tracking for the same wind energy conversion systems. Conventional methods like P&O was characterized by slow dynamic response and susceptibility to steady-state oscillations around the maximum power point, especially under rapidly changing wind conditions. The customized MARL approach overcame these limitations by utilizing multiple agents that collaborated, improving energy output and responsiveness to wind speed variations. The meta-learned discount factor optimized the MARL algorithm, reducing learning time and enhancing convergence. Extensive simulations, along with a 1000W prototype implementation, demonstrated the MARL strategy's superiority over traditional MPPT methods, confirming its practical advantages and reliability in real-world applications.

Lastly, the thesis explored power prediction, management, and energy scheduling in a microgrid. The MG integrated renewable energy sources such as WT, PV systems, and BESS, alongside combined cooling, heating, and power units. Traditional forecasting models, often failed to capture the complex temporal dependencies and variability of renewable energy sources, leading to inaccurate predictions. A multi-layer recurrent neural network was developed to provide accurate 24-hour forecasts for renewable PV and WT energy generation, in the first phase of study. A grid-search method was used to optimally tune the number of RNN layers and the optimizer learning rate. The predicted values were then used as the maximum WT and PV output capacities to optimize power management within the MG using the proposed MARL method, in the second phase. This approach minimized fuel and CO_2 emissions costs, enhanced coordination among MG components, and ensured efficient power distribution, resource utilization, and BESS scheduling.

Traditional centralized control methods often proved computationally intensive and less responsive to real-time changes, whereas the decentralized control offered by MARL reduced computational burden and improved response quality, demonstrating its effectiveness in maintaining optimal MG performance. Comparative analyses and optimally tunning the number of agents and DF value confirmed the effectiveness of the proposed approach.

Hence, this thesis provided a comprehensive and innovative framework for enhancing wind energy conversion, renewable energy forecasting, and microgrid power management using a unified MARL-based approach. Python programming and MATLAB Simulink were utilized for coding and simulation, respectively. The findings highlighted the potential of MARL applications to improve the efficiency, reliability, and environmental sustainability of renewable energy systems and WECS.

References

[1] F. Blaabjerg, R. Teodorescu, M. Liserre, and A. V. Timbus, “Overview of control and grid synchronization for distributed power generation systems,” IEEE Transactions on Industrial Electronics, vol. 53, no. 5, pp. 1398–1409, 2006

[2] B. Li, R. Yang, D. Xu, G. Wang, W. Wang, and D. Xu, “Analysis of the phase-shifted carrier modulation for modular multilevel converters,” IEEE Transactions on Industrial Electronics, vol. 30, no. 1, pp. 297–310, 2015.

[3] GWEC's 16th flagship Global Wind Report 2021.

[4] M. Vimala, “Harmonic Analysis of Multilevel Inverter driven by Variable Speed Wind Electric Generator,” Proceedings of the 2013 International Conference on Green Computing, Communication and Conservation of Energy, ICGCE, no. 1, pp. 409–414, 2013.

[5] J. a Baroudi, V. Dinavahi, and a M. Knight, “A Review of Power Converter Topologies for Wind Generators,” IEEE International Conference on Electric Machines and Drives, pp. 458–465, 2005.

[6] M. M. Hussein, T. Senjyu, M. Orabi, M. A. Wahab, and M. M. Hamada, “Control of a Variable Speed Stand Alone Wind Energy Supply System,” IEEE International Conference on Power and Energy, Kota Kinabalu Sabah, Malaysia December, no pp. 2–5, 2012.

[7] Z. Chen, J. M. Guerrero, F. Blaabjerg, and S. Member, “A Review of the State of the Art of Power Electronics for Wind Turbines,” IEEE Transaction on Power Electronics, vol. 24, no. 8, pp. 1859–1875, 2009.

[8] F. Blaabjerg, Z. Chen, and S. B. Kjaer, “Power electronics as efficient interface in dispersed power generation systems,” IEEE Transaction on Power Electronics, vol. 19, no. 5, pp. 1184–1194, 2004.

[9] K. Ma, F. Blaabjerg, and D. Xu, “Power devices loading in multilevel converters for 10 MW wind turbines,” IEEE International Symposium on Industrial Electronics, pp. 340–346, 2011.

[10] Sixing Du, Apparao Dekka, Bin Wu and Navid Zargari, “Modular Multilevel Converter: Analysis Control, and Application,” IEEE, 2018, 2018, pp. 1-36, doi: 10.1002/9781119367291.

[11] S. Kouro et al., “Recent Advances and Industrial Applications of Multilevel Converters,” IEEE Transactions on Industrial Electronics, vol. 57, no. 8, pp. 2553–2580, 2010.

[12] A. Uehara, S. Member, A. Pratap, S. Member, T. Goya, and S. Member, “A Coordinated Control Method to Smooth Wind Power Fluctuations of a PMSG-Based WECS,” IEEE Transactions on Energy Conversion, vol. 26, no. 2, pp. 550–558, 2011.

[13] K. S. M. Raza, “Maximum power point tracking control and voltage regulation of a DC grid-tied wind energy conversion system based on a novel permanent magnet reluctance,” International Conference on Electrical Machines and Systems (ICEMS), pp. 1533–1538, 2007.

[14] C. Xia, X. Gu, T. Shi, and Y. Yan, “Neutral-point potential balancing of three-level inverters in direct-driven wind energy conversion system,” IEEE Transactions on Energy Conversion, vol. 26, no. 1, pp. 18–29, 2011.

[15] P. Tenca, A. A. Rockhill, T. A. Lipo, and P. Tricoli, “Current source topology for wind turbines with decreased mains current harmonics, further reducible via functional minimization,” IEEE Transactions on Power Electronics, vol. 23, no. 3, pp. 1143–1155, 2008.

[16] M. A. Perez, S. Bernet, J. Rodriguez, S. Kouro, and R. Lizana, “Circuit topologies, modeling, control schemes, and applications of modular multilevel converters,” IEEE Transactions on Power Electronics, vol. 30, no. 1, pp. 4–17, 2015.

[17] S. F. F. Deng, Z. Chen, D. Karwatzki, A. Mertens, M. Parker, “Converter designs based on new components and modular multilevel topologies,” Innovative Wind Conversion Systems (INNWIND.EU), Europe, Tech. Report, November, 2017.

[18] N. A. Rahim and M. F. M. Elias, W. P. Hew, “Transistor-clamped H-bridge based cascaded multilevel inverter with new method of capacitor voltage balancing”, IEEE Transaction on Industrial Electronics, vol. 60, no. 8, pp. 2943–2956, 2013.

[19] Y. S. Lai and F. S. Shyu, "Topology for hybrid multilevel inverter," in IEEE Proceedings - Electric Power Applications, vol. 149, no. 6, pp. 449-458, 2002.

[20] A.B. Ponniran, K. Orikaw, J. Itoh, “Minimum flying capacitor for N-Level capacitor DC/DC boost converter”, IEEE Trans Industrial Applications, vol. 52, no. 4, pp. 3255–3266, 2016.

[21] B. P. McGrath and D. G. Holmes, “Analytical modeling of voltage balance dynamics for a flying capacitor multilevel converter,” IEEE Transaction on Power Electronics., vol. 23, no. 2, pp. 543–550, 2008.

[22] M. Wang, Y. Hu, W. Zhao, Y. Wang, G. Chen, “Application of modular multilevel converter in medium voltage high power permanent magnet synchronous generator wind energy conversion systems”, IET Renewable Power Generation, vol.10, no. 6, pp. 824–833, 2016.

[23] V. Najmi, W. Jun, R. Burgos, D. Boroyevich, “Reliability-oriented switching frequency analysis for modular multilevel converter (MMC)”, In: Energy conversion congress and exposition (ECCE) IEEE; 2015.
[24] J. Rodriguez, S. Bernet, P. K. Steimer and I. E. Lizama, "A Survey on Neutral-Pointclamped Inverters," in IEEE Transactions on Industrial Electronics, vol. 57, no. 7, pp. 2219- 2230, 2010.
[25] C. Alejandro, S. Alepuz, J. Bordonau, P. Cortes and J. Rodriguez, “Predictive control of a back-to-back NPC converter-based wind power system”, IEEE Transaction on Industrial Electronics, vol. 63, no. 7, pp. 4615–4627, 2016.
[26] A. Méndez, R. Cardoner, A. Barrado, and A. Lázaro, "Analysis of Switching Frequency Harmonic Impact in Modular Multilevel and Cascaded H-Bridge Converters," IEEE Transactions on Power Electronics, vol. 35, no. 11, pp. 11676-11687, Nov. 2020.
[27] Y. Liu, X. Duan, P. Davari, and F. Blaabjerg, "An Overview of Multi-Carrier PWM Methods for Two-Level Three-Phase Converters," IEEE Transactions on Power Electronics, vol. 36, no. 7, pp. 8201-8216, July 2021.
[28] B. Bose, Power Electronics and Motor Drives: Recent Advances and Trends. Academic Press, 2006.
[29] Z. Wang, X. Wang, J. Cao, M. Cheng and Y. Hu, "Direct Torque Control of T-NPC Inverters-Fed Double-Stator-Winding PMSM Drives With SVM," in IEEE Transactions on Power Electronics, vol. 33, no. 2, pp. 1541-1553, Feb. 2018.
[30] A. Mora, R. Cardenas, R. P. Aguilera, A. Angulo, P. Lezana and D. D. -C. Lu, "Predictive Optimal Switching Sequence Direct Power Control for Grid-Tied 3L-NPC Converters," in IEEE Transactions on Industrial Electronics, vol. 68, no. 9, pp. 8561-8571, Sept. 2021.
[31] F. Fuchs, J. Dannehl, “Discrete sliding mode current control of grid-connected three-phase PWM converters with LCL filter”, in IEEE Int. Symp. On Ind. Electron. (ISIE), pp. 779–785, 2010.
[32] X. Shen et al., "High-Performance Second-Order Sliding Mode Control for NPC Converters," in IEEE Transactions on Industrial Informatics, vol. 16, no. 8, pp. 5345-5356, Aug. 2020.
[34] M. Habibullah, “Simplified finite-state predictive torque control strategies for induction motor drives”, PhD Dissertation, The University of Sydney, Australia, 2016.
[35] X. Wei et al., "Robust Multilayer Model Predictive Control for a Cascaded Full-Bridge NPC Class-D Amplifier with Low Complexity," in IEEE Transactions on Industrial Electronics, vol. 68, no. 4, pp. 3390-3401, April 2021.
[36] Y. Zhang, J. Jiao, D. Xu, D. Jiang, Z. Wang and C. Tong, "Model Predictive Direct Power Control of Doubly Fed Induction Generators Under Balanced and Unbalanced Network Conditions," in IEEE Transactions on Industry Applications, vol. 56, no. 1, pp. 771-786, Feb. 2020.
[37] Yaramasu and B. Wu, "Predictive Control of a Three-Level Boost Converter and an NPC Inverter for High-Power PMSG-Based Medium Voltage Wind Energy Conversion Systems," in IEEE Transactions on Power Electronics, vol. 29, no. 10, pp. 5308-5322, Oct. 2014.
[38] A. Calle-Prado, S. Alepuz, J. Bordonau, P. Cortes and J. Rodriguez, "Predictive Control of a Back-to-Back NPC Converter-Based Wind Power System," in IEEE Transactions on Industrial Electronics, vol. 63, no. 7, pp. 4615-4627, July 2016.
[39] Z. Zhang, C. M. Hackl and R. Kennel, "Computationally Efficient DMPC for Three-Level NPC Back-to-Back Converters in Wind Turbine Systems With PMSG," in IEEE Transactions on Power Electronics, vol. 32, no. 10, pp. 8018-8034, Oct. 2017.
[40] Y. Guo, H. Gao, Q. Wu, H. Zhao, J. Østergaard and M. Shahidehpour, "Enhanced Voltage Control of VSC-HVDC-Connected Offshore Wind Farms Based on Model Predictive Control," in IEEE Transactions on Sustainable Energy, vol. 9, no. 1, pp. 474-487, Jan. 2018.
[41] Z. Zhang, Z. Li, M. P. Kazmierkowski, J. Rodríguez and R. Kennel, "Robust Predictive Control of Three-Level NPC Back-to-Back Power Converter PMSG Wind Turbine Systems with Revised Predictions," in IEEE Transactions on Power Electronics, vol. 33, no. 11, pp. 9588-9598, Nov. 2018.
[42] M. Nasir Uddin, and Yazdan H. Tabrizi, “Machine Learning Based Control Strategy of a Three-Phase Neutral-Point Clamped Back-to-Back Power Converter for WECS with Ensured Power Quality,” 2022 IEEE Industry Applications Society Annual Meeting (IAS), 2022.
[43] M. Babaie, M. Sharifzadeh and K. Al-Haddad, "Three-Phase Grid-Connected NPC Inverter Based on a Robust Artificial Neural Network Controller," 2020 IEEE Power & Energy Society General Meeting (PESGM), pp. 1-5, 2020.
[44] G. G. Koch, C. R. D. Osório, H. Pinheiro, R. C. L. F. Oliveira and V. F. Montagner, "Design Procedure Combining Linear Matrix Inequalities and Genetic Algorithm for Robust Control of Grid-Connected Converters," in IEEE Transactions on Industry Applications, vol. 56, no. 2, pp. 1896-1906, March-April 2020.

