AN AGENT-BASED DEEP REINFORCEMENT LEARNING APPROACH FOR NETWORKED MICROGRID SCHEDULING TO IMPROVE RESILIENCE

by

Sujay A. Kaloti

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Electrical Engineering

Charlotte

2024

Approved by:

_____
Dr. Badrul Chowdhury

_____
Dr. Valentina Cecchi

_____
Dr. Rob Cox

_____
Dr. Churlzu Lim

_____
Dr. Linquan Bai

ABSTRACT

SUJAY A. KALOTI. An Agent-Based Deep Reinforcement Learning Approach for Networked Microgrid Scheduling to Improve Resilience. (Under the direction of DR. BADRUL CHOWDHURY)

Because of climate change, the ageing power system infrastructure is under threat due to the ever-increasing intensity and frequency of high-impact, low-probability (HILP) events. Although, in most cases, these events are area-specific, the impact of such events, if unaddressed, can lead to cascading failures. Therefore, it is vital for the grid of tomorrow to not only be reliable but also be resilient in view of the broad inter-dependencies. Despite being a widely researched topic, the applicability of the concept of resilience, especially in power systems terms, is not a straightforward task due to the lack of consensus on a consistent definition, or a set of robust metrics. Therefore in this dissertation, an analysis of different definitions, frameworks, metrics, and enhancement techniques related to resilience proposed by multiple researchers and research organizations are discussed. Together, the aforementioned concepts set up the fundamental basis of this dissertation.

Strengthening the existing system to withstand the extreme weather events (infrastructure resilience) or improving the operability of the system under emergency conditions (operational resilience) are the two important aspects of power system resilience improvement. While the infrastructure resilience improvement techniques are effective, due to the inherent characteristic of extreme weather events (to be spatio-temporal in nature), the benefits these improvement techniques provide are unevenly distributed which hinders the applicability of such techniques. Therefore, improving the operational resilience aspect of the distribution system under normal and emergency conditions is the primary focus of this dissertation.

Microgrids (MGs) have emerged as one the solutions for improving operational resilience. By supplying the critical loads using localized generating resources under

emergency operating conditions, MGs can improve the survivability aspect of system's resilience. MGs that are closely located geographically can be interconnected with each other forming a network of MGs called Networked Microgrids (NMGs). These clusters of MGs further enhances the operational resilience by improving critical load pickup through resources sharing under abnormal conditions.

With an aim to improve the operability under uncertainties (during normal operating condition) and resilience by supplying critical loads (during emergencies), we introduce a novel Dual Agent-Based framework for optimizing the scheduling of DERs and loads within a NMG that leverages the field of advanced machine learning. A deep reinforcement learning (DRL) framework that aims to minimize operational and environmental costs during normal operations while enhancing critical load supply indices (CSI) under emergency conditions is developed in this dissertation. Modeling a robust reward function that provides a feedback to the agent regarding the best control actions is pivotal in DRL frameworks. Therefore, a multi-temporal dynamic reward shaping structure along with the incorporation of an error coefficient to enhance the learning process of the agents is designed.

To appropriately manage loads during emergencies, we propose a load flexibility classification system that categorizes loads based on its criticality index. The scalability of the proposed approach is demonstrated through running multiple case-studies on a modified IEEE 123-node benchmark distribution network. We also test the proposed method with different DRL algorithms to demonstrate its compatibility and ease of application, whereas for validation we compared the results with the meta-heuristic algorithms: particle swarm optimization (PSO) and genetic algorithm (GA).

To gain a deeper understanding of the developed model, we conducted a sensitivity study. The key findings from this study align with the mathematical foundation of the approach outlined in this dissertation, providing further support.

DEDICATION

I dedicate this dissertation to my family and friends.

ACKNOWLEDGEMENTS

I am profoundly grateful to the many individuals who have played a significant role in molding my academic journey. To begin with, I would like to extend my sincere gratitude to my advisor, Dr. Badrul Chowdhury, whose constant support and guidance have been priceless. The path has certainly had some challenges, and there were times when I struggled to contemplate key aspects of my doctoral work. Even during those moment, Dr. Chowdhury was incredibly patient with me, dedicating his precious time and energy to address my questions and concerns. Working under his tutelage has been a privilege that enabled me to grow both professionally and personally.

I would also like to express my heartfelt appreciation to my doctoral dissertation committee members: Dr. Valentina Cecchi, Dr. Rob Cox, Dr. Churlzu Lim, and Dr. Linquan Bai for their time and invaluable feedback, which greatly contributed towards the development of my dissertation.

Additionally, I am thankful for the financial support I received as a part of the Graduate Assistant Support Plan (GASP) from the Graduate School of the University of North Carolina at Charlotte. Without this support, pursuing my academic aspirations would have been an arduous task.

Again, I would like to acknowledge Dr. Chowdhury for involving me in advance research opportunities (that involved being part of the Center for Advance Power Engineering Research: CAPER) as his research assistant. Working on these projects immensely helped me develop my research and analytical skills. Furthermore, I am grateful to receive numerous teaching assistantship opportunities from the Department of Electrical and Computer Engineering. These opportunities allowed me to develop my teaching and communications skill thereby enhancing my overall academic experience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BES	Battery Energy Storage
CSI	Critical Load Supply Index
DER	Distributed Energy Resource
DG	Diesel Generator
DNN	Deep Neural Networks
DRL	Deep Reinforcement Learning
MG	Microgrid
NMG	Networked Microgrid
PV	Photovoltaic
RL	Reinforcement Learning

Indices

j	Index for the DG systems
k	Index for the BES systems
l	Index for the loads
m	Index of the MG systems
n	Index for the voltage nodes
p	Index for the PV systems

t	Index for the time step
x	Index for the load tier

Parameters

β^{BES}	Dynamic reward variable for BES
$\eta^{\text{BES,ch}}$	Charging efficiency of the BES
$\eta^{\text{BES,disch}}$	Discharging efficiency of the BES
$\gamma_{\text{CO}_2}^{\text{factor}}$	Fuel to CO_2 emissions conversion factor
λ^{BES}	Cost of BES power in \$/kWh
λ^{DG}	Cost of DG power in \$/kWh
λ^{G}	Cost to import/export grid power (\$/kWh)
λ^{load}	Inverse criticality of the load
λ^{over}	Over load allocation penalty
λ^{under}	Under load allocation penalty
$\lambda_{\text{SC}-\text{CO}_2}^{\text{cost}}$	Social cost of CO_2 in \$ per ton of CO_2
ν	Voltage violation penalty
ϕ^{DG}	Operating power factor of DG
C^{load}	Criticality index of the load
$E^{\text{BES,max\_cap}}$	Maximum capacity of the BES
$E^{\text{BES,min\_cap}}$	Minimum allowable capacity of the BES

$f^{\text{DG\_fuel}}$	Average diesel consumed in gallons/kWh
NLS^{load}	No load supply penalty
$P^{\text{actual\_load}}$	Total actual load available in kW
$P^{\text{BES,max\_ch\_rate}}$	Maximum charge rate of the BES
$P^{\text{BES,max\_disch\_rate}}$	Maximum discharge rate of the BES
$P^{\text{BES,min\_ch\_rate}}$	Minimum charge rate of the BES
$P^{\text{BES,min\_disch\_rate}}$	Minimum discharge rate of the BES
$P^{\text{DER\_available}}$	Total DER power available in kW
$P^{\text{DG,max}}$	Maximum DG power output
$P^{\text{DG,min}}$	Minimum DG power output
$P^{\text{G,max\_export}}$	Maximum power exported to the grid
$P^{\text{G,max\_import}}$	Maximum power imported from the grid
$P^{\text{load\_max}}$	Maximum allocated load (kW)
$P^{\text{load\_min\_allowed}}$	Minimum kW required for essential operation
P^{load}	Total load available in kW
P^{loss}	System loss in kW
P^{M}	Mismatch value
P^{PV}	Total PV power available in kW
v^{max}	Maximum allowable voltage in p.u.
v^{min}	Minimum allowable voltage in p.u.

Variables

α^{BES}	Action vector for BES operation
α^{DG}	Action vector to supply DG power
α^{load}	Action to allocate load power
E^{cap}	Energy stored in the BES
$f(P^{\mathrm{M}})$	Mismatch penalty function
$P^{\mathrm{allocated\_load}}$	Power supplied to the allocated loads
$P^{\mathrm{BES,ch}}$	Power utilized by the BES to charge
$P^{\mathrm{BES,disch}}$	Power supplied by the BES to discharge
P^{BES}	Power supplied by the BES
P^{DG}	Power supplied by the DG
V^{count}	Number of voltage nodes in violation

CHAPTER 1: INTRODUCTION

An electric power system is the heart of today's modern society as it is inextricably interconnected with a multitude of the critical infrastructure sectors. Stable operation of the electrical power system is essential, particularly during external disruptive events, for the societal well-being because of the interdependencies. Even minor interruptions in the electricity supply may result in a considerable material as well as economical losses. While modern electric power systems are designed to withstand short duration power disruptions, it's the longer duration disturbance that is of primary concern. The concept of power system resilience has gained significant attention in recent years due to the increase in the amount of extreme weather-induced long duration outages. Such extreme weather events that challenge the power system resilience are called high-impact, low-probability (HILP) events for obvious reasons. As climate change becomes more prevalent, the frequency and the severity of such events will be more compelling [1].

The reliability concept is well established and is being widely used to define the system performance during planned and unplanned events. The reliability standards defined by NERC and IEEE, can be partially applied to quantify resilience. However these standards are not sufficient to get a comprehensive view of a network's resilience as there are inherent differences between the concept of reliability and resilience. Some of the distinctions between these two concepts are highlighted in Table 1.1.

In the past two decades, extreme weather events have caused major disruptions to power system infrastructure and operations that have led to widespread social and economic losses. The most recent Texas grid failure as a result of the winter storm Uri caused about 4.5 million people to lose power [3]. The northeastern region of the

Table 1.1: Reliability vs Resilience

Reliability	**Resilience**
Applicable to low impact, high probability events	Applicable to high impact, low probability events
Deals with the ability to provide power during the normal operating conditions (blue sky days)[2]	Deals with the ability to operate fully or under reduced form during abnormal operating conditions (black sky days) [2]
The events that deal with the reliability of a system are spread across the network area	The events that deal with resilience of a system are area and time specific (spatiotemporal).
Outages/interruptions last for minutes to hours	Outages/interruptions last for hours to days
Lower interruption costs	Higher interruption costs
Systems are usually designed to have a certain level of reliability based on the widely accepted standards	Currently, there are no standards for designing system that need to fulfill a certain level of resiliency

United States was hit by hurricane Irene in 2011 and hurricane Sandy in 2012 that resulted in around 6.69 million and 8.66 million people losing power respectively [4]. The southern parts of the United States experienced loss of power to about 2 million customers due to the landfall of category 4 hurricane Harvey in 2017 [5]. The tsunami which was a consequence of the great earthquake in the eastern part of Japan caused loss of power to roughly 8.5 million customers [6]. The California wildfire of 2018 also known as the "Camp fire" caused due to the negligence of aging transmission infrastructure damaged approximately 18,804 structures and around 84 individuals lost their lives [7]. This wildfire was a consequence of prolonged draught situation in the area and human error. Reference [8] presents all the billion-dollar disaster events that affected the United States between 1980 to 2021. Statistics shows that frequency of the weather-related events was about 6.7 events/year in the 2000s (2000-2009) which increased to about 18.7 events/year in the last three years (2019-2021). During the period from 1980 to 2021, severe storms caused the highest number of billion-dollar events ($\sim 160\ events$) while tropical cyclones brought about the most damages ($\sim \$\ 1,194.4\ billion$).

Although power system resilience is a widely studied concept, due the lack of general consensus on the standard power system resilience definitions, framework, and

metrics, the applicability of the resilience enhancement techniques is not straight forward [9]. Hence, the subsequent subsection provides a comprehensive overview of the concept of power system resilience, power system resilience assessment frameworks, power system resilience evaluation metrics, and power system resilience enhancement techniques.

1.1 Power System Resilience

1.1.1 Concept of Power System Resilience

In the United States' Presidential Policy Directive-21 (PPD-21) [10] the term resilience is defined from the perspective of critical infrastructure where a resilient critical infrastructure is able to adapt to and withstand the changing conditions and/or recover promptly from a disrupted state during any contingency event. This resilience definition can intentionally be applied to power systems since, apart from being a critical infrastructure itself, a power system enables functionalities for other critical infrastructures. The U.S. Department of Energy (DOE) proposed a model called the North American Energy Resilience Model (NAERM) [11] which incorporates long-term energy planning and real-time situational awareness capabilities to ensure reliable and resilient energy delivery. This framework adopts the resilience definition proposed in the PPD-21. The Sandia National Laboratory (SNL) also used the resilience definition proposed in PPD-21 to quantify and develop enhancement strategies for power system resilience [12].

The Pacific Northwest National Laboratory (PNNL) has adopted a similar definition where resilience is defined as the ability to prepare for and adapt to changing conditions, withstand and recover rapidly from disruptions for a number of disruptive events [13]. Likewise, the authors in [14] from NREL have quoted the definition proposed in PPD-21 in their work that involves improving distribution system resilience using Model Predictive Controlled (MPC) critical load restoration. It should be noted that the above-mentioned national laboratories (in addition to a few others)

are members of the Grid Modernization Laboratory Consortium (GMLC) [15] which was established as a strategic partnership between the DOE and the national labs for collaborative work on grid modernization. The resilience definition followed by the labs, which are part of this consortium, is standard in all the studies performed under the Grid Modernization Initiative (GMI) [16] and can be stated as "*the ability to anticipate, prepare for, and adapt to changing conditions and withstand, respond to, and rapidly recover from disruption through adaptable and holistic planning and technical solutions*" [17].

A resilience definition applicable to the distribution system resilience was proposed by the Electric Power Research Institute (EPRI) that constitute three components: prevention, recovery, and survivability [18]. The United Kingdom Energy Research Center (UKERC) [19] regards resilience as the capability of a system to tolerate and continue to deliver affordable services during an extreme event. It further emphasizes the recoverability aspect of resilience and highlights the importance of alternative means to provide post-disastrous event services. In a report created for the National Association of Regulatory Utilities Commissioners (NARUC) [20], the importance of the robustness and recoverability characteristics is highlighted for a resilient utility infrastructure operation that help avoid or minimize service interruptions. The International Council of Large Electric Systems (CIGRE) [21] defines power system resilience as the ability to limit the extent, severity, and duration of system degradation after an extreme event.

A generalized definition for resilience which is the ability to absorb, adapt to, and/or recover rapidly from a degraded state was provided by the National Infrastructure Advisory Council (NIAC) [22]. When it comes to critical infrastructures this definition can be extended to be the ability to maintain critical functions and operations, prepare, respond, and manage resources during a crisis event, and to return to normal operating conditions as quickly and efficiently as possible. The North American

Reliability Corporation (NERC) considers resilience as the time-dependent component of reliability as defined in the Adequate Level of Reliability (ALR). The ALR performance is determined by the stable operation of the Bulk Electric System (BES) during normal and predefined disturbances [23]. The objective of the ALR assesses the BES over four time horizons: 1) steady-state; 2) transient state; 3) operations state; 4) recovery and system restoration state. These four states corresponds to the four resilient power system characteristics defined by the NIAC in [22]. Hence, the ALR definition filed by NERC is consistent with the NIAC resilience framework and the FERC definition [24] for resilience that addresses the robustness, resourcefulness, rapid recovery, and adaptability of the bulk power system.

In [25], Haimes introduced resilience as the flexibility of the grid to restore its operation, with little or no human intervention, to a normal and reliable operating state. This definition was adopted in [26] to quantify the resilience improvement measures. Another definition for resilience was proposed by the North American Transmission Forum (NATF) which is the ability of the system and its components to minimize damage and improve recovery from a non-routine disruptions in a reasonable timeframe [27]. From the power system's standpoint, the authors of [28] have defined resilience as the system's ability to resist HILP events and rapidly recover from such events and adapt its operation and structure to mitigate impact of such events in the future.

Since there is no universally accepted definition for power system resilience, its applicability largely depends on the type of problem being tackled. Nevertheless, based on the review of the literature that solely focuses on defining power system resilience, resilience features can be standardized to form the building blocks of power system resilience definition. These building blocks can be stated as: the ability to anticipate and sustain a disruptive event or adapt to and recover efficiently after a disruptive event (anticipate and sustain or adapt and recover). An important point

to note is by standardizing power system resilience definition (characteristics), the possibility of standardizing power system resilience metrics for quantifying resilience improvement techniques and driving resilience oriented investments, increases drastically. Nonetheless, developing a “one size fits all" resilience metric is an arduous task due to the inherent characteristics of resilience-oriented studies that largely depends on the predefined set of resilience goals.

1.1.2 Power System Resilience Analysis Framework

Due to the increased importance of the concept of resilience as part of grid modernization operations and planning efforts, it is vital to develop a robust resilience framework and quantification approaches. A resilience framework would help provide a set of instructions to analyze the system's resilience. The results of the analysis will form the basis for the resilience-oriented system operations and planning decisions. Metrics to quantify resilience improvement are required in order to weigh certain techniques against others for supporting investment strategies. In this section, different power system resilience assessment framework are discussed.

The authors of the report presented in [29] developed a method for assessing baseline resilience and evaluating resilience improvement measures called the Resilience Analysis Process (RAP). The RAP is a risk-based decision making process for stakeholders and decision-makers that contains six steps (seven steps if the resilience improvement evaluation is included) for assessing system performance. The RAP processes begins by defining high-level resilience goals which sets the foundation for the following steps. Defining the system and resilience metrics that involves setting the scope of the analysis is performed in the second step. Information from the stakeholders regarding the type of consequences to be considered in the analysis is considered in this step. Threat characterization is performed in step three. The extent of damage to the system due to a specific threat (threat determined in step three) is estimated in step four. Information related to the disrupted components is then used as an input

to the system models for system's state evaluation (in step five). In step six, the results obtained from the system models are quantified and mapped to the resilience metrics defined in step two. The evaluation of resilience improvement techniques is performed in step seven, if the goals defined in step one involves proposing resilience enhancement strategies. Figure 1.1 represents the steps involved in the RAP.

Organizations seeking to improve the resilience of the system can use the RAP to streamline their resilience-oriented studies. First, the baseline level of the system's resilience to a specific threat can be estimated following the aforementioned six steps of the RAP. The resilience goal defined by an organization can be to improve the recovery of the system after an extreme event. Metrics that can quantify such improvement would then be defined to estimate the effects of such events on the system at its current state (the measure of the consequences can be to determine the duration for which customers were out of power). Once the baseline resilience is quantified, improvement techniques can be applied and evaluated to ascertain the advancement in the resilience metric (step seven). To summarize, the metrics defined in the RAP can be used for two purposes: first, to provide the system's baseline resilience performance index, and second, to evaluate the improvement in the system's resilience after an improvement technique is applied by providing a means to compare the improved performance vs the baseline performance.

A framework for power system resilience evaluation was proposed in [30] where the system resilience evaluation was grouped into two categories: 1) Qualitative framework 2) Quantitative framework. Qualitative framework can be used to evaluate the power system resilience and other interdependent systems where capabilities such as emergency preparedness, mitigation strategies, rapid response and recovery, etc. are studied. Resilience evaluation methods include surveys and questionnaires matrix development; a two-dimensional framework used to quantify improvement in energy-related attributes due to measures taken in an interdependent sector [31],

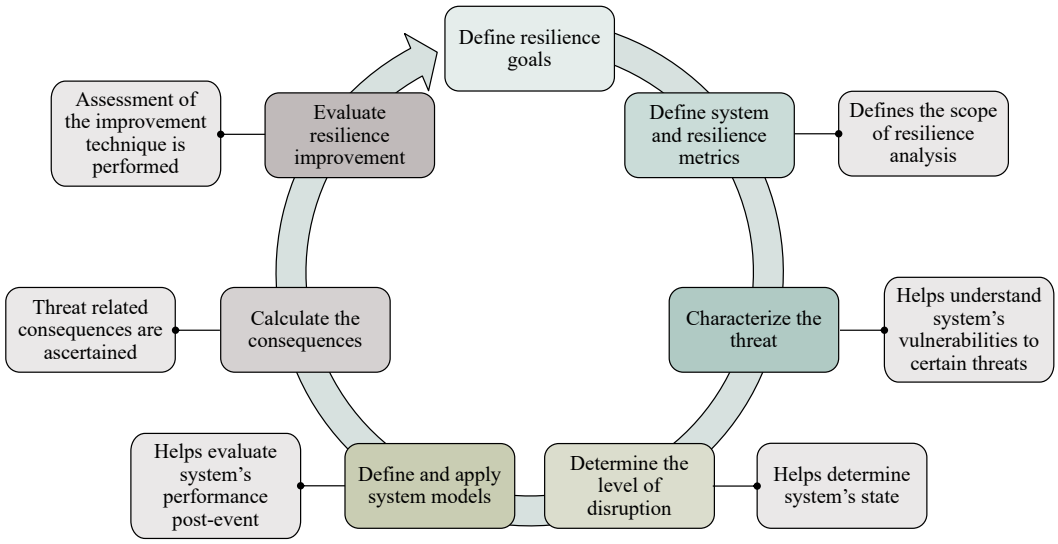

Figure 1.1: The Resilience Analysis Process

etc. Quantitative framework depends on the quantification of the system's performance attributes. Resilience metrics developed using the quantitative framework are event-specific and provide a basis for decision making [29]. The approaches used for resilience evaluation include simulations-based, analytical-based, and statistical-based approaches.

Based on the above literature review, the main step in any resilience-oriented studies is to define a proper set of resilience goals. As mentioned earlier, it is unrealistic to consider all resilience related issues to be addressed in these studies. Therefore, it is crucial to have clearly defined resilience goals followed by the steps and a robust set of metrics not only to achieve these goals but also to justify the applicable resilience enhancement strategies, as the implementation of such strategies involves a considerable amount of initial investments. Moreover, having disaster preparedness guidelines also helps the system operators to be ready for a certain set of consequences by appropriately planning their system's event response.

1.1.3 Power System Resilience Metrics

Reliability and resilience oriented enhancement strategies are largely governed by the extent of the benefits its implementation render to the society. To assess the benefits of reliability improvement techniques, well-defined reliability metrics has been defined for power distribution systems, eg. System Average Interruption Duration Index (SAIDI), System Average Interruption Frequency Index (SAIFI), Customer Average Interruption Duration Index (CAIDI), Momentary Average Interruption Frequency Index (MAIFI) [32] as well as for the transmission systems, eg. Loss Of Load Probability (LOLP) and Loss Of Load Expectation (LOLE) [33] which forms the basis for reliability improvements. On the other hand, to assess the benefits of resilience improvement techniques, currently there are no standard well-defined set of metrics that can guide resilience investments. Although, in many cases, a more reliable system can be considered as a more resilient system and vice versa, that is not true in every situation [34]. Moreover, application of reliability metrics to justify resilience improvement techniques might fall short of evaluating certain key factors associated with resilience (event impact, outage duration, etc.). Hence, knowing what factors are important for developing resilience metrics is crucial in the metric development process.

In [29] a set of recommendations for developing resilience metrics are presented. These include:

1. Metrics should be defined considering a specific type of the HILP event.
2. Metrics should appropriately quantify the performance of the system under study.
3. Metrics should capture the threat level associated with each extreme event.
4. Metrics used to quantify resilience must account for the uncertainties associated with the HILP

5. Metrics should effectively capture the resilience attributes like the ability to anticipate, prepare, withstand, adapt, and recover

Additional guidelines/recommendations towards developing resilience metrics were proposed in [9] where the authors emphasized that the metrics used should be able to enumerate the system's resilience for a particular category of consequences. From [35], resilience metrics must also address the geographical and time-varying aspect that an extreme event has on the system's resilience. Several other desirable properties of the resilience metrics were presented in [36] which includes ease of application, comprehension, and interpretation. Although, the specified recommendations will help develop resilience metrics, it is not a requirement for resilience metrics to include all the above points as the metrics would depend on the resilience goals. Thus for different applications/improvement strategies, different metrics can be proposed or developed.

The authors in [37] proposed an extended version of the concept of resilience triangle called a multi-phase resilience trapezoid which was used to develop the resilience metrics. Figure 1.2 shows the proposed multi-phase resilience trapezoid. The different phases of the multi-phase resilience trapezoid that characterize power system states during an extreme event are described below:

1. Phase I indicates the disturbance phase ($t \in [t_{oe}, t_{ee}]$) with two key elements of the resilience metrics linked to this phase. The first element describes how quickly the systemâs resilience decreases (from $[P_0$ to $P_{pd}]$) during the extreme event whereas the second element gives the magnitude of the drop in resilience $[P_0 - P_{pd}]$.

2. Phase II indicates the disturbed or degraded state (post extreme event) of the power system. During this state the time, for which the system remains in the degraded state, is considered as the resilience measure ($t \in [t_{ee}, t_{re}]$).

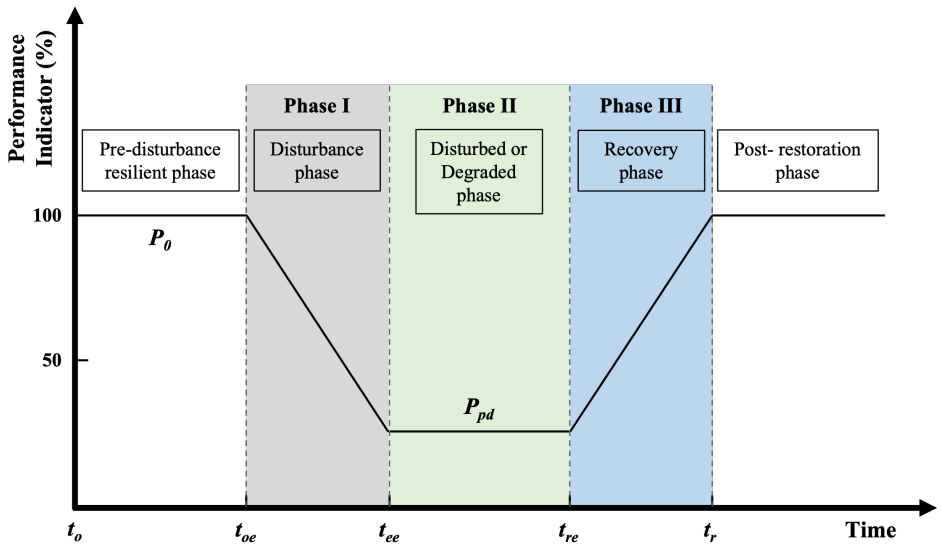

Figure 1.2: Multi-phase Resilience Trapezoid

3. Phase III is associated with the recovery phase post extreme event ($t \in [t_{re}, t_r]$). The measure of resilience depends on how quickly the system recovers from the degraded state to the pre-fault or acceptable level (state).

On the basis of the multi-phase resilience trapezoid concept, the authors developed a $\Phi\Lambda E\Pi$ ($FLEP$) metric. Φ and Λ metric measures how fast and how low the resilience drops, E represents how long the post-event degradation lasts and the Π metric quantifies the promptness of the network recovery. In addition to the four metrics, an area metric was also proposed that essentially considers the integral of the trapezoid for the event duration. Based on the applicability, the standard multi-phase resilience trapezoid concept cane be extended to consider the operational and infrastructure aspects of system resilience.

The mathematical expression associated with each of the aforementioned metrics is presented below. The Φ metrics is calculated using Equation (1.1) where the measuring unit could be the MW/hours lost or Number of lines tripped/hour.

$$\Phi = \frac{P_{pd} - P_0}{t_{ee} - t_{oe}} \tag{1.1}$$

The mathematical expression for the Λ metrics, which measures the total MW or Number of lines tripped during an event, is given by Equation (1.2)

$$\Lambda = P_0 - P_{pd} \tag{1.2}$$

The metric E that measures the hours for which the system remains in the degraded state can be expressed by Equation (1.3)

$$E = t_{re} - t_{ee} \tag{1.3}$$

The MW (load) or the number of lines restored per hour is quantified by the Π metric whose mathematical expression is given by Equation (1.4)

$$f_b(x) = \frac{P_0 - P_{pd}}{t_r - t_{re}} \tag{1.4}$$

Equation (1.5) provides the *Area* metrics' mathematical expression, which is used to determine the performance of the system during an extreme event

$$Area = \int_{t_{oe}}^{t_r} P(t) \tag{1.5}$$

A similar approach was used in [38] where a standard resilience trapezoid was considered and the system resilience was quantified as the reciprocal of the systemâs loss of performance. The loss of performance was determined using the largest deviation from the normal level of performance and the integration of the relative deviation during the degradation phase. The metrics also considers the rapid recovery aspect by considering performance degradation duration. The authors in [28], developed a Severity Risk Index (SRI) as a metric to determine whether the proposed resilience enhancement technique should be implemented. The SRI depends on the probability

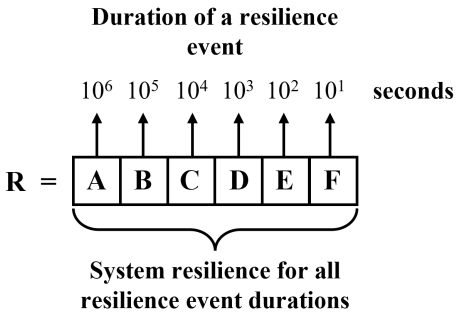

Figure 1.3: Code-based Resilience Metric

of an extreme event and the consequences associated with a specific extreme event.

A majority of extreme events has a spatiotemporal aspect associated with it. To address this aspect, the authors of [39] have proposed a time-series analysis approach to assess current and future system resilience. The metrics used for the analysis include time to repair (TTR) which is a function of the severity of the event, and reliability-based metrics like the loss of load frequency (LOLF), expected energy not served (EENS), and loss of load expectation (LOLE). To analyze the improvement in the power system resilience due to the use of microgrids, the authors in [40] proposed four indices that are combined to form a power grid resilience metric âΘâ. The indices include an index K for expected number of line outages, an index for loss of load probability (LOLP) to measure load loss probability, an index for energy demand not served (EDNS) to enumerate expected demand that was not satisfied, and an index G for measuring the level of difficulty in grid recovery.

A code-based resilience metric was proposed in [36] where the measure of the networkâs resilience was governed by an empirical equation which is designed to capture the impact of an unfavorable event. The authors proposed six variables A, B, ..., F that correspond to the event's time duration in 10^i secs (where $i = 0, 1, 2\ldots, 6$) of an extreme event shown in Figure 1.3 and has a resilience value between 1 to 9 associated with it.

$$m' = c(\alpha + e^f)(1 + f) \tag{1.6}$$

$$f = \frac{\text{Load unaffected by an extreme event (kW)}}{\text{Total load (kW)}} \tag{1.7}$$

The unscaled resilience value is calculated using Equation (1.6), where c is the binary indicator for extreme event occurrence, α is the event duration time, and f is the fraction of unaffected loads given by Equation (1.7). The calculated unscaled resilience value is appropriately scaled using Table 1.2 to get a resilience value between 1 to 9.

The authors of [41] proposed a resilience vector that included five resilience indices which were used to quantify the resilience of the network. The first index was associated with the load shedding cost saved ($), whereas the second index considered the cost saved during the restoration process ($/hr$). The next two indicators were graph theory-based indices which were the weighted algebraic connectivity and weighted betweenness centrality. The last index was a function of the first two indices and was termed as the adaptability index. In [42], the authors developed a multi-temporal resilience metrics that quantifies the anticipate, withstand, and recovery aspects of power system resilience. In that paper, each aspect of power system resilience has its own set of indices/impacting factors that are used to develop the resilience score. The anticipate metric score relies on the weighted sum of the three domains namely, threat and vulnerability, power delivery and loads, and restoration and recovery. For the withstand aspect, the resilience score R_w depends on the critical loads not served, total available generation, critical load demand, topological robustness, and threat impact (here topological robustness was determined using graph theory concepts). Lastly, the score of recovery metrics depends on the critical load restored, path redundancy, generation redundancy, switching operations, and switching time.

Table 1.2: Scaling of code-based resilience metrics

m'	1.00-3.71	3.72-6.42	6.43-9.13	9.14-11.85	11.86-14.56	14.57-17.27	17.28-19.98	19.99-22.70	22.71-25.41
m	1	2	3	4	5	6	7	8	9
	Low resilience			Moderate resilience				High resilience	

In [26], the authors used three metrics to define the system's operational resilience. The first metric was system flexibility index that measured the demand served after each recovery iteration given by Equation (1.8).

$$R^{\lambda}_{i,n,d,t} = \frac{\sum_{i\in I}\sum_{n\in N} P^{t|\epsilon}_{d_n,i}}{P^{T}_{d}} \tag{1.8}$$

The second metric was the outage cost recovery which is the amount of customer costs regained after each corrective action - Equation (1.9)

$$R^{\mu}_{i,n,d,t} = \sum_{i\in I}\sum_{n\in N} C_{d_n}\ (P^{t|\epsilon}_{d_n,i+1} - P^{t|\epsilon}_{d_n,i}) \tag{1.9}$$

The percentage of demand recovered in each recovery step compared to the total demand lost was the last metric named as the outage recovery capacity metric given by Equation (1.10).

$$R^{\vartheta}_{i,n,d,t} = \sum_{i\in I}\sum_{n\in N} \frac{(P^{t|\epsilon}_{d_n,i} - P^{t_d|\epsilon}_{d_n})}{(P^{T}_{d} - P^{t_d|\epsilon}_{d_n})} \times 100 \tag{1.10}$$

Here, the $R^{\lambda}_{i,n,d,t}$, $R^{\mu}_{i,n,d,t}$, and $R^{\vartheta}_{i,n,d,t}$ are the flexibility, recovery capacity, and outage cost recovery metrics of load demand d ($\forall d \in D$: System demands) at load node n ($\forall n \in N$: System buses) after the adoption of the i^{th} ($\forall i \in I$: Iteration count for recovery process) network reconfiguration plan at time t ($\forall t \in T$: Time step). Also, C_{d_n} is the value of lost load d at node n , $P^{t|\epsilon}_{d_n,i}$ is the active power demand at bus n after the i^{th} recovery action for ϵ extreme event, $P^{t_d|\epsilon}_{d_n}$ is the active power demand at node n when the extreme event ϵ ends, and P^{T}_{d} is the total active power demand at node n during normal conditions.