[45] H. Lin et al., "Fuzzy Logic System-based Sliding-Mode Control for Three-Level NPC Converters," in IEEE Transactions on Transportation Electrification, doi: 10.1109/TTE.2021.3134279.
[46] H. Lin et al., "Fuzzy Logic System-based Sliding-Mode Control for Three-Level NPC Converters," in IEEE Transactions on Transportation Electrification, doi: 10.1109/TTE.2021.3134279.
[47] S. Zhao, F. Blaabjerg and H. Wang, "An Overview of Artificial Intelligence Applications for Power Electronics," in IEEE Transactions on Power Electronics, vol. 36, no. 4, pp. 4633-4658, April 2021.
[48] S. Wang, T. Dragicevic, G. F. Gontijo, S. K. Chaudhary and R. Teodorescu, "Machine Learning Emulation of Model Predictive Control for Modular Multilevel Converters," in IEEE Transactions on Industrial Electronics, vol. 68, no. 11, pp. 11628-11634, Nov. 2021.
[49] H. Soliman, H. Wang, and F. Blaabjerg, "A review of the condition monitoring of capacitors in power electronic converters," IEEE Trans. Ind. Appl., vol. 52, no. 6, pp. 4976–4989, Nov./Dec. 2016.
[50] X. X. Zheng and P. Peng, "Fault diagnosis of wind power converters based on compressed sensing theory and weight constrained AdaBoost-SVM," J. Power Electron., vol. 19, no. 2, pp. 443–453, Mar. 2019.
[51] X. Fu and S. Li, "Control of Single-Phase Grid-Connected Converters with LCL Filters Using Recurrent Neural Network and Conventional Control Methods," in IEEE Transactions on Power Electronics, vol. 31, no. 7, pp. 5354-5364, July 2016.
[51] J. Monteiro, J. F. Silva, S. F. Pinto and J. Palma, "Linear and Sliding-Mode Control Design for Matrix Converter-Based Unified Power Flow Controllers," in IEEE Transactions on Power Electronics, vol. 29, no. 7, pp. 3357-3367, July 2014.
[52] Y. Wang, X. Ruan, Y. Leng and Y. Li, "Hysteresis Current Control for Multilevel Converter in Parallel-Form Switch-Linear Hybrid Envelope Tracking Power Supply," in IEEE Transactions on Power Electronics, vol. 34, no. 2, pp. 1950-1959, Feb. 2019.
[53] M. Narimani, Bin Wu, V. Yaramasu, Zhongyuan Cheng and N. R. Zargari, "Finite Control-Set Model Predictive Control (FCS-MPC) of Nested Neutral Point-Clamped (NNPC) Converter," in IEEE Transactions on Power Electronics, vol. 30, no. 12, pp. 7262-7269, Dec. 2015.
[54] M. Hajihosseini, M. Andalibi, M. Gheisarnejad, H. Farsizadeh and M. Khooban, "DC/DC Power Converter Control-Based Deep Machine Learning Techniques: Real-Time Implementation," in IEEE Transactions on Power Electronics, vol. 35, no. 10, pp. 9971-9977, Oct. 2020.
[55] Mahdavinejad M.S. et al., "Machine learning for Internet of Things data analysis: A survey", Digital Communications and Networks, 2017.
[56] W. Cai. M. Zhang and Y. Zhang, "Batch Mode Active Learning for Regression with Expected Model Change," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 7, pp. 1668-1681, July 2017.
[57] S. Nan, L. Sun, B. Chen, Z. Lin and K. Toh, "Density-Dependent Quantized Least Squares Support Vector Machine for Large Data Sets," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 1, pp. 94-106, Jan. 2017.
[58] X. Zhao, F. Nie, R. Wang and X. Li, "Robust Fuzzy K-Means Clustering with Shrunk Patterns Learning," IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2021.3116257.
[59] B. Wang, L. Gao and Z. Juan, "Travel Mode Detection Using GPS Data and Socioeconomic Attributes Based on a Random Forest Classifier," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 5, pp. 1547-1558, May 2018.
[60] R. Low, L. Cheah and L. You, "Commercial Vehicle Activity Prediction with Imbalanced Class Distribution Using a Hybrid Sampling and Gradient Boosting Approach," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 3, pp. 1401-1410, March 2021.
[61] L. Grippo, A. Manno and M. Sciandrone, "Decomposition Techniques for Multilayer Perceptron Training," in IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 11, pp. 2146-2159, Nov. 2016.
[62] G. I. Webb, "Encyclopedia of Machine Learning", Springer, Boston, MA, 2011.
[63] P. Andras, "Function Approximation Using Combined Unsupervised and Supervised Learning," in IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 3, pp. 495-505, March 2014.
[64] M. Al-Gabalawy, H. S. Ramadan, M. A. Mostafa and S. A. Hussien, "Power Curve Estimation of a Wind Turbine Considering Different Weather Conditions using Machine Learning Algorithms," 23rd International Middle East Power Systems Conference (MEPCON), Cairo, Egypt, pp. 1-8, 2022.
[65] D. Pinna, R. Toso, R. Coutinho, A. Isabel Pereira and D. Brandão, "Fault Identification in Wind Turbines: A Data-Centric Machine Learning Approach," 2022 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, pp. 565-568, 2022.
[66] W. Kim, A. Kanezaki and M. Tanaka, "Unsupervised Learning of Image Segmentation Based on Differentiable Feature Clustering," in IEEE Transactions on Image Processing, vol. 29, pp. 8055-8068, 2020.

[67] D. Chen, R. Lu, X. Wang and Y. Wang, "Power System Operation Mode Identification Method Based on Improved Clustering Algorithm," 2022 IEEE Sustainable Power and Energy Conference, Australia, pp. 1-5, 2022.
[68] M. Wang, W. Fu, S. Hao, D. Tao and X. Wu, "Scalable Semi-Supervised Learning by Efficient Anchor Graph Regularization," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 7, pp. 1864-1877, 1 July 2016.
[69] F. P. G. de Sá, D. N. Brandão, E. Ogasawara, R. d. C. Coutinho and R. F. Toso, "Wind Turbine Fault Detection: A Semi-Supervised Learning Approach with Automatic Evolutionary Feature Selection," 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Brazil, pp. 323-328, 2020.
[70] C. H. Yu, F. Gao and Q. -Y. Wen, "An Improved Quantum Algorithm for Ridge Regression," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 3, pp. 858-866, 1 March 2021.
[71] Z. Zhang, F. Li, M. Zhao, L. Zhang and S. Yan, "Robust Neighborhood Preserving Projection by Nuclear/L2,1-Norm Regularization for Image Feature Extraction," in IEEE Transactions on Image Processing, vol. 26, no. 4, pp. 1607-1622, April 2017.
[72] X. Luo et al., "Short-Term Wind Speed Forecasting via Stacked Extreme Learning Machine with Generalized Correntropy," in IEEE Transactions on Industrial Informatics, vol. 14, no. 11, pp. 4963-4971, Nov. 2018.
[73] M. Mohammadi Amiri and D. Gündüz, "Machine Learning at the Wireless Edge: Distributed Stochastic Gradient Descent Over-the-Air," in IEEE Transactions on Signal Processing, vol. 68, pp. 2155-2169, 2020.
[74] X. Peng, L. Li and F. -Y. Wang, "Accelerating Minibatch Stochastic Gradient Descent Using Typicality Sampling," in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 4649-4659, Nov. 2020.
[75] C. T. Kelley, "Iterative Methods for Optimization", Society for Industrial and Applied Mathematics publisher, 1999.
[76] L. N. Egidio, A. Hansson and B. Wahlberg, "Learning the Step-size Policy for the Limited-Memory Broyden Fletcher Goldfarb Shanno Algo," 2021 International Joint Conference on Neural Networks, pp. 1-8, 2021.
[77] Z. Zhang, D. Wang, and J. Gao, “Learning automata-based multiagent reinforcement learning for optimization of cooperative tasks,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 10, pp. 4639–4652, Oct. 2021.
[78] W. Du, S. Ding, C. Zhang and Z. Shi, "Multiagent Reinforcement Learning with Heterogeneous Graph Attention Network," in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 10, pp. 6851-6860, Oct. 2023.
[79] M. N. Uddin and Y. H. Tabrizi, "A Model-Free Multi-Agent Reinforcement Learning Approach to Reach a Robust, Optimal, and Environment-Friendly Power Management in a Micro-Grid," 2023 IEEE Industry Applications Society Annual Meeting (IAS), Nashville, TN, USA, pp. 1-9, 2023.
[80] Y. Zhou, L. Zhao and W. -J. Lee, "Robustness Analysis of Dynamic Equivalent Model of DFIG Wind Farm for Stability Study," in IEEE Transactions on Industry Applications, vol. 54, no. 6, pp. 5682-5690, Nov.-Dec. 2018.
[81] M. Jahanpour-Dehkordi, S. Vaez-Zadeh and J. Mohammadi, "Development of a Combined Control System to Improve the Performance of a PMSG-Based Wind Energy Conversion System Under Normal and Grid Fault Conditions," in IEEE Transactions on Energy Conversion, vol. 34, no. 3, pp. 1287-1295, Sept. 2019.
[82] J. Chen, J. Chen and C. Gong, "On Optimizing the Aerodynamic Load Acting on the Turbine Shaft of PMSG-Based Direct-Drive Wind Energy Conversion System," in IEEE Transactions on Industrial Electronics, vol. 61, no. 8, pp. 4022-4031, Aug. 2014.
[83] R. Wang, L. Ai and C. Liu, "A Novel Three-Phase Dual-Output Neutral-Point-Clamped Three-Level Inverter," in IEEE Transactions on Power Electronics, vol. 36, no. 7, pp. 7576-7586, July 2021.
[84] R. Selvaraj, K. Desingu, T. R. Chelliah, D. Khare and C. Bharatiraja, "Fault Tolerant Operation of Parallel-Connected 3L-Neutral-Point Clamped Back-to-Back Converters Serving to Large Hydro-Generating Units," in IEEE Transactions on Industry Applications, vol. 54, no. 5, pp. 5429-5443, Sept.-Oct. 2018.
[85] Y. Yin et al., "Observer-Based Adaptive Sliding Mode Control of NPC Converters: An RBF Neural Network Approach," in IEEE Transactions on Power Electronics, vol. 34, no. 4, pp. 3831-3841, April 2019.
[86] A. Bahrami, M. Norambuena, M. Narimani and J. Rodriguez, "Model Predictive Current Control of a Seven-Level Inverter with Reduced Computational Burden," in IEEE Transactions on Power Electronics, vol. 35, no. 6, pp. 5729-5740, June 2020.
[87] P. Qashqai, M. Babaie, R. Zgheib and K. Al-Haddad, "A Model-Free Switching and Control Method for Three-Level Neutral Point Clamped Converter Using Deep Reinforcement Learning," in *IEEE Access*, vol. 11, pp. 105394-105409, 2023, doi: 10.1109/ACCESS.2023.3318264.

[88] M. A. Mahmud, T. K. Roy, S. Saha, M. E. Haque, and H. R. Pota, ‘‘Robust nonlinear adaptive feedback linearizing decentralized controller design for islanded DC microgrids,’’ IEEE Trans. Ind. Appl., vol. 55, no. 5, pp. 5343–5352, Sep. 2019.
[89] Y. Zhang, J. Jiao, D. Xu, D. Jiang, Z. Wang and C. Tong, "Model Predictive Direct Power Control of Doubly Fed Induction Generators Under Balanced and Unbalanced Network Conditions," in IEEE Transactions on Industry Applications, vol. 56, no. 1, pp. 771-786, Feb. 2020.
[90] Yaramasu and B. Wu, "Predictive Control of a Three-Level Boost Converter and an NPC Inverter for High-Power PMSG-Based Medium Voltage Wind Energy Conversion Systems," in IEEE Transactions on Power Electronics, vol. 29, no. 10, pp. 5308-5322, Oct. 2014.
[91] Z. Zhang, C. M. Hackl and R. Kennel, "Computationally Efficient DMPC for Three-Level NPC Back-to-Back Converters in Wind Turbine Systems With PMSG," in IEEE Transactions on Power Electronics, vol. 32, no. 10, pp. 8018-8034, Oct. 2017.
[92] A. Laka, J. A. Barrena, J. Chivite-Zabalza, M. Á. Rodríguez Vidal and P. Izurza-Moreno, "New Hexagonal Three-Phase Voltage-Source Converter Topology for High-Power Applications," in IEEE Transactions on Industrial Electronics, vol. 62, no. 1, pp. 30-39, Jan. 2015.
[93] H. Obeid, L. M. Fridman, S. Laghrouche, and M. Harmouche, “Barrier function-based adaptive sliding mode control,” Automatica, vol. 93, pp. 540–544, 2018.
[94] V. Utkin, “Discussion aspects of high-order sliding mode control,” IEEE Trans. Autom. Control, vol. 61, no. 3, pp. 829–833, Mar. 2016.
[95] C. Lascu, A. Argeseanu, and F. Blaabjerg, “Supertwisting sliding-mode direct torque and flux control of induction machine drives,” IEEE Trans. Power Electron., vol. 35, no. 5, pp. 5057–5065, May 2020.
[96] H. Lin et al., "Fuzzy Logic System-based Sliding-Mode Control for Three-Level NPC Converters," in IEEE Transactions on Transportation Electrification, doi: 10.1109/TTE.2021.3134279.
[97] Z. Zhang, Wei Tian, Wanyi Xiong and R. Kennel, "Predictive torque control of induction machines fed by 3L-NPC converters with online weighting factor adjustment using Fuzzy Logic," 2017 IEEE Transportation Electrification Conference and Expo (ITEC), USA, pp. 84-89, 2017.
[98] H. Soliman, H. Wang, and F. Blaabjerg, “A review of the condition monitoring of capacitors in power electronic converters,” IEEE Trans. Ind. Appl., vol. 52, no. 6, pp. 4976–4989, Nov./Dec. 2016.
[99] X. Fu and S. Li, "Control of Single-Phase Grid-Connected Converters with LCL Filters Using Recurrent Neural Network and Conventional Control Methods," in IEEE Transactions on Power Electronics, vol. 31, no. 7, pp. 5354-5364, July 2016.
[100] S. Wang, T. Dragicevic, G. F. Gontijo, S. K. Chaudhary and R. Teodorescu, "Machine Learning Emulation of Model Predictive Control for Modular Multilevel Converters," in IEEE Transactions on Industrial Electronics, vol. 68, no. 11, pp. 11628-11634, Nov. 2021.
[101] M. N. Uddin and Y. H. Tabrizi, "Artificial Intelligence Based Control Strategy of a Three-Phase Neutral-Point Clamped Back-to-Back Power Converter with Ensured Power Quality for WECS," 2022 IEEE Industry Applications Society Annual Meeting (IAS), Detroit, MI, USA, pp. 1-8, 2022.
[102] M. Gheisarnejad, H. Farsizadeh and M. H. Khooban, "A Novel Nonlinear Deep Reinforcement Learning Controller for DC–DC Power Buck Converters," in IEEE Transactions on Industrial Electronics, vol. 68, no. 8, pp. 6849-6858, Aug. 2021.
[103] Y. Tang et al., "Reinforcement Learning Based Efficiency Optimization Scheme for the DAB DC–DC Converter with Triple-Phase-Shift Modulation," in IEEE Transactions on Industrial Electronics, vol. 68, no. 8, pp. 7350-7361, Aug. 2021.
[104] C. Cui, N. Yan, B. Huangfu, T. Yang and C. Zhang, "Voltage Regulation of DC-DC Buck Converters Feeding CPLs via Deep Reinforcement Learning," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 69, no. 3, pp. 1777-1781, March 2022.
[105] E. ON Netz GmbH, Grid Code: High and Extra High Voltage, 2006. Available: www.eon-netz.com
[106] H. H. Mousa, A.-R. Youssef, and E. E. Mohamed, “State of the art perturb and observe MPPT algorithms-based wind energy conversion systems: A technology review,” International Journal of Electrical Power & Energy Systems, vol. 126, p. 106598, 2021.
[107] H. T. Do, T. D. Dang, H. V. A. Truong and K. K. Ahn, "Maximum Power Point Tracking and Output Power Control on Pressure Coupling Wind Energy Conversion System," in IEEE Transactions on Industrial Electronics, vol. 65, no. 2, pp. 1316-1324, 2018.
[108] Z. Cui, L. Song and S. Li, "Maximum Power Point Tracking Strategy for a New Wind Power System and Its Design Details," in IEEE Transactions on Energy Conversion, vol. 32, no. 3, pp. 1063-1071, 2017.