Resilience of the distribution system is measured in terms the critical load restoration capability in [43]. It is governed by the integral of the performance function $F(t)$ given in Equation (1.11) which is proportional to total power supplied to the critical

loads weighted by their priority.

$$R = \int_{t_r}^{t_r+T^0} F(t)dt \tag{1.11}$$

The resilience metric R is defined for the restoration period $[t_r,\, t_r + T^0]$ where T^0 is the duration of the outage and t_r is the time at which the first restoration action is taken.

Authors of[37] and[38] proposed a way to quantify the resilience during the Phase II of the multi-phase trapezoid, where[37] used the duration of Phase II as a metric while[38] calculated the area under the curve bounded by the duration of Phase II to quantify resilience. The aforementioned works consider the recovery phase (Phase III) of power system resilience as one of the most important phase of the resilience trapezoid.[37] and[38] used a recovery rate function to determine how quickly the system returns to the normal or acceptable operating limits. In[39] the time to repair (TTR) is used to quantify the resilience during this phase. This TTR was expressed as a function of wind speed ($f_w(w(s))$) and normal time required for a power system component to repair (TTR_{norm}). A grid recovery index was proposed in[40] which essentially provides information regarding the severity associated with the recovery efforts. Here, the authors assigned weights (w_i) and values (η_i) to each (i^{th}) factor affecting the recovery of the system. The metric proposed to quantify Phase III of the resilience trapezoid in[41] deals with evaluating the restoration cost savings (ΔC_T^R). The higher the restoration cost saving, the quicker is the system recovery (shorter Phase III). Different researchers use different metrics to quantify resilience based on their research question. Nonetheless, these metrics can be categorized based on the respective resilience phases to which it is being applied to. For example: the Phase I metric might include reliability based metric, or a probability based metric whereas the Phase III metric might consist of a cost-based or a time-based metric. Considering

an approach that uses resilience phases to standardize a set of metrics for each of these phases can help regulate how resilience is quantified.

Additional work on developing resilience metrics was conducted by the Grid Modernization Laboratory Consortium (GMLC) Metrics team [44] who proposed a resilience metric comprising of two main categories. The first category is the multi-criteria decision analysis (MCDA) which provides a baseline understanding of the networkâs resilience in the form of a resilience index (RI) and facilitates improvement options consideration. The second category is a performance-based metric that quantitatively describes the effect of a certain extreme event on the network. Key indicators for the performance-based metric include cumulative customer-hours of outage, time to recovery, loss of utility revenue, etc. In [34], the authors present two metrics where the first metric focuses on the recovery aspect during the first 12 hours of a storm, while the second metric quantified the robustness and ability to withstand the event.

Although performance and attribute-based quantification of power system resilience is important, it is not sufficient to capture holistic significance of resilience enhancement strategies. One of the key factors according to the resilience enhancement circle [35] is the benefit/cost analysis for selecting the appropriate enhancement strategy. Typically, there are two approaches for appraising resilience, namely the bottom-up approach and the economy-wide approach [45]. The bottom-up approach uses customer preferences, responses, or behavior in determining the value of resilience whereas the economy-wide approach estimates resilience value by considering the effects of a power outage on regional economies using appropriate indicators. The chart presented in Figure 1.4 shows the subcategories and different models used in each of these subcategories for valuing resilience.

One of the concepts (which is based on the bottom-up approach) used to perform economical evaluations of power systems is the concept of Value of Lost Load (VoLL).

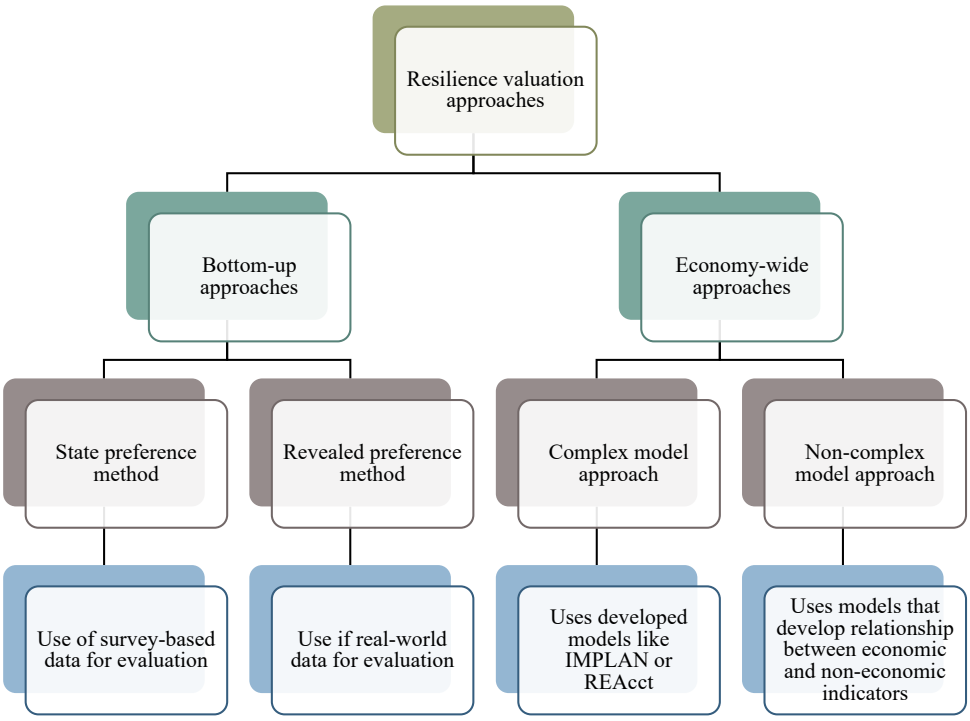

Figure 1.4: Resilience Valuation Approaches

Monetizing the value that represents the importance of electricity continuity helps in informed decisions-making. VoLL is the "perceived" value that a consumer places on the convenience of having uninterrupted supply of electricity [46]. The VoLL is generally estimated using the stated preference approach where customer response to certain outage situations are recorded via. customer surveys and converted into a monetary metric. However, using the VoLL calculated for a short-term outage cannot be used to justify the resilience-oriented investments as it does not capture the compounding effects of a long-term outage on the customer damage cost estimation [47]. Moreover, in a majority of studies that incorporate VoLL for resilience decision assessment, VoLL has been assumed to be a static or a constant value [48] which is not an accurate representation of customer outage cost as it over-simplifies the damage cost.

In [49] a new way of calculating the customer damage cost is presented which considers the effect of a long-term power interruption called the customer damage

function. The customer damage function provides the damage cost as a function of outage duration. The factors/characteristics that contribute toward estimating the customer damage cost are the outage characteristics, customer characteristics, and some other factors [50]. Outage characteristics include the elements that account for the outage duration, frequency, time of day, day of week, season of the year, etc. The customers characteristics that influence the customer damage cost are the type of customer (commercial, residential, etc,), number of customers, in each of these types, affected and the criticality index of the affected customers.

Although this subsection provides a comprehensive review of the resilience evaluation metrics, the application of these metrics highly depends on the type of enhancement technique being used. As the concept of power system is multi-phased, a standard metric that can quantify the resilience improvements is a challenging task. At the same time, standardized metric might not offer the most accurate measure of resilience improvement compared to a customized metric tailored for a specific enhancement technique. Nonetheless it is essential to understand what are the different metrics that are being used to quantify resilience as it can provide invaluable insights into the development of these custom resilience metrics.

1.1.4 Power System Resilience Enhancement Techniques

There is a wide range of literature published that highlights the resilience improvement techniques. In this subsection, some of the most commonly-used and upcoming techniques for resilience improvement are discussed. The resilience enhancement strategies can be applied in regards to the two aspects of resilience as suggested in [51] which are the infrastructure resilience improvement aspect and the operational resilience improvement aspect. As the name suggests infrastructure resilience improvement deals with boosting the robustness of the system components whereas the operational resilience improvement deals with maintaining secure supply to the loads during an imminent disaster. The decisions to improve infrastructure resilience are

mainly planning-based decisions which require a broader time horizon for implementation. On the other hand, use of smart grid technologies like the use of automatic switches for network reconfiguration which isolates the faulty sections of the network thus improving operational resilience can be treated as a real-time decisions.

Equipment hardening is the most common and effective (not a cost-effective solution) way to improve infrastructure resilience by bolstering network's robustness and resistance to change. In [37], hardening techniques were applied on the transmission network to reduce network degradation during an event. As these measures are not cost-effective, targeted hardening is performed using the fragility curves (which gives the probability of failure as a function of the weather element) for the components to be hardened. Additionally, the importance of undergrounding the distribution lines, upgrading poles and structures, and increasing redundancy to improve infrastructure resilience is presented in [52, 53, 54, 55]. Majority of distribution system outages during an extreme event are vegetation-induced outages. A report published by EPRI [56] provides an insight into current utility practices for vegetation management and how it helps in improving overall resilience. As a part of the wildfire damage mitigation strategy, PG&E an investor-owned utility company, has proposed an enhanced vegetation management (EVM) program [57] that specify standards for minimum power line clearance as well as enable conducting additional inspections beyond routine patrols to remove hazardous vegetation. Coastal substations, especially the ones on the east coast of the United States, are vulnerable to storm surges and flooding. To address this issue, alternative design approaches for modeling substation are presented in [58, 59].

One of the most common techniques to improve operational resilience is the network reconfiguration that makes use of automatic switches, microgrids (MGs), demand response programs, mobile emergency resources and priority-based load restoration strategies [60]. In [26, 61] optimal topology control was presented that enhances

network's emergency operation while [28] discusses a defensive islanding strategy to avoid cascading failure. A framework to improve critical load pick-up using MGs was developed in [43, 62, 63]. MGs can also be used to reduce the load restoration times [64], with optimal remote-controlled switch placement [65], being an alternate approach. Efficient use of Mobile Emergency Resources (MERs) [66], and energy storage devices [67] are some of the other techniques discussed in the literature.

A Networked MG (NMG) can be developed by interconnecting closely located MGs to form a reliable cluster of MGs. This enables resource sharing and critical loads support during emergency conditions [68, 69, 38], thus improving overall reliability and resiliency. An energy management strategy for day-ahead scheduling of the NMG is presented in [70] where the scheduling problem is formulated as a mixed integer linear programming problem to minimize the operational cost. [71] developed a bi-level optimization model for sizing the energy storage system that can be used in a NMG setup. As more and more renewable energy resources (like solar PV systems, distributed wind energy systems, etc.) come online on the distribution side, the challenge of addressing the uncertainties associated with such resources arises. In [72] consideration was given to these uncertainties associated not only with renewable generation but also to the unintentional islanding scenario while formalizing a NMG scheduling problem.

1.2 Rationale

The majority of operational resilience improvement techniques employ mathematical optimization frameworks that require strict adherence to the specifications of the optimization algorithm, which in turn requires a detailed model of the network under analysis [73]. The complexity of the formulation increases as the network becomes more intricate. Moreover, to ensure optimal results under uncertainties, running a large number of scenarios using mathematical frameworks can lead to a considerable amount of computational load. Therefore, certain assumptions or relaxations

are needed to strike a balance between the number of scenarios considered vs the required computation time. Methods like reinforcement learning (RL) and/or deep reinforcement learning (DRL) offer an alternative approach to addressing such complexities. Various research works explore the use of such techniques in the power system operations [73] and controls [74, 75] domain.

Such techniques can also be applied to the field of energy management and resource scheduling of a single or multi-MGs, especially in situations that involve uncertainties. Table 1.3 some of the works with discrete and continuous state/action spaces that employ the RL/DRL frameworks for applications like demand response, demand-side energy management, and MG energy management. Majority of these works falls short as the actions space selected does not necessarily represent real-world scenarios. Moreover, approaches presented in [76, 77] considers a single MG setup where only the normal operating conditions are modeled. In today's day and age, it is important for a framework to focus not only on the normal operating conditions but also on the emergency operating conditions. Therefore, in our work, we aimed to address these concerns.

Table 1.3: Literature review: Use of RL/DRL framework for Power System applications

Ref. No.	**Application**	**Learning Algorithm**	**Control Actions**
[78]	Demand Response	Q-learning	Discrete
[79, 80]	Energy Management (Demand-side)	Deep Q-learning	Discrete
[81, 82]	Energy Management (Demand-side)	Policy Gradient-based	Continuous
[76]	Energy Management (MG Control)	Proximal Policy Gradient	Continuous
[77]	Energy Management (MG Control)	Deep Q-learning	Discrete

In this dissertation, we introduce a Dual Agent-Based (DA) framework that utilizes the principles of DRL to enhance the operational resilience. The proposed DA

framework aims at minimizing the operational costs and environmental impacts under normal operating conditions, while maximizing the critical load supply indices (CSI) under emergency operating conditions. Several research works utilize a multi agent-based approach to manage resources in multi-MGs, such as in [83]. The concept of multi-agent deep reinforcement learning (MA-DRL) framework requires that agents to operate within a shared environment. However, the environment dynamics, observations, and actions associated with the normal and emergency operating conditions, respectively, are different. Therefore, in this approach, each agent within the DA set-up will reside in its own environment with its own set of state and action vectors.

In DRL frameworks, a reward function plays a pivotal role. This dissertation introduces a dynamic reward shaping approach to develop the reward function that integrates the environment's multi-temporal fluctuations. This approach enhances the DA's decision-making process by providing supplementary information regarding the actions taken in specific states. We also developed a training architecture that utilizes day-ahead forecasts combined with an error coefficient that resembles real-world forecasting errors. This enables the DAs to explore a wide range of scenarios during training allowing it to develop a robust policy. The DAs are trained using the day-ahead forecasts and deployed online to provide real-time control. Therefore, to assist with the real-time application, we propose a one-shot episode termination approach (diverging from the conventional episode termination approach based on time-series) which additionally enables faster convergence with relatively lower computational burden.

During emergency conditions, more specifically when the clustered microgirds are operating in an islanded mode, prioritizing power supply to the critical loads is of utmost importance. However, due to constrained resource availability (due to DER disconnection and islanded operation) a thorough understanding of load flexibility is

required. Determining the flexibility of each load individually in real-world systems would be an arduous task. Hence, this dissertation introduces a multi-tier classification system for load categorization based on the load criticality index enabling efficient load allocation during emergency scenarios. The contributions of this work include:

- A novel Dual Agent-based Deep Reinforcement Learning (DA-DRL) framework that automates real-time resource scheduling and load allocation during normal and emergency operating conditions, respectively.
- A multi-temporal dynamic reward shaping approach that enhances the agent's decision-making process.
- A training paradigm that integrates an error coefficient into the day-ahead forecasts to prepare the DAs for a multitude of scenarios.
- A one-shot episode termination method that enable faster convergence, reduce computational load, and ease real-time application.
- A multi-tier classification system for load categorization based on the load criticality index that aids in critical load allocation.

1.3 Dissertation Organization

The rest of the dissertation is structured as presented in Figure 1.5. An overview on MGs, NMGs, and Agent-based modeling is presented in Chapter 2. This chapter discusses the basic structure and control architectures of MGs and NMGs along with the mathematical basis of RL and DRL methods which enables the agent's learning. Chapter 3 offers an insight in to the methodology of the proposed framework. It further highlights the MDP modeling and the problem formulation of the respective agents. Additionally, this chapter expands on the DA's training paradigm developed in this work along with the DG's testing structure. To test the DA framework,

multiple case studies were run whose results are depicted in Chapter 4. Comparison of the agent's learning with different DRL algorithms and the validation of the results with the metahueristics techniques is also included in this chapter. Moreover, a section that deals with evaluating the sensitivity of the model with respect to certain learning parameters, is also detailed in this chapter. Lastly, Chapter 5 encompasses the concluding remarks accompanied by the discussion on major contributions of the proposed work, and possible future works.

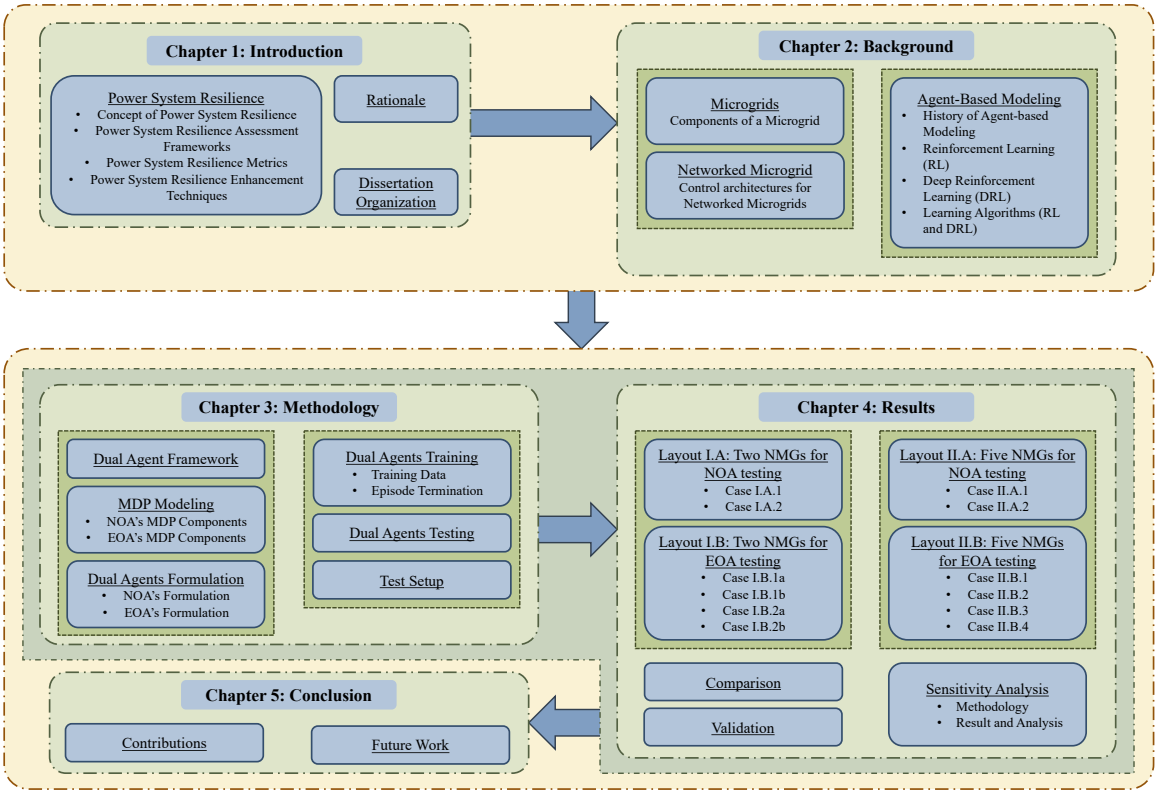

Figure 1.5: Dissertation Organization

CHAPTER 2: BACKGROUND

2.1 Microgrids

This section provides a high-level overview of MGs and its components. A MG is a localized energy system consisting of a group of interconnected loads and Distributed Energy Resources (DERs) that operate as a single controllable entity [84]. MGs have the ability to operate in both a grid-connected mode and an islanded mode (where the MG is disconnected from the main grid). Due to proliferation of DERs at the distribution level in recent years, MGs provide a flexible and an efficient way to integrate these growing resource interconnections. Additional advantages of the MGs include [85]:

- MGs reduces power losses due to it using the localize energy resource for load delivery.
- MGs can enhance the network's reliability and resilience especially under extreme conditions.
- MGs improves local energy management thereby reducing the operational and the resource planning costs.
- MGs provide platform to integrate carbon-zero resources thus advocating the carbon neutral energy generation initiative.

MGs have the capability to transform the traditional paradigm of energy generation and consumption that incorporates local DERs and advance energy management systems. The Department of Energy (DOE) Office of Electricity (OE) considers MGs

to be the building blocks of the electricity delivery system of the future [86] making it a key player in the grid modernization efforts.

2.1.1 Components of Microgrids

A standard MG layout is demonstrated in Figure 2.1 that highlights the crucial components of a MG. These components include:

- Distributed Energy Resources (DERs)
- Network Loads
- Point of Common Coupling (PCC)
- Energy Management System (EMS)

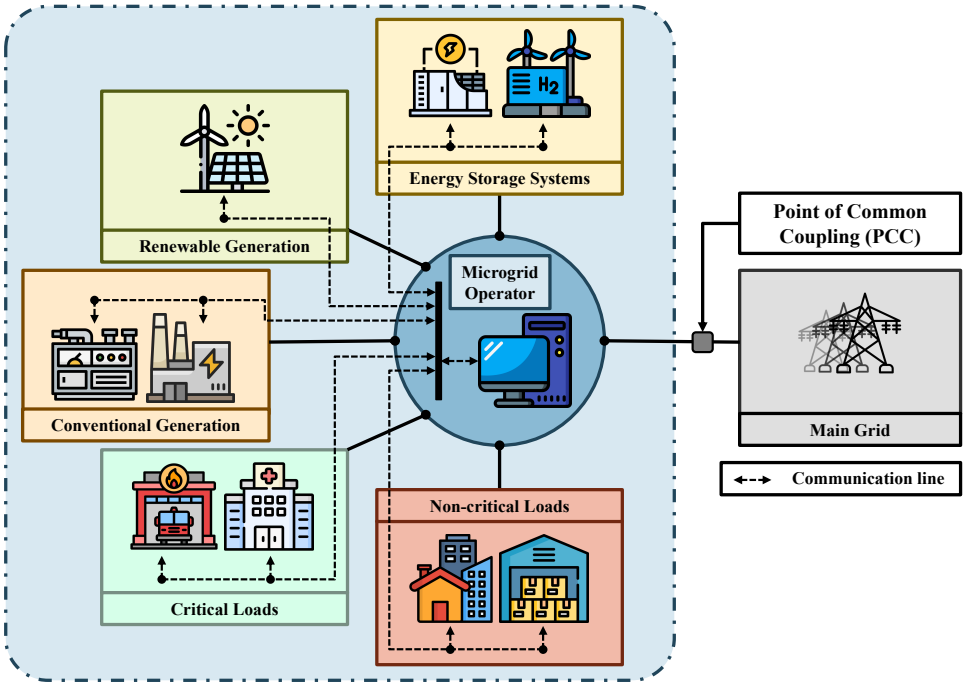

Figure 2.1: Standard MG structure

2.1.1.1 Distributed Energy Resources (DERs)

Distributed Energy Resources (DERs) can be considered as small-scale energy generating resources that are interconnected at the distribution level. These resources

can be further categorized into conventional DERs, renewable DERs, and storage systems.

Conventional DERs can also be called as controllable/dispatchable DERs as the output of such units can be controlled by the operator. These set of DERs include a micro gas turbines (MTs), diesel generators (DGs), combined heat and power systems (CHP), fuel cells, etc. MTs are small combustion gas turbines that uses ignition of a mixture of compressed air and fuel creating combustion gases which are used to drive the turbines [87]. Typically the output of the MTs ranges between 30 kW to 200 kW. DGs on the other hand use a diesel engine to convert diesel fuel energy into mechanical energy which is coupled to an alternator that generates electrical energy [88]. CHP systems are also called cogeneration systems due to its ability to generate heat and electricity from a single energy source and reuse the excess heat to further improve its operational efficiency [89]. Fuel cells are energy conversion devices that combines the gaseous fuel (hydrogen) with oxidant gas (oxygen) to generate heat and electricity [90]. Availability of controllable/dispatchable DERs in the MG structure helps minimize the operational cost in grid-connected mode coupled with an improvement in MG's overall reliability and resilience in an islanded mode.

Most common renewable DERs include the solar photovoltaic (PV) systems and wind turbines. As the output of such DERs depend on the external intermittent factor (solar irradiance for solar PV systems and wind energy for wind turbines), these resources are considered to be non-controllable/non-dispatchable resources. Moreover, some of the controllable/dispatchable renewable DERs include small scale steam power plants that use geothermal energy or biomass energy to generate electricity. Micro (Small) hydropower systems that converts the energy of running water into electricity can also be considered as a controllable renewable DERs [91].

In a MG comprising of only renewable DERs, storage systems provides the necessary controllability. Moreover, any excess generation from these renewable DERs can

be stored and used in later hours of the day to provide peak shaving or serve as a back-up source under emergencies. Different energy storage technologies are discussed in [92] out of which the ones suitable to use in a MG setup are [93]:

- Battery Energy Storage (BES) System: A system that uses an assembly of electrochemical cells connected in series and parallel to generate electricity through an electrochemical reaction.

- Hydrogen Storage System: A type of chemical energy storage system that uses electricity to produce and store hydrogen through the electrolysis of water [94]. The stored hydrogen can then be used in a fuel cell or can be burnt directly to generate electricity [95].

- Flywheel Energy Storage: A type of mechanical energy storage that stores the rotational kinetic energy which is transferred in and out of the flywheel through an electrical machine operating as a motor or a generator depending on charge/discharge operation [96].

- Supercapacitor Energy Storage System: An electrochemical energy storage system that stores energy through electrostatic charge separation and provides a higher power density option compared to other electrochemical storage systems, like BESs [97].

- Thermal Energy Storage System: A system that enables storage of energy in the form of heat or cold for later dispatch [98].

Out of all the aforementioned energy storage types, BES systems are the most commonly used technology in a MG setup that provides necessary services. Different types of BES systems are discussed in [99] which includes Li-ion batteries, Lead-acid batteries, Flow batteries, etc. Lead-acid batteries were traditionally used for MG

operations. However due to the higher efficiency and energy density, Li-ion batteries are mostly preferred.

2.1.1.2 Network Loads

In a MG, network loads comprises of the facilities or infrastructure that require electrical energy to function appropriately. Network loads can be broadly classified into critical and non-critical loads [100]. Each of the load classes consist of controllable, deferrable, and non-controllable load types. Controllable loads means the MG operator has the ability to alter the behavior of such loads. Deferrable loads can be scheduled to be served in a different operating window. Controllable and deferrable loads provide operators with an option to manipulate the load's consumption profile unlike non-controllable loads. These loads need to be served (if in critical load class) or to be shed (if in non-critical load class) if at all required, to maintain load and generation balance.

Examples of the critical non-controllable loads can be hospitals, emergency lighting, data centers, etc. Majority of the critical loads can be considered as non-critical loads as the load power required by such loads must be served all the time. Instances of the non-critical controllable loads are HVAC systems of non-essential facilities, industrial equipment that can be temporarily turned off, etc. Moreover, an updated water heater or electric vehicle charging schedule can be considered as an example of a deferrable load.

2.1.1.3 Point of Common Coupling (PCC)

An electrical link that serves as an interfacing point between a MG and the main grid is called as a Point of Common Coupling (PCC) [101]. The PCC enables power exchange between the main grid and the MG during grid-connected mode. Additionally, it acts as an isolation switch for the MGs to transition from grid-connected to islanded mode.

2.1.1.4 Energy Management System (EMS)

Energy Management System (EMS) can be considered at the central processing unit of a MG where information is collected and control tasks are initiated. EMS is a control system that is responsible for optimizing the overall operation of a MG by appropriately adjusting the DERs power output to economically supply the network loads [102]. EMS also monitors the operating conditions and addresses the challenging task of balancing the loads with intermittent generation by efficiently deploying dispatchable generating units during islanded operations [103]. The EMS control architectures for a typical MG can be classified into a Centralized Control Architecture and a Decentralized Control Architecture.

Centralized Control Architecture

In this type of control architecture, a main central controller (MCC) controls the operation of all the components of a MG. The MCC collects the necessary information regarding the DER units and loads along with other relevant information that contains forecast entities, network parameter limits, operation modes, etc. [104]. A typical centralized control architecture is presented in Figure 2.7. A local controller (LC) interacts with the MCC by sharing the necessary data that is required to optimize the overall operation of the MG. In centralized control, the job of local controller is only to share the localize DER/load information with the MCC.

Optimization of a MG through a centralized control architecture is computationally extensive due to the exchange of high volume of real-time data. However, it has a simple structure which is enables ease of implementation.

Decentralized Control Architecture

In a decentralized control system, an optimized operation is obtained through the autonomous behavior of the DERs and loads in which each of these entities seek to maximize the overall objective [100]. Figure 2.8 represents the decentralized control

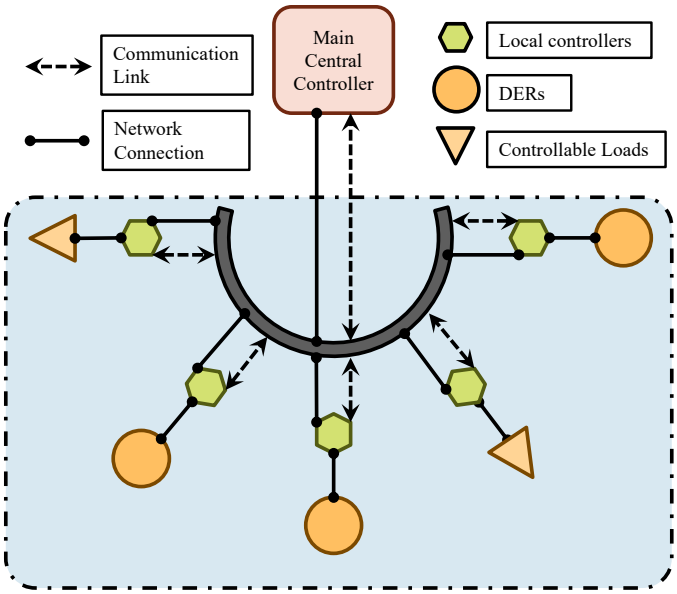

Figure 2.2: Centralized Control Architecture

architecture, where the LC (of the DERs/loads) has the ability to make intelligent operational decisions independently. The LCs communicate with the neighboring LCs, exchanging information to reach a consensus on the optimal control actions under a given situation. Although MCC does not govern any control actions in decentralized control architecture, it can still provide electricity pricing signal and under contingencies can take over control if required [102].

Advantage of using a decentralized control architecture is its computational effective. However, it can be potentially inefficient when comes to resource allocation due to a lack of global optimizing entity.

Communication in MG

To exchange the necessary information between the essential components of a MG, EMS has to rely on a fast, secure, and reliable communication system. IEC 61850 can be used for satisfying the communication requirements for a MG operation [103]. Although this standard was originally developed to regulate the communication protocol in substation automation systems (SAS) [105], due to its flexibility it can be easily adopted for MG operation and communication.

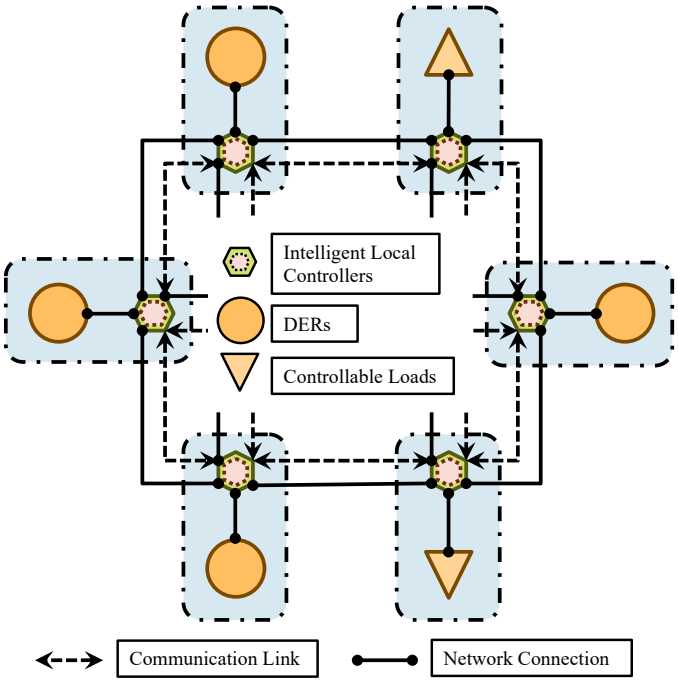

Figure 2.3: Decentralized Control Architecture

The objective of IEC 61850 standard is to provide a framework that enables interoperability between intelligent electronic devices (IEDs) [106]. Moreover, it characterizes the way each of the devices must organize the information to maintain data exchange consistency thus making it a viable solution to the communication problem within a multi-device MG set-up.

2.2 Networked Microgrids

Another efficient approach to effectively use the DERs is to interconnect multiple MGs to form a Networked Microgrid (NMG) [38]. NMGs are a cluster of self-regulating MGs that are responsible for providing robust infrastructure for DER integration thereby aim to improve system's reliability and resilience [107]. Depending on the operating conditions, NMGs can operate either independently optimizing individual MG's objectives or in a cooperative manner where each MG operates collectively to achieve the global objective.

NMGs bolsters grid modernization by enabling interoperability of various smart

grid technologies [108]. It ensures reliable power supply to the critical loads during emergencies, while maintaining a high power quality standard. The interconnected nature of NMG provides a high level of network redundancy which significantly enhances network's reliability and resilience.

In general there are three basic interconnection architectures that essentially standardize different MGs connection with each other (in a NMG) or with the Grid [109, 110].

- Serial Interconnection Architecture:

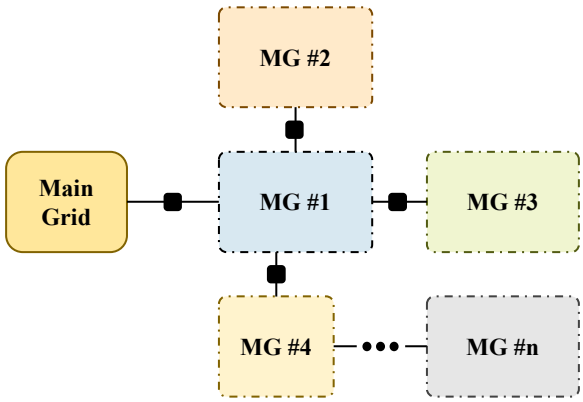

Figure 2.4: Serial Interconnection Architecture

  In this architecture, two or more MGs are networked in series with each others. Figure 2.4 demonstrates the serial interconnection architecture where there is only one MG (from the NMG setup) that has a direct connection with the grid.

- Parallel Interconnection Architecture:

  The parallel interconnection architecture enables multiple MGs of the NMGs to have a direct connection with the grid, separately, as shown in Figure 2.5. A connection between these MGs can be achieved through an interconnection tie.