[109] M. A. Abdullah, T. Al-Hadhrami, C. W. Tan and A. H. Yatim, "Towards Green Energy for Smart Cities: Particle Swarm Optimization Based MPPT Approach," in IE EE Access, vol. 6, pp. 58427-58438, 2018
[110] R. Alhamdawee and M. M. Hussain, "A study of conventional and modern algorithms employed for MPPT in wind energy conversion systems: A review," 2024 3rd International conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC), Mathura, India, pp. 14-23, 2024.
[111] B. Kumar, N. Agrawal, S. K. Singh and A. Agarwal, "A Comparative Analysis of Wind Energy Conversion Systems Based on PMSG for Maximum Power Extraction," 2022 IEEE 10th Power India International Conference (PIICON), New Delhi, India, pp. 1-6, 2022.
[112] A. J. Balbino, B. d. S. Nora and T. B. Lazzarin, "An Improved Mechanical Sensorless Maximum Power Point Tracking Method for Permanent-Magnet Synchronous Generator-Based Small Wind Turbines Systems," in IEEE Transactions on Industrial Electronics, vol. 69, no. 5, pp. 4765-4775, May 2022.
[113] C. Wei, Z. Zhang, W. Qiao and L. Qu, "An Adaptive Network-Based Reinforcement Learning Method for MPPT Control of PMSG Wind Energy Conversion Systems," in IEEE Transactions on Power Electronics, vol. 31, no. 11, pp. 7837-7848, 2016.
[114] J. Mishra, M. Pattnaik and S. Samanta, "Drift-Free Perturb and Observe MPPT Algorithm with Improved Performance for SEIG-Based Stand-Alone Wind Energy Generation System," in IEEE Transactions on Power Electronics, vol. 35, no. 6, pp. 5842-5849, June 2020.
[115] N. Priyadarshi, M. S. Bhaskar and D. Almakhles, "A Novel Hybrid Whale Optimization Algorithm Differential Evolution Algorithm-Based Maximum Power Point Tracking Employed Wind Energy Conversion Systems for Water Pumping Applications: Practical Realization," in IEEE Transactions on Industrial Electronics, vol. 71, no. 2, pp. 1641-1652, 2024.
[116] D. Yang et al., "Adaptive Frequency Droop Feedback Control-Based Power Tracking Operation of a DFIG for Temporary Frequency Regulation," in IEEE Transactions on Power Systems, vol. 39, no. 2, pp. 2682-2692, 2024.
[117] M. M. Mahmoud, M. M. Aly, H. S. Salama, and A.-M.-M. Abdel-Rahim, ‘‘A combination of an OTC based MPPT and fuzzy logic current control for a wind-driven PMSG under variability of wind speed,’’ Energy Syst., Aug. 2021.
[118] A. A. Salem, N. A. N. Aldin, A. M. Azmy, and W. S. E. Abdellatif, ‘‘Implementation and validation of an adaptive fuzzy logic controller for MPPT of PMSG-based wind turbines,’’ IEEE Access, vol. 9, pp. 165690–165707, 2021.
[119] S. Marmouh, M. Boutoubat and L. Mokrani, "MPPT fuzzy logic controller of a wind energy conversion system based on a PMSG," 2016 8th International Conference on Modelling, Identification and Control (ICMIC), Algiers, Algeria, pp. 296-302, 2016.
[120] N. T. -T. Vu, H. D. Nguyen and A. T. Nguyen, "Reinforcement Learning-Based Adaptive Optimal Fuzzy MPPT Control for Variable Speed Wind Turbine," in IEEE Access, vol. 10, pp. 95771-95780, 2022.
[121] M. C. Di Piazza and M. Pucci, "Induction-Machines-Based Wind Generators With Neural Maximum Power Point Tracking and Minimum Losses Techniques," in IEEE Transactions on Industrial Electronics, vol. 63, no. 2 , pp. 944-955, Feb. 2016.
[122] T. K. Mahmoud, Z. Y. Dong and J. Ma, "A Developed Integrated Scheme Based Approach for Wind Turbine Intelligent Control," in IEEE Transactions on Sustainable Energy, vol. 8, no. 3, pp. 927-937, 2017.
[123] H. Chojaa et al., "Robust Control of DFIG-Based WECS Integrating an Energy Storage System with Intelligent MPPT Under a Real Wind Profile," in IEEE Access, vol. 11, pp. 90065-90083, 2023.
[124] M. Alzayed, H. Chaoui, E. Elhaji and C. Zhang, "Universal Maximum Power Extraction Controller for Wind Energy Conversion Systems Using Deep Belief Neural Network," in IEEE Transactions on Sustainable Energy, vol. 14, no. 1, pp. 630-641. 2023.
[125] G. Mayilsamy, S. R. Lee, Y. H. Joo, “An improved model predictive control of back-to-back three-level NPC converters with virtual space vectors for high power PMSG-based wind energy conversion systems”, ISA Transactions, vol. 143, pp. 503-524, 2023.
[126] Y. Arias-Esquivel, R. Cárdenas, M. Diaz and L. Tarisciotti, "Continuous Control Set Model Predictive Control of a Hybrid Modular Multilevel Converter for Wind Energy Applications," in IEEE Transactions on Industrial Electronics, 2024, doi: 10.1109/TIE.2024.3370982.
[127] C. Wei, Z. Zhang, W. Qiao and L. Qu, "Reinforcement-Learning-Based Intelligent Maximum Power Point Tracking Control for Wind Energy Conversion Systems," in IEEE Transactions on Industrial Electronics, vol. 62, no. 10, pp. 6360-6370, 2015.

[128] A. Kushwaha, M. Gopal and B. Singh, "Q-Learning based Maximum Power Extraction for Wind Energy Conversion System with Variable Wind Speed," in IEEE Transactions on Energy Conversion, vol. 35, no. 3, pp. 1160-1170, Sept. 2020.
[129] R. Venkateswaran, B. Natesan, S. R. Lee and Y. H. Joo, "Maximum Power Extraction for PMVG-Based WECS Using Q-Learning MPPT Algorithm with Finite-Time Control Scheme," in IEEE Transactions on Sustainable Energy, vol. 14, no. 1, pp. 516-524, 2023.
[130] Y. H. Tabrizi and M. N. Uddin, "Multi-Agent Reinforcement Learning-Based Maximum Power Point Tracking Approach to Fortify PMSG-Based WECSs," 2023 IEEE Industry Applications Society Annual Meeting (IAS), Nashville, TN, USA, pp. 1-8, 2023.
[131] GWEC's 16th flagship Global Wind Report 2021.
[132] F. Delfino, G. Ferro, M. Robba and M. Rossi, "An Energy Management Platform for the Optimal Control of Active and Reactive Powers in Sustainable Microgrids," in IEEE Transactions on Industry Applications, vol. 55, no. 6, pp. 7146-7156, 2019.
[133] F. Yang, X. Feng and Z. Li, "Advanced Microgrid Energy Management System for Future Sustainable and Resilient Power Grid," IEEE Trans. on Industry Applications, vol. 55, no. 6, pp. 7251-7260, Nov.-Dec. 2019.
[134] W. Shi, X. Xie, C. -C. Chu and R. Gadh, "Distributed Optimal Energy Management in Microgrids," in IEEE Transactions on Smart Grid, vol. 6, no. 3, pp. 1137-1146, May 2015.
[135] C. Cecati, C. Citro and P. Siano, "Combined Operations of Renewable Energy Systems and Responsive Demand in a Smart Grid," in IEEE Transactions on Sustainable Energy, vol. 2, no. 4, pp. 468-476, Oct. 2011.
[136] S. A. Pourmousavi, M. H. Nehrir, C. M. Colson and C. Wang, "Real-Time Energy Management of a Stand-Alone Hybrid Wind-Microturbine Energy System Using Particle Swarm Optimization," in IEEE Transactions on Sustainable Energy, vol. 1, no. 3, pp. 193-201, Oct. 2010.
[137] P. Siano, C. Cecati, H. Yu and J. Kolbusz, "Real Time Operation of Smart Grids via FCN Networks and Optimal Power Flow," in IEEE Transactions on Industrial Informatics, vol. 8, no. 4, pp. 944-952, Nov. 2012.
[138] G. Mohy-ud-din, K. M. Muttaqi and D. Sutanto, "Adaptive and Predictive Energy Management Strategy for Real-Time Optimal Power Dispatch from VPPs Integrated with Renewable Energy and Energy Storage," in IEEE Transactions on Industry Applications, vol. 57, no. 3, pp. 1958-1972, May-June 2021.
[139] P. Zeng, H. Li, H. He and S. Li, "Dynamic Energy Management of a Microgrid Using Approximate Dynamic Programming and Deep Recurrent Neural Network Learning," in IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 4435-4445, July 2019.
[140] C. Liu, C. Lee, H. Chen and S. Mehrotra, "Stochastic Robust Mathematical Programming Model for Power System Optimization," in IEEE Transactions on Power Systems, vol. 31, no. 1, pp. 821-822, Jan. 2016.
[141] F. Hafiz, M. A. Awal, A. R. d. Queiroz and I. Husain, "Real-Time Stochastic Optimization of Energy Storage Management Using Deep Learning-Based Forecasts for Residential PV Applications," in IEEE Transactions on Industry Applications, vol. 56, no. 3, pp. 2216-2226, May-June 2020.
[142] Y. Du and F. Li, "Intelligent Multi-Microgrid Energy Management Based on Deep Neural Network and Model-Free Reinforcement Learning," in IEEE Trans. on Smart Grid, vol. 11, no. 2, pp. 1066-1076, March 2020.
[143] Y. Zheng, S. Li and R. Tan, "Distributed Model Predictive Control for On-Connected Microgrid Power Management," in IEEE Transactions on Control Systems Technology, vol. 26, no. 3, pp. 1028-1039, May 2018.
[144] F. Delfino, G. Ferro, M. Robba and M. Rossi, "An Energy Management Platform for the Optimal Control of Active and Reactive Powers in Sustainable Microgrids," in IEEE Transactions on Industry Applications, vol. 55, no. 6, pp. 7146-7156, Nov.-Dec. 2019.
[145] Z. Zhao et al., "Distributed Robust Model Predictive Control-Based Energy Management Strategy for Islanded Multi-Microgrids Considering Uncertainty," in IEEE Transactions on Smart Grid, vol. 13, no. 3, pp. 2107-2120, May 2022.
[146] M. Marzband, M. Ghadimi, A. Sumper, and J. L. Domínguez-Garcíac, "Experimental validation of a real-time energy management system using multi-period gravitational search algorithm for microgrids," Appl. Energy, vol. 128, pp. 164–174, 2014.
[147] K. Roy, K. K. Mandal, and A. C. Mandal, "Modeling and managing of micro grid connected system using improved artificial bee colony algorithm," Int. J. Elect. Power Energy Syst., vol. 75, pp. 50–58, 2016.
[148] T. Sadeq, M. Rezkallah, A. Chandra, Z. Feger, H. Ibrahim and J. -F. Savard, "Real-Time Power Management Strategy based on Fuzzy Logic Controller and Human-Computer Interface for DC Off-Grid System," 2020 IEEE Industry Applications Society Annual Meeting, Detroit, MI, USA, pp. 1-6, 2020.
[149] M. G. M. Abdolrasol, R. Mohamed, M. A. Hannan, A. Q. Al-Shetwi, M. Mansor and F. Blaabjerg, "Artificial Neural Network Based Particle Swarm Optimization for Microgrid Optimal Energy Scheduling," in IEEE Transactions on Power Electronics, vol. 36, no. 11, pp. 12151-12157, Nov. 2021.

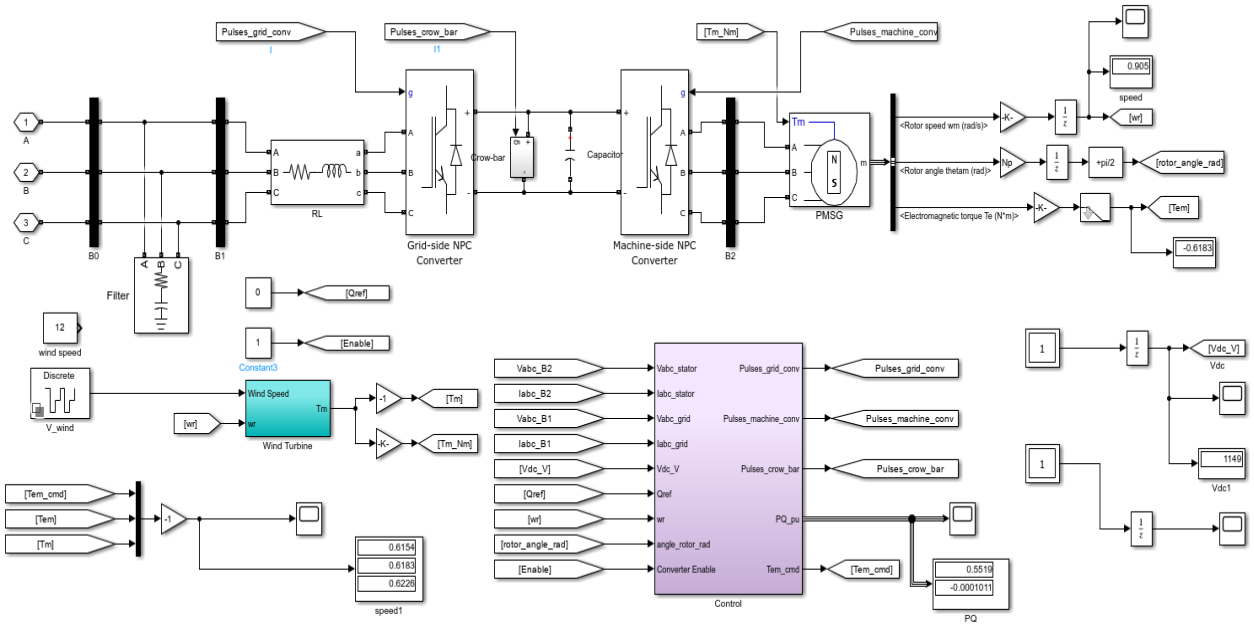
[150] X. Hu, D. Cao, and B. Egardt, "Condition monitoring in advanced battery management systems: moving horizon estimation using a reduced electrochemical model," IEEE/ASME Transactions on Mechatronics, vol. 23, pp. 167-178, 2018.
[151] F. Brahman, M. Honarmand, and S. Jadid, "Optimal electrical and thermal energy management of a residential energy hub, integrating demand response and energy storage system," Energy and Buildings, vol. 90, pp. 65-75, 2015.
[152] M. N. Uddin and Y. H. Tabrizi, "Artificial Intelligence Based Control Strategy of a Three-Phase Neutral-Point Clamped Back-to-Back Power Converter with Ensured Power Quality for WECS," 2022 IEEE Industry Applications Society Annual Meeting (IAS), Detroit, MI, USA, 2022.
[153] N. Ruan, Y. Xu, W. K. Chan and Z. Wei, "Real-time Energy Management of a Low-Carbon Micro-Grid Based on Deep Reinforcement Learning," 2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Shanghai, China, pp. 1102-1107, 2022.
[154] Y. Ye, D. Qiu, X. Wu, G. Strbac and J. Ward, "Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning," IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 3068- 3082, July 2020.
[155] M. N. Uddin and Y. H. Tabrizi, "A Model-Free Multi-Agent Reinforcement Learning Approach to Reach a Robust, Optimal, and Environment-Friendly Power Management in a Micro-Grid," 2023 IEEE Industry Applications Society Annual Meeting (IAS), Nashville, TN, USA, pp. 1-9, 2023.
[156] X. Peng, L. Li and F. -Y. Wang, "Accelerating Minibatch Stochastic Gradient Descent Using Typicality Sampling," in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 4649-4659, Nov. 2020.
[157] Y. H. Tabrizi and M. N. Uddin, "Optimum Cost and Eco-Friendly Power Management in a Micro-Grid, Based on Multi-Agent Reinforcement Learning," 2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG), Wollongong, Australia, pp. 1-6, 2023.
[158] Y. Wei, F. R. Yu, M. Song and Z. Han, "User Scheduling and Resource Allocation in HetNets with Hybrid Energy Supply: An Actor-Critic Reinforcement Learning Approach," in IEEE Transactions on Wireless Communications, vol. 17, no. 1, pp. 680-692, 2018.

Publications

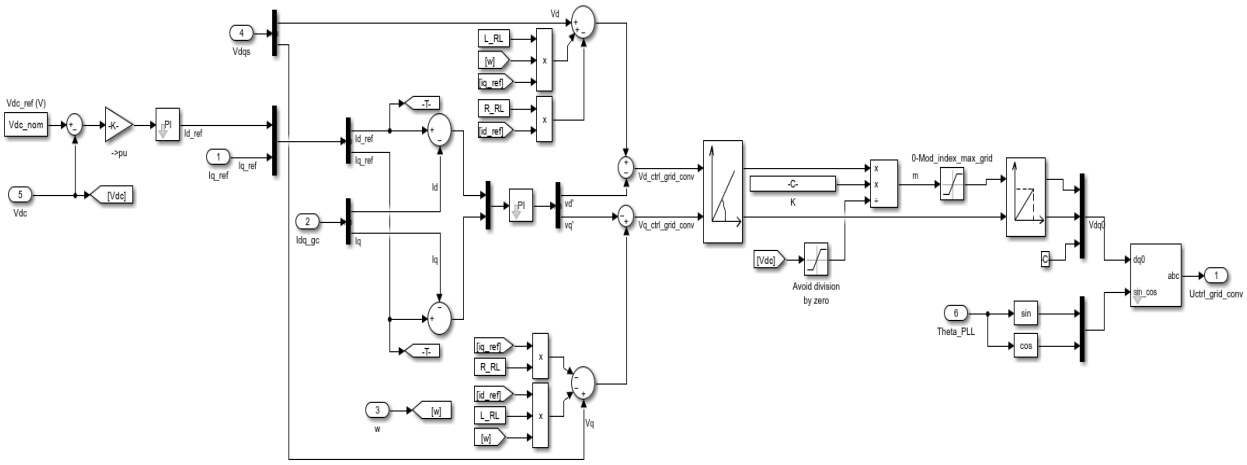
1. Y. H. Tabrizi and M. N. Uddin, "Multi-Agent Reinforcement Learning-Based Maximum Power Point Tracking Approach to Fortify PMSG-Based WECSs", IEEE Transactions on Industry Applications, 2024.
2. Y. H. Tabrizi and M. N. Uddin, A Model-Free Multi-Agent Reinforcement Learning Approach to Reach a Robust, Optimal, and Environment-Friendly Power Management in a Micro-Grid", IEEE Transactions on Industry Applications, 2024 (under review).
3. M. N. Uddin and Y. H. Tabrizi, "Artificial Intelligence Based Control Strategy of a Three-Phase Neutral-Point Clamped Back-to-Back Power Converter with Ensured Power Quality for WECS," *2022 IEEE Industry Applications Society Annual Meeting (IAS)*, Detroit, MI, USA, 2022, pp. 1-8, doi: 10.1109/IAS54023.2022.9939815.
4. Y. H. Tabrizi and M. N. Uddin, "Optimum Cost and Eco-Friendly Power Management in a Micro-Grid, Based on Multi-Agent Reinforcement Learning," *2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG)*, Wollongong, Australia, 2023, pp. 1-6, doi: 10.1109/ETFG55873.2023.10407492.
5. Y. H. Tabrizi and M. N. Uddin, "Multi-Agent Reinforcement Learning-Based Maximum Power Point Tracking Approach to Fortify PMSG-Based WECSs," *2023 IEEE Industry Applications Society Annual Meeting (IAS)*, Nashville, TN, USA, 2023, pp. 1-8, doi: 10.1109/IAS54024.2023.10406511.
6. M. N. Uddin and Y. H. Tabrizi, "A Model-Free Multi-Agent Reinforcement Learning Approach to Reach a Robust, Optimal, and Environment-Friendly Power Management in a Micro-Grid," *2023 IEEE Industry Applications Society Annual Meeting (IAS)*, Nashville, TN, USA, 2023, pp. 1-9, doi: 10.1109/IAS54024.2023.10406423.
7. Y. H. Tabrizi and M. N. Uddin, "A Cost-Effective and Eco-Friendly Micro-Grid by Employing a Model-Free Multi-Agent Reinforcement Learning Strategy for Power Management", *2024 IEEE Industry Applications Society Annual Meeting (IAS),* USA, 2024.
8. Y. H. Tabrizi and M. N. Uddin, "Multi-Agent Reinforcement Learning Control Strategy for a Neutral-Point Clamped Power Converter, Maintaining Reliable Wind Energy Conversion System", 2024 IEEE Energy Conversion Congress and Exposition, USA, 2024.

Appendix

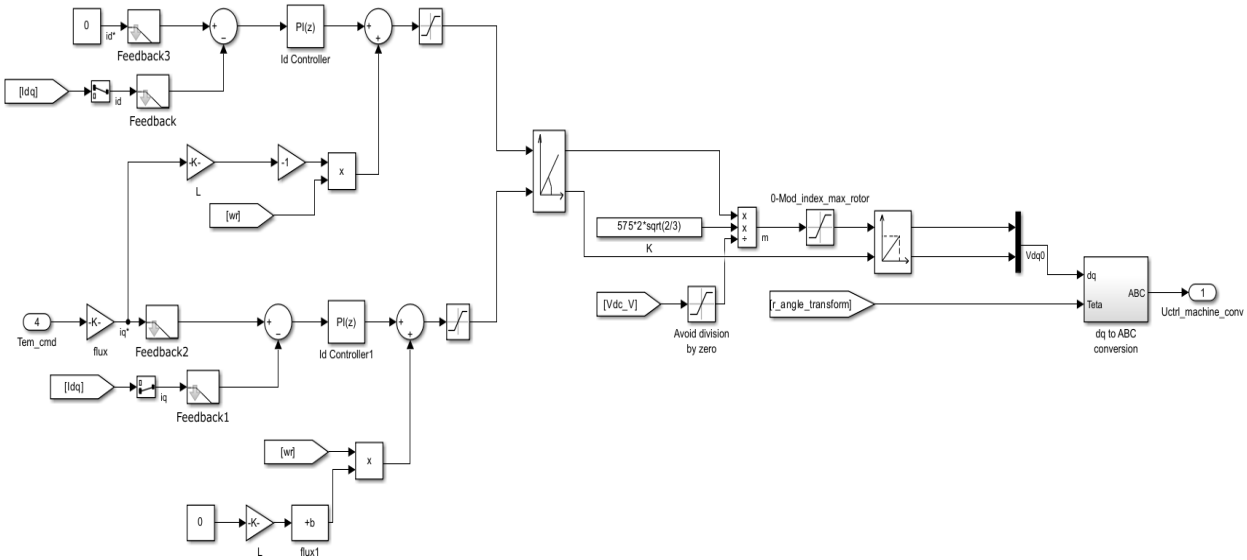
Simulink MATLAB diagram of the direct-drive PMSG-based wind energy conversion system.