- Interconnected MG Architecture:

  In an interconnected MG architecture, multiple MGs are connected to multiple feeders, separately, as depicted in the Figure 2.6. Similar to a parallel intercon-

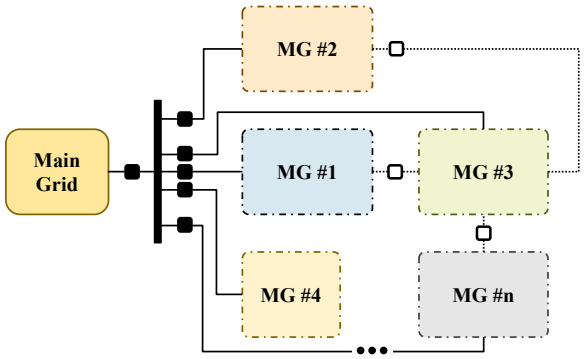

Figure 2.5: Parallel Interconnection Architecture

nection architecture, an interconnection tie can be used to provide a possible interconnection between these MGs.

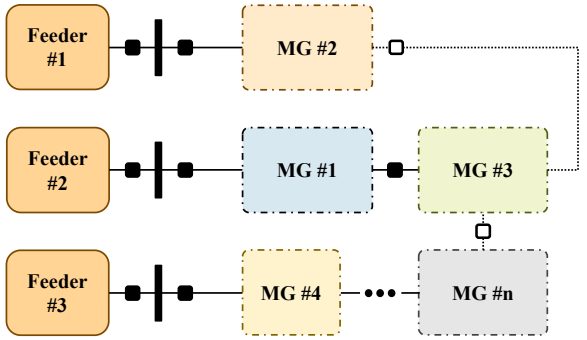

Figure 2.6: Interconnected MG Architecture

2.2.1 Control Architectures for NMGs

2.2.1.1 Centralized Control Architecture

Centralized control architecture used for the NMGs shares the same operating characteristics as the centralized control architecture that controls a single MG. As in single MG control structure, a MCC has the authority to develop control strategies by collecting the necessary information from various NMG's components [111]. Figure 2.7 represents the centralized control architecture from a NMG standpoint.

The MCC determines optimal control actions that achieve the overall objective and communicates those to the individual MG controller. In this structure, the primary task of the individual MG controller is to locally implement the control actions by

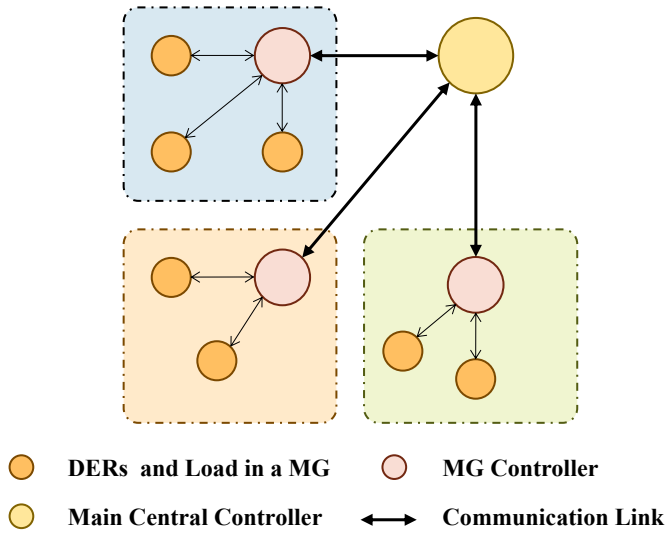

Figure 2.7: Centralized Control Architecture for NMGs

accordingly adjusting the DERs or loads set-points (it cannot operate autonomously except under contingencies). A centralized structure provides a cohesive and integrated approach to solving the energy management problem.

The advantages of a centralized control architecture includes ease of implementation due to a simplistic structure and a relatively low computational burden. However, one of the major concerns with centralized control is the its susceptibility to a single-point failure.

2.2.1.2 Decentralized Control Architecture

In a decentralized control architecture, the authority of decision-making is assigned to the individual components like individual MG controllers. These individual MG controllers can have its own structure (including communication) that would enhance the decision making process.

The basic structure of a decentralized control architecture is shown in Figure 2.8 where a MG controller can communicate and exchange appropriate information with the individual DERs to provide localized optimal control. Additionally, depending on the structure, these individual DERs can also communicate with neighboring DERs

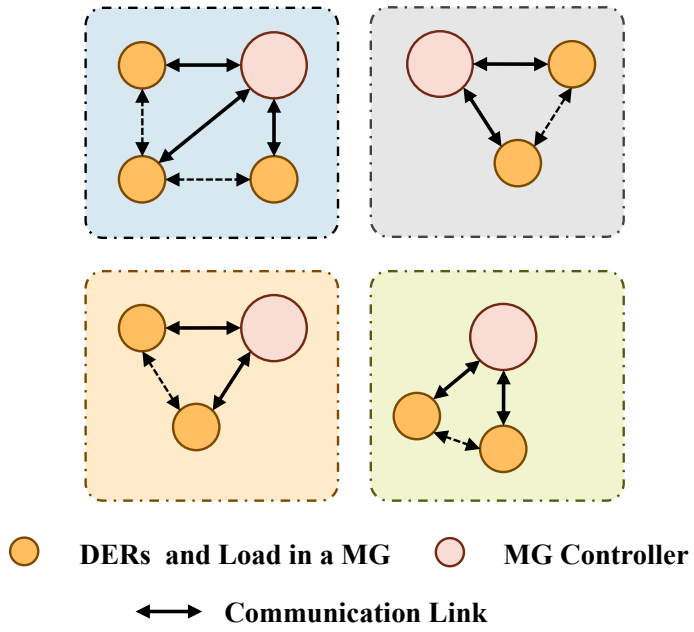

Figure 2.8: Decentralized Control Architecture for NMGs

(as explained in previous the section under decentralized control architecture for single MG) to provide a balance between coordination and autonomy.

One on the major advantages of a decentralized control architecture is it eliminates the vulnerability to single-point failure. On the top of that this control architecture has an improved adaptability and scalability. However, achieving global optimality at a system-level is a major concern.

2.2.1.3 Distributed Control Architecture

Distributed control architecture is an extension of the decentralized control architecture that enables parallel data processing and control [112]. A robust communication between the local DERs bolsters the decision-making process by enhanced coordination at an individual MG level. Moreover, distributed control architecture enables communication with the neighboring MGs that allows regulation of power exchanges amongst them.

Figure 2.9 presents the distributed control architecture for NMG operation. Individual DERs can communicate with each other as well as with the individual MG

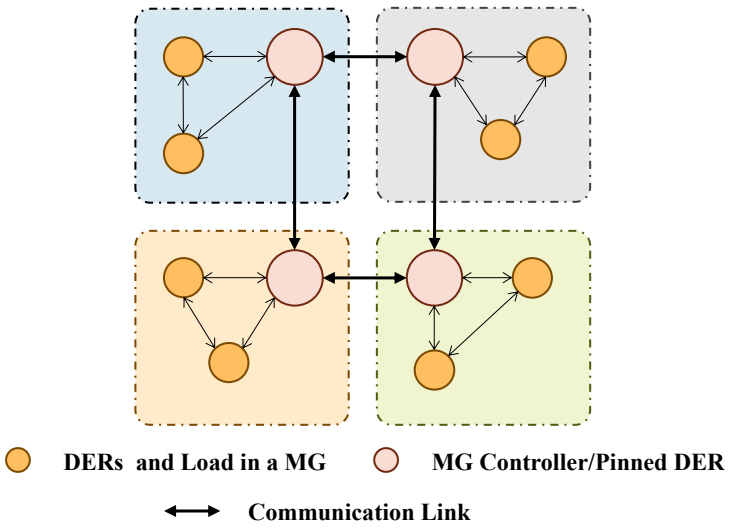

Figure 2.9: Distributed Control Architecture for NMGs

operator (or a pinned DER as per [112]). These individual DERs follow respective individual MG operator for appropriate control and operation.

Although distributed control architecture enhances the coordination and increases the likelihood of achieving global optimality, its implementation remains relatively complex due to its dependence on advance algorithms and sophisticated coordination requirement.

2.3 Agent-Based Modeling

Another important concept that forms the basis of this dissertation is the use of Agent-based DRL framework for solving the NMG's scheduling problem. This section provides an overview of Agent-based modeling from its inception to the present day. Moreover, it includes the details regarding the mathematical background of modern Agent-based RL framework. Lastly, this section discusses the advancements in the applicability of the RL frameworks to real-world problems by introducing DRL paradigm that uses deep neural networks (DNN) for performance enhancement and scalability improvement of the agent-based systems.

2.3.1 History of Agent-Based Modeling

Agent-based Modeling (ABM) is a multi-disciplinary computational simulation technique that relies on an autonomous agent's interactions to evaluate a complex system [113, 114]. An agent is a self-regulating decision-making entity which collectively forms an ABM system. These agents follow a set of predefined rules/objectives to assess its environment and perform necessary tasks.

The first application of an ABM system was observed in the field of social sciences. An ABM was developed in [115] to study racial segregation in an urban set-up that showcases the correlation between individual preferences and the extent of segregation. Although this work did not exclusively use the ABM terminology or computational tools, it laid the groundwork for future ABM advancements by demonstrating how individual decisions can influence overall system's behavior. Another work that provided the foundational grounds for social simulation is presented in [116] which uses ABM for studying social experiences through agents interaction.

Furthermore, ABM was also used in the field of economics and finance to simulate financial systems, replicate trading behaviors, or model markets. An agent-based market model was developed in [117] that evaluates observed parameters by the use of interacting agents to aggregate macroeconomics with the financial data. ABM also finds it use in the health sector especially to ascertain disease progression as a consequence of individual interactions. In [118] an ABM approach was used to develop a model that leverages geographical information to study spread of communicable diseases in an urban environment.

Technological advancement in the field of computer science led to a paradigm shift in the way agents in an ABM are viewed. As computational power increases, incorporation of adaptable and intelligent agents that learn from their own experiences by interacting with a dynamic environment became more manageable. Emerging approaches that combine ABM with advance machine learning techniques like RL can

help develop complex systems. Here the ABM provides a platform for the agents to interact and gather experiences while RL provides learning algorithms to learn from these experiences. This framework enables development of intelligent agents that exhibit dynamic and adaptive behavior, which is particularly important in the systems that are evolving in nature (like an electric power system).

2.3.2 Reinforcement Learning

Reinforcement Learning (RL) is a branch of a machine learning framework that teaches agents to take intelligent decision that maximizes the overall returns [119]. The agent learns to take optimal actions through trial-and-error interactions with the dynamically changing environment [120]. At the beginning of the learning process, the agent explores the environment by taking random actions. However, as learning progresses, it tends to prefer actions that yield higher returns providing a testament to the knowledge it has acquired throughout the training process.

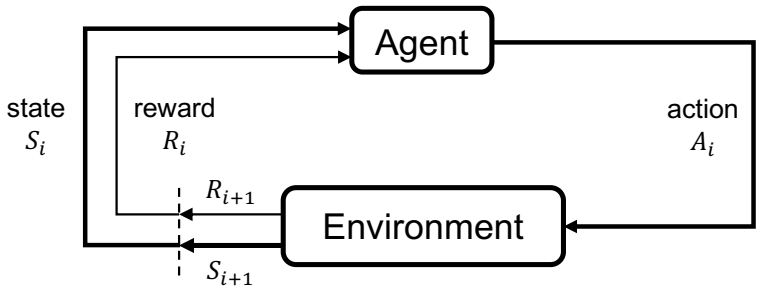

Figure 2.10: Reinforcement Learning Framework

A diagrammatic representation of a RL framework are depicted in Figure 2.10. There is an agent that, after encountering state $S_i \in \mathcal{S}$: where $\mathcal{S}$ is set of possible states, interacts with an environment by taking action $A_i \in \mathcal{A}(S_i)$: where $\mathcal{A}(S_i)$ is a set of action available in state S_i, for which it receives reward $R_i \in \mathcal{R} \subset \mathbb{R}$. Upon taking the action A_i, the states of the environment transition from S_i to a transitioned state S_{i+1}. The main components of a RL framework are:

- Agent: An agent is software program that is modeled to learn and take intelli-

gent decisions.

- Environment: An environment is a set-up where the agent resides and with the agent interacts. An environment is modeled in such a way that it captures the change in the states space dynamics when an agent takes a specific action a. It is also responsible to provide a feedback in the form of reward r to the agent.

- State space (S): A set of state values (s) or observations encountered by the agent that contains essential information about the environment's characteristics.

- Action space (A): A set of action values (a) that the agent can choose to interact with the environment.

- Reward: A scalar value (r) received by the agent as a feedback from the environment that indicates how good the agent's action a was given a state s.

- Value function: A value function provides a value of an agent's specific state. It essentially provides a measure of future expected rewards that agent will receive when in state s. There are two type of value functions:

  - State value function: Provides the measure of expected returns when at state s following a policy π. State value function is represented by $V_\pi(s)$. It quantifies the value of being in a specific state s.

  - Action value function or Q-value function: Provides the measure of expected returns when at state s following policy π and taking action a. Action value function is represented by $Q_\pi(s, a)$. It quantifies the value of an action a taken given a state s.

- Policy: Policy π of an agent governs the action a the agent takes given a state s. A policy of an agent directly maps a specific state s to an action a: $\pi(s_i) \rightarrow a_i$.

An optimal policy selects an action that would provide the maximum reward r: $\pi_*(s_i) \rightarrow a_i$ where: $a_i = \arg\max_{a_i \in A} Q_{\pi^*}(s_i, a_i)$. Here $Q_{\pi^*}(s_i, a_i)$ is the optimal action value or Q-value which will be discussed in the subsequent sections.

Another aspect of a RL environment is an episode. An episode is the Agent-Environment interactions from the initial state to the terminal state. Therefore, the goal of an agent is to learn a policy π that maximizes the overall reward received over span of an episode.

2.3.2.1 Markov Decision Process

The basis to formulate a RL problem must comply with the mathematical framework used to model sequential decision making problem called the Markov Decision Process (MDP) [121]. The basic assumption of an MDP is that the process must satisfy the Markov property. A Markov property suggests that the future state s_{i+1} only depends on the current state s_i and the action a_i taken in that state s_i and not the preceding sequences [122]. The dynamics of the system modeled as MDP can be considered as memoryless where the current state values essentially captures the information of the past states.

In RL, the learning problem is modeled as an MDP. An MDP consists of a tuple of five components: (S, A, P, r, γ). Again, S and A represents the state space and the action space of the agent. P is the conditional probability expressed as $P(s_{i+1}|s_i, a_i)$ that signifies the probability to transition to state s_{i+1} taking an action a_i given a state s_i. Reward function $r \rightarrow r_i$, which can be expanded as $r(s_i, a_i, s_{i+1})$, represents the reward or incentive received by the agent after taking action a_i given a state s_i. The discount factor $\gamma \in [0, 1]$ characterizes the inclusion of future states the agent must consider to compute its returns.

The aim of the agent is to develop a policy ($\pi \rightarrow \pi(s)$) that maximizes the cumulative returns. These returns are given by Equation 2.1 which is the sum of discounted rewards.

$$G_i = r_{i+1} + \gamma.r_{i+2} + \gamma^2.r_{i+3} + \gamma^3.r_{i+4} + \cdots = \sum_{k=0}^{\infty} \gamma^k.r_{i+k+1} \tag{2.1}$$

Using the concept of MDP and leveraging on the above equation, two important function of RL (as described earlier) i.e. the equations for the state-value function ($V_\pi(s)$) and the action-value function ($Q_\pi(s, a)$) can be derived. State-value function that provides a measure of how valuable being is certain state is, is given by Equation 2.2 (i.e. the expected returns when in state $s_i = s$).

$$V_\pi(s) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k.r_{i+k+1} \middle| s_i = s\right] \tag{2.2}$$

The action-value function that signify the value of an action taken in a specific state is given by Equation 2.3 (i.e. the expected returns when action $a_i = a$ is taken in state $s_i = s$).

$$Q_\pi(s, a) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k.r_{i+k+1} \middle| s_i = s, a_i = a\right] \tag{2.3}$$

The action-value function or the Q-value function has been used in many learning algorithms like Q-learning that focuses on developing the policy by maximizing the Q-value rather than the policy itself.

2.3.2.2 Bellman Equations

The value functions defined in Equations (2.2) and (2.3) must satisfy the recursive relationship called the Bellman Equations [123], which forms the cornerstone of many RL algorithms. It enables recursive decomposition of the value function of a policy by expressing the value of a state in terms of immediate and discounted future rewards associated with the successive states.

For a policy π and state s, the value of the possible successive states can be ascertained by Equation 2.7.

$$V_\pi(s) = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k.r_{i+k+1}\middle|S_i = s\right] \tag{2.4}$$

$$= \mathbb{E}_\pi\left[R_{i+1} + \gamma.\sum_{k=0}^{\infty}\gamma^k.r_{i+k+2}\middle|S_i = s\right] \tag{2.5}$$

$$= \sum_a \pi(a|s)\sum_{s_{i+1}}\sum_r p(s_{i+1}|s,a)\left[r + \gamma.\mathbb{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k.r_{i+k+2}\middle|S_{i+1} = s_{i+1}\right]\right] \tag{2.6}$$

$$= \sum_a \pi(a|s)\sum_{s_{i+1},r} p(s_{i+1}|s,a)\Big[r + \gamma.V_\pi(s_{i+1})\Big] \tag{2.7}$$

Equation 2.7 provides a relationship between the value of the current state and that of its successor states. The term $\pi(a|s).p(s_{i+1}|s,a)$ computes the probability for each a, s_{i+1}, and $r \rightarrow (a, s_{i+1}, r)$. By weighing each $\rightarrow a, s_{i+1}, r$; with its respective probability, Equation 2.7 determines the state value by averaging over all the possibilities.

Optimal Value Functions

Again, the end goal of a RL agent is to develop a policy π that would maximize it overall expected returns. If the expected returns of a policy π are greater than that of another policy π' then and only then it can be considered that policy π will perform better than another π'. Therefore, $\pi \geq \pi'$ if and only if $V_\pi(s) \geq V_{\pi'}(s)\ \forall\ s \in \mathcal{S}$. A policy that is at least better than or equal to all other policies is called an Optimal policy denoted by π_*. The optimal state-value function associated with an optimal policy is given by Equation 2.8.

$$V_*(s) = \max_\pi V_\pi(s), \quad \forall\ s \in \mathcal{S} \tag{2.8}$$

Similarly, the optimal action-value function can be computed using Equation 2.9.

$$Q_*(s,a) = \max_{\pi} Q_\pi(s,a), \quad \forall\ s \in \mathcal{S} \text{ and } a \in \mathcal{A}(s) \tag{2.9}$$

Equation 2.9 quantifies the expected return when the agent take action a in a state s following an optimal policy. The Bellman optimality equation supports the idea that the state value under an optimal policy must equal the expected returns for the best action taken in that state. This equation in terms of state-value function is expressed in Equations (2.13) and (2.14).

$$V_*(s) = \max_{a \in \mathcal{A}(s)} Q_{\pi_*}(s,a) \tag{2.10}$$

$$= \max_a \mathbb{E}_{\pi_*}\left[\sum_{k=0}^{\infty} \gamma^k . r_{i+k+1} \middle| S_i = s_i, A_i = a\right] \tag{2.11}$$

$$= \max_a \mathbb{E}_{\pi_*}\left[r_{i+1} + \gamma . \sum_{k=0}^{\infty} \gamma^k . r_{i+k+2} \middle| S_i = s_i, A_i = a\right] \tag{2.12}$$

$$= \max_a \mathbb{E}[r_{i+1} + \gamma . V_*(s_{i+1}) | S_i = s, A_i = a] \tag{2.13}$$

$$= \max_{a \in \mathcal{A}(s)} \sum_{s_{i+1}, r} p(s_{i+1}|s,a)\Big[r + \gamma . V_*(s_{i+1})\Big] \tag{2.14}$$

Similarly, the Bellman optimality equation for optimal action-value function is given by Equations (2.15) and (2.16).

$$Q_*(s,a) = \mathbb{E}\Big[r_{i+1} + \gamma . \max_{a_{i+1}} Q_*(s_{i+1}, a_{i+1}) | S_i = s, A_i = a\Big] \tag{2.15}$$

$$= \sum_{s_{i+1}, r} p(s_{i+1}|s,a)\Big[r + \gamma . \max_{a_{i+1}} Q_*(s_{i+1}, a_{i+1})\Big] \tag{2.16}$$

The Bellman optimality equations especially Equations (2.15) and (2.16) forms the basis for the development of numerous RL and DRL algorithms.

2.3.3 Deep Reinforcement Learning

Environment with low dimensional state and action spaces are well-suited for traditional RL techniques to work effectively. However, as the complexity of the tasks increases with state and action spaces requiring high-dimensional or continuous values, the traditional RL techniques tend to struggle [124, 125]. Most famous RL algorithms like the Q-learning or SARSA can only operate or help develop an agent's policy when the state and action spaces are discrete in nature [126]. The reason being these techniques employ a tabular approach to map the states with value functions of all the possible actions. This approach is impractical when it comes to state and action spaces that are continuous in nature, which is the case with a majority of real-world problems.

DRL provides a solution to the aforementioned limitations that the tradition RL techniques exhibit. DRL leverages DNN [127] to handle complex, high-dimensional state and action spaces enabling the base RL techniques to be applied to more complicated real-world problems. DNNs in DRL framework are used as a function approximator meaning the DNNs are trained to approximate the optimal policy (π_*) and/or the optimal value functions (V_*, Q_*, and A_*) [128]. A detailed explanation of how the DNNs are trained and how the model parameters are optimized is provided in the Appendix D

2.3.4 Learning Algorithms (RL and DRL)

There are numerous RL and DRL algorithms that can be used to train the agent. These algorithms can be classified into two categories: 1) Model-free algorithm, 2) Model-based algorithms. In Model-free algorithms the agent directly learns the value function or the policy by interacting with the environment. These algorithms use the experience it gained through trial and error interaction with the environment to maximize the rewards. Conversely, Model-based algorithms relies on creating a

model of the environment that predicts the next state and reward value given the current state and actions. The goal of these algorithms is to build an internal model of the environment dynamics and reward function which is used for decision making. Due to the complexity involved in explicitly modeling the environment dynamics of a real-world system, model-free algorithms are preferred over model-based algorithms.

Model-free algorithms can be further classified as two sub categories, based on the entity they are trying to optimize. An algorithm that directly optimizes the policy of the agent are the Policy Optimization algorithms. These algorithms use policy gradient updates to optimize the model's parameter. These updates are almost always done on-policy, meaning the updates only use the data that was collected while acting based on the latest updated policy. Another subcategory of model-free algorithms include the algorithms that rely on the approximation of the action-value (Q-value) of the agent's actions also called as Q-learning algorithms. The updates in the Q-learning algorithms are almost always performed off-policy, meaning the agent can use samples of collected data at any point during the learning process regardless of when that data was gathered.

There are another set of algorithms called the Actor-critic algorithms, that combines the advantages of: 1) Policy optimization algorithms $\rightarrow$ in the form of policy updates that essentially governs agents actions (also termed as Actor network), 2) Q-learning algorithms $\rightarrow$ in the form of Q-value updates that essentially evaluates the actions taken by the policy or actor network by ascertaining its Q-value (also known as the Critic network). The Actor-Critic group of algorithms also perform updates using off-policy approach essentially using the experience it gathered to optimize the networks (actor and critic) parameters.

To train the agent and to perform comparative analysis, in this dissertation, actor-critic based algorithms are used. There are three actor-critic algorithms that are most commonly used for continuous state and action spaces which are:

- Deep Deterministic Policy Gradient (DDPG) [129]
- Twin-Delayed Deep Deterministic Policy Gradient (TD3) [130]
- Soft Actor-Critic (SAC) [131]

Details of each of these algorithms are presented in the Appendix C.

CHAPTER 3: METHODOLOGY

In this chapter, the complete methodology used to develop the DA framework is discussed. This includes introduction to the DA framework followed by the MDP model of the DAs. The DA's environment dynamics and reward function modeling is presented in the problem formulation section. The training paradigm developed for the DAs along with the mathematical formulation for the one-shot episode termination approach is explained subsequently. Next, a brief explanation of the real-time testing structure for evaluating the DAs is provided. Lastly, the test set-up information is highlighted in the concluding section of this chapter.

3.1 Dual Agent Framework

To manage the NMGs, a centralized control strategy, as described in [132], is implemented. A central controller interacts with all the interconnected MGs by collecting and transferring vital information which it then uses to perform necessary control tasks. It is also equipped with a backup unit to provide redundancy in case of primary control failure. Moreover, an additional layer of individual MG controllers are provided that enable each MG's independent operation in case of backup controller failure. Figure 3.1 demonstrates the centralized control structure for NMGs.

Depending on the operating conditions, the objective of a MG operator varies. Normal Operating Agent (NOA) is the term (as the name suggests) used to describe the agent that is deployed under normal operating conditions whereas Emergency Operating Agent (EOA) is the agent that is deployed during an emergency operating phase. The objectives of each of these agents are:

- Normal Operating Agent (NOA): To minimize the operational cost of the NMG

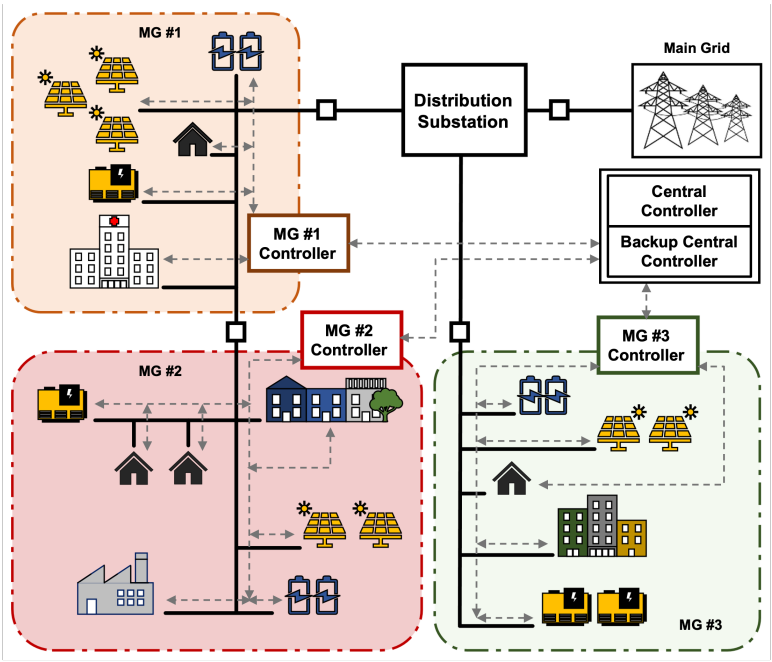

Figure 3.1: Centralized control strategy for NMGs

by determining appropriate set-points for the dispatchable DERs in the NMGs.

- Emergency Operating Agent (EOA): To maximize the critical load supply index by allocating power to the essential loads of the NMG.

The state space, action space, and reward function are the major components of any agent's design. While the reward function for each of the DAs will be discussed in the problem formulation section, the state space and action space for the NOA and EOA are discussed in the following (MDP modeling) section.

3.2 MDP Modeling

3.2.1 NOA's MDP components

Under normal operating conditions, the MG operator will try to minimize the operational cost by appropriately deploying the resources. To achieve the aforementioned task, the MG operator or (in this case) the NOA must have information about the states of these resources (along with some additional information). Therefore, the NOA's observation or state space consists of the information given by Equation (3.1).

For every hour or a time step t, the state values are updated to replicate the varying nature of these values over a 24-hr period.

$$S_t = \{P^{\text{PV}}_{t,m,p}, P^{\text{DG}}_{t,m,j}, \lambda^{\text{DG}}_{t,m,j}, E^{\text{cap}}_{t,m,k}, P^{\text{BES}}_{t,m,k}, \lambda^{\text{BES}}_{t,m,k}, P^{\text{load}}_{t}, P^{\text{G}}_{t}, P^{\text{loss}}_{t}, \lambda^{\text{G}}_{t}, V^{\text{count}}_{t}, \nu_t\} \tag{3.1}$$

The action space of NOA includes the control actions that operates the DG set-points and the charging/idling/discharging of the BESs as presented in Equation (3.2).

$$A_t = \{\alpha^{\text{DG}}_{t,m,j}, \alpha^{\text{BES}}_{t,m,k}\} \tag{3.2}$$

These action variables governs the amount power dispatched by the DGs and supplied to/by the BESs at any given time step by the NOA. Figure 3.2 shows the complete framework of the NOA. This framework also demonstrates how the transition between state S to S' is governed when the NOA takes a specific action.

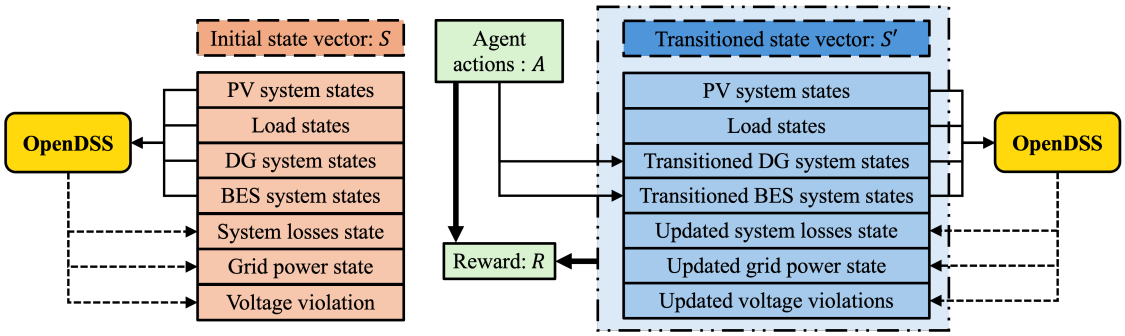

Figure 3.2: NOA's Framework

The initial states values for $P^{\text{PV}}_{t,m,p}$, $\forall\ m, p$ and P^{load}_{t} are obtained from the day-ahead forecasts. In order to account for the forecasting errors, these forecasts are integrated with the an error coefficient ϵ similar to the typical error values. This inclusion of ϵ into the day-ahead forecasts enhances the learning process of the NOA by providing a range from which the state values for the PV power output and load power can be sampled. One point to note is the policy learnt by the agent highly

depends on the states (training sample) it encounters during the training process. A biased training sample will point the agent towards developing a biased policy which will perform accurately for more frequent states whereas it will under-perform for less frequent states. Therefore as Equation (3.3) suggests, a uniform random value generator is used to sample these states.

$$S^{\text{PV,load}} = U\big([1-\epsilon]\ .S^{\text{forecast}}_{\text{PV,load}}, [1+\epsilon]\ .S^{\text{forecast}}_{\text{PV,load}}\big) \tag{3.3}$$

The DG power output and the BES status are kept at zero and idle, respectively. These initial values of the resources are the input to OpenDSS for power flow calculations. These calculations provide the P_t^{loss}, V_t^{count}, and P_t^{G}. Combining all these state values results in the initial state S. Once the agent takes actions A, the next state S' is achieved where the transition between S to S' is governed by the transition equations Equations (3.10), (3.18) and (3.19) which will be discussed in the problem formulation section. During normal operation, the states that demonstrates transition directly due to the actions are the DG power state and the states associated with BES. Similarly, these new state values are sent to OpenDSS engine for power flow calculations which in turn returns P_t^{loss}, V_t^{count}, and P_t^{G}. These state values along with the updated state values for the DG and BES form S' which is the transitioned state. Along with transitioned state, another important component that essentially determines how good the action taken by the agent is, is the reward value R. Again, a detail discussion about the reward function is presented in the problem formulation section.

3.2.2 EOA's MDP components

During the emergency conditions, caused due the extreme weather events, the NMGs might have to be operated in an islanded mode. Moreover, these conditions increase the possibility of multiple NMG resources to be disconnected from the net-

work due infrastructure failure, thereby creating a deficiency in the availability of power generating sources. Initiating rolling blackouts can be a possible solution, however it is pivotal to understand the criticality aspect of the loads being shed. Hence, the EOA is designed so that it can allocate power to the loads based on its criticality index.

Therefore, the state space for the EOA is given by Equation (3.4). One important state value that is unique to the EOA is the generation unit's status value. Each of the generating resources available in the NMG is accompanied by this status value that determines whether a specific resource is connected to the network or not.

$$S_t = \{u^{\mathrm{PV}}_{t,m,p}, P^{\mathrm{PV}}_{t,m,p}, u^{\mathrm{DG}}_{t,m,j}, P^{\mathrm{DG}}_{t,m,j}, u^{\mathrm{BES}}_{t,m,k}, E^{\mathrm{cap}}_{t,m,k}, P^{\mathrm{BES}}_{t,m,k}, P^{\mathrm{load}}_{t}, P^{\mathrm{allocated\_load}}_{t,x}, P^{\mathrm{M}}_{t}, P^{\mathrm{loss}}_{t}, V^{\mathrm{count}}_{t}, \nu_t\} \tag{3.4}$$

A typical distribution network contains a large number of loads and hence it would be an arduous task to allocate individual load. Hence, in this work, a multi-tier classification system is developed that categorizes the loads based on its criticality index. The loads will be categorized into multiple tiers which will take into account how critical a specific load is for the well-being of the community. The classification includes:

- Tier I loads: Loads that are critical for the community's safety and functionality
- Tier II loads: Loads that are critical and can provide emergency accommodation services
- Tier III loads: Medium/Large loads that are non-critical from the community service perspective
- Tier IV loads: Residential loads

A hospital load or a police station load can be considered as a Tier I loads whereas a supermarket or a warehouse load can be considered to be Tier II loads. An indus-

trial load which is relatively non-essential from the standpoint of community welfare, can be considered as Tier III loads. Tier IV loads mostly will consist of single or a multi-family homes or apartment complexes which has a lower criticality during emergency conditions. Thus the action space for the EOA can be represented by the Equations (3.5) and (3.6). In a distribution network, the number of residential loads is considerably higher than all other categories of loads. Hence the Tier IV loads can be further sub-divided into different zones depending on the topological location of these loads on the network, given by Equation (3.7).

$$A_t = \{\alpha_{t,x}^{\text{load}}\} \tag{3.5}$$

$$\alpha_{t,x}^{\text{load}} = \{\alpha_{t,\text{TierI}}^{\text{load}}, \alpha_{t,\text{TierII}}^{\text{load}}, \alpha_{t,\text{TierIII}}^{\text{load}}, \alpha_{t,\text{TierIV}}^{\text{load}}\} \tag{3.6}$$

$$\alpha_{t,\text{TierIV}}^{\text{load}} = \{\alpha_{t,\text{TierIV,ZONE\_1}}^{\text{load}}, \alpha_{t,\text{TierIV,ZONE\_2}}^{\text{load}}, \ldots, \alpha_{t,\text{TierIV,ZONE\_z}}^{\text{load}}\} \tag{3.7}$$

Each of the these tiers will have a minimum and maximum of load power that signifies the load flexibility of that tier. To determine load flexibility, we used the load consumption data published in [133]. We looked at different building type for example a hospital, or a warehouse and identified the amount of power consumed by each of the load types such as cooling load, fan load, lighting load etc. An educated assumption about which load types would be essential and non-essential during an emergency period was made. Equation (3.8) was used to get load flexibility factor which when multiplied with the base load of the specific building type will provide the load flexibility range. An example of estimating the load flexibility factor and load flexibility range of a specific building type is included in Appendix A.

$$LFI_{t,o,c} = \sum_{o=1}^{O} l_o.\varepsilon_o \tag{3.8}$$

Here, $LFI_{t,o,c}$ is the load flexibility index of c^{th} building type with l_o as the decimal

equivalent of portion of total load consumed by the o^{th} load type ($l_o \in [0, 1]$) and ε_o as the decimal equivalent of percent of essential loads present in the o^{th} load type ($\varepsilon_o \in [0, 1]$). So, the load flexibility range for a specific tier is essentially the sum of all the load flexibility ranges of the building types in that specific tier.