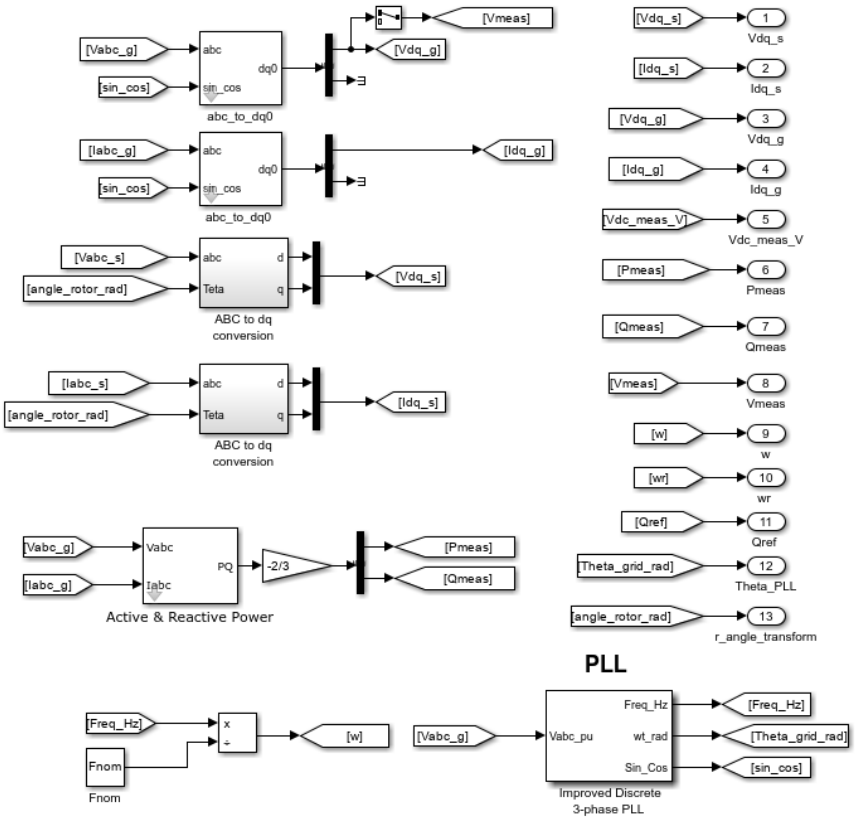
Grid-side power converter controller:



Machine-side power converter controller:



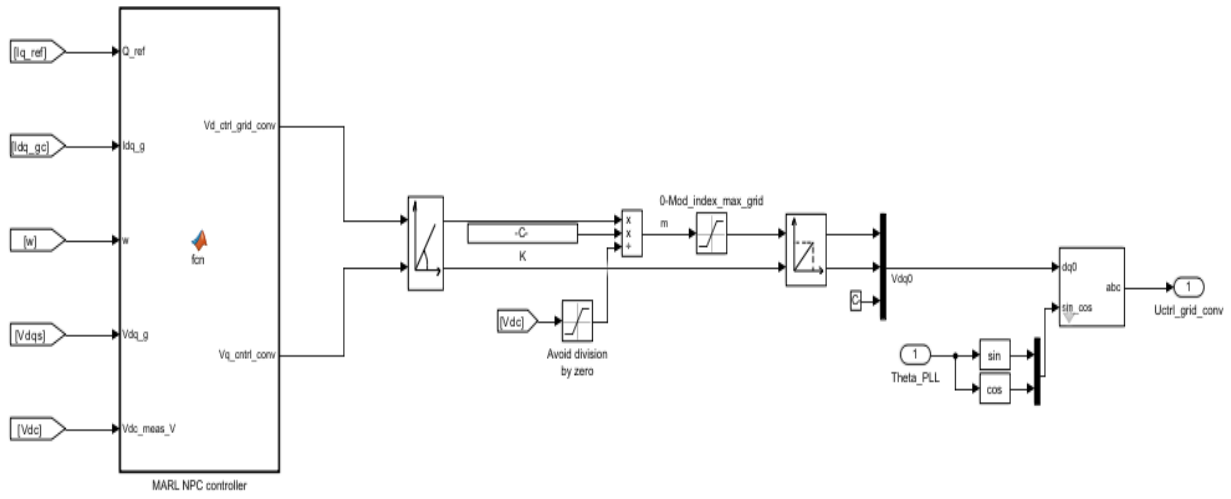
Measurements and transformations



MATLAB function all_marl_thd_optimization.m to interface with the Python script for the proposed grid-side NPC converter control method:

```
function [optimal_discount_factor, Vd_ctrl_grid_conv, Vq_cntrl_conv] = call_marl_thd_optimization(Q_ref, Idq_g, w, Vdq_g, Vdc_meas_V)
    if count(py.sys.path,'') == 0
        insert(py.sys.path,int32(0),'');
    end
    pyModule = py.importlib.import_module('marl_thd_optimization');
    state = [Q_ref, Idq_g, w, Vdq_g, Vdc_meas_V];
    next_state = state;
    reward = 0;
    done = false;
    result = pyModule.marl_thd_optimization(state, next_state, reward, done);
    optimal_discount_factor = double(result{1});
    Vd_ctrl_grid_conv = double(result{2});
    Vq_cntrl_conv = double(result{3});
end
```

Proposed grid-side NPC converter control method diagram based on MARL equipped with meta-learning approach to tune the DF:



The MATLAB function block code:

```
function [Vd_ctrl_grid_conv, Vq_cntrl_conv] = fcn(Q_ref, Idq_g, w, Vdq_g, Vdc_meas_V)
    % Define the state and next_state based on inputs
    state = [Q_ref, Idq_g, w, Vdq_g, Vdc_meas_V];
    next_state = state;
    reward = 0;
    done = false;

    % Call the THD optimization function
    [optimal_discount_factor, Vd_ctrl_grid_conv, Vq_cntrl_conv] = call_marl_thd_optimization(state, next_state, reward, done)
end
```

The python code interfaced with Simulink MATLAB for grid-side NPC power converter control strategy based on proposed MARL approach:

```
import numpy as np
import tensorflow as tf
from tensorflow.keras import models, layers, optimizers
import optuna

class MARLAgent:
    def __init__(self, state_size, action_size, discount_factor):
        self.state_size = state_size
        self.action_size = action_size
        self.discount_factor = discount_factor
        self.model = self._build_model()

    def _build_model(self):
        model = models.Sequential()
        model.add(layers.Dense(24, input_dim=self.state_size, activation='relu'))
        model.add(layers.Dense(24, activation='relu'))
        model.add(layers.Dense(self.action_size, activation='linear'))
        model.compile(optimizer=optimizers.Adam(learning_rate=0.001), loss='mse')
        return model

    def act(self, state):
        q_values = self.model.predict(state)
        return np.argmax(q_values[0])

    def train(self, state, action, reward, next_state, done):
        target = reward
        if not done:
            target += self.discount_factor * np.amax(self.model.predict(next_state)[0])
        target_f = self.model.predict(state)
        target_f[0][action] = target
        self.model.fit(state, target_f, epochs=1, verbose=0)

class MetaLearning:
    def __init__(self):
        self.study = optuna.create_study(direction='maximize')

    def objective(self, trial):
        discount_factor = trial.suggest_float('discount_factor', 0.9, 0.999)
        cumulative_reward = train_marl(discount_factor)
        return cumulative_reward
```

```
    def tune_discount_factor(self):
        self.study.optimize(self.objective, n_trials=100)
        return self.study.best_params['discount_factor']

def train_marl(discount_factor):
    import gym
    env = gym.make('CartPole-v1')
    state_size = env.observation_space.shape[0]
    action_size = env.action_space.n
    agent = MARLAgent(state_size, action_size, discount_factor)
    episodes = 100
    rewards = []

    for e in range(episodes):
        state = env.reset()
        state = np.reshape(state, [1, state_size])
        episode_reward = 0
        for time in range(500):
            action = agent.act(state)
            next_state, reward, done, _ = env.step(action)
            next_state = np.reshape(next_state, [1, state_size])
            agent.train(state, action, reward, next_state, done)
            state = next_state
            episode_reward += reward
            if done:
                break
        rewards.append(episode_reward)

    return np.mean(rewards)
```

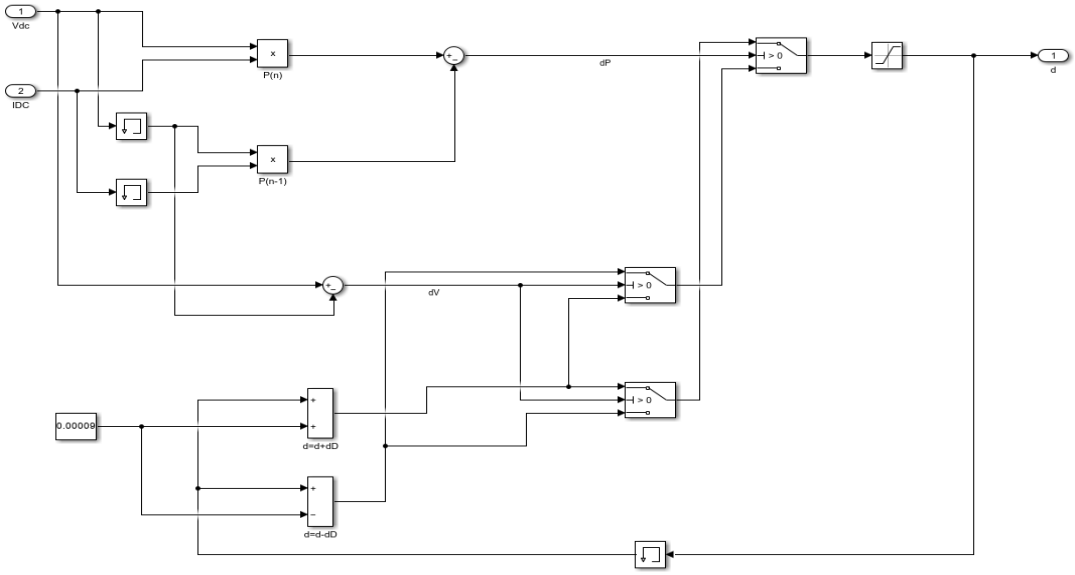
```
def marl_thd_optimization(state, next_state, reward, done):
    state = np.array(state).reshape(1, -1)
    next_state = np.array(next_state).reshape(1, -1)
    state_size = state.shape[1]
    action_size = 2
    meta_learning = MetaLearning()
    optimal_discount_factor = meta_learning.tune_discount_factor()

    agent = MARLAgent(state_size, action_size, optimal_discount_factor)
    action = agent.act(state)
    agent.train(state, action, reward, next_state, done)


    return float(optimal_discount_factor), action[0], action[1]

if __name__ == '__main__':
    state = [0.0, 0.0, 0.0, 0.0]
    next_state = [0.1, 0.1, 0.1, 0.1]
    reward = 1.0
    done = False
    optimal_discount_factor, Vd, Vq = marl_thd_optimization(state, next_state, reward, done)
    print(f'Optimal Discount Factor: {optimal_discount_factor}, Vd: {Vd}, Vq: {Vq}')
```

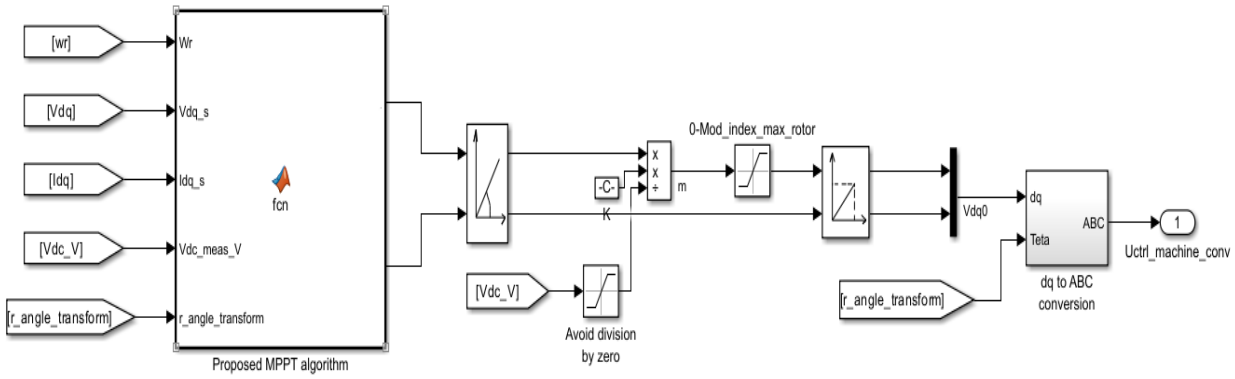
General P&O MPPT algorithm:



MATLAB function call_marl_mppt.m to interface with the Python script for the proposed MPPT method:

```
function [optimal_discount_factor, I_d, I_q] = call_marl_mppt(state, next_state, reward, done)
    if count(py.sys.path,'') == 0
        insert(py.sys.path,int32(0),'');
    end
    pyModule = py.importlib.import_module('marl_mppt');
    result = pyModule.marl_mppt(state, next_state, reward, done);
    optimal_discount_factor = double(result{1});
    I_d = double(result{2});
    I_q = double(result{3});
end
```

Proposed MPPT method based on MARL equipped with meta-learning approach to tune the DF:



The MATLAB function block code:

```
function [I_d, I_q] = fcn(Wr, Vdq_s, Idq_s, Vdc_meas_V, r_angle_transform)
    state = [Wr, Vdq_s, Idq_s, Vdc_meas_V, r_angle_transform];
    next_state = state;
    reward = 0; %
    done = false;
    % Call the MPPT function
    [optimal_discount_factor, I_d, I_q] = call_marl_mppt(state, next_state, reward, done);
end
```

The python code interfaced with Simulink MATLAB:

```python
import numpy as np
import tensorflow as tf
from tensorflow.keras import models, layers, optimizers
import optuna

class MARLAgent:
    def __init__(self, state_size, action_size, discount_factor):
        self.state_size = state_size
        self.action_size = action_size
        self.discount_factor = discount_factor
        self.model = self._build_model()

    def _build_model(self):
        model = models.Sequential()
        model.add(layers.Dense(24, input_dim=self.state_size, activation='relu'))
        model.add(layers.Dense(24, activation='relu'))
        model.add(layers.Dense(self.action_size, activation='linear'))
        model.compile(optimizer=optimizers.Adam(learning_rate=0.001), loss='mse')
        return model

    def act(self, state):
        q_values = self.model.predict(state)
        return np.argmax(q_values[0])

    def train(self, state, action, reward, next_state, done):
        target = reward
        if not done:
            target += self.discount_factor * np.amax(self.model.predict(next_state)[0])
        target_f = self.model.predict(state)
        target_f[0][action] = target
        self.model.fit(state, target_f, epochs=1, verbose=0)

class MetaLearning:
    def __init__(self):
        self.study = optuna.create_study(direction='maximize')

    def objective(self, trial):
        discount_factor = trial.suggest_float('discount_factor', 0.9, 0.999)
        cumulative_reward = train_marl(discount_factor)
        return cumulative_reward
```

```python
    def tune_discount_factor(self):
        self.study.optimize(self.objective, n_trials=100)
        return self.study.best_params['discount_factor']

def train_marl(discount_factor):
    import gym
    env = gym.make('CartPole-v1')  # Replace with your actual environment
    state_size = env.observation_space.shape[0]
    action_size = env.action_space.n
    agent = MARLAgent(state_size, action_size, discount_factor)
    episodes = 100
    rewards = []

    for e in range(episodes):
        state = env.reset()
        state = np.reshape(state, [1, state_size])
        episode_reward = 0
        for time in range(500):
            action = agent.act(state)
            next_state, reward, done, _ = env.step(action)
            next_state = np.reshape(next_state, [1, state_size])
            agent.train(state, action, reward, next_state, done)
            state = next_state
            episode_reward += reward
            if done:
                break
        rewards.append(episode_reward)

    return np.mean(rewards)

def marl_mppt(state, next_state, reward, done):
    state = np.array(state).reshape(1, -1)
    next_state = np.array(next_state).reshape(1, -1)
    state_size = state.shape[1]
    action_size = 3  # Define based on your specific action space
    meta_learning = MetaLearning()
    optimal_discount_factor = meta_learning.tune_discount_factor()
```

```
    agent = MARLAgent(state_size, action_size, optimal_discount_factor)
    action = agent.act(state)
    agent.train(state, action, reward, next_state, done)

    return float(optimal_discount_factor), int(action)

if __name__ == '__main__':
    state = [0.0, 0.0, 0.0, 0.0]
    next_state = [0.1, 0.1, 0.1, 0.1]
    reward = 1.0
    done = False
    optimal_discount_factor, action = marl_mppt(state, next_state, reward, done)
    print(f'Optimal Discount Factor: {optimal_discount_factor}, Action: {action}')
```

Python code for PV and WT output power prediction using multi-layer RNN equipped grid-searched hyperparameters of number of layers and learning rates.

```
# Import libraries
import numpy as np
import pandas as pd
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import SimpleRNN, Dense
from tensorflow.keras.optimizers import SGD
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import mean_squared_error
from tensorflow.keras.wrappers.scikit_learn import KerasRegressor
import matplotlib.pyplot as plt

# Load dataset
data = pd.read_csv('your_dataset.csv')

# Function to create sequences from the data
def create_sequences(data, seq_length):
    xs, ys = [], []
    for i in range(len(data) - seq_length):
        x = data[i:i+seq_length]
        y = data[i+seq_length]
        xs.append(x)
        ys.append(y)
    return np.array(xs), np.array(ys)

# Sequence length set to 24 hours
seq_length = 24
X_pv, y_pv = create_sequences(data['PV_power'].values, seq_length)
X_wt, y_wt = create_sequences(data['WT_power'].values, seq_length)

# Split the dataset into training, validation, and testing sets
X_train_pv, X_temp_pv, y_train_pv, y_temp_pv = train_test_split(X_pv, y_pv, test_size=0.3, random_state=42)
X_val_pv, X_test_pv, y_val_pv, y_test_pv = train_test_split(X_temp_pv, y_temp_pv, test_size=0.6667, random_state=42)

X_train_wt, X_temp_wt, y_train_wt, y_temp_wt = train_test_split(X_wt, y_wt, test_size=0.3, random_state=42)
X_val_wt, X_test_wt, y_val_wt, y_test_wt = train_test_split(X_temp_wt, y_temp_wt, test_size=0.6667, random_state=42)
```

```
# Define the RNN model
def create_model(layers=1, learning_rate=0.01):
    model = Sequential()
    for _ in range(layers - 1):
        model.add(SimpleRNN(50, activation='relu', return_sequences=True))
    model.add(SimpleRNN(50, activation='relu'))
    model.add(Dense(1))
    optimizer = SGD(learning_rate=learning_rate)
    model.compile(optimizer=optimizer, loss='mean_squared_error')
    return model

# Wrapper for KerasRegressor
model = KerasRegressor(build_fn=create_model, epochs=30, batch_size=1, verbose=0)


# Grid search for PV power prediction hyperparameters
learning_rates = np.arange(0.001, 10.01, 0.01)
param_grid_pv = { 'layers': [1, 2, 3, 4, 5, 6, 7, 8], 'learning_rate': learning_rates}

# Grid search for WT power prediction hyperparameters
param_grid_wt = { 'layers': [1, 2, 3, 4, 5, 6, 7, 8], 'learning_rate': learning_rates}

# Perform grid search for PV power prediction
grid_pv = GridSearchCV(estimator=model, param_grid=param_grid_pv,
cv=3, scoring='neg_mean_squared_error')
grid_result_pv = grid_pv.fit(X_train_pv, y_train_pv, validation_data=(X_val_pv, y_val_pv))

# Perform grid search for WT power prediction
grid_wt = GridSearchCV(estimator=model, param_grid=param_grid_wt, cv=3,
scoring='neg_mean_squared_error')
grid_result_wt = grid_wt.fit(X_train_wt, y_train_wt, validation_data=(X_val_wt, y_val_wt))

# Display best hyperparameters
print("Best parameters for PV power prediction: ", grid_result_pv.best_params_)
print("Best parameters for WT power prediction: ", grid_result_wt.best_params_)
```

```
# Train the final model with the best parameters for PV power
best_model_pv = create_model(layers=grid_result_pv.best_params_['layers'],
learning_rate=grid_result_pv.best_params_['learning_rate'])
history_pv = best_model_pv.fit(X_train_pv, y_train_pv, epochs=30, batch_size=1,
validation_data=(X_val_pv, y_val_pv))

# Train the final model with the best parameters for WT power
best_model_wt = create_model(layers=grid_result_wt.best_params_['layers'],
learning_rate=grid_result_wt.best_params_['learning_rate'])
history_wt = best_model_wt.fit(X_train_wt, y_train_wt, epochs=30, batch_size=1,
validation_data=(X_val_wt, y_val_wt))

# Predictions
y_pred_pv = best_model_pv.predict(X_test_pv)
y_pred_wt = best_model_wt.predict(X_test_wt)

# Calculate nRMSE
nrmse_pv = np.sqrt(mean_squared_error(y_test_pv, y_pred_pv)) / np.mean(y_test_pv)
nrmse_wt = np.sqrt(mean_squared_error(y_test_wt, y_pred_wt)) / np.mean(y_test_wt)

print("nRMSE for PV power prediction: ", nrmse_pv)
print("nRMSE for WT power prediction: ", nrmse_wt)

# Plotting the results
plt.figure(figsize=(12, 6))

# PV power prediction plot
plt.subplot(2, 1, 1)
plt.plot(y_test_pv, label='True PV Power')
plt.plot(y_pred_pv, label='Predicted PV Power')
plt.title('PV Power Prediction')
plt.xlabel('Time (hours)')
plt.ylabel('Power')
plt.legend()
```

```
# WT power prediction plot
plt.subplot(2, 1, 2)
plt.plot(y_test_wt, label='True WT Power')
plt.plot(y_pred_wt, label='Predicted WT Power')
plt.title('WT Power Prediction')
plt.xlabel('Time (hours)')
plt.ylabel('Power')
plt.legend()

plt.tight_layout()
plt.show()

# Plot training history for PV power
plt.figure(figsize=(12, 4))
plt.plot(history_pv.history['loss'], label='Train Loss')
plt.plot(history_pv.history['val_loss'], label='Validation Loss')
plt.title('Training History for PV Power Prediction')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()

# Plot training history for WT power
plt.figure(figsize=(12, 4))
plt.plot(history_wt.history['loss'], label='Train Loss')
plt.plot(history_wt.history['val_loss'], label='Validation Loss')
plt.title('Training History for WT Power Prediction')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()
```