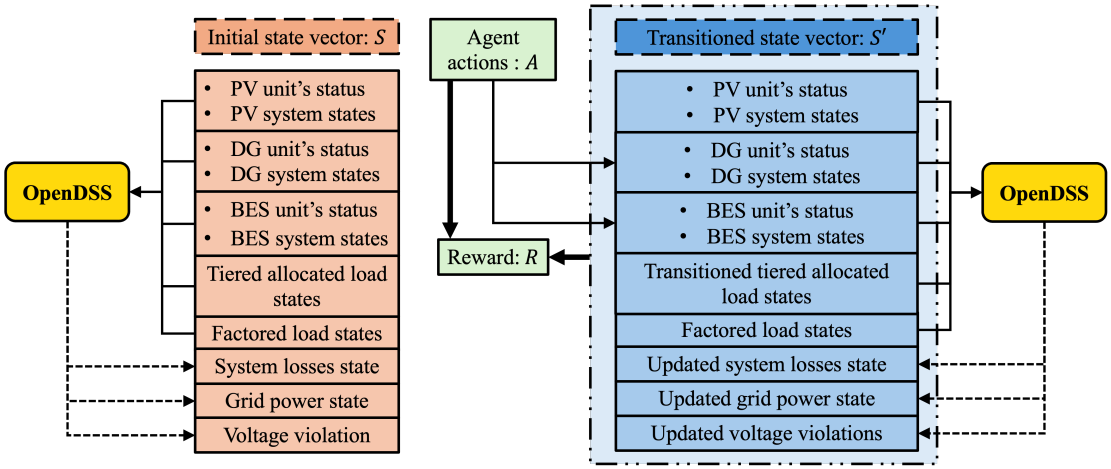

Figure 3.3: EOA Framework

The complete framework for the EOA is shown in Figure 3.3. The program flow of getting from initial state S to the transitioned state S' after taking action A is similar to that of the NOA. The difference is in some of the initial state values and the actions that the EOA take to ensure improvement in the CSI. As the action space for the EOA is different from that of the NOA the corresponding state transition equations are different as well. These transition Equations (3.29) and (3.30) will be discussed in the problem formulation section along with EOA's reward function.

3.3 Dual Agents Formulation

3.3.1 NOA's Formulation

As mentioned in the earlier section, the objective of the NOA is to minimize the operating of the interconnected MGs. Therefore the NOA's optimization problem can be formulates as a minimization problem given by Equation (3.9).

$$f_t^{\mathrm{NOA}} = \mathrm{minimize}\Big(P_t^{\mathrm{G}}.\lambda_t^{\mathrm{G}} + \sum_m \sum_j \left[P_{t,m,j}^{\mathrm{DG}}\left(\lambda_{t,m,j}^{\mathrm{DG}} + \lambda_{t,m,j}^{\mathrm{ENV-CO_2}}\right)\right] + \sum_m \sum_k \left[P_{t,m,k}^{\mathrm{BES}}.\lambda_{t,m,k}^{\mathrm{BES}} - \beta_{t,m,k}^{\mathrm{BES}}\right] + \sum_m \sum_n V_{t,m,n}^{\mathrm{count}}.\nu_t\Big) \tag{3.9}$$

In RL/DRL framework, the agent end goal is to maximize its reward. Hence, the objective function can be converted into the reward by simply taking a negative of the minimization problem i.e. $r_t = -f_t^{\mathrm{NOA}}$. A key component of the NOA's reward function is the time-dependent dynamic reward shaping variable β^{BES}. This variable is modeled in such a way that it can help the agent take the most appropriate BES actions at a given hour of the day. Moreover, this variable can capture the variation in the day-ahead forecasts and the BES states by appropriately adjusting its value, thus assisting agent's policy development. Use of such dynamically varying scalar quantity can help agent consider the actions that can have an impact on the future states as well in addition to the next states.

Equation (3.10) represents how the DG power is calculated while Equations (3.11) and (3.12) specify the operating limits of the DG units.

$$P_{t,m,j}^{\mathrm{DG}} = \alpha_{t,m,j}^{\mathrm{DG}}.P_{t,m,j}^{\mathrm{DG,max}} \tag{3.10}$$

$$P_{t,m,j}^{\mathrm{DG,min}} \leq P_{t,m,j}^{\mathrm{DG}} \leq P_{t,m,j}^{\mathrm{DG,max}} \tag{3.11}$$

$$\phi_{t,m,j}^{\mathrm{DG}} = 0.85 \tag{3.12}$$

Equation (3.13) estimates the social or the environmental cost associated with running a DG.

$$\lambda_j^{\mathrm{ENV-CO_2}} = f_j^{\mathrm{DG\_fuel}}.\gamma_{\mathrm{CO_2}}^{\mathrm{factor}}.\lambda_{\mathrm{SC-CO_2}}^{\mathrm{cost}} \tag{3.13}$$

One important point to note is that the learning algorithms (used for agent's training in this approach) require the action values to be within a range from $[-1, 1]$. Therefore, the raw action space for the DG actions is given by Equation (3.14). This requirement of the learning algorithm can be changed by using appropriate activation function in the output layer of the neural network. However, after multiple trial and errors it was determined that keeping the raw action space between $[-1, 1]$ is the best approach.

$$\alpha_{t,m,j}^{\text{DG\_raw}} \in [-1, 1] \tag{3.14}$$

However, during the actual deployment of the DGs these action values need to be converted into the ones ranging between $[0, 1]$. Thus, Equation (3.15) provides a way to rescale the action space of the DG to be between $[0, 1]$ as given in Equation (3.16) where a and b are the rescaling coefficients.

$$\begin{aligned} \alpha_{t,m,j}^{\text{DG\_raw}} \cdot a + b &= 0 \\ \alpha_{t,m,j}^{\text{DG\_raw}} \cdot a + b &= 1 \end{aligned} \tag{3.15}$$

$$\alpha_{t,m,j}^{\text{DG\_rescaled}} \in [0, 1] \tag{3.16}$$

So, Equation (3.10) becomes Equation (3.17) which is then used in the objective function to calculate the operating cost.

$$P_{t,m,j}^{\text{DG}} = \alpha_{t,m,j}^{\text{DG\_rescaled}} . P_{t,m,j}^{\text{DG,max}} \tag{3.17}$$

The amount of power supplied by the BES whether it will be for charging or discharging is given by Equation (3.18) whereas the BES capacity is updated using Equation (3.19). The raw action space for a BES system is given by Equation (3.20). For BES operation, the raw action space is the actual action space as it captures the charging and discharging aspect of a BES.

$$P^{\text{BES}}_{t,m,k} = \alpha^{\text{BES}}_{t,m,k} . P^{\text{BES,max\_rate}}_{t,m,k} \tag{3.18}$$

$$E^{\text{cap}}_{t,m,k} = E^{\text{cap}}_{t-1,m,k} + P^{\text{BES}}_{t,m,k} \cdot \Delta t \tag{3.19}$$

$$\alpha^{\text{BES}}_{t,m,k} \in [-1, 1] \tag{3.20}$$

In the formulation of the BES, we are considering 3 operating mode that are charging, discharging, and idling. In order for the agent to appropriately select the BES's operating mode, semi-discretization of the action space was performed which is given by Equation (3.21). For charging or discharging of the BES Equation (3.18) is used.

$$P^{\text{BES,max\_rate}}_{t,m,k} = \begin{cases} P^{\text{BES,max\_ch\_rate}}_{t,m,k}, & \text{if } \alpha^{\text{BES}}_{t,m,k} \geq 0.5 \\ 0, & \text{if } -0.5 < \alpha^{\text{BES}}_{t,m,k} < 0.5 \\ P^{\text{BES,max\_disch\_rate}}_{t,m,k}, & \text{if } \alpha^{\text{BES}}_{t,m,k} \leq -0.5 \end{cases} \tag{3.21}$$

Equations (3.22) and (3.23) provide the charging and discharging limit based on the available BES capacity or state of charge (SOC).

$$P^{\text{BES,max\_ch\_rate}}_{t,m,k} = \min\left(P^{\text{BES,max\_ch\_rate}}_{t,m,k}, \frac{E^{\text{max\_cap}}_{m,k} - E^{\text{cap}}_{t,m,k}}{\eta^{\text{BES,ch}}_{m,k}} \right) \tag{3.22}$$

$$P^{\text{BES,max\_disch\_rate}}_{t,m,k} = \min\Big(P^{\text{BES,max\_disch\_rate}}_{t,m,k}, \Big| \left(E^{\text{min\_cap}}_{m,k} - E^{\text{cap}}_{t,m,k} \right) \Big| . \eta^{\text{BES,disch}}_{m,k} \Big) \tag{3.23}$$

Other constraints that guide agent to develop a policy that provides a valid solution are given in Equations (3.24) and (3.25) which are the power balance and the voltage violation constraints, respectively.

$$P_t^{\mathrm{G}} + \sum_m \Big[\sum_j P_{t,m,j}^{\mathrm{DG}} + \sum_p P_{t,m,p}^{\mathrm{PV}} + \sum_k P_{t,m,k}^{\mathrm{BES,disch}} \Big] = \sum_m \Big[\sum_l P_{t,m,l}^{\mathrm{load}} + \sum_k P_{t,m,k}^{\mathrm{BES,ch}} \Big] + P_t^{\mathrm{loss}} \tag{3.24}$$

$$v_{t,m,n}^{\min} \leq v_{t,m,n} \leq v_{t,m,n}^{\max}$$
$$v_{t,m,n}^{\min} = 0.95\, p.u.; \quad v_{t,m,n}^{\max} = 1.05\, p.u. \tag{3.25}$$

The voltage constraints are enforced as soft constraints where if the solution provided by the agent violates the voltage limits, it will receive a penalty which is directly integrated into its reward function. This will ensure agent provides a solution that satisfies the network's operational requirements.

3.3.2 EOA's Formulation

The goal of the EOA is to maximize the critical load supplied during the emergency operating conditions. In addition to the critical load supplied, the EOA must also ensure the network operation constraints are met. Therefore, the optimization objective of the EOA is modeled by Equation (3.26) whose negative will give the reward function for the EOA, i.e. $r_t = -f_t^{\mathrm{EOA}}$.

$$f_t^{\mathrm{EOA}} = \mathrm{minimize}\Big(f(P_t^{\mathrm{M}}) + \sum_m \sum_x \lambda_{t,m,x}^{\mathrm{load}} . P_{t,m,x}^{\mathrm{allocated\_load}} + \sum_m \sum_n V_{t,m,n}^{\mathrm{count}} . \nu_t + NLS_t^{\mathrm{load}} \Big) \tag{3.26}$$

The function $f(P_t^{\mathrm{M}})$ is the mismatch penalty function which given by Equation (3.27). The mismatch power P_t^{M} is calculated using the Equation (3.28)

$$f(P_t^{\mathrm{M}}) = -\Big| \frac{\left(\lambda_t^{\mathrm{under}} - \lambda_t^{\mathrm{over}}\right)}{2} . \left|P_t^{\mathrm{M}}\right| + \frac{\left(\lambda_t^{\mathrm{under}} + \lambda_t^{\mathrm{over}}\right)}{2} . P_t^{\mathrm{M}} \Big| \tag{3.27}$$

$$P_t^{\mathrm{M}} = \min\Big(\sum_m \sum_s P_{t,m,x}^{\text{actual\_load}}, \sum_m \sum_d P_{t,m,d}^{\text{DER\_available}}\Big) - \sum_m \sum_x P_{t,m,x}^{\text{allocated\_load}} \tag{3.28}$$

NLS_t^{load} is the no load supply penalty which gets allocated if the EOA doesn't allocate any load even when there is enough generation to at least supply minimum allowed load of a specific tier. Another important aspect of EOA's formulation is it's ability to consider load criticality while allocating the loads. Therefore in the objective function formulation, a load supply penalty factor $\lambda^{\text{load}} = \frac{1}{C^{\text{load}}}$ is introduced which is essentially an inverse of the criticality of a specific load tier.

Like NOA's BES action values, the EOA's action space is semi-discretized using Equation (3.29). The load allocation function given in Equation (3.30) is used to allocate power to the tiered loads.

$$P_{t,m,s}^{\text{allocated\_load}} = \begin{cases} P_{t,m,x}^{\text{load\_min}}, & \text{if } \alpha_{t,m,x}^{\text{load}} < 0.25 \\ P_{t,m,x}^{\text{load\_min\_allowed}}, & \text{if } 0.25 \leq \alpha_{t,m,x}^{\text{load}} < 0.5 \\ f\left(P_{t,m,x}^{\text{load\_limit}}\right), & \text{if } \alpha_{t,m,x}^{\text{load}} \geq 0.5 \end{cases} \tag{3.29}$$

$$f\left(P_{t,m,x}^{\text{load\_limit}}\right) = \Big[P_{t,m,x}^{\text{load\_min\_allowed}} + \alpha_{t,m,x}^{\text{load}} \cdot \left(P_{t,m,x}^{\text{load\_max}} - P_{t,m,x}^{\text{load\_min\_allowed}}\right) \Big] \tag{3.30}$$

Additional constraint to which EOA is subjected to is the voltage constraint given by Equation (3.25). The power that is allocated to the load either lies between a specified range for that particular tier or it is set to zero. This may lead to higher mismatch values under scenarios where there is generation available but not enough to meet the minimum allowable load of any of the load tiers. Therefore CSI, which is given by Equation (3.31), along with the mismatch is used to evaluate the EOA's

performance.

$$CSI_t = \frac{\sum_x C_{t,x}^{\text{load}} . P_{t,x}^{\text{allocate\_load}}}{\sum_x C_{t,x}^{\text{load}} . P_{t,x}^{\text{load\_max}}} \tag{3.31}$$

CSI serves as a metric to quantify the critical load supply during an emergency condition that weighs the load allocated with respect its criticality.

3.4 Dual Agents Training

3.4.1 Training Data

The proposed framework implements a decoupled training approach where each DA is trained independently as the operating environment of each of these DAs vary. The training of each of the DAs is performed for every hour of the day (or time step) where the end of the training checkpoints are stored in a local directory. Each of these checkpoints contain information about the trained model's weights and structure. These checkpoints are then recalled at the time of real-time application during a specific hour of the day. The complete training paradigm is presented in the flowchart represented by 3.4.

For instance: Let T ($T = 24$h) be the total number of hours in a day with t being the index for each hour of day. Consider f_a^t and f_e^t to be the checkpoints stored at t at the end of the training for the respective agents (index a for NOA and index e for EOA) where $f_a^t \in F_a$ and $f_e^t \in F_e$. Here F_a and F_e are the set of checkpoints for the NOA and EOA respectively, stored in the directory. During the real-time application at hour t, depending on the operating conditions, appropriate checkpoint (either f_a^t or f_e^t) will be called. Operating conditions or states observed at t will be provided as an input to the model called from the checkpoint to initiate necessary control actions.

3.4.2 Episode Termination

A one shot episode termination approach was developed which allows for training the agent for every time-step. Therefore, at the time of training the learning algorithm

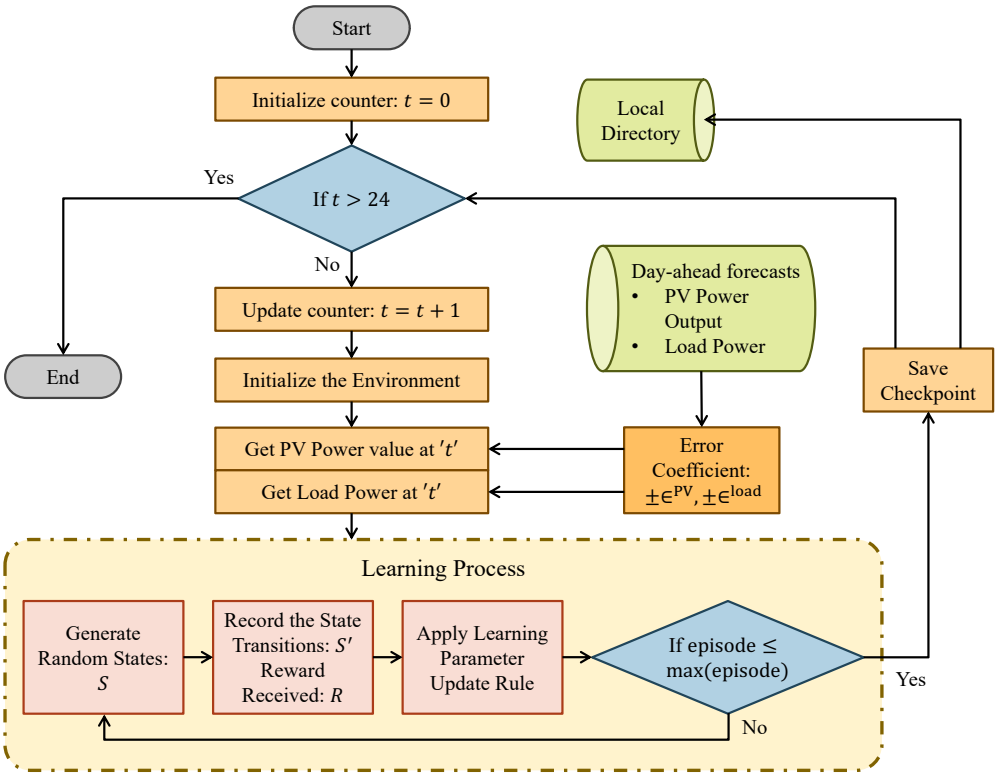

Figure 3.4: Training Flowchart

will collect samples that will consist of initial state s, action a, transitioned state s', and the reward r generating a tuple (s, a, s', r). The learning algorithm used to train the agent uses Q-value estimation to determine whether a specific action is better than the other at given state. This Q-value is essentially ascertained by Equation (3.32).

$$Q(s,a) \leftarrow Q(s,a) + \alpha.[r(s,a) + \gamma. \max_{a'} Q(s',a') - Q(s,a)] \tag{3.32}$$

Here, α is the learning rate whereas $Q(s,a)$ is the Q-value associated with the state-action pair. As discussed earlier, in DRL framework this value is estimated by DNNs. Since, in the proposed MDP a one shot episode termination criteria is developed, the transitioned state is the terminal state of that episode. If the transitioned state is the terminal state, the Q-value associated with that state is considered to be zero i.e. $\max_{a'} Q(s',a') = 0$. Therefore, Equation (3.32) can be modified as Equations (3.33)

and (3.34) which signify preference to the immediate returns.

$$Q(s,a) \leftarrow Q(s,a) + \alpha.[r(s,a) - Q(s,a)] \tag{3.33}$$

$$Q(s,a) \leftarrow (1-\alpha).Q(s,a) + \alpha.r(s,a) \tag{3.34}$$

Let us look at how the one-shot episode termination MDP works from an actor-critic setup perspective as those are the set of algorithms used to train the agents in this dissertation. In an actor-critic set-up, actor network takes actions and the critic network criticizes those actions by specifying how good those actions are. Additionally there are two other network namely, target actor and target critic networks used that essentially provide stability to the training process.

The critic network updates its network parameters by minimizing the loss function given in Equation (3.35)

$$L(\theta^Q) = \Big(y - Q(s,a|\theta^Q)\Big)^2 \tag{3.35}$$

Here, θ^Q are the critic network parameters and y is given by Equation (3.36).

$$y = r(s,a) + \gamma.Q(s',a'|\theta^{Q'}) \tag{3.36}$$

The $Q(s',a'|\theta^{Q'})$ is determined by the target critic network where $\theta^{Q'}$ are the target critic network parameters. However as the transitioned state s' is the terminal state, $Q(s',a'|\theta^{Q'}) = 0$ (as explained earlier). Therefore, Equation (3.36) essentially contains the $r(s,a)$ value which is the immediate rewards the agent receives by taking action a given a state s. Now, the updated loss function is given by Equation (3.37).

$$L(\theta^Q) = \Big(r(s,a) - Q(s,a|\theta^Q)\Big)^2 \tag{3.37}$$

The actual update of the critic network parameters is performed using gradient descent on the loss function which is given by Equation (3.38).

$$\theta^Q \leftarrow \theta^Q - \alpha_{\mathrm{Q}}.\nabla_{\theta^Q} L(\theta^Q) \tag{3.38}$$

Thus, the critic network update is not only influenced by the immediate reward which is reflected in the loss function (Equation (3.37)) but also the critic learning rate α_{Q} (which also represented as α_{critic} in the later sections).

Similarly, the actor network which updates its policy gradient through Equation (3.39) incorporates the gradient of the Q-value ($Q(s, a|\theta^Q)$) estimated and subsequently updated by the critic network using Equations (3.35) to (3.38).

$$\nabla_{\theta^\mu} J \approx \nabla_a Q(s, a|\theta^Q).\nabla_{\theta^\mu} \mu(s|\theta^\mu) \tag{3.39}$$

Here, $\mu(s|\theta^\mu)$ is the action value generated by the actor network given a state s and θ^μ are the actor network parameters which are updated by using gradient ascent due reward maximization as given in Equation (3.40).

$$\theta^\mu \leftarrow \theta^\mu + \alpha_\mu.\nabla_{\theta^\mu} J \tag{3.40}$$

Again, it can be observed that α_μ which is the actor learning rate (represented by α_{actor} in later sections) along with α_{Q} plays a key role in the model's parameter update and consequently the agent's policy development (this can be verified in the Sensitivity Analysis section).

So without loss of generality, it can be said that the one-shot episode termination criteria prioritizes the immediate returns as it eliminates the discount factor γ and the Q-value of future states ($Q(s', a')$). However, the mathematical basis of this approach may vary depending on the type of learning algorithm used. Nonetheless, the idea that

immediate returns are prioritized remains unchanged. Thus, this episode termination criteria is suitable for the DA-framework and the training paradigm being proposed in this dissertation.

3.5 Dual Agents Testing

To test the DAs, the saved training checkpoint (for a specific hour of day) is loaded. Real-time operating conditions are sent as an input to the trained actor network model. These operating condition are treated as the initial state values that the agent encounters $\rightarrow S$. Observing these values the agent instantly outputs the best action values that it has learnt during the learning/training process. These action values either include the DERs set-points or the load allocation values, depending on the agent being deployed. Figure 3.5 shows the complete interaction between real-time states and the control action generation.

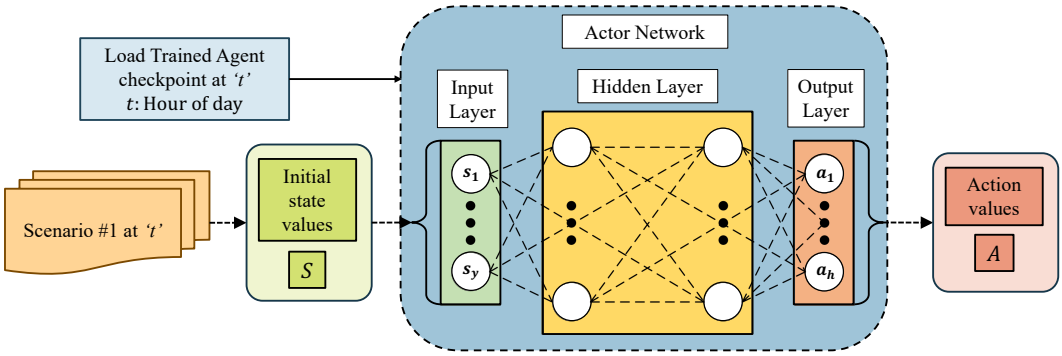

Figure 3.5: Testing Framework

Additionally, the transitioned state values $\rightarrow S'$ after the agent takes necessary actions, although not mentioned in the Figure 3.5, are tracked and sent to OpenDSS as well, to verify the satisfaction of the operational constraints.

3.6 Test Setup

A modified IEEE-123 node test feeder equipped with multiple interconnected MGs was used as the test setup. Each of these MGs consist of multiple DERs. Figure 3.6 shows the boundary of the five NMGs and the location and type of the DERs

connected to the network. In total there are 27 DERs (when all five MGs are interconnected) out of which 14 are solar PV systems, 6 are BES systems, and 7 are the DG systems. Additionally, the Tier IV zones for EOA's testing are also presented.

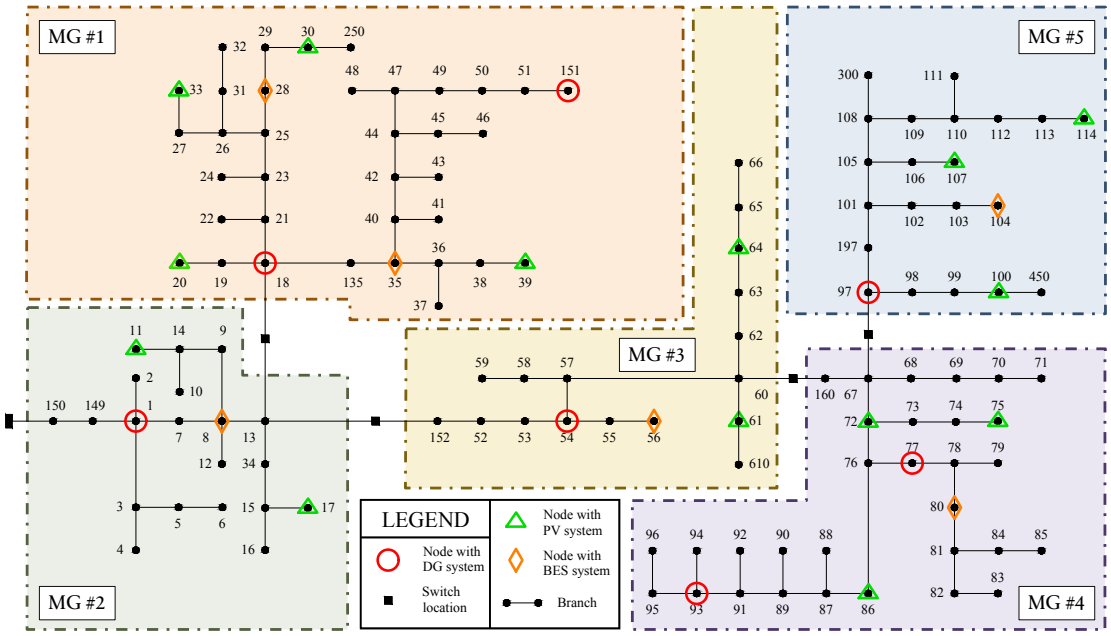

Figure 3.6: Modified IEEE-123 Node Test system

The ratings of each of the DERs are presented in Tables 3.1 and 3.2. The discharge rate (equal to charge rate) of the BESs is determined in such a way that the BES is (at its rated capacity) able to supply power continuously for 12-hrs. The day-ahead forecasts for the PV system output were generated using NREL PVWatts tool, whereas (as mentioned in the previous section) [133] was used to develop the load forecasts.

The operating cost for the DG #1 of MG #1, and all DG units of MG #2, #4 and #5 is considered to be 0.3 \$/kWh while that of DG #2 of MG #1 and DG unit of MG #3 is assumed to be 0.34 \$/kWh. To determine the levelized cost of energy (LCOE) for PV systems, NREL System Advisor Model (SAM) was used. The LCOE was calculated to be 0.0859 \$/kWh.

To provide a time varying cost signal for grid power import, a time-of-use rate

Table 3.1: DER Ratings (MG #1 - #3)

MG #1				**MG #2**		**MG #3**	
DER	**Rating**	**DER**	**Rating**	**DER**	**Rating**	**DER**	**Rating**
PV #1	400 kW	DG #1	750 kW	PV #1	400 kW	PV #1	200 kW
PV #2	300 kW	DG #2	500 kW	PV #2	350 kW	PV #2	150 kW
PV #3	250 kW	BES #1	480 kWh	DG #1	500 kW	DG #1	500 kW
PV #4	150 kW	BES #2	600 kWh	BES #1	720 kWh	BES #1	600 kWh

Table 3.2: DER Ratings (MG #4 and #5)

MG #4				**MG #5**			
DER	**Rating**	**DER**	**Rating**	**DER**	**Rating**	**DER**	**Rating**
PV #1	200 kW	DG #2	300 kW	PV #1	100 kW	BES #1	600 kWh
PV #2	250 kW	BES #1	480 kWh	PV #2	300 kW		
PV #3	100 kW			PV #3	250 kW		
DG #1	500 kW			DG #1	300 kW		

(TOU) was considered. The TOU period and the rates used for the case-studies are presented in Table 3.3 which were obtained from the local utility company. These TOU rates and periods along with load and PV forecasts used for the case studies were for the city located at the east coast of the United States.

Table 3.3: TOU period and rates

Summer Months			
	Discount period	**Off-peak period**	**On-peak period**
TOU period	1AM-6AM	6AM-6PM; 9PM-1AM	6PM-9PM
TOU Rate	0.06814 $/kwh	0.10467 $/kWh	0.27653 $/kWh
Winter Months			
	Discount period	**Off-peak period**	**On-peak period**
TOU period	1AM-3AM; 11AM-4PM	3AM-6AM; 9PM-11AM; 4PM-1AM	6AM-9AM
TOU Rate	0.06814 $/kwh	0.10467 $/kWh	0.27653 $/kWh

CHAPTER 4: RESULTS

Multiple case studies were run using the test setup described in the previous section. First two case studies involve two interconnected MGs, namely MG #1 and #2. In the subsequent case studies, all five MGs are interconnected which encompasses the complete IEEE 123 node test network. To provide varying operating conditions, a TOU rate and period for summer and winter months is used for NOA's performance evaluation. For the EOA's testing, it was assumed that the NMGs are operating in an islanded mode as consequence of a transmission level event that essentially disconnected the distribution feeder from the main grid. Additionally, we considered the NMG to be either in an abnormal condition or an extreme condition. An abnormal condition occurs when majority of the DERs are available while an extreme condition occurs when most of the DERs are unavailable. The unavailability of the DERs can be a result of infrastructure damage caused due to extreme weather events. This can be ascertained by using fragility curves or prediction based algorithms, however in this work, these scenarios are generated randomly (using uniform distribution) as it helps the EOA learn and develop a policy for a multitude of scenarios.

4.1 Layout I.A: Two NMGs for NOA testing

In this section, Layout I.A described in the Figure 4.1 was used to test the NOA. Only MG #1 and #2 are interconnected for this case study. For the evaluation of the NOA, the time varying grid power cost signal used for the summer and winter months is presented in Table 3.3. Therefore the case studies that test the NOA performance are:

1. NOA Operation on a Summer Day

2. NOA Operation on a Winter Day

There are 12 DERs connected in Layout I.A which includes 6 PV systems, 3 BES systems, and 3 DG units. The information regarding the location of these DERs and its rating is provided in Figure 4.1 and Table 3.1 (column MG #1 and MG #2), respectively.

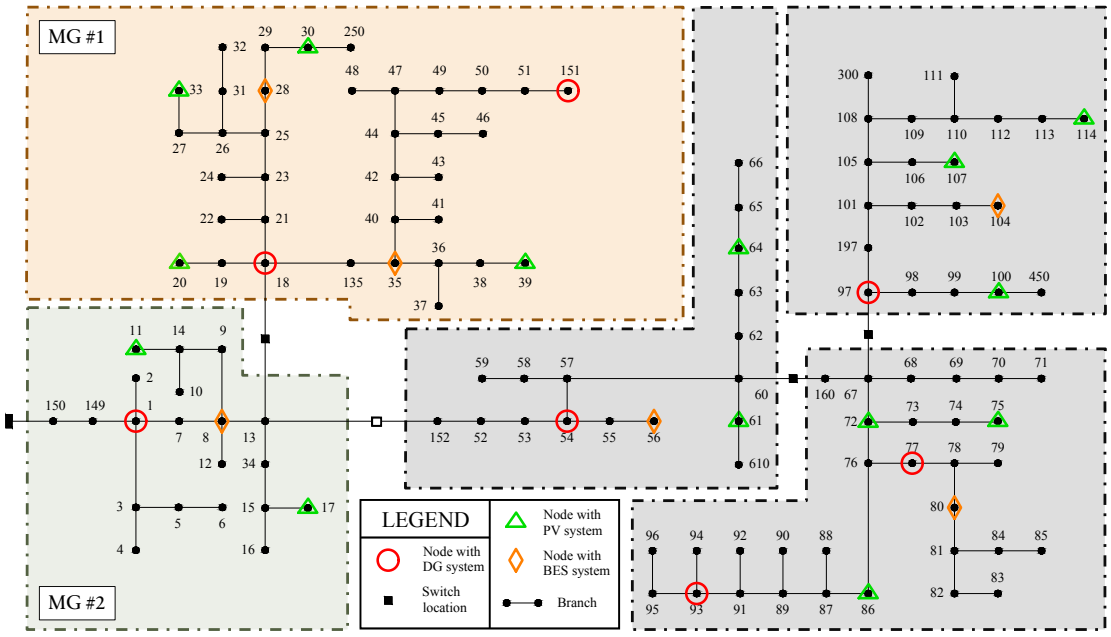

Figure 4.1: NMG #1 - #2 setup for NOA

4.1.1 Case I.A.1: NOA Operation on a Summer Day

The performance of the NOA for a day in a summer month is presented in this section. Figure 4.2 shows the overall operation of the N-MGs over the 24-hr period. The NOA is responsible to charge/discharge/idle the BES and/or deploy the DG units in an economical way. The BES capacity variation over a day is also shown in Figure 4.2. It can be seen that the NOA charges the BES only during the discount period or when the total PV generation is higher than the load whereas it discharges the BES only during the on-peak period. For the off-peak period the BES remains in an idle state.

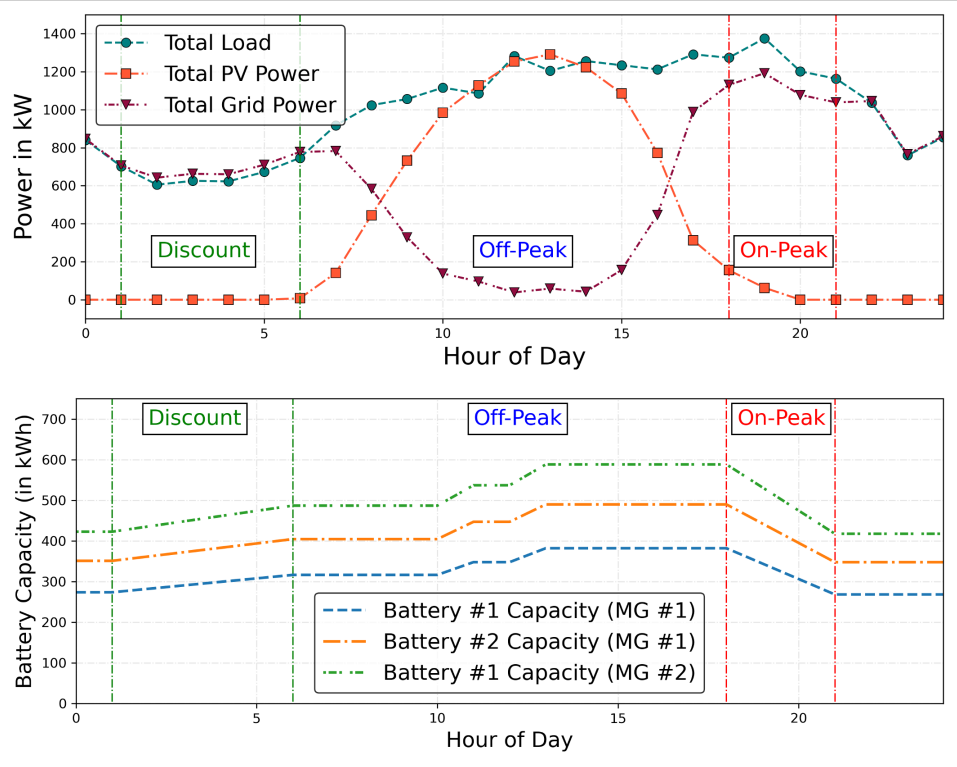

Figure 4.2: NMG #1 - #2 Operation (Summer Month)

From the NOA's action values plot in Figure 4.3, the aforementioned behavior can be confirmed. At around 11AM, the total PV generation is higher than the total load during off-peak period. Therefore, the NOA decides to charge (even in off-peak period) the BES while at 12PM the total PV generation becomes lower than the total load which is addressed by setting the BES in an idle mode. Similarly, at 1PM, when the total PV generation again becomes higher than the total load, the NOA charges the BES. The charging of the BES during the discount period is done at the minimum allowable charge rate signified by the action value between 0.5 to 0.75. However if the charging action occurs when the total PV generation exceeds the total load, the BES is charged at the maximum charging rate to minimize the power export to the grid. Furthermore, it was observed that at no point in time over the 24-hr period does the NOA deploys any DG units.

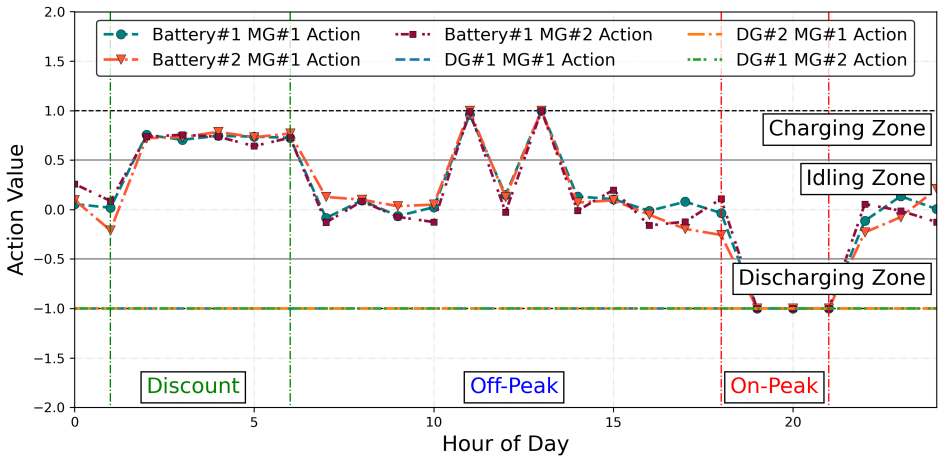

Figure 4.3: Agent's actions NMG #1 - #2 (Summer Month)

4.1.2 Case I.A.2: NOA Operation on a Winter Day

To further test the NOA's performance, it was deployed online to provide necessary control during a day in a winter month. The NOA's overall operation, presented in Figure 4.4 and its action values, presented in Figure 4.5 demonstrates a similar behavior as observed during the summer months of testing. The difference is the period during which the NOA is charging/discharging/idlying the BES as the TOU periods are different during winter months. Nonetheless, the NOA charges the BES (at its minimum charge rate) only during discount period, discharges during on-peak period and idles during off-peak period.

During the winter months of testing, the total PV generation was consistently lower than the total load. Hence the NOA only charges the BES during the discount period. Important point to note is that the reward function of the NOA is integrated with a dynamic variable that can appropriately provide a reward/penalty signal that enables BES charging at it's rated capacity (even during the discount period) under low state of charge conditions.

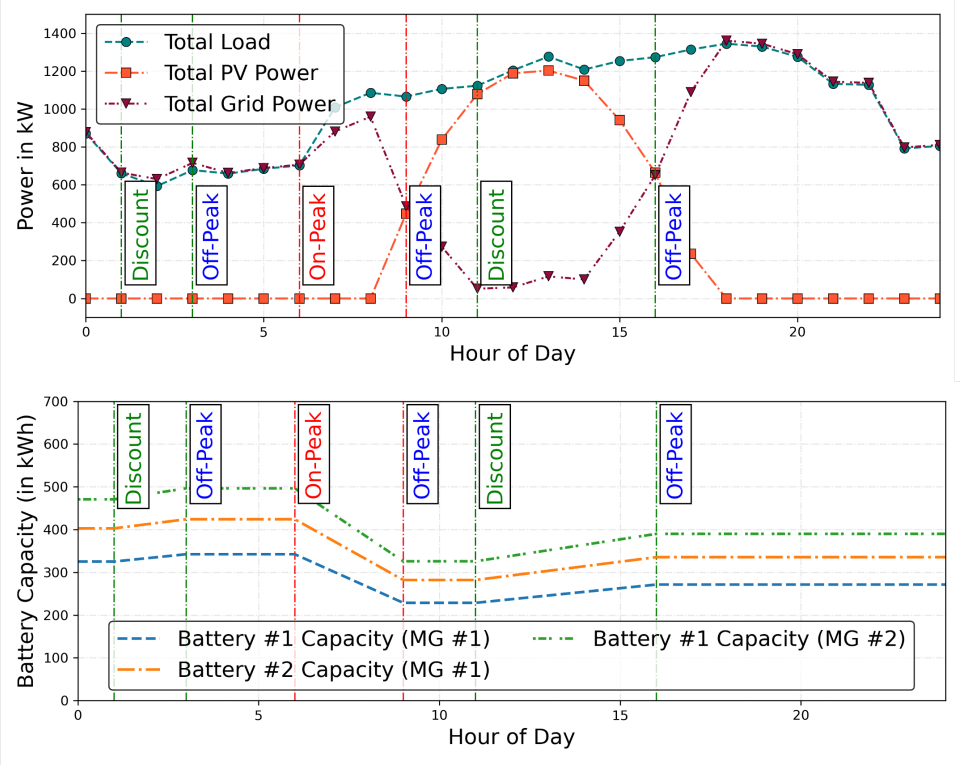

Figure 4.4: NMG #1 - #2 Operation (Winter Month)

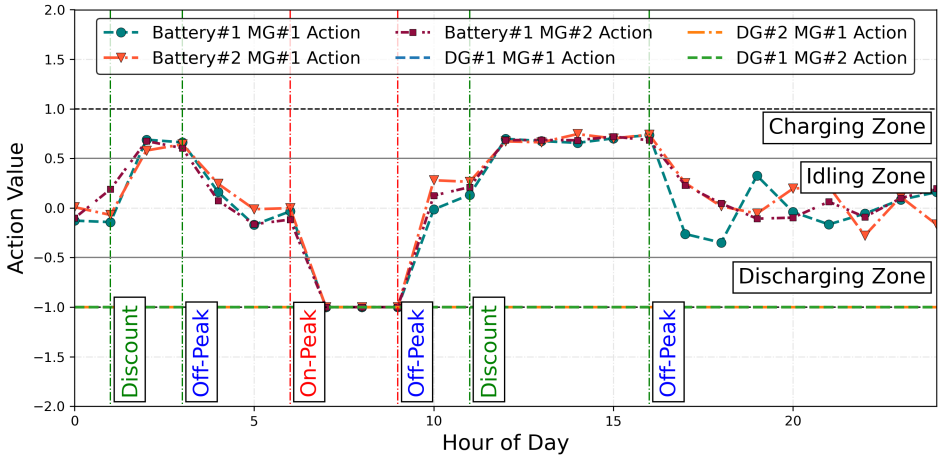

Figure 4.5: Agent's actions NMG #1 - #2 (Winter Month)

4.2 Layout I.B: Two NMGs for EOA testing

In this section, the testing of the EOA is performed for a 10-hr outage starting at 7AM and lasting till 5PM during which the NMGs are operating in an islanded mode. As described in the earlier sections, in order to aid in the load allocation process, the NMG setup is divided into multiple tiers based the load criticality. Moreover, the Tier IV loads mainly residential loads, are further divided into different zones depending on its topological location. Figure 4.6 shows the different Tier IV zones that were considered in this part of the case study. In total there are four Tier IV zones with an action value associated with each zone.

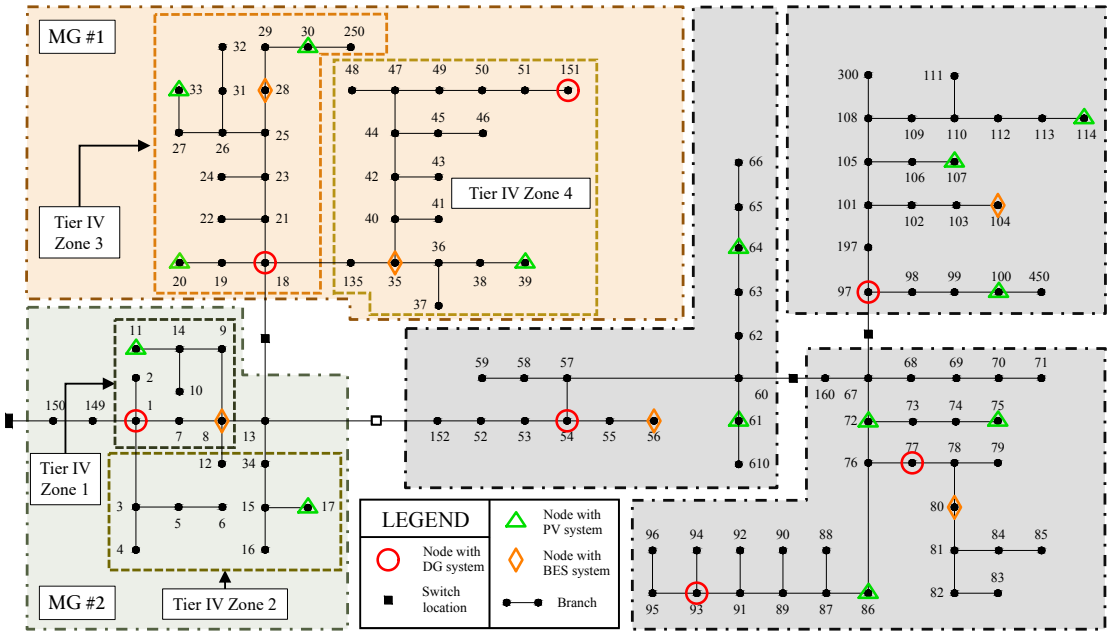

Figure 4.6: NMG #1 - #2 setup for EOA

The islanded NMGs were operated under abnormal and extreme conditions for a summer and a winter day, respectively. The list of case studies performed to test the EOA's critical load allocation ability are:

- EOA Operation on a Summer Day under Abnormal Conditions
- EOA Operation on a Summer Day under Extreme Conditions

- EOA Operation on a Winter Day under Abnormal Conditions
- EOA Operation on a Winter Day under Extreme Conditions

The available DER power depends on the operating conditions. Under abnormal conditions, majority of the time the available generator power will be more than the total available load (even though some of the DERs will be out of service) whereas under extreme conditions, the available generator power will almost always be less than the total available load.

4.2.1 Case I.B.1a: EOA Operation on a Summer Day under Abnormal Conditions

The available DERs during abnormal condition are highlighted in Figure 4.7 that suggests PV #1, #4, DG #1 and BES #1 from MG #1 and PV #2 and DG #1 from MG#2 are available to supply power to the loads.

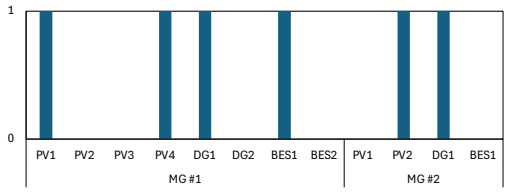

Figure 4.7: NMG #1 - #2: DER status information under abnormal conditions (Summer Month)

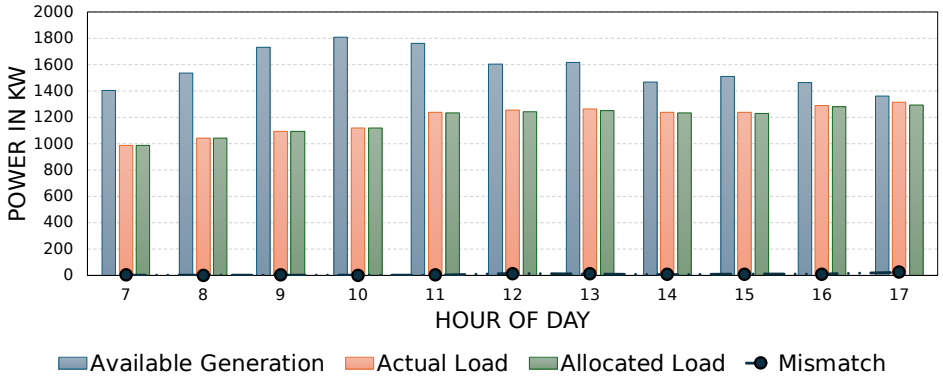

Figure 4.8: NMG #1 - #2: Load and generation information under abnormal conditions (Summer Month)

The load and generation information is provided in Figure 4.8. Since majority

of the generating units are available (i.e. generation is higher than the load), the mismatch value (given by Equation (3.27)) is almost equal to zero. Figure 4.9 shows the tiered allocated load and the CSI information. The CSI (given by Equation (3.31)) observed during the abnormal conditions is almost always 1 which means that all the loads (including the critical and non-critical loads) are being supplied by the EOA.

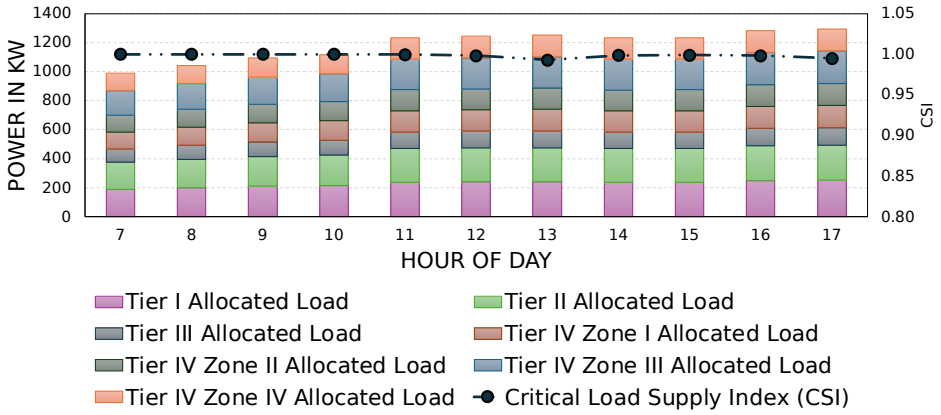

Figure 4.9: NMG #1 - #2: Tiered allocated load and CSI information under abnormal conditions (Summer Month)

4.2.2 Case I.B.1b: EOA Operation on a Summer Day under Extreme Conditions

Under extreme conditions on summer day, fewer generating units are available to supply the load which creates a generation deficit. Due to this deficit, the EOA will have to shed some loads while maintaining a higher level of CSI. A higher CSI means more critical loads are being supplied during the outage. Figure 4.10 shows the status of the available DERs. Here, only PV #1, DG #1, BES #1 and BES 2 from MG 1 and PV 2 from MG 2 are online.

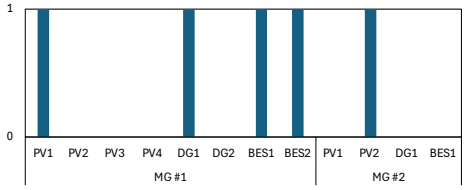

Figure 4.10: NMG #1 - #2: DER status information under extreme conditions (Summer Month)

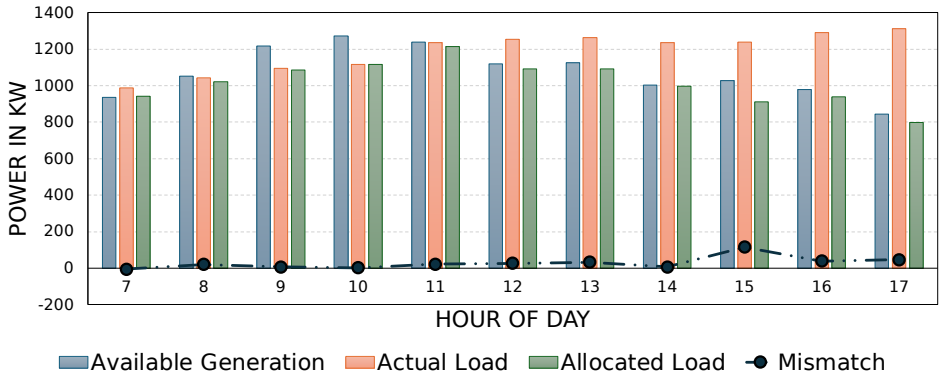

Figure 4.11: NMG #1 - #2: Load and generation information under extreme conditions (Summer Month)

The generation and load information along with the mismatch information is provided in the Figure 4.11. It can be seen that the mismatch value is relatively higher under extreme conditions. As explained earlier (in EOA's problem formulation section), this can be attributed to a situation where partial load power can be supplied, however, this partial load power is lower than the minimum allowable load power for that specific load tier.

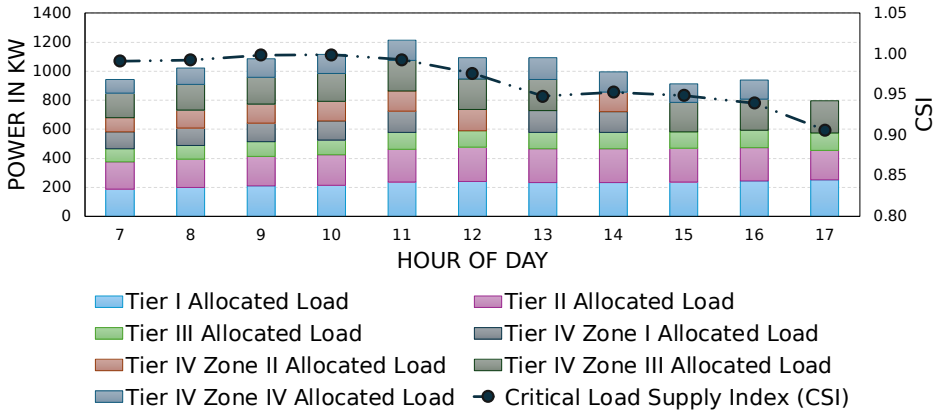

Figure 4.12: NMG #1 - #2: Tiered allocated load and CSI information under extreme conditions (Summer Month)

The tiered allocated load and the CSI information is presented in Figure 4.12. The total load and generation value is almost equal at the beginning of the outage. Hence the CSI value is close to 1 with the highest value observed to be 0.9987 at 10AM. The lowest CSI of 0.9063 was observed at around 5PM when the PV generation

is considerable low and the total available load is high. Nonetheless, Figure 4.12 highlights the fact that the EOA only sheds the Tier IV loads which have the least criticality.

4.2.3 Case I.B.2a: EOA Operation on a Winter Day under Abnormal Conditions

Similar to the Case I.B.1a, an abnormal condition was simulated for an outage that occurred on a winter day. The available DERs for supplying power to the load are presented in Figure 4.13. MG #1 has PV #3, PV #4, DG #1, DG #2, and BES #1 available whereas MG #2 has PV #1, PV #2, and BES #1 connected to the network.

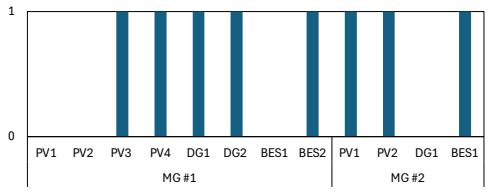

Figure 4.13: NMG #1 - #2: DER status information under abnormal conditions (Winter Month)

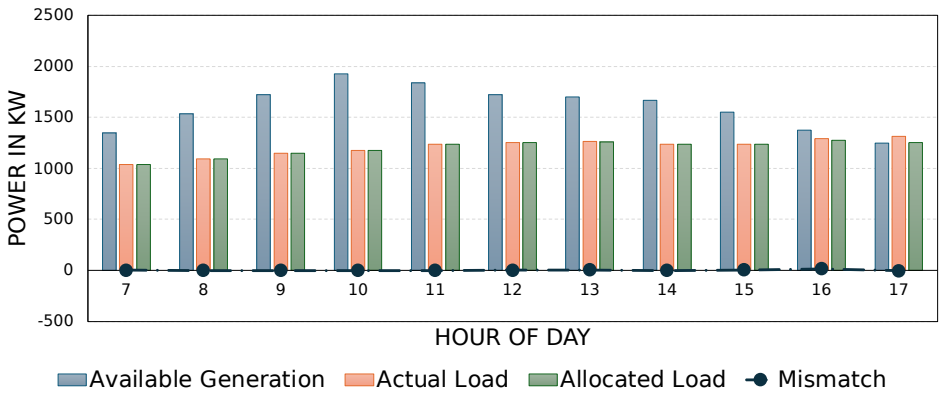

Figure 4.14: NMG #1 - #2: Load and generation information under abnormal conditions (Winter Month)

The load and generation information can be seen in Figure 4.14 which depicts the mismatch value equal to zero over the outage period (similar to what was observed in Case I.B.1a). Likewise, the CSI was observed (from Figure 4.15) to close to 1 for

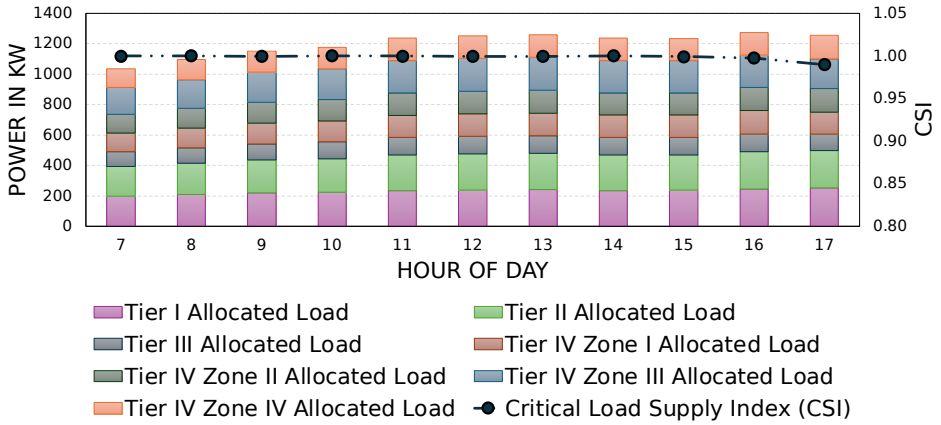

Figure 4.15: NMG #1 - #2: Tiered allocated load and CSI information under abnormal conditions (Winter Month)

all the hours except one which is at 5PM. During this hour, the total generation was slightly lower than the load hence the EOA shed some loads making the CSI equal to 0.9896. Nonetheless, this value is still close to one signifying the power allocation to all the critical loads.

4.2.4 Case I.B.2b: EOA Operation on a Winter Day under Extreme Conditions

In this case, only PV #1, DG #1, and BES #1 from MG #1 and PV #1 from MG #2 are available to supply the loads as seen from Figure 4.16. The generation and load information plot (Figure 4.17) shows difference between the total generation power and the total load available over the outage period. Similar to the Case I.B.1b, because of the power deficit and the tiered load classification structure, the mismatch value is slightly higher compared to the abnormal conditions case (winter day).

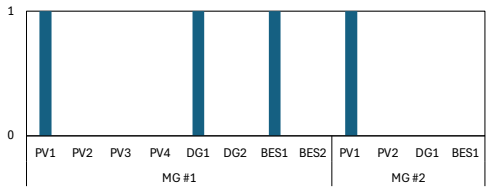

Figure 4.16: NMG #1 - #2: DER status information under extreme conditions (Winter Month)

From Figure 4.18 it can be seen that initially, the CSI value was lower. However

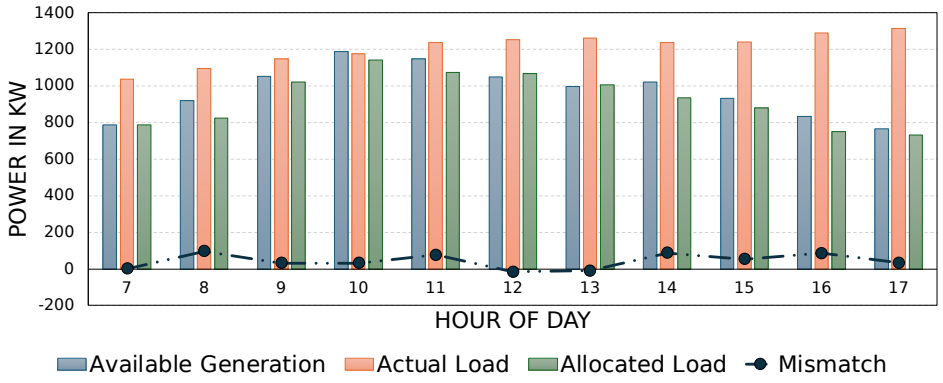

Figure 4.17: NMG #1 - #2: Load and generation information under extreme conditions (Winter Month)

as the total PV generation starts to increase, the CSI value increases reaching its maximum of 0.9934 at 10AM over the outage period. Moreover, as the total PV generation lowers in the later stages of the outage, the CSI value decrease highlighting load shedding initiated by the EOA. Despite that, the EOA is able to maintain a high CSI value over 10-hr outage period with the lowest CSI of 0.9109 observed at 5PM.

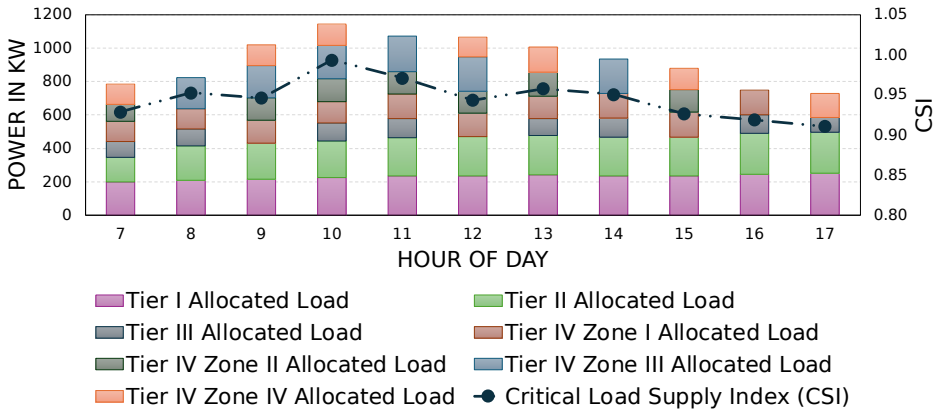

Figure 4.18: NMG #1 - #2: Tiered allocated load and CSI information under extreme conditions (Winter Month)

4.3 Layout II.A: Five NMGs for NOA testing

In this section of the case study, all five MGs are interconnected as shown in Figure 4.19 for the NOA's testing. The cases run in this section similar to that run for NOA's testing with two NMGs setup.

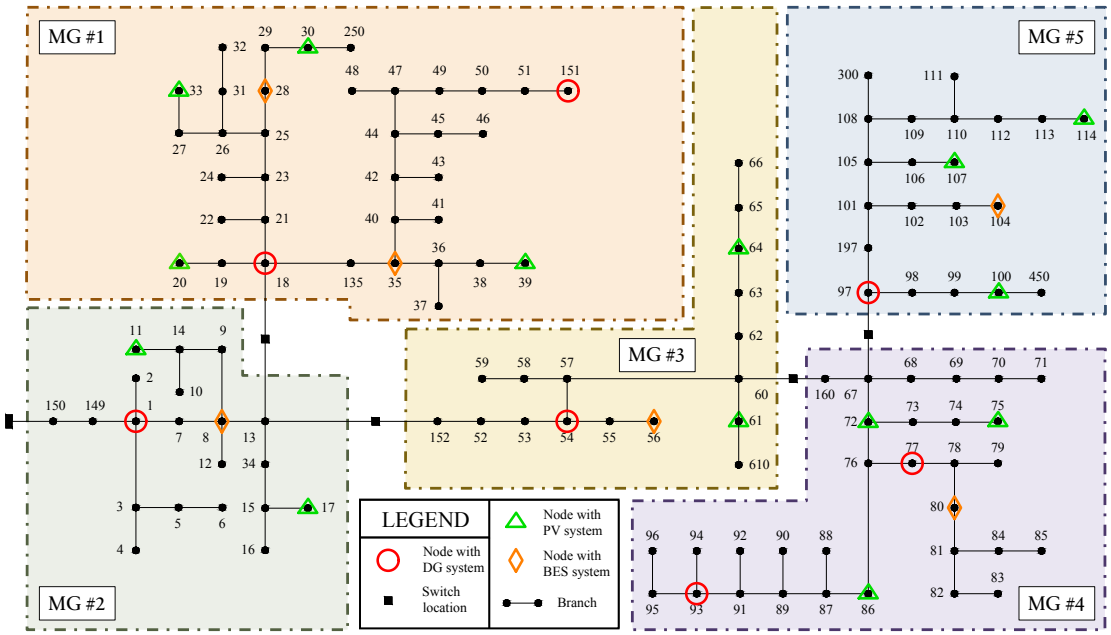

Figure 4.19: NMG #1 - #5 setup for NOA

The difference being in the number of DERs considered. In this layout, there are in total 27 DERs out of which 14 are PV systems, 6 are BES systems, and 7 are DG units. The location of the DERs can be seen in Figure 4.19 as well while the details regarding the ratings of each of these units are provided in Tables 3.1 and 3.2.

4.3.1 Case II.A.1: NOA Operation on a Summer Day

The NOA's operation along with the BES system's capacity variation in a five NMG setup for summer day is presented in Figure 4.20. It was observed (similar to the two NMGs setup) that the TOU govern the NOA's action where the BES charging, discharging, and idling are a consequence of the discount, on-peak, and off-peak TOU periods, respectively.

Important point to note is that due to numerous voltage violations observed in later hours of the day (4PM to 8PM), the NOA decides to dispatch DGs to eliminate these violations. Table 4.1 provides information about the number of voltage violations before and after the NOA takes actions (for 6PM and 7PM). The detailed table showing the voltage violations for all hours from 4PM to 8PM is presented in

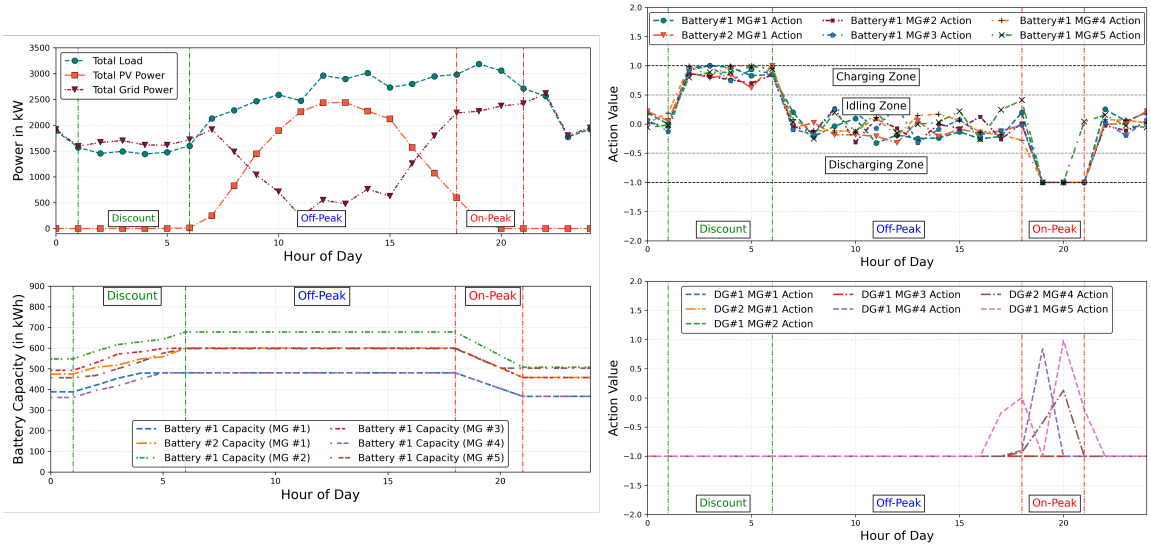

Figure 4.20: NMG #1 - #5 Operation and Agent's actions (Summer Month)

Appendix B (Table B.1). These violations can be attributed to the increase in load while the PV generation decrease. Additionally, apart from the main grid, the DGs are the only source of reactive power support in the N-MG setup. Hence, the NOA decides to deploy the DGs to satisfy the operational requirements of the network.