Python code for MG power management and BESS energy scheduling based on the proposed MARL method.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import random
from tqdm import tqdm

# Constants and Parameters
Agents = 4
agent = Agents
C_NG = 0.157  # Natural gas cost ($/m^3)
ρ_WT = 0.1    # Penalty factor for wind turbine power difference ($/kWh)
ρ_PV = 0.09   # Penalty factor for PV power difference ($/kWh)

# Maximum and minimum power values
P_MT_max = 100  # kW
P_MT_min = 10   # kW
Q_boiler_max = 80  # kW
Q_AC_max = 100     # kW
P_EC_max = 100     # kW
P_WT_max = 100     # kW
P_PV_max = 100     # kW

# Efficiency and other parameters
η_e = 0.5
η_b = 0.8
β = 0.09043  # m^3/kWh
COP_AC = 0.7
COP_EC = 3
η_hu = 0.8
μ_CO2_ρ_CO2 = 0.062  # ($/m^3)
q_BESS_max = 20      # kWh
P_BESS_dch_max = 6   # kW
q_BESS_min = 6       # kWh
η_ch = 0.85
P_BESS_ch_max = 6    # kW
η_dch = 0.98
```

```
# Read actual power data for PV and WT
P_wind_actual = pd.read_csv('wind_actual.csv')['power'].values
P_wind_forecasted = pd.read_csv('wind_forecasted.csv')['power'].values
P_PV_actual = pd.read_csv('PV_actual.csv')['power'].values
P_PV_forecasted = pd.read_csv('PV_forecasted.csv')['power'].values

# Read load data
D_elec = pd.read_csv('electrical_load.csv')['load'].values
D_heating = pd.read_csv('heating_load.csv')['load'].values
D_cooling = pd.read_csv('cooling_load.csv')['load'].values

# Number of states and actions
num_states = 100
num_actions = 8

# Q-tables for each agent
Q_tables = {
    'MT': np.zeros((num_states, num_actions)),
    'boiler': np.zeros((num_states, num_actions)),
    'AC': np.zeros((num_states, num_actions)),
    'EC': np.zeros((num_states, num_actions)),
    'WT': np.zeros((num_states, num_actions)),
    'PV': np.zeros((num_states, num_actions)),
    'BESS_ch': np.zeros((num_states, num_actions)),
    'BESS_dch': np.zeros((num_states, num_actions))
}

# Define action space for each agent
actions = {
    'MT': np.linspace(P_MT_min, P_MT_max, num_actions),
    'boiler': np.linspace(0, Q_boiler_max, num_actions),
    'AC': np.linspace(0, Q_AC_max, num_actions),
    'EC': np.linspace(0, P_EC_max, num_actions),
    'WT': np.linspace(0, P_WT_max, num_actions),
    'PV': np.linspace(0, P_PV_max, num_actions),
    'BESS_ch': np.linspace(0, P_BESS_ch_max, num_actions),
    'BESS_dch': np.linspace(0, P_BESS_dch_max, num_actions)
}
```

```
def reward_function(P_MT, Q_boiler, Q_AC, P_EC, P_WT, P_PV, P_BESS_ch, P_BESS_dch, t):
    """Reward function considering penalties for PV and WT discrepancies, and constraint violations."""
    fuel_cost = (P_MT + Q_boiler) * C_NG
    penalty_wind = ρ_WT * abs(P_wind_actual[t] - P_WT)
    penalty_PV = ρ_PV * abs(P_PV_actual[t] - P_PV)

    # Equality constraints penalties
    elec_constraint = abs(P_WT + P_PV + P_MT + P_BESS_dch - P_BESS_ch - P_EC - D_elec[t])
    heat_constraint = abs(Q_boiler - Q_AC + (P_MT * η_e / η_hu) - D_heating[t])
    cool_constraint = abs(Q_AC * COP_AC + P_EC * COP_EC - D_cooling[t])

    return -(fuel_cost + penalty_wind + penalty_PV + elec_constraint + heat_constraint + cool_constraint)

def select_action(Q_table, state, epsilon):

    if np.random.uniform(0, 1) < epsilon:
        return random.choice(range(num_actions))
    else:
        return np.argmax(Q_table[state])

# Q-learning parameters
alpha = 0.1  # Learning rate
gamma = 0.9  # Discount factor
epsilon = 0.1  # Exploration factor

# Training loop
for episode in tqdm(range(1000)):
    state = 0  # initial state
    for t in range(24):
        # Select actions for each agent
        actions_selected = {agent: select_action(Q_tables[agent], state, epsilon) for agent in Q_tables}
```

```
        # Retrieve actions
        P_MT = actions['MT'][actions_selected['MT']]
        Q_boiler = actions['boiler'][actions_selected['boiler']]
        Q_AC = actions['AC'][actions_selected['AC']]
        P_EC = actions['EC'][actions_selected['EC']]
        P_WT = actions['WT'][actions_selected['WT']]
        P_PV = actions['PV'][actions_selected['PV']]
        P_BESS_ch = actions['BESS_ch'][actions_selected['BESS_ch']]
        P_BESS_dch = actions['BESS_dch'][actions_selected['BESS_dch']]

        reward = reward_function(P_MT, Q_boiler, Q_AC, P_EC, P_WT, P_PV, P_BESS_ch, P_BESS_dch, t)
        new_state = state + 1  # For simplicity, assume the new state is incremented by 1

        # Update Q-tables
        for agent in Q_tables:
            Q_tables[agent][state, actions_selected[agent]] += alpha * (reward + gamma * np.max(Q_tables[agent][new_state]) - Q_tables[agent][state,
                                                                        actions_selected[agent]])

        state = new_state

# Extract optimal actions
optimal_actions = {agent: np.argmax(Q_tables[agent], axis=1) for agent in Q_tables}

# Extract results from optimal actions
results = {agent: actions[agent][optimal_actions[agent][:24]] for agent in optimal_actions}

# Calculate final overall cost
final_cost = sum(reward_function(results['MT'][t], results['boiler'][t], results['AC'][t], results['EC'][t], results['WT'][t], results['PV'][t],
                                 results['BESS_ch'][t], results['BESS_dch'][t], t) for t in range(24))

print("Final overall cost: ", final_cost)
```

```
# Visualize results
# Plot power generation from PV, WT, micro turbine
plt.subplot(2, 2, 1)
for agent in ['WT', 'PV', 'MT']:
    plt.plot(range(24), results[agent], label=f'{agent} Power (kW)')
plt.xlabel('Time (hours)')
plt.ylabel('Power (kW)')
plt.legend()
plt.title('Wind Turbine, PV, and Micro Turbine Power Generation')

# Plot BESS charge and discharge power
plt.subplot(2, 2, 2)
plt.plot(range(24), results['BESS_ch'], label='BESS Charge Power (kW)')
plt.plot(range(24), results['BESS_dch'], label='BESS Discharge Power (kW)')
plt.xlabel('Time (hours)')
plt.ylabel('Power (kW)')
plt.legend()
plt.title('BESS Charge and Discharge Power')

# Plot other power components
plt.subplot(2, 2, 3)
for agent in ['boiler', 'AC', 'EC']:
    plt.plot(range(24), results[agent], label=f'{agent} Power (kW)')
plt.xlabel('Time (hours)')
plt.ylabel('Power (kW)')
plt.legend()
plt.title('Boiler, AC, and EC Power')

# Plot all power components together
plt.subplot(2, 2, 4)
for agent in results:
    plt.plot(range(24), results[agent], label=f'{agent} Power (kW)')
plt.xlabel('Time (hours)')
plt.ylabel('Power (kW)')
plt.legend()
plt.title('All Power Components')
plt.tight_layout()
plt.show()
```