Table 4.1: NOA's (Summer Month) voltage violation check (6PM and 7PM)

At 6PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
147.15	3182.89	3109.84	44
At 6PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
147.15	3182.89	2269.81	0
At 7PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3054.47	3128.31	34
At 7PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3054.47	2370.25	0

4.3.2 Case II.A.2: NOA Operation on a Winter Day

Similar behavior (as that in Case I.A.1) was observed in this case as well. Figure 4.21 shows the NOA's operation and action values, respectively.

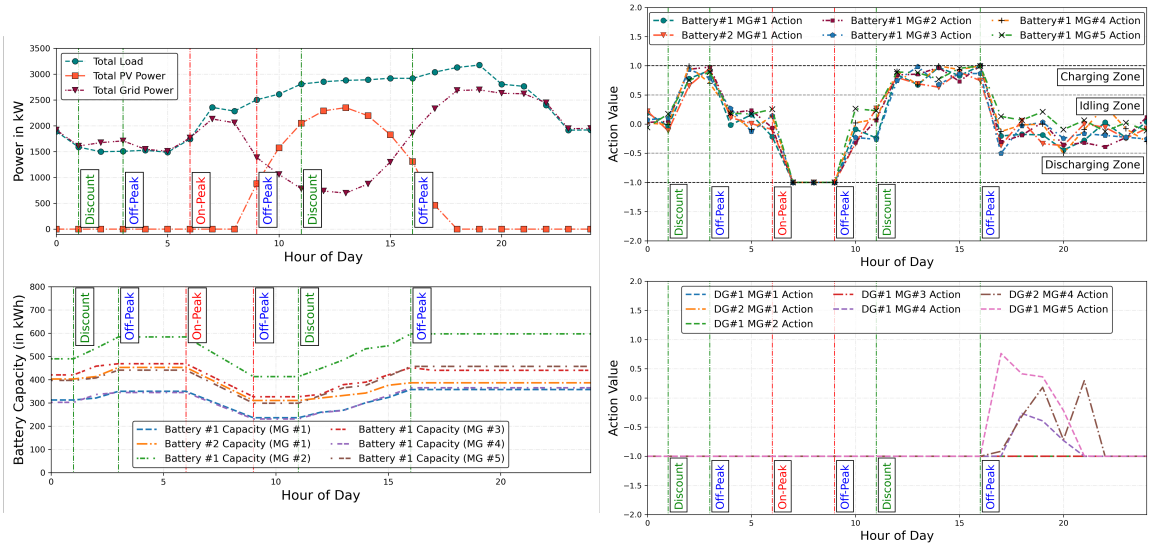

Figure 4.21: NMG #1 - #5 Operation and Agent's actions (Winter Month)

Voltage violation were also observed as a part of the initial conditions in the NOA's observation space. Like in previous case, the NOA appropriately deploys the DGs to satisfy the network constraints by eliminating these voltage violations. Details of these voltage violation before and after NOA took actions are presented in Table 4.2 (for 6PM and 7PM). (A detailed table is presented in Appendix B: Table B.2)

4.4 Layout II.B: Five NMGs for EOA testing

The five NMGs setup used for testing the EOA's capabilities is shown in Figure 4.22. The Tier IV zones are also highlighted on this figure. In total there are 10 Tier IV zones making the total action vector length to 13 (Three base tiers i.e. Tier I, Tier II and Tier III along with 10 zones of Tier IV loads).

Similar outage of 10-hrs starting at 7AM lasting till 5PM was considered, however the operating conditions were different. It was considered that the outage occurred on an overcast summer day thereby simulating low PV generation conditions. The testing

Table 4.2: NOA's (Winter Month) voltage violation check (6PM and 7PM)

At 6PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3174.16	3253.97	44
At 6PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3174.16	2694.33	0
At 7PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2804.22	2866.29	6
At 7PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2804.22	2627.96	0

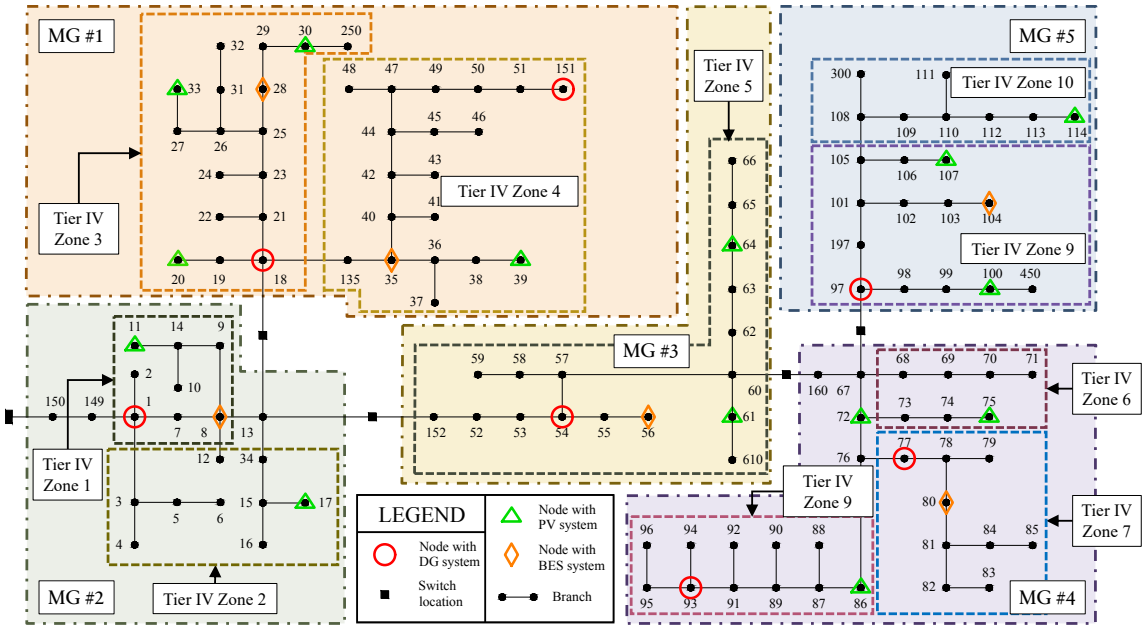

Figure 4.22: NMG #1 - #5 setup for EOA

scenarios were randomly generated where each scenario varies in the availability of different DERs to supply power (similar to EOA testing with two NMGs setup).

4.4.1 Case II.B.1: EOA Operation on a Low PV production day (Scenario #1)

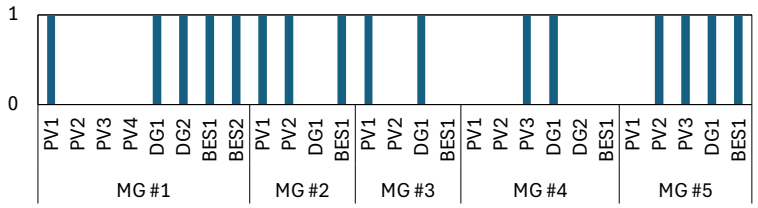

Figure 4.23: NMG #1 - #5: DER status information for Scenario #1

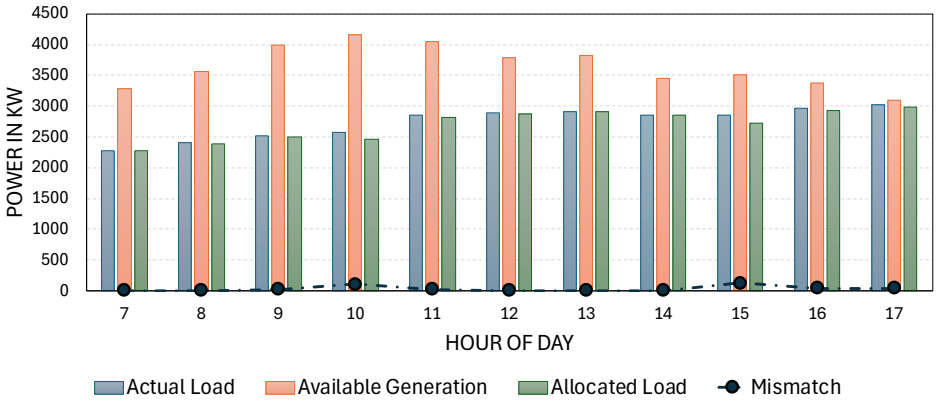

Figure 4.24: NMG #1 - #5: Load and generation information for Scenario #1

The available DERs under Scenario #1 are depicted in Figure 4.23. It can be seen that majority of DERs are connected to the network and hence this scenario can be considered as an abnormal conditions. Based on the aforementioned results, under abnormal conditions the mismatch value observed is almost zero over the outage period, which can be verified in Figure 4.24. Similarly, the CSI was observed to be close to 1 which evident from Figure 4.25. To re-emphasize, higher proportion of the critical loads are supplied as the CSI value reaches 1.

4.4.2 Case II.B.2: EOA Operation on a Low PV production day (Scenario #2)

Figure 4.26 represents the DERs available under Scenario #2. Under this scenario, the actual load is always higher than the available generation (seen in Figure 4.27).

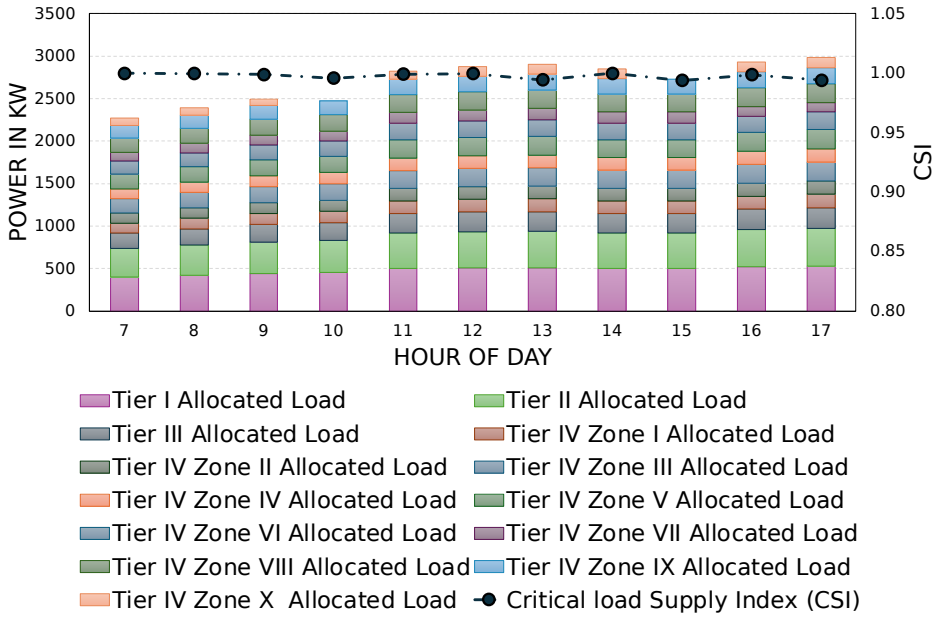

Figure 4.25: NMG #1 - #5: Tiered allocated load and CSI data for Scenario #1

Hence this scenario resembles extreme condition where due to generation deficit, the EOA has to shed some loads at the same time the EOA ensures that the most critical loads are still being supplied.

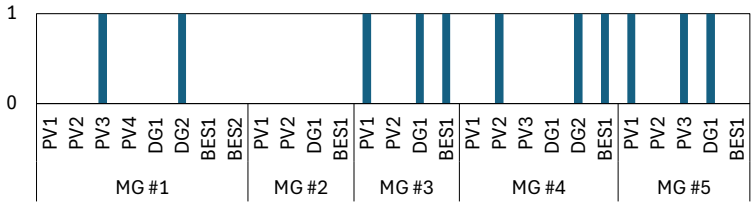

Figure 4.26: NMG #1 - #5: DER status information for Scenario #2

Due the nature of the operating conditions a fluctuation in the CSI value was observed in Figure 4.28 with the lowest CSI of 0.9196 was observed at 5PM. Nonetheless, Figure 4.28 confirms that the EOA is only shedding the Tier IV loads.

4.4.3 Case II.B.3: EOA Operation on a Low PV production day (Scenario #3)

The DER status information presented in Figure 4.29 for Scenario #3 again signifies extreme operating conditions (available generation lower than the actual load as in Figure 4.30).

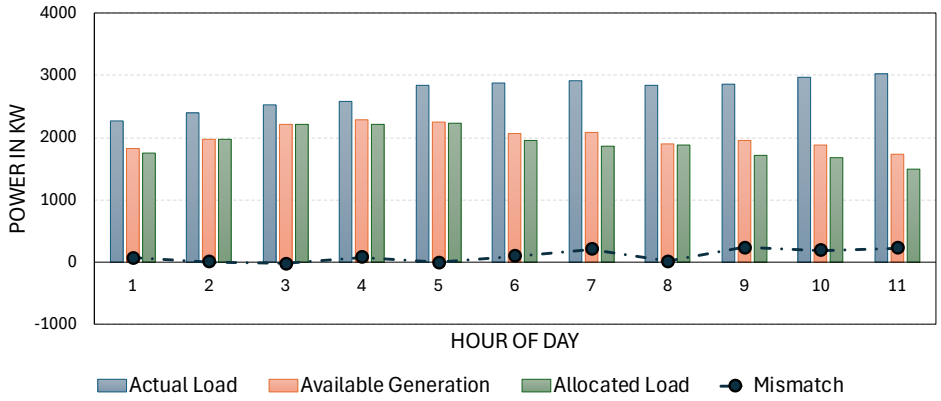

Figure 4.27: NMG #1 - #5: Load and generation information for Scenario #2

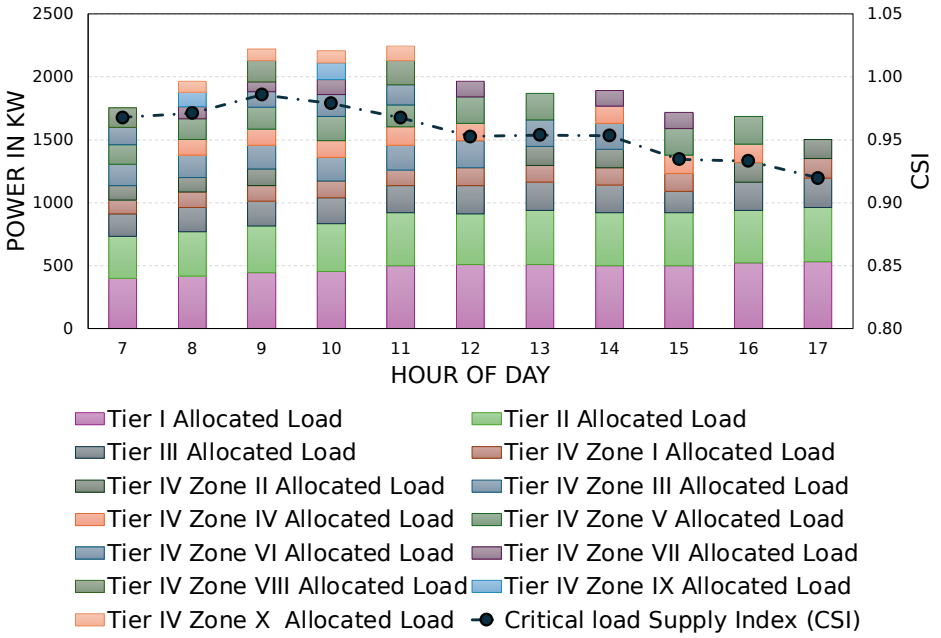

Figure 4.28: NMG #1 - #5: Tiered allocated load and CSI data for Scenario #2

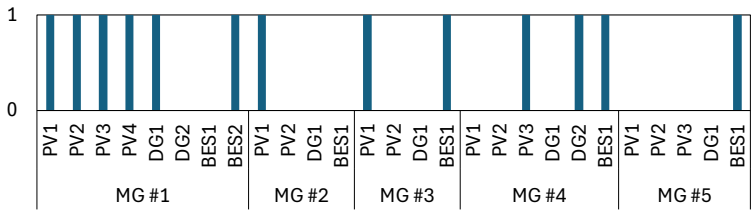

Figure 4.29: NMG #1 - #5: DER status information for Scenario #3

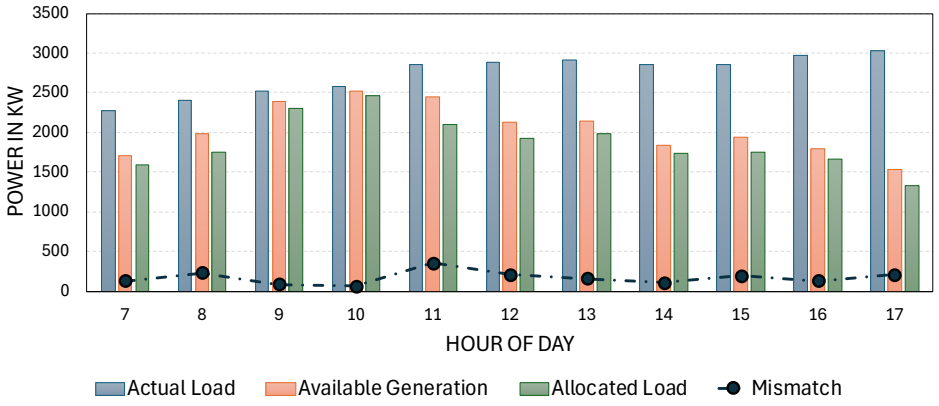

Figure 4.30: NMG #1 - #5: Load and generation information for Scenario #3

This scenario provided the worst CSI observed across all the scenarios. The lowest CSI of 0.8864 was observed at around 12PM. This low CSI can be attributed to a sudden drop in total PV generation at this hour, as majority of the power (in this scenario) is generate by the PV systems. Again, considering the tiered load allocation plot, it is visible that even at lowest CSI point (at 12PM), the most critical tiered loads i.e. Tier I and Tier II loads, are being supplied by the EOA.

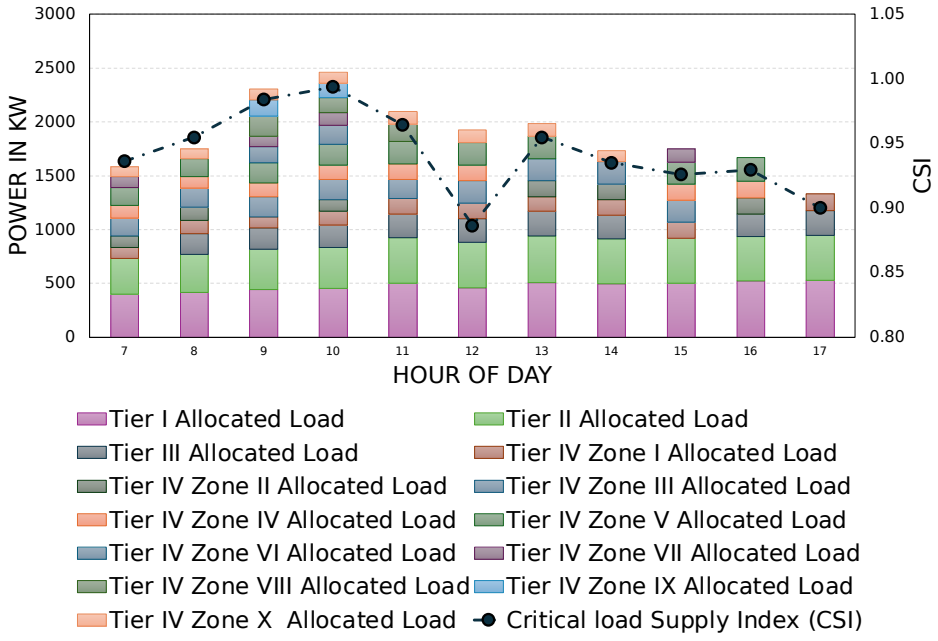

Figure 4.31: NMG #1 - #5: Tiered allocated load and CSI data for Scenario #3

4.4.4 Case II.B.4: EOA Operation on a Low PV production day (Scenario #4)

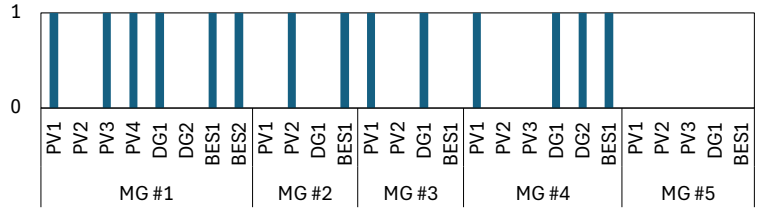

Figure 4.32: NMG #1 - #5: DER status information for Scenario #4

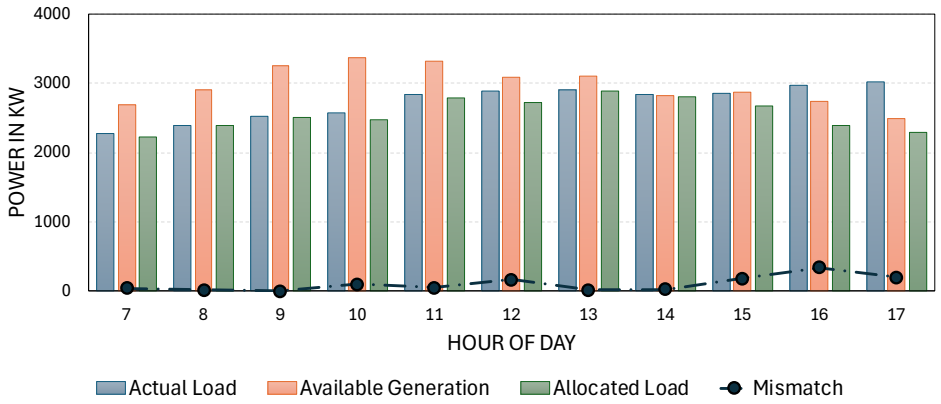

Figure 4.33: NMG #1 - #5: Load and generation information for Scenario #4

From Figure 4.32 it can be seen that Scenario #4 is an example of mixed conditions where at certain hours the available generation is higher than the actual load an vice versa. Hence when the the available generation is higher than the actual load a low mismatch was observed whereas in converse situations, a relatively higher CSI was observed (as seen in Figure 4.33).

Similarly, the EOA performs load shedding during the later stages of the outage to maintain balance between load and generation as seen from Figure 4.34. In this scenario, the lowest CSI of 0.9664 was observed at the final hour of the outage.

4.5 Comparison with other training algorithms

To verify the versatility and efficacy the proposed framework, a comparative study was performed where we used three different training algorithms to observe how the

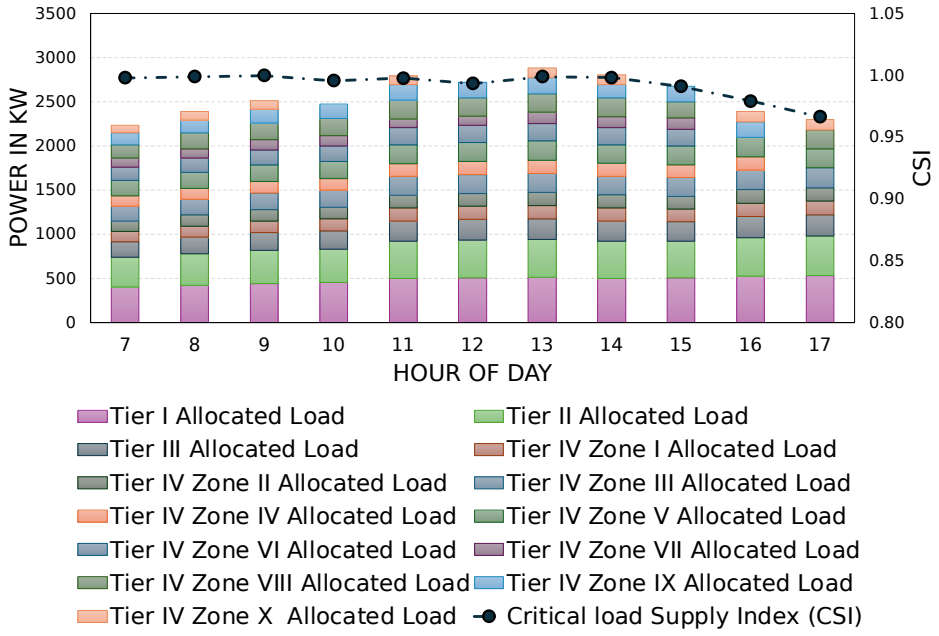

Figure 4.34: NMG #1 - #5: Tiered allocated load and CSI data for Scenario #4

agent performs. We specifically looked at differentiating the operating cost observed with different algorithms. Off-policy model-free DRL algorithms that combines policy optimization and Q-learning in an actor-critic set-up were used for the comparative analysis. The algorithms that are in this class include, Deep Deterministic Policy Gradient (DDPG), Twin-Delayed DDPG (TD3), and Soft Actor-Critic (SAC) algorithm. The cost comparison of a day's operation of the N-MG for a summer and a winter month under normal operating conditions was performed.

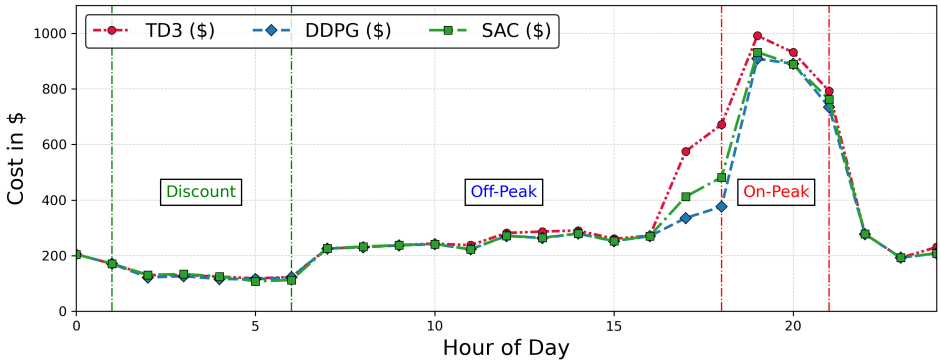

Figure 4.35: NMG #1 - #5: Operation cost comparison (Summer Month)

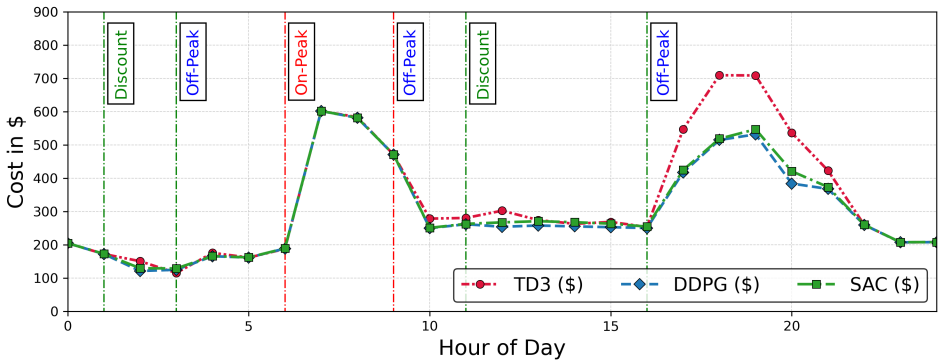

Figure 4.36: NMG #1 - #5: Operation cost comparison (Winter Month)

The results of the summer month's cost comparison are presented in Figure 4.35 while Figure 4.36 shows the results of the winter month's cost comparison. The best performing algorithm in terms of the operating cost values was the DDPG. Hence, the results presented in the Case Study and Results section were the ones obtained using this algorithm. However, the difference between the operating cost observed for SAC and DDPG is not much. Also, the training time required for SAC is much lesser than the DDPG ($\approx$ 2-3 mins/per time-steps for SAC compared to $\approx$ 7-8 mins/per time-step). So, from the stand point of computational time and getting comparable results, SAC will be a better option. The TD3 algorithm performed the worst out of all the algorithms that were tested. The majority of the differences were observed during the discount period in the BES operation and during the later hours from 4PM-8PM in the DG operation.

TD3 is the improved version of the DDPG algorithm, so in theory it should outperform DDPG. However, TD3 is highly sensitive to its hyperparameters which makes hyperparameter tuning the most important task while using the TD3 algorithm. The TD3 algorithm's results can be improved by running the hyperparameter tuning for every operational hour of the day and choosing the best hyperparameters for training. However, this was not the case with the other algorithms. To ensure the comparison of the results of all three algorithms follow similar tuning, training and test structure,

the aforementioned approach was not adopted for TD3 algorithm which may cause it to under-perform.

4.6 Validation with metaheuristic algorithms

We have validated the operating cost results obtained from DDPG and SAC against two well-known metaheuristic algorithms, namely, particle swarm optimization (PSO) and genetic algorithm (GA). Similar to the DRL algorithms comparison, a summer and a winter day's operation were considered. Figure 4.37 and Figure 4.38 show the comparison of the cost results for the two DRL algorithms along with the two metaheuristic algorithms. It can be observed that the cost values are comparable with the PSO algorithm providing the best results.

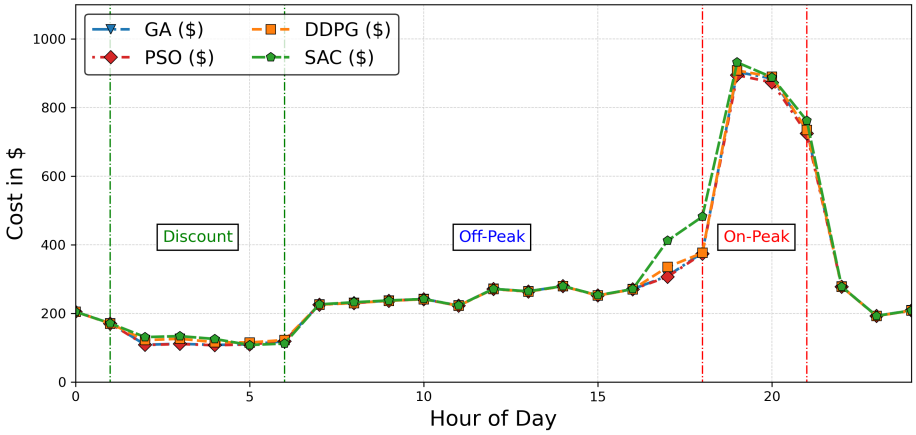

Figure 4.37: NMG #1 - #5: Cost function validation (Summer Month)

One important point to note is the emphasis on the application of the proposed DA framework in real-time. The results obtained from the PSO or GA are valid only for the initial states/inputs provided to the model. If results for any other state/initial conditions are required, then the PSO or GA needs to be re-run which would consume invaluable time. In contrast, once a DRL agent is trained using the day-ahead forecasts, it can be easily deployed online to provide real-time control where the control actions are instantaneously received as soon as the initial/input

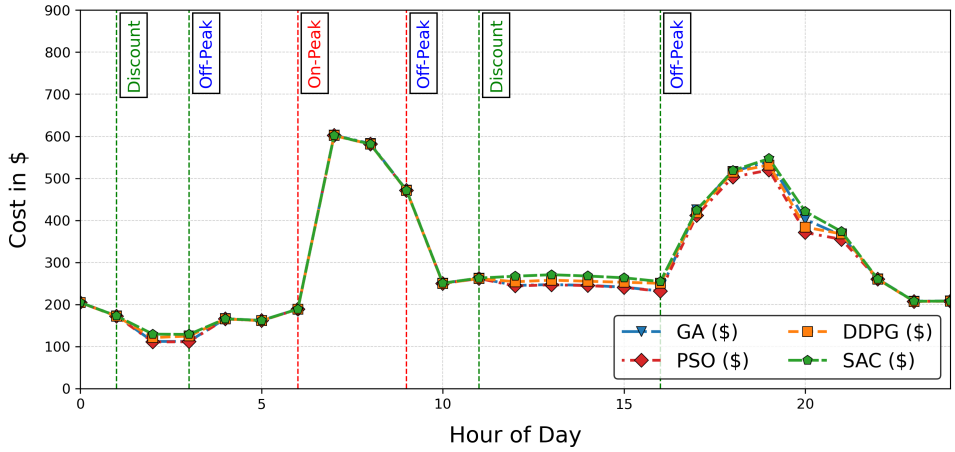

Figure 4.38: NMG #1 - #5: Cost function validation (Winter Month)

states are provided to the agent. The main intention for validation step is to ensure the control actions (to minimize or maximize its objective function) that the agent take are comparable to the more traditional techniques, which is proven from the results presented in this section.

4.7 Sensitivity Analysis

Learning of the RL/DRL models highly depends on the hyperparameter values being used. Therefore, it is imperative to determine whether the hyperparameters being used for training the model generates the most optimal results. Moreover knowing which hyperparameters provides the best training statistics further validates the model's accuracy and performance. Hence, in this section, a sensitivity analysis is performed to ascertain which hyperparameters can be considered most crucial to the agent's learning process along with its actual values.

While sensitivity analysis usually refers to a study that assesses how changes in the model's input or certain assumptions affect the model's outcome, in the context of machine learning, especially in this work, sensitivity analysis predominantly involves understanding how variation in hyperparameters impact model's performance. In other words, the sensitivity analysis performed in this section can be considered synonymous to hyperparameter tuning which is an essential aspect of any machine

learning modeling framework.

4.7.1 Methodology for sensitivity analysis

Tuning of the hyperparameters involve meticulously searching for different parameter values and evaluating its effect on the model performance. There are essentially two ways to perform the hyperparameter tuning:

1. Manual trial and error: A manual trial and error method, as the name suggests, involves manually changing each of the parameter values which is not ideal as a multitude of permutations and combinations of the parameter values are involved.

2. Systematic search: A systematic search provides an automatic way to search the best possible values for the hyperparameters. Hence, this approach is used in this work to determine the appropriate hyperparameter values.

Figure 4.39 demonstrates the complete hyperparameter tuning process that uses a systematic search approach. The central piece of the hyperparameter tuning is a study: that controls the experiments that will be performed to discover the best hyperparameters. A typical study aims to optimize the objective function by running multiple trials and evaluating the objective function during each of the trials. A trial consists of hyperparameters to be tested whose values are generated through a sampler. For each of these values, the objective function is evaluated to ascertain which values provides the best results. Moreover these values are tracked and relayed to the study so that at the end of each experiment/trial, a report that contains objective function value and the associated hyperparameter values can be generated. The tuning process also consists of a pruner that basically stops a specific trial prematurely if it is underperforming.

There are multiple hyperparameter tuning libraries available that provides easy to implement APIs for performing hyperparameter tuning. However in this work,

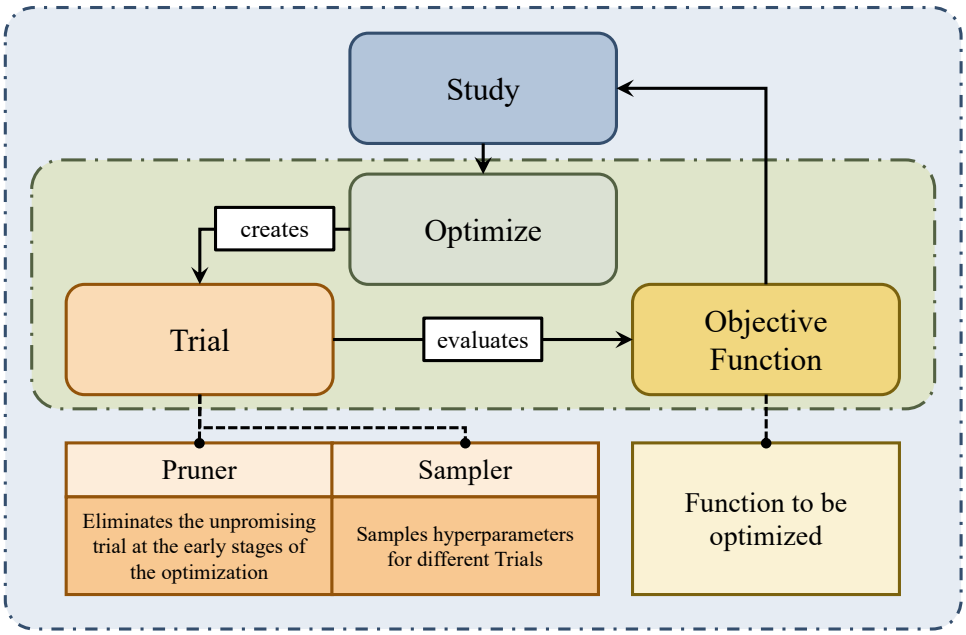

Figure 4.39: Hyperparameter Tuning Process

we used Optuna [134] which is an open source hyperparameter optimization framework that automates hyperparameter search. Optuna is compatible with most of the modern machine learning tools and can be easily integrated with the state-of-the-art RL/DRL algorithms.

One of the main attributes of hyperparameters tuning is the process of sampling the parameter values. Optuna provides multiple ways to sample the parameter as given below:

1. Grid search
2. Random search
3. Tree-structured Parzen Estimator (TPE)
4. Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

In this work, we use the TPE for generating samples as it can efficiently handle complex search spaces, compatible with continuous hyperparameter spaces, prioritizes

promising search areas by adapting over time, and is less computationally expensive (compared to CMA-ES and/or Grid search) [135, 136].

4.7.2 Result and Analysis of the sensitivity analysis

The study that was set up for tuning the hyperparameter includes training statistics for an hour of day on a winter day under normal operating conditions. The objective function for the hyperparameter optimization is the reward function of the NOA. Following is the list of hyperparameters that were optimized:

- Learning rate (α)
- Discount factor (γ)
- Neural Network's (NN) hidden layers
- Epsilon End value (ϵ_{end})

Since the learning algorithm used for training the agents were actor-critic based approaches, the learning rate was further classified as the Actor learning rate (α_{actor}) and the Critic learning rate (α_{critic}). The actor learning rate controls the policy updates. If this rate is smaller, the policy updates will be slower leading to a slow yet stable policy convergence. On the other hand if this rate is high, an unstable policy might be learned by the agent. The critic learning rate (α_{critic}) signifies the Q-value update rate. The range from which the sampler can sample α_{actor} and α_{critic} was set to be between 10^{-5} to 10^{-1}, respectively. In practice, α_{actor} must be set lower than $\alpha_{\text{critic}}$for a stable policy learning.

Discount factor, γ provides the significance of the future rewards in relation to the immediate rewards. The range for the sampling of γ was considered to be from 0 to 1. Four different NN hidden layer sizes: $\{32, 64, 128, 256\}$; were examined during the hyperparameter tuning. Lastly, Epsilon End value, ϵ_{end} which is incorporated into the noise function applied to the actor's output layer, that balances the exploration

and exploitation aspect was assessed. The range for this value was considered to be between 0 to 0.3.

A total of 20 trials were run where the reward function (objective function) for different set of hyperparameter was tracked. Out of 20, we are presenting results of the 16 trials which are classified in to three categories: 1) Unstable training statistics 2) Early Convergence training statistics, and 3) Best training statistics.

Figure 4.40 shows how the objective function value for Trial #2, #3, #6, and #9 varies over the training period. Moreover, the hyperparameters associated with each of the trials are presented in Table 4.3. It is clear that the training of the agent using these hyperparameters led to an unstable policy development. This behavior can be attributed to the fact that in all the trials (except Trial #9), α_{actor} is higher than α_{critic}. Because of these values, the actor network undertake large policy changes before the critic network gets to learn the value of the actions taken, thus leading to an unstable policy. In Trial #9 the hyperparamters used appear to be ineffective as the agent demonstrate no sign of learning, as seen in the Figure 4.40.

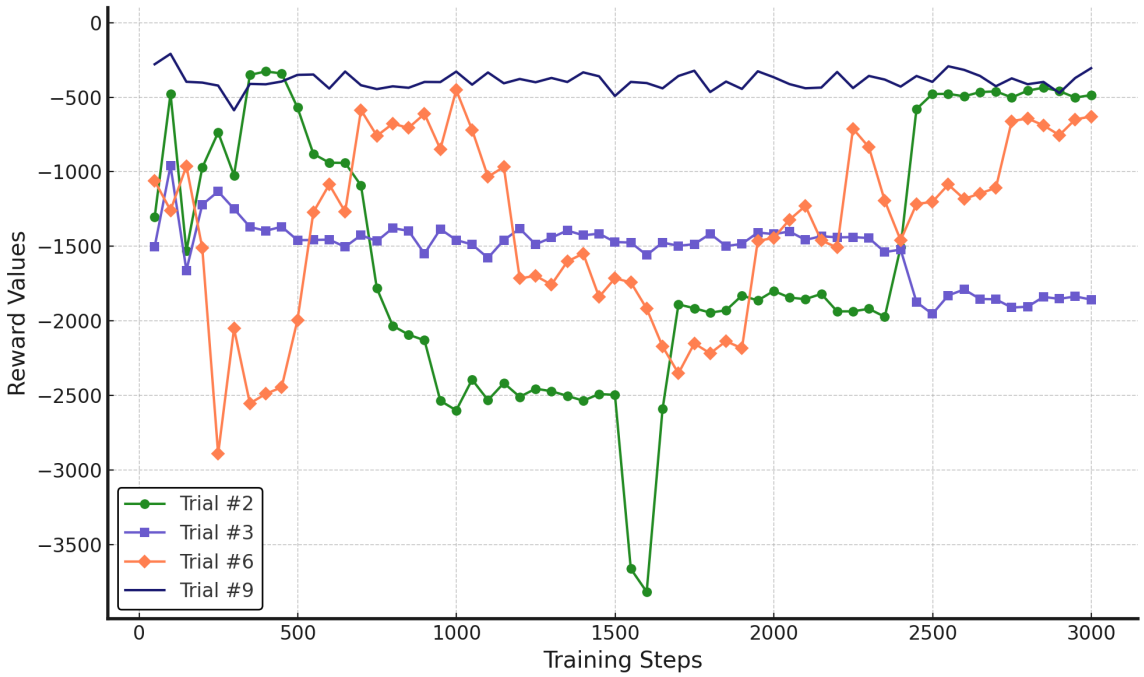

Figure 4.40: Unstable Training Statistics

Table 4.3: Unstable Training Statistics

Parameter	Trial #2	Trial #3	Trial #6	Trial #9
$\boldsymbol{\alpha_{\text{actor}}}$	0.000182	0.00454	0.000374	0.00185
$\boldsymbol{\alpha_{\text{critic}}}$	0.0000222	0.000101	0.0000305	0.0544
$\boldsymbol{\gamma}$	0.5981	0.000770	0.529	0.150
NN hidden layer size	32	128	128	256
$\boldsymbol{\epsilon_{\text{end}}}$	0.1027	0.0625	0.2278	0.2244
Reward	-485.983	-1856.582	-629.473	-305.711

Similarly, as $\alpha_{\text{actor}} > \alpha_{\text{critic}}$ (from Table 4.4) in Trial #1, #4, #5, and #19, an early policy convergence is observed as demonstrated in Figure 4.41. The critic network needs time to develop an understanding of the values associated with each of the action in a specific state. If the actor updates the policy at a higher rate, the policy may converge to a suboptimal value becuase of not having sufficient information about the environment dynamics. Early convergence can also be attributed to the agent not exploring the environment enough. However, in these trials, a low value of ϵ_{end} signifies that enough weight is assigned to the exploration of the environment thereby eliminating that to be a contributing factor towards early policy convergence.

Table 4.4: Early Convergence Training Statistics

Parameter	Trial #1	Trial #4	Trial #5	Trial #19
$\boldsymbol{\alpha_{\text{actor}}}$	0.00152	0.000676	0.0182	0.0959
$\boldsymbol{\alpha_{\text{critic}}}$	0.0000932	0.000647	0.000324	0.0390
$\boldsymbol{\gamma}$	0.637	0.0159	0.489	0.912
NN hidden layer size	256	256	32	64
$\boldsymbol{\epsilon_{\text{end}}}$	0.0203	0.0354	0.1023	0.2008
Reward	-327.733	-312.269	-819.723	-249.563

Table 4.5 provides the hyperparameters for the trials that provided the best training results. In all these trials, $\alpha_{\text{actor}} < \alpha_{\text{critic}}$ thus providing us a stable and near optimal policy convergence as depicted in Figure 4.42. One point to mention is that γ (discount factor) does not play any significant role. The reason being the one-shot episode termination approach developed in this work. Additionally, the size of the

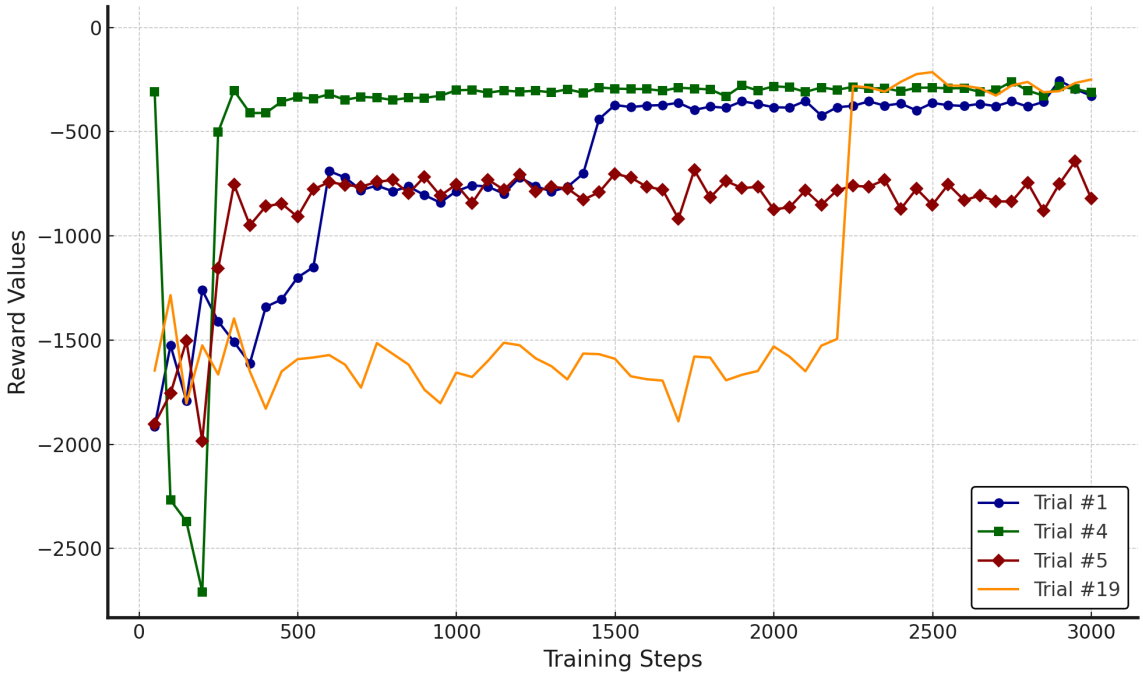

Figure 4.41: Early Convergence Training Statistics

Table 4.5: Best Training Statistics

Parameter	**Trial #10**	**Trial #11**	**Trial #15**	**Trial #18**
α_{actor}	0.0000157	0.0000160	0.0000714	0.000154
α_{critic}	0.0232	0.0188	0.0809	0.0234
γ	0.9407	0.9701	0.8196	0.7679
NN hidden layer size	256	256	256	256
ϵ_{end}	0.1634	0.1681	0.2899	0.1297
Reward	-38.946	-64.348	-37.177	-67.399

hidden layers for the neural networks that provide the best results was observed to contain 256 neurons (again from Table 4.5).

In majority of RL/DRL models it is very difficult to discern how the input and output relationship is developed due to the use of DNN. However, this analysis provides valuable insights into how certain parameters influence the model's behavior, strengthening the overall understanding of the model. Furthermore, the results of the sensitivity analysis bolster the accuracy and validity of the proposed model as it runs additional scenarios (as part of trails) with varied sets of model parameters,

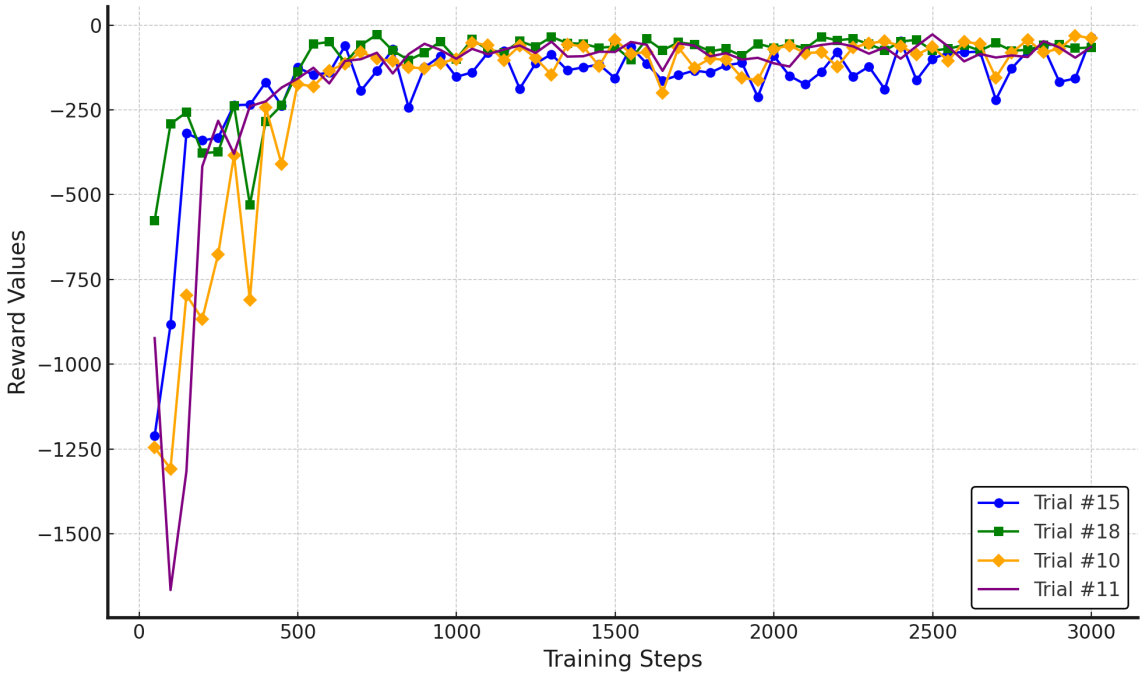

Figure 4.42: Best Training Statistics

providing an additional layer of verification.

CHAPTER 5: CONCLUSION AND FUTURE WORK

The concept of power system resilience has become more and more prevalent due to the evolving weather patterns. Even though it is a widely researched topic, there is no common consensus on some of the core components of power system resilience. This dissertation bridged the gap by presenting an extensive study of the concept of power system resilience that includes the proposed power system resilience definitions, analysis frameworks, and quantification metrics. The resilience definitions and frameworks used by some of the prominent research organizations were highlighted which demonstrated encouraging indications regarding standardizing the resilience definition. As more and more research organizations adopt a certain resilience definition for conducting studies, it is just a matter of time before an accord is reached on a standard power system resilience definition that encompasses all aspects.

Quantifying resilience improvement is one of the vital aspects of resilience assessment. Hence, some of the key factors to be considered during the development of resilience metrics were documented. Although building a standard resilience metric is exceedingly difficult, additional efforts are needed towards developing a set of guidelines that would guide researchers through a process of resilience assessment. There are some general guidelines that help in the metrics development process; however, these guidelines cannot address the variety of the resilience improvement objectives. Moreover, metrics development will become much easier once a consensus on the power system resilience definition is reached.

A brief review of some of the resilience enhancement techniques was presented. This literature survey provided insights into one of the core rationales of this dissertation i.e. to primarily focus on the enhancement of operational resilience of the system.

Thus, an amalgamation of all the preceding ideas pertaining to the concept of power system resilience, its evaluation frameworks, quantifying metrics and enhancement techniques formed the foundational basis for the remainder of the dissertation.

The core essence of this dissertation is the proposed unique Dual Agent based DRL framework that performs resource scheduling during normal operating conditions and load allocation during emergency operating conditions thereby improving the operational resilience of the network. This framework incorporated a time-dependent dynamic reward shaping variable, that bolsters agent's policy development. To further improve the agent's policy development, a training structure integrated with an error coefficient was developed in this work. Moreover, diverging from the conventional episode termination methods based on time series, we proposed a one-shot episode termination approach that facilitates the real-time application of our framework. To effectively allocate loads under emergency operating conditions, a distinctive approach that helps with determining load flexibility along with a multi-tier load classification system that categorizes loads into tiers based on the criticality index was also discussed.

To test the proposed framework, multiple case studies that considered varying NMG layouts and the operating conditions were performed. Based on the results of the case studies, the following conclusions were drawn:

1. Under normal operating conditions, the agent is able to appropriately schedule resources by minimizing the operational cost at the same time satisfying the operational limits of the network.

2. Under emergency operating conditions, the agent successfully allocates the most critical loads in the network which is highlighted by a high CSI value, by that means enhancing the survivability aspect of the operational resilience.

The concluding remarks were further supported by the comparative studies per-

formed using multiple DRL algorithms. These studies substantiate that ease of applicability of the proposed framework to any DRL algorithm. Furthermore, a result validation step was implemented to differentiate the DRL approach results with a more traditional metaheuristic optimization algorithm based approach. The consequence of this step supplemented the aforementioned remarks by providing exceptionally comparable outcomes.

From the one-shot episode termination proof it was evident that the most important component affecting agent's learning is the reward function. However, there are other parameters specifically associated with the DNN model that, if not modeled appropriately, would largely hinder agent's learning. Therefore, in order to identify and accordingly tune these parameters, sensitivity analysis was performed. As a part of this study, learning rates, neurons count in the DNN hidden layers, discount factor, and parameters addressing the exploration/exploitation dilemma were evaluated. All the trials that demonstrated the best returns had the following things in common:

- The Actor learning rate α_{actor} was lower than the Critic learning rate α_{critic}.
- The neurons count in each of the hidden layer of the DNN were 256.
- The parameter ϵ_{end} the govern exploration/exploitation dilemma was observed to be between 0.1 to 0.3.

Despite the fact that DNN are considered to be black-boxes especially when it comes to ascertaining the input-output relationship, studies (comparison, validation, and sensitivity analysis) performed in this dissertation bolster the reliability and accuracy of the model, and by doing so further endorses the proposed framework.

5.1 Contributions

The major contributions of this dissertation are highlighted below:

- A novel Dual Agent-based DRL framework designed to automate resource scheduling and load allocation, addressing both normal operational conditions and

emergency scenarios is proposed. The following are the modeling objectives of the respective DAs:

  - Normal Operating Agent (NOA): Objective to minimize the operational cost under standard operating conditions.
  - Emergency Operating Agent (EOA): Objective to maximize the CSI during emergency operating conditions.

- A reward function that utilizes a multi-temporal dynamic reward shaping approach is developed. This approach enhances the agent's decision-making process by capturing the time-varying dynamics of a power system environment.

- A training paradigm that integrates an error coefficient into the day-ahead forecasts that replicates forecasting errors was introduced. Incorporating the forecasting error coefficients into the training phase prepares the DAs for a multitude of scenarios thereby enhancing its learning capabilities.

- A one-shot episode termination MDP was designed for ease of real-time application. Furthermore, this developed approach enables faster convergence, and reduces computational load.

- A multi-tier classification system was created for load categorization based on the load criticality index that aids in efficient critical load allocation.

To summarize, this research provides a progressive approach, that leverages the advancements in the field of advanced machine learning, to solving two of the most critical problems of the modern power systems which are: 1) the problem of resource scheduling under uncertainty, and 2) the problem of resilience improvement by supplying critical loads under emergencies.

5.2 Future Work

This research opens new avenues for state-of-art research opportunities by uniting the field of power system analysis with the ever-evolving field of advanced machine learning. This research work is especially important in today's era, where a modern power grid can capitalize on the advancements in artificial intelligence (AI) to enhance its operational intelligence.

The load allocation problem addressed by the EOA can be further enhanced by adding supplementary load control capability. Currently, the proposed multi-tier load classification system has simplified the task of critical load allocation or load allocation in general to some extent. However, under certain situations, a higher mismatch has been observed, meaning that some of the non-critical tier loads were shed despite the availability of the generation capacity to support partial loads in these tiers. As a part of future work, a methodology that can consider such scenarios and develop a supplementary load allocation strategy to power these partial loads would greatly enhance the EOA's operating framework.

Furthermore, it will be important to analyze how the proposed framework operates when ran on a system with higher computing capabilities. At present, the training, testing, and validation of the DAs has been performed on a 2020 M1 MacBook Pro with 8 GB of RAM without the MPS (GPU) support. Although the training process was relatively quick, it will be interesting to study how faster (or slower) the model can train when it is deployed on a system with GPU or TPU support. Additionally, it would be fascinating to observe the hyperparameters variations, especially the number of training epoch required for convergence and training samples required to develop a robust policy, when a parallel computing approach is used.

As the NMGs become more prevalent and their operational structure grow increasingly sophisticated, with multiple entities interconnecting their MGs together to form a larger network of MGs, alternative paradigms, such as game-theory based multi-

agent systems (GT-MAS), need to be explored. The concept of game theory can help set-up a transactional system between the competing (under normal operating conditions) and cooperating (under emergency operating conditions) entities that builds on the dual agent environment set-up developed in this dissertation. The resultant framework can be termed as the Dual-Agent-based Multi-Agent system (DA-MAS) that uses the current DA set-up and modify its inner operations with multiple agents representing different MGs operated by different organizations.

Lastly, an agent-based DRL approach can be applied to the MG restoration problem that utilizes the available DERs with black start capabilities. The main objective of the agent would be to exploit the multi-tiered load classification introduced in this dissertation to restore Tier I through Tier IV loads. This approach can extend the EOA's environment to account for line outages, adding an extra layer of uncertainty to the formulation. These outages can be modeled using fragility curves or historical data, if available. The state space will include similar parameters as the EOA with inclusion of a marker that indicates the black start units. Additionally, we can leverage the concepts of graph theory to develop supplementary states that allow the agent to interpret the connectivity or the adjacency matrix of the complete network thereby enhancing its restoration planning. The agent's action space will be similar to that of the EOA which is focused on allocating the most critical loads in the network. However, this approach would place a special emphasis on line outages, making it pivotal to consider the availability of black start capable DER units in the network while allocating or restoring the loads.

REFERENCES

[1] Executive Office of the President. Council of Economic Advisers, "Economic benefits of increasing electric grid resilience to weather outages," 2013.

[2] M. Kintner-Meyer, "Towards Metrics for Resilience Characterization and Challenges in Valuing Distribution System Resilience Improvements," 2021.

[3] K. Watson, R. Cross, and M. Jones, "The winter storm of 2021. hobby school of public affairs, university of houston," 2021.

[4] P. Hoffman and W. Bryan, "Comparing the impacts of northeast hurricanes on energy infrastructure," *Office of Electricity Delivery and Energy Reliability, US Dept. of Energy, Washington, DC*, 2013.

[5] NERC, "Hurricane harvey event analysis report," tech. rep., North American Electric Reliability Corporation, March 2018.

[6] H. Aki, "Demand-side resiliency and electricity continuity: experiences and lessons learned in japan," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1443–1455, 2017.

[7] B. C. D. Attorney, "The camp fire public report: A summary of the camp fire investigation," *Oroville, CA*, 2020.

[8] N. NCEI, "Us billion-dollar weather and climate disasters (2022)," 2022.

[9] E. D. Vugrin, A. R. Castillo, and C. A. Silva-Monroy, "Resilience metrics for the electric power system: A performance-based approach.," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2017.

[10] B. Obama, "Presidential policy directive 21: Critical infrastructure security and resilience," *Washington, DC*, 2013.

[11] Office of Electricity, "North american energy resilience model," 2019.

[12] R. F. Jeffers, M. M. Hightower, N. S. Brodsky, M. J. Baca, A. Wachtel, S. Walsh, M. S. Aamir, J. Gibson, W. E. Fogleman, W. J. Peplinski, *et al.*, "A Grid Modernization Approach for Community Resilience: Application to New Orleans, LA," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2017.

[13] D. M. Anderson, S. Ganguli, A. L. Cooke, and M. L. Moore, "Grid modernization: Metrics analysis (gmlc1. 1)-affordability," tech. rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2019.

[14] A. T. Eseye, X. Zhang, B. Knueven, and W. Jones, "Enhancing distribution grid resilience through model predictive controller enabled prioritized load restoration strategy," in *2020 52nd North American Power Symposium (NAPS)*, pp. 1–6, IEEE, 2021.

[15] M. S. Lave, "Grid Modernization Laboratory Consortium Testing Network (GMLC 1.2. 3).," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2018.

[16] A. Ellis, "DOE Grid Modernization Initiative and Sandia R&D.," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2019.

[17] J. Torres and N. D. Laws, "Energy resilience through grid modernization and renewables integration," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2018.

[18] Electric Power Research Institute (EPRI), "Enhancing distribution resiliency: Opportunities for applying innovative technologies," 2013.

[19] M. Chaudry, P. Ekins, K. Ramachandran, A. Shakoor, J. Skea, G. Strbac, X. Wang, and J. Whitaker, "Building a resilient uk energy system," 2011.

[20] P. Stockton, "Resilience for black sky days," *Report prepared for the National Association of Regulatory Utility Commissioners and the US Department of Energy, February*, 2014.

[21] E. Ciapessoni, D. Cirio, A. Pitto, M. Panteli, M. Van Harte, and C. Mak, "on behalf of cigre wg c4. 47 (2019): Defining power system resilience," *Published in electra*, vol. 306.

[22] National Infrastructure Advisory Council (US), *Critical infrastructure resilience: Final report and recommendations*. National Infrastructure Advisory Council, 2009.

[23] NERC, "Definition of âadequate level of reliabilityâ," 2007.

[24] K. Zitelman, "Advancing electric system resilience with distributed energy resources: A review of state policies about sein," 2020.

[25] Y. Y. Haimes, "On the definition of resilience in systems," *Risk Analysis: An International Journal*, vol. 29, no. 4, pp. 498–501, 2009.

[26] P. Dehghanian, S. Aslan, and P. Dehghanian, "Quantifying power system resiliency improvement using network reconfiguration," in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 1364–1367, IEEE, 2017.

[27] NATF, "Transmission resilience overview," *North American Transmission Forum*, 2021.

[28] M. Panteli, D. N. Trakas, P. Mancarella, and N. D. Hatziargyriou, "Boosting the power grid resilience to extreme weather events using defensive islanding," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2913–2922, 2016.

[29] J.-P. Watson, R. Guttromson, C. Silva-Monroy, R. Jeffers, K. Jones, J. Ellison, C. Rath, J. Gearhart, D. Jones, T. Corbet, *et al.*, "Conceptual framework for developing resilience metrics for the electricity oil and gas sectors in the united states," *Sandia national laboratories, albuquerque, nm (united states), tech. rep*, 2014.

[30] Z. Bie, Y. Lin, G. Li, and F. Li, "Battling the extreme: A study on the power system resilience," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1253–1266, 2017.

[31] P. E. Roege, Z. A. Collier, J. Mancillas, J. A. McDonagh, and I. Linkov, "Metrics for energy resilience," *Energy Policy*, vol. 72, pp. 249–256, 2014.

[32] I. S. Association *et al.*, "IEEE 1366-2012-IEEE Guide for Electric Power Distribution Reliability Indices," *New York, USA: IEEE Standards Association*, 1998.

[33] M. R. Milligan, *Methods to Model and Calculate Capacity Contributions of Variable Generation for Resource Adequacy Planning (IVGTF1-2)*. North American Electric Reliability Corporation, 2011.

[34] B. Chiu and A. Bose, "Resilience framework methods and metrics for the electricity sector," *Resource Center, IEEE Power & Energy Society*, 2020.

[35] M. Panteli and P. Mancarella, "The grid: Stronger, bigger, smarter?: Presenting a conceptual framework of power system resilience," *IEEE Power and Energy Magazine*, vol. 13, no. 3, pp. 58–66, 2015.

[36] S. Chanda, A. K. Srivastava, M. U. Mohanpurkar, and R. Hovsapian, "Quantifying power distribution system resiliency using code-based metric," *IEEE Transactions on Industry Applications*, vol. 54, no. 4, pp. 3676–3686, 2018.

[37] M. Panteli, D. N. Trakas, P. Mancarella, and N. D. Hatziargyriou, "Power systems resilience assessment: Hardening and smart operational enhancement strategies," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1202–1213, 2017.

[38] Z. Li, M. Shahidehpour, F. Aminifar, A. Alabdulwahab, and Y. Al-Turki, "Networked microgrids for enhancing the power system resilience," *Proceedings of the IEEE*, vol. 105, no. 7, pp. 1289–1310, 2017.

[39] G. Fu, S. Wilkinson, R. J. Dawson, H. J. Fowler, C. Kilsby, M. Panteli, and P. Mancarella, "Integrated approach to assess the resilience of future electricity infrastructure networks to climate hazards," *IEEE Systems Journal*, vol. 12, no. 4, pp. 3169–3180, 2017.

[40] X. Liu, M. Shahidehpour, Z. Li, X. Liu, Y. Cao, and Z. Bie, "Microgrids for enhancing the power grid resilience in extreme conditions," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 589–597, 2016.

[41] S. Abbasi, M. Barati, and G. J. Lim, "A parallel sectionalized restoration scheme for resilient smart grid systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1660–1670, 2017.

[42] G. Kandaperumal, S. Pandey, and A. Srivastava, "AWR: Anticipate, Withstand, and Recover Resilience Metric for Operational and Planning Decision Support in Electric Distribution System," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 179–190, 2021.

[43] H. Gao, Y. Chen, Y. Xu, and C.-C. Liu, "Resilience-oriented critical load restoration using microgrids in distribution systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2837–2848, 2016.

[44] F. Petit, V. Vargas, J. Kavicky, M. Kintner-Meyer, and J. Eto, "Grid modernization: Metrics analysis (gmlc1. 1)–resilience," *Pacific Northwest National Laboratory*, 2020.

[45] W. Rickerson, J. Gillis, and M. Bulkeley, "The value of resilience for distributed energy resources: An overview of current analytical practices," *Prepared for National Association of Regulatory Utility Commissioners (NARUC). Washington DC: Converge Strategies (CSL). https://pubs. naruc. org/pub/531AD059-9CC0-BAF6-127B-99BCB5F02198*, 2019.

[46] A. Longo, S. Giaccaria, T. Bouman, and T. Efthimiadis, "Societal appreciation of energy security: Volume 1: Value of lost load–households (ee, nl and pt)," 2019.

[47] National Academies of Sciences, Engineering, and Medicine and others, *Enhancing the resilience of the nation's electricity system*. National Academies Press, 2017.

[48] K. Anderson, X. Li, S. Dalvi, S. Ericson, C. Barrows, C. Murphy, and E. Hotchkiss, "Integrating the value of electricity resilience in energy planning and operations decisions," *IEEE Systems Journal*, vol. 15, no. 1, pp. 204–214, 2020.

[49] G. Wacker and R. Billinton, "Customer cost of electric service interruptions," *Proceedings of the IEEE*, vol. 77, no. 6, pp. 919–930, 1989.

[50] A. Ratha, E. Iggland, and G. Andersson, "Value of lost load: How much is supply security worth?," in *2013 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2013.

[51] M. Panteli, P. Mancarella, D. N. Trakas, E. Kyriakides, and N. D. Hatziargyriou, "Metrics and quantification of operational and infrastructure resilience in power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4732–4742, 2017.

[52] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies," *Electric Power Systems Research*, vol. 127, pp. 259–270, 2015.

[53] EPRI, "Distribution grid resiliency: Undergrounding," Tech. Rep. 3002006782, Electr. Power Res. Institute, Palo Alto, CA, USA, 2015.

[54] S. Ma, B. Chen, and Z. Wang, "Resilience enhancement strategy for distribution systems under extreme weather events," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1442–1451, 2016.

[55] E. O. of the President. Council of Economic Advisers, "Economic benefits of increasing electric grid resilience to weather outages," tech. rep., 2013.

[56] EPRI, "Distribution grid resiliency: Vegetation management," Tech. Rep. 3002006781, Electr. Power Res. Institute, Palo Alto, CA, USA, 2015.

[57] PG&E, "Enhanced vegetation management," tech. rep., Pacific Gas and Electric Company, San Francisco, CA, 2021.

[58] J. Boggess, G. Becker, and M. Mitchell, "Storm & flood hardening of electrical substations," in *2014 IEEE PES T&D Conference and Exposition*, pp. 1–5, IEEE, 2014.

[59] J. W. Baker, "Eliminating hurricane-induced storm surge damage to electric utilities via in-place elevation of substation structures and equipment," in *2014 IEEE PES T&D Conference and Exposition*, pp. 1–5, IEEE, 2014.

[60] N. Bhusal, M. Abdelmalak, M. Kamruzzaman, and M. Benidris, "Power system resilience: Current practices, challenges, and future directions," *IEEE Access*, vol. 8, pp. 18064–18086, 2020.

[61] P. Dehghanian, S. Aslan, and P. Dehghanian, "Maintaining electric system safety through an enhanced network resilience," *IEEE Transactions on Industry Applications*, vol. 54, no. 5, pp. 4927–4937, 2018.

[62] C. Chen, J. Wang, F. Qiu, and D. Zhao, "Resilient distribution system by microgrids formation after natural disasters," *IEEE Transactions on smart grid*, vol. 7, no. 2, pp. 958–966, 2015.

[63] Y. Xu, C.-C. Liu, K. P. Schneider, F. K. Tuffner, and D. T. Ton, "Microgrids for service restoration to critical load in a resilient distribution system," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 426–437, 2016.

[64] J. Li, X.-Y. Ma, C.-C. Liu, and K. P. Schneider, "Distribution system restoration with microgrids using spanning tree search," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 3021–3029, 2014.

[65] Y. Xu, C.-C. Liu, K. P. Schneider, and D. T. Ton, "Placement of remote-controlled switches to enhance distribution system restoration capability," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1139–1150, 2015.

[66] L. Che and M. Shahidehpour, "Adaptive formation of microgrids with mobile emergency resources for critical service restoration in extreme conditions," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 742–753, 2018.

[67] B. Zhang, P. Dehghanian, and M. Kezunovic, "Optimal allocation of pv generation and battery storage for enhanced resilience," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 535–545, 2017.

[68] G. Liu, M. R. Starke, B. Ollis, and Y. Xue, "Networked microgrids scoping study," *ORNL, TN.[Online]. Available: https://info. ornl. gov/sites/publications/files/Pub68339. pdf*, pp. 0093–9994, 2016.

[69] M. Shahidehpour, Z. Li, S. Bahramirad, Z. Li, and W. Tian, "Networked microgrids: Exploring the possibilities of the iit-bronzeville grid," *IEEE Power and Energy Magazine*, vol. 15, no. 4, pp. 63–71, 2017.

[70] A. Hussain, V.-H. Bui, and H.-M. Kim, "A resilient and privacy-preserving energy management strategy for networked microgrids," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2127–2139, 2016.

[71] H. Xie, X. Teng, Y. Xu, and Y. Wang, "Optimal energy storage sizing for networked microgrids considering reliability and resilience," *IEEE Access*, vol. 7, pp. 86336–86348, 2019.

[72] S. Teimourzadeh, O. B. Tor, M. E. Cebeci, A. Bara, and S. V. Oprea, "A three-stage approach for resilience-constrained scheduling of networked microgrids," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 4, pp. 705–715, 2019.

[73] Y. Huang, G. Li, C. Chen, Y. Bian, T. Qian, and Z. Bie, "Resilient distribution networks by microgrid formation using deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4918–4930, 2022.

[74] Z. Yan and Y. Xu, "A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4599–4608, 2020.

[75] T. Yu, H. Wang, B. Zhou, K. W. Chan, and J. Tang, "Multi-agent correlated equilibrium q (λ) learning for coordinated smart generation control of interconnected power grids," *IEEE transactions on power systems*, vol. 30, no. 4, pp. 1669–1679, 2014.

[76] B. Zhang, Z. Chen, and A. M. Ghias, "Deep reinforcement learning-based energy management strategy for a microgrid with flexible loads," in *2023 International*

Conference on Power Energy Systems and Applications (ICoPESA), pp. 187–191, IEEE, 2023.

[77] Y. Shu, W. Dong, Q. Yang, and Y. Wang, "Microgrid energy management using improved reinforcement learning with quadratic programming," in *2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 2015–2020, IEEE, 2021.

[78] S.-J. Chen, W.-Y. Chiu, and W.-J. Liu, "User preference-based demand response for smart home energy management using multiobjective reinforcement learning," *IEEE Access*, vol. 9, pp. 161627–161637, 2021.

[79] T. Xu, T. Chen, C. Gao, and H. Hui, "Intelligent home energy management strategy with internal pricing mechanism based on multiagent artificial intelligence-of-things," *IEEE Systems Journal*, 2023.

[80] A. Kahraman and G. Yang, "Home energy management system based on deep reinforcement learning algorithms," in *2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pp. 1–5, IEEE, 2022.

[81] R. Lu, Z. Jiang, H. Wu, Y. Ding, D. Wang, and H.-T. Zhang, "Reward shaping-based actor–critic deep reinforcement learning for residential energy management," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 3, pp. 2662–2673, 2022.

[82] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2751–2762, 2019.

[83] S. Li, D. Cao, W. Hu, Q. Huang, Z. Chen, and F. Blaabjerg, "Multi-energy management of interconnected multi-microgrid system using multi-agent deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, 2023.

[84] K. Prabakar, A. Valibeygi, S. A. R. Konakalla, B. Miller, R. A. De Callafon, A. Pratt, M. Symko-Davies, and T. Bialek, "Remote hardware-in-the-loop approach for microgrid controller evaluation," in *2020 Clemson University Power Systems Conference (PSC)*, pp. 1–8, IEEE, 2020.

[85] D. Vine, D. Attanasio, and E. Shittu, "Microgrid momentum: Building efficient, resilient power," *Center for Climate and Energy Solutions*, 2017.

[86] S. F. POC, M. Baggu, and V. Donde, "Doe oe 2021 strategy white papers on microgrids: Program vision, objectives, and r&d targets in 5 and 10 years–topic area," 2021.

[87] R. Boukhanouf, "Small combined heat and power (chp) systems for commercial buildings and institutions," in *Small and micro combined heat and power (CHP) systems*, pp. 365–394, Elsevier, 2011.

[88] H. Shayeghi and M. Alilou, "Distributed generation and microgrids," in *Hybrid Renewable Energy Systems and Microgrids*, pp. 73–102, Elsevier, 2021.

[89] M. Ozturk and I. Dincer, "4.30 district energy conversion systems," *Comprehensive Energy Systems*, 2018.

[90] A. B. Stambouli and E. Traversa, "Fuel cells, an alternative to standard sources of energy," *Renewable and sustainable energy reviews*, vol. 6, no. 3, pp. 295–304, 2002.

[91] N. Nachman-Hunt, "Small hydropower systems: energy efficiency and renewable energy clearinghouse," tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2001.

[92] X. Luo, J. Wang, M. Dooner, and J. Clarke, "Overview of current development in electrical energy storage technologies and the application potential in power system operation," *Applied energy*, vol. 137, pp. 511–536, 2015.

[93] S. Choudhury, "Review of energy storage system technologies integration to microgrid: Types, control strategies, issues, and future prospects," *Journal of Energy Storage*, vol. 48, p. 103966, 2022.

[94] P. Breeze, "Hydrogen energy storage," *Power system energy storage technologies*, pp. 69–77, 2018.

[95] M. Aneke and M. Wang, "Energy storage technologies and real life applications–a state of the art review," *Applied Energy*, vol. 179, pp. 350–377, 2016.

[96] S. Koohi-Fayegh and M. A. Rosen, "A review of energy storage types, applications and recent developments," *Journal of Energy Storage*, vol. 27, p. 101047, 2020.

[97] S. De, S. Acharya, S. Sahoo, and G. C. Nayak, "Present status of biomass-derived carbon-based composites for supercapacitor application," in *Nanostructured, Functional, and Flexible Materials for Energy Conversion and Storage Systems*, pp. 373–415, Elsevier, 2020.

[98] B. K. Choure, T. Alam, and R. Kumar, "A review on heat transfer enhancement techniques for pcm based thermal energy storage system," *Journal of Energy Storage*, vol. 72, p. 108161, 2023.

[99] K. C. Divya and J. Østergaard, "Battery energy storage technology for power systemsâan overview," *Electric power systems research*, vol. 79, no. 4, pp. 511–520, 2009.

[100] F. Katiraei, R. Iravani, N. Hatziargyriou, and A. Dimeas, "Microgrids management," *IEEE power and energy magazine*, vol. 6, no. 3, pp. 54–65, 2008.

[101] M. Hannan, S. Y. Tan, A. Q. Al-Shetwi, K. P. Jern, and R. Begum, "Optimized controller for renewable energy sources integration into microgrid: Functions, constraints and suggestions," *Journal of Cleaner Production*, vol. 256, p. 120419, 2020.

[102] W. Su and J. Wang, "Energy management systems in microgrid operations," *The Electricity Journal*, vol. 25, no. 8, pp. 45–60, 2012.

[103] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi, C. A. Cañizares, R. Iravani, M. Kazerani, A. H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke, *et al.*, "Trends in microgrid control," *IEEE Transactions on smart grid*, vol. 5, no. 4, pp. 1905–1919, 2014.

[104] S. Sen and V. Kumar, "Microgrid control: A comprehensive survey," *Annual Reviews in control*, vol. 45, pp. 118–151, 2018.

[105] A. Ruiz-Alvarez, A. Colet-Subirachs, F. A.-C. Figuerola, O. Gomis-Bellmunt, and A. Sudria-Andreu, "Operation of a utility connected microgrid using an iec 61850-based multi-level management system," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 858–865, 2012.

[106] R. E. Mackiewicz, "Overview of iec 61850 and benefits," in *2006 IEEE Power Engineering Society General Meeting*, pp. 8–pp, IEEE, 2006.

[107] M. N. Alam, S. Chakrabarti, and A. Ghosh, "Networked microgrids: State-of-the-art and future perspectives," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1238–1250, 2018.

[108] E. Trinklei, G. Parker, W. Weaver, R. Robinett, L. B. Gauchia, C.-W. Ten, W. Bower, S. F. Glover, and S. Bukowski, "Scoping study: networked microgrids," tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2014.

[109] M. R. Starke, "Networked microgrids scoping study," tech. rep., Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2016.

[110] F. Kamal and B. Chowdhury, "Model predictive control and optimization of networked microgrids," *International Journal of Electrical Power & Energy Systems*, vol. 138, p. 107804, 2022.

[111] J. A. Rodriguez-Gil, E. Mojica-Nava, D. Vargas-Medina, M. F. Arevalo-Castiblanco, C. A. Cortes, S. Rivera, and J. Cortes-Romero, "Energy management system in networked microgrids: An overview," *Energy Systems*, pp. 1–32, 2024.

[112] Q. Zhou, M. Shahidehpour, A. Paaso, S. Bahramirad, A. Alabdulwahab, and A. Abusorrah, "Distributed control and communication strategies in networked microgrids," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2586–2633, 2020.

[113] C. M. Macal and M. J. North, "Tutorial on agent-based modeling and simulation," in *Proceedings of the Winter Simulation Conference, 2005.*, pp. 14–pp, IEEE, 2005.

[114] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the national academy of sciences*, vol. 99, no. suppl_3, pp. 7280–7287, 2002.

[115] T. C. Schelling, "Dynamic models of segregation," *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.

[116] J. M. Epstein and R. Axtell, *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, 1996.

[117] B. LeBaron, "Empirical regularities from interacting long-and short-memory investors in an agent-based stock market," *Ieee transactions on evolutionary computation*, vol. 5, no. 5, pp. 442–455, 2001.

[118] L. Perez and S. Dragicevic, "An agent-based approach for modeling dynamics of contagious disease spread," *International journal of health geographics*, vol. 8, pp. 1–17, 2009.

[119] R. S. Sutton, "Reinforcement learning: An introduction," *A Bradford Book*, 2018.

[120] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[121] M. Van Otterlo and M. Wiering, "Reinforcement learning and markov decision processes," in *Reinforcement learning: State-of-the-art*, pp. 3–42, Springer, 2012.

[122] D. Elavarasan, D. R. Vincent, V. Sharma, A. Y. Zomaya, and K. Srinivasan, "Forecasting yield by integrating agrarian factors and machine learning models: A survey," *Computers and electronics in agriculture*, vol. 155, pp. 257–282, 2018.

[123] R. Grosse, C. Maddison, J. Bae, and S. Pitis, "Csc 311: Introduction to machine learning," 2020.

[124] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, J. Pineau, *et al.*, "An introduction to deep reinforcement learning," *Foundations and Trends® in Machine Learning*, vol. 11, no. 3-4, pp. 219–354, 2018.

[125] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

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

[127] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[128] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[129] T. Lillicrap, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[130] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*, pp. 1587–1596, PMLR, 2018.

[131] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*, pp. 1861–1870, PMLR, 2018.

[132] F. Kamal, B. H. Chowdhury, and C. Lim, "Networked microgrid scheduling for resilient operation," *IEEE Transactions on Industry Applications*, 2023.

[133] S. Ong and N. Clark, "Commercial and residential hourly load profiles for all tmy3 locations in the united states," tech. rep., DOE Open Energy Data Initiative (OEDI); National Renewable Energy Lab.(NREL âŠ, 2014.

[134] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[135] S. Watanabe, "Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance," *arXiv preprint arXiv:2304.11127*, 2023.

[136] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," *Advances in neural information processing systems*, vol. 24, 2011.

APPENDIX A: Load Flexibility Estimation

In this appendix section, an example that provides insight into the process of ascertaining the load flexibility index and the load flexibility range for a specific building type is presented. Lets consider a hospital load (building) type (load data was obtained from [133]) to determine its flexibility factor and range. Ideally, all the loads in a hospital building need to be considered as essential, hence must be supplied continuously. For the purpose of conceptual explanation, we are considering a flexible demand for a hospital load. The load value associated with each of the load types in a hospital building is specified in Table A.1.

Table A.1: Hospital Building Loads in kW (Hourly)

Time	**Fan**	**Cooling**	**Heating**	**Light**	**Equipment**	**Miscellaneous equipment**
01:00:00	93.63	469.64	0	46.10	113.31	136.96

Leveraging Equation (3.8), the l_o values and the ε_o values computed for the hospital buildings, are presented in Table A.2. Again l_o is the decimal equivalent of portion of total load consumed by the o^{th} load type ($l_o \in [0, 1]$) and ε_o as the decimal equivalent of percent of essential loads present in the o^{th} load type ($\varepsilon_o \in [0, 1]$).

It can be seen that the fan load (l_{fan}) is 11%, cooling load (l_{cooling}) is 54%, light load (l_{light}) is 5%, interior equipment load (l_{intEquip}) is 14% and miscellaneous equipment ($l_{\text{miscEquip}}$) load a 16% of the total load. The values associated with ε_o represents the criticality of a specific load type, with 1 indicating the highest level of criticality. A value of 1 also means that specific load type must be served in its entirety. However, certain load types, such as miscellaneous equipment loads, come with a

Table A.2: Rescaled values for Load Flexibility Estimation

	Fan	**Cooling**	**Heating**	**Light**	**Equipment**	**Miscellaneous equipment**
l_o	0.11	0.54	0	0.05	0.14	0.16
ε_o	1	1	0	1	1	0.3

degree of flexibility. For instance, some equipment within this load category may not be as critical and can therefore contribute to determining the overall load flexibility for that particular building type. In this example, the percentage of operationally critical miscellaneous equipment was considered to be 30% making its $\varepsilon_{\text{miscEquip}}$: 0.3. Therefore, the following Equations (A.1) to (A.3) provides the load flexibility index of a hospital load.

$$LFI_{1,o,hosp} = \sum_{o=1}^{O} l_o.\varepsilon_o \tag{A.1}$$

$$= 0.11 \cdot 1 + 0.54 \cdot 1 + 0 \cdot 0 + 0.05 \cdot 1 + 0.14 \cdot 1 + 0.16 \cdot 0.3 \tag{A.2}$$

$$= 0.889 \tag{A.3}$$

To estimate the range $[P_t^{\text{load\_min\_allowed}}, P_t^{\text{load\_max}}]$ for a specific building type (hospital), Equation (A.4) can be used which essentially determines the $P_t^{\text{load\_min\_allowed}}$ by making use of the load flexibility index calculated previously.

$$P_1^{\text{load\_min\_allowed}} = LFI_{1,o,hosp} \cdot P_1^{\text{load\_max}} \tag{A.4}$$

Important point to note is that the ε_o associated with a specific load type in a building can be computed by conducting surveys based on conjoint analysis. This approach would yield a more accurate ε_o estimations thereby improving the estimated load flexibility index.

APPENDIX B: Additional Results

B.1 Case II.A.1: Complete Voltage Violation Table

The complete voltage violations associated with Case II.A.1 are presented in the Table B.1 below:

Table B.1: NOA's (Summer Month) voltage violation check (4PM - 8PM)

At 4PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
1073.58	2946.46	1910.02	3
At 4PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
1073.58	2946.46	1794.65	0
At 5PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
598.14	2980.19	2431.68	7
At 5PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
598.14	2980.19	2236.29	0
At 6PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
147.15	3182.89	3109.84	44
At 6PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
147.15	3182.89	2269.81	0
At 7PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3054.47	3128.31	34
At 7PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3054.47	2370.25	0
At 8PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2709.09	2766.96	5
At 8PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2709.09	2419.11	0

B.2 Case II.A.2: Complete Voltage Violation Table

The complete voltage violations associated with Case II.A.1 are presented in the Table B.2 below:

Table B.2: NOA's (Winter Month) voltage violation check (4PM - 8PM)

At 4PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
461.64	3037.75	2632.11	12
At 4PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
461.64	3037.75	2330.94	0
At 5PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3127.86	3205.33	44
At 5PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3127.86	2681.56	0
At 6PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3174.16	3253.97	44
At 6PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	3174.16	2694.33	0
At 7PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2804.22	2866.29	6
At 7PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2804.22	2627.96	0
At 8PM (Before NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2762.41	2822.60	6
At 8PM (After NOA's Actions)			
Total PV power	**Total load**	**Total grid power**	**Voltage violation count**
0	2762.41	2618.80	0

APPENDIX C: DRL Algorithms Pseudo-code

C.1 Deep Deterministic Policy Gradient (DDPG) Algorithm

Algorithm 1 Deep Deterministic Policy Gradient (DDPG) Algorithm

Randomly initialize the critic network $Q(s, a|\theta^Q)$ and actor network $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ
Initialize the target networks Q' and μ' with the weights $\theta^{Q'}$ and $\theta^{\mu'}$
Initialize the replay buffer R
for episode = 1, M **do**
    Initialize a random process $\mathcal{N}$ for exploration of the action space
    Get initial state values s_1
    for $t = 1, T$ **do**
        Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_\sqcup$
        Execute action a_t and observe a_{t+1}, r_{t+1}
        Insert transition (s_t, a_t, r_t, s_{t+1}) into the replay buffer R
        Sample random batch of N transitions from R
        Set $y_i = r_i + \gamma.Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
        Update critic network by minimizing the loss:

$$L = \frac{1}{N}\sum_{i=1}(y_i - Q(s_i, a_i|\theta^Q))^2$$

        Update actor policy using sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_{i=1} \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)}.\nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s=s_i}$$

        Update the target network:

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

    end for
end for

In DDPG algorithm [129] there are four DNNs: 1) Actor Network, 2) Critic Network, 3) Target Actor Network, and 4) Target Critic Network. Actor and Critic networks are updated as specified in the Algorithm 1 whereas a delayed parameter update which slowly maps the Actor-Critic NN parameters to the target networks is applied.

C.2 Twin-Delayed DDPG (TD3) Algorithm

Algorithm 2 Twin-Delayed DDPG (TD3) Algorithm

Randomly initialize the critic networks $Q(s, a|\theta_1^Q)$, $Q(s, a|\theta_2^Q)$ and actor network $\mu(s|\theta^\mu)$ with weights θ_1^Q, θ_2^Q and θ^μ
Initialize the target networks $\theta_1^{Q'} \leftarrow \theta_1^Q$, $\theta_2^{Q'} \leftarrow \theta_2^Q$ and $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize the replay buffer R
for episode = 1, M **do**
    Initialize a random process $\mathcal{N}$ for exploration of the action space
    Get initial state values s_1
    for $t = 1, T$ **do**
        Select action with exploration noise $a \sim \mu(s) + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma)$
        Observe reward r and next state s_{t+1}
        Insert transition (s_t, a_t, r_t, s_{t+1}) into the replay buffer R
        Sample random mini-batch of N transitions from R
        $\tilde{a} \leftarrow \mu'(s_{t+1}) + \epsilon, \quad \epsilon \sim \text{clip}(\mathcal{N}(0, \tilde{\sigma}), -c, c)$
        $y \leftarrow r + \gamma . \min_{i=1,2} Q'(s_{t+1}, \tilde{a}|\theta_i^{Q'})$
        Update critic network by:

$$\theta_i \leftarrow \arg\min_{\theta_i} \frac{1}{N} \sum (y - Q(s, a|\theta_i^Q))^2$$

        if t mod d **then**
            Update actor policy by deterministic policy gradient:

$$\nabla_{\theta^\mu} J = \frac{1}{N} \sum \nabla_a Q(s, a|\theta_1^Q)|_{a=\mu(s|\theta^\mu)} . \nabla_{\theta^\mu} \mu(s|\theta^\mu)$$

            Update the target network:

$$\theta_i^{Q'} \leftarrow \tau . \theta_i^Q + (1 - \tau) . \theta_i^{Q\prime}$$
$$\theta^{\mu'} \leftarrow \tau . \theta^\mu + (1 - \tau) . \theta^{\mu\prime}$$

        end if
    end for
end for

TD3 algorithm [130] is modified version of DDPG which uses six DNN: 1) Actor Network, 2) Two Critic Networks, 3) Target Actor Network, and 4) Two Target Critic Networks. However, while computing the loss, only one Critic network's estimated Q-value is used (minimum between the two critic networks) to eliminate the over estimation issues observed with DDPG (Algorithm 2).

C.3 Soft Actor-Critic (SAC) Algorithm

Algorithm 3 Soft Actor-Critic (SAC) Algorithm

Randomly initialize the critic networks $Q(s,a|\theta_1^Q)$, $Q(s,a|\theta_2^Q)$ and actor network $\mu(s|\theta^\mu)$ with weights θ_1^Q, θ_2^Q and θ^μ
Initialize the target networks $\theta_1^{Q'} \leftarrow \theta_1^Q$, $\theta_2^{Q'} \leftarrow \theta_2^Q$ and $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize the replay buffer R
for episode = 1, M **do**
    Observe the state s and select an action $a \sim \mu_{\theta^\mu}(\cdot|s)$
    Execute a in the environment and observe the next state and reward
    Store (s_t, a_t, r_t, s_{t+1}) in the replay buffer R
    for $t = 1, T$ **do**
        Sample random batch of N transitions from R
        Compute the targets for the Q functions:

$$y(r, s_{t+1}, d) = r + \gamma.(1-d)\Big(\min_{i=1,2} Q'(s_{t+1}, \tilde{a}_{t+1}|\theta_i^{Q'}) - \alpha.\log\mu'(\tilde{a}_{t+1}|s_{t+1})|_{\theta=\theta^{\mu'}}\Big),$$
$$\tilde{a}_{t+1} \sim \mu'(\cdot|s_{t+1})|_{\theta=\theta^{\mu'}}$$

        Update Q-functions by one step gradient descent using::

$$\nabla_{\theta_i^Q} \frac{1}{|N|} \sum_{(s_t,a_t,r_t,s_{t+1})\in R} (Q(s,a|\theta_i^Q) - y(r, s_{t+1}, d))^2$$

        Update the the policy by one step gradient ascent using

$$\nabla_{\theta^\mu} \frac{1}{|N|} \sum_{s\in B} \Big(\min_{i=1,2} Q(s, \tilde{a}_{\theta^\mu}(s)|\theta^\mu) - \alpha.\log\mu(\tilde{a}_{\theta^\mu}(s)|s)|_{\theta=\theta^\mu}\Big)$$

        Update target networks with:

$$\theta_i^{Q'} \leftarrow \tau.\theta_i^Q + (1-\tau).\theta_i^{Q\prime}$$
$$\theta^{\mu'} \leftarrow \tau.\theta^\mu + (1-\tau).\theta^{\mu\prime}$$

    end for
end for

SAC algorithm [131] is slightly different from DDPG and SAC in the sense that instead of developing a deterministic policy, it develops a probabilistic policy, meaning it outputs probabilities of taking a specific actions. However the Actor and Critic structure is similar to that of TD3 algorithm with network update equations provided in Algorithm 3.

APPENDIX D: Neural Networks

Neural Network (NN) is a computational model that derives its inspiration from the inner structure and working of a human brain. It is a type of machine learning framework that comprises of multiple layers of interconnected nodes called as artificial neurons (neurons). These neurons are responsible to process and transform the input data that can be used to generate output predictions [127].

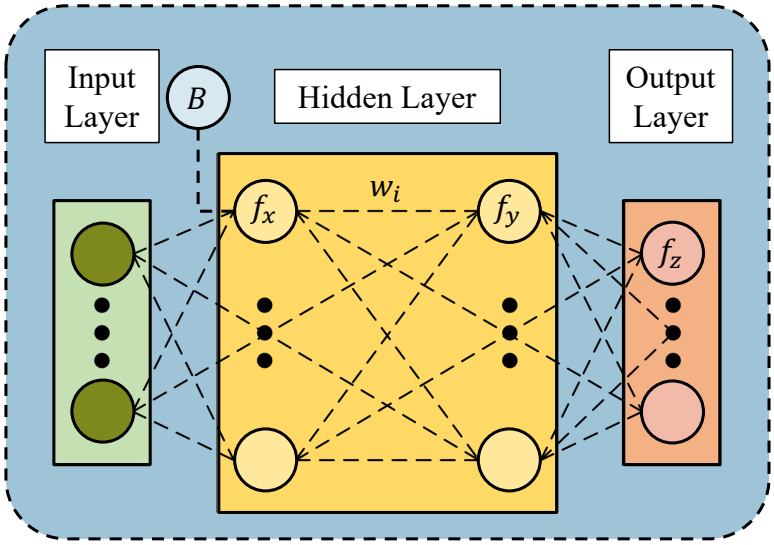

Figure D.1: Typical Neural Network

A typical NN is presented in Figure D.1 highlighting its structure that comprises of an input layer, multiple hidden layers, and an output layer. Key components of a NN are enlisted below:

- Neurons: Neurons are the basic building blocks of a NN that processes the input data using an activation function and propagates the processed data to the next layer.

- Layers: Layers in a NN provide a structure to interconnect neurons. There are three types of layers:

    1. Input layer: The layer that directly receives the input data.

2. Hidden layer: Intermediate layers that performs the computation to transform input data to a more appropriate representation.
3. Output layer: The layer that generates the final prediction value.

- Weights: A weight (w_i) in a NN is a component that signifies the importance of a specific neuron in the model development. Each of the connections between the neuron has a weight component associated with it which is optimized during the training process.
- Biases: Biases are constant values that are added to the computation to offset the results and to ensure zero value processing.
- Activation Function: An activation function (f_x, f_y, f_z) is a non-linear function applied at each neuron of the hidden and output layer to transform input data in a format that enables the network to recognize patterns and make appropriate decisions. The most commonly used activation function are:
  - ReLU function (Rectified Linear Unit)
  - Sigmoid function
  - Tanh function (Hyperbolic Tangent)

D.1 Training Process

The training of NN involve a multi-step process where the weights and the biases of the NN are adjusted to minimize the loss functions. A trained NN is capable of recognizing relationship between the input and output data that enables accurate predictions.

Step #1 of the training process involves random initialization of the NN where the weights of the NN are stochastically (or using certain methods) assigned. In Step #2, a forward propagation step is performed where the input data is passed through all the layers of the NN. Figure D.2 shows the forward propagation step of NN. The

values x_1, x_2, x_3 are the inputs whereas w_1, w_2, w_3 are the randomly assigned weights of the NN. At a specific neuron in a layer, the inputs are multiplied with weights and summed together (z from the Figure D.2). This sum-product is then passed through an activation function which provides a rescaled non-linear output (y) that is further forwarded to the other neurons which has a similar structure.

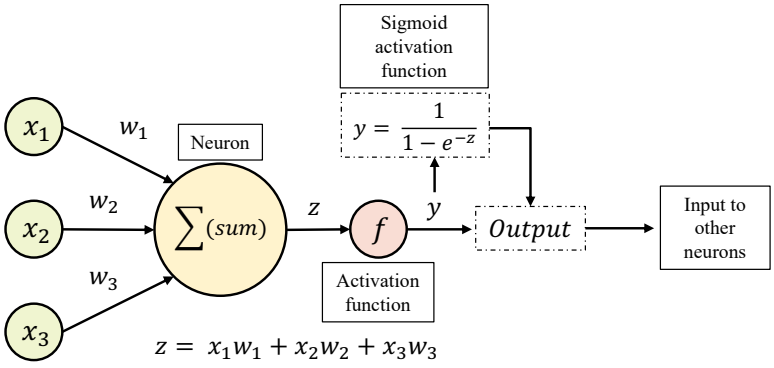

Figure D.2: Forward Propagation Step

Step #3 includes loss minimization where the output generated through the forward propagation is compared with actual value to evaluate how far off the predicted value is using a loss function. There are numerous loss functions out of which the most commonly used loss functions are:

- Mean Squared Error (MSE): $MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
- Cross-Entropy Loss (CE): $CE = -\frac{1}{N}\sum_{i=1}^{N}(y_i.\log(\hat{y}_i) + (1 - y_i).\log(1 - \hat{y}_i))$

To minimize the loss function, an algorithm called Back propagation is used which calculates the gradient of the loss function with respect to the individual weights of the NN by using the chain rule. Figure D.3 demonstrates different gradients that are computed along with the application of chain rule.

Let's assume the loss function used is MSE which is given by Equation (D.1). Here $S(y_i^k)$ is the rescaled output from the activation function activated neuron predicted by the NN; d_i^k is the actual value to be predicted; ϵ_k is the MSE value; N total

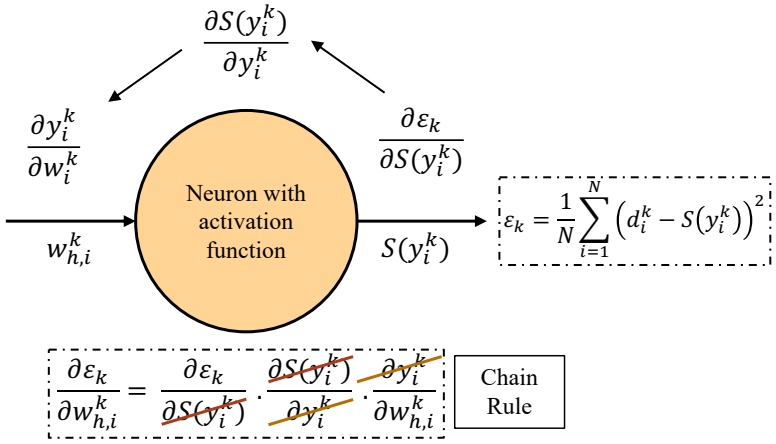

Figure D.3: Back Propagation and Chain Rule

number of values to be predicted (essentially the count of neurons); k is the index for the training epochs and h is the index for the sending end neuron ($w_{1,2}$: mean weight of the connection between Neuron #1 from previous layer and Neuron #2 of the current layer).

$$\epsilon_k = \frac{1}{N}\sum_{i=1}^{N}\left(d_i^k - S(y_i^k)\right)^2 \tag{D.1}$$

The goal of Step #4 is to ascertain the discrepancy between the predicted value and the actual value, which is done by calculating the gradient of the loss function with respect to the rescaled output, given by Equation (D.2) (the first term in the chain rule equation).

$$\frac{\partial \epsilon_k}{\partial S(y_i^k)} = -\frac{2}{N}(d_i^k - S(y_i^k)) \tag{D.2}$$

The second term is the gradient of the output from the neuron $S(y_i^k)$ with respect to the raw output value y_i^k. Equation (D.3) computes the aforementioned entity which is basically a partial derivative of the activation function. It is assumed that the activation function used is the Sigmoid function whose equation is represented as y in Figure D.2.

$$\frac{\partial S(y_i^k)}{\partial y_i^k} = -\frac{e^{-y_i^k}}{(1+e^{-y_i^k})^2} \tag{D.3}$$

The third term signifies how much the raw input y_i^k is changing with respect to the

weights $w_{h,i}^k$, where this relationship is expressed as Equation (D.4). Taking the partial derivative of Equation (D.4) with respect to weights provides us the interdependence between the NN's weights and the input values (Equation (D.5)).

$$y_i^k = \sum_{h=1}^{Q} w_{h,i}^k . S(z_h^k) \tag{D.4}$$

$$\frac{\partial y_i^k}{\partial w_{h,i}^k} = S(z_h^k) \tag{D.5}$$

Here, $S(z_h^k)$ is the output from the h^{th} neuron (activation function included). Using chain rule on Equations (D.2), (D.3) and (D.5), the relation between the change in loss function with reference to the NN's weights can be established as seen in Figure D.3 and Equation (D.6).

$$\frac{\partial \epsilon_k}{\partial w_{h,i}^k} = \frac{\partial \epsilon_k}{\partial S(y_i^k)} \cdot \frac{\partial S(y_i^k)}{\partial y_i^k} \cdot \frac{\partial y_i^k}{\partial w_{h,i}^k} \tag{D.6}$$

Once these values are computed the final step of the training process can be initiated which involves updating the weights of the NN. The weight update of a neuron is performed using gradient descent (if minimizing the loss function) or gradient ascent (if maximizing the total gains). As in the previous discussion, in this example, loss minimization is the aim. Therefore, Equation (D.8) provides the gradient descent equation to update the NN's weights.

$$w_{h,i}^{k+1} = w_{h,i}^k - \Delta w_{h,i}^k \tag{D.7}$$

$$= w_{h,i}^k - \alpha . \Big(\frac{\partial \epsilon_k}{\partial w_{h,i}^k} \Big) \tag{D.8}$$

Where, α is the learning rate for weight update rule that signifies how fast we update the weights. Thus, setting an appropriate α is crucial as it defines the NN optimization step size.