QATAR UNIVERSITY

COLLEGE OF ENGINEERING

TWO-TIMESCALE MULTI-OBJECTIVE VOLT/VAR OPTIMIZATION CONSIDERING DISTRIBUTED ENERGY RESOURCES IN ACTIVE DISTRIBUTION NETWORKS

BY

TASSNEEM AHMED ZAMZAM

A Thesis Submitted to

the College of Engineering

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Electrical Engineering

June 2021

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Tassneem Zamzam defended on 07/04/2021.

Dr. Ahmed Massoud
Thesis Supervisor

Dr. Khaled Bashir Shaban
Thesis Co-Supervisor

Dr. Adel Gastli
Committee Member

Dr. Mohamed Shaaban
Committee Member

Dr. Amr Mohamed
Committee Member

Approved:

Dr. Khalid Kamal Naji, Dean, College of Engineering

ABSTRACT

ZAMZAM, TASSNEEM, A., Masters: June:[2021:], Masters of Science in Electrical Engineering

Title: Two-Timescale Multi-Objective Volt/Var Optimization Considering Distributed Energy Resources in Active Distribution Networks

Supervisor of Thesis: Ahmed, M, Massoud.

The high penetration of distributed energy resources (DERs) introduces several challenges to the power network. They may cause a high level of voltage variation, sudden over/under-power generation, high power losses, and negatively impacted distribution assets. Thus, there is a vital need for volt/var optimization (VVO) schemes that integrate utility-owned assets with inverter-interfaced resources to overcome these challenges. This thesis addresses the above-mentioned challenges by proposing a comprehensive two-timescale multi-objective VVO algorithm. The slow timescale utilizes utility-owned assets to minimize system losses and maximize asset lifetime in a three-step methodology. This stage incorporates the utility operator's direct input for setting the utility-owned assets. At the faster scale, the algorithm optimizes the reactive power of DERs to minimize voltage variations and system losses. The proposed VVO is solved using conventional optimization and reinforcement learning algorithms. The IEEE 33-bus system is modified and used to demonstrate the effectiveness of the proposed algorithms.

DEDICATION

To mom and dad

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to Prof. Ahmed Massoud and Dr. Khaled Shaban. This thesis would not have been possible without their guidance, motivation, and great supervision.

This thesis was made possible by NPRP11S-1202-170052 grant from Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

TABLE OF CONTENTS

DEDICATION iv

ACKNOWLEDGMENTS v

LIST OF TABLES xi

LIST OF FIGURES xiii

NOMENCLATURE xvi

Chapter 1 : Introduction 1

1.1 Motivations 1

1.2 Distributed Energy Resources 2

1.2.1 Interface-Inverter of DERs 4

1.2.2 Challenges and Benefits Associated with DERs 4

1.3 Active Distribution Networks 6

1.4 Volt/Var Optimization in Active Distribution Networks 8

1.4.1 Volt/Var Optimization Utilizing DERs 9

1.5 Reinforcement Learning 11

1.6 Thesis Objectives 14

1.7 Thesis Contributions 17

1.8 Thesis Outline 18

Chapter 2 : Review of Related Work 20

2.1 Historical View on Volt/Var Optimization 20

2.2 Volt/Var Management Market 22
2.3 Studies on Solving the Volt/Var Optimization Problem 23
2.3.1 The Utilization of Utility-Owned Assets in VVO Algorithms 24
2.3.2 The Utilization of DERs VVO based Algorithms 25
2.3.3 Multi-Timescale VVO based Algorithms 26
2.4 Studies on Solving Volt/Var Optimization using Reinforcement Learning .27
2.5 Summary 31
Chapter 3 : Volt/Var Optimization Topology and Resources 33
3.1 Distribution Network Structure 33
3.2 VVO Main Regulating Components 35
3.2.1 Voltage Controlling Components 36
3.2.2 Reactive Power Controlling Components 38
3.2.3 Distributed Energy Resources 39
3.3 Summary 46
PART I: Conventional Optimization-Based Volt/Var Optimization Scheme 48
Chapter 4 : The Two-Timescale Volt/Var Optimization Model 49
4.1 Proposed VVO Problem Formulation 49
4.1.1 Stage I: h Time-Resolution Multi-Objective Formulation 49
4.1.2 Stage II: h/n Time-Resolution Single Objective Formulation 52
4.2 Solutions for the VVO Problem 53

4.2.1 Genetic Algorithm Features and Components 55
4.2.2 Multi-objective Genetic Algorithm 57
4.2.3 Interior Point Method 59
4.3 Proposed VVO Algorithm Framework 60
4.3.1 Stage I: Within the h Time-Resolution 61
4.3.2 Stage II: At the h/n Time Resolution 63
4.4 Summary 63
Chapter 5 : The Two-Timescale Volt/Var Optimization Solution 64
5.1 Test System Description and Parameter Settings 64
5.1.1 Case Study A 64
5.1.2 Case Study B 67
5.2 Simulated Cases 69
5.3 Code Implementation 70
5.3.1 Stage I: Hourly Basis 71
5.3.2 Stage II: 15-min Basis 74
5.4 Proposed two-timescale results and discussions 74
5.4.1 Stage I: Hourly basis 74
5.4.2 Stage II: 15-minute basis 81
5.5 Evaluation and Validation 86
5.6 Summary 89

5.7 Limitations 90

PART II: Reinforcement Learning-Based Volt/Var Optimization Scheme 91

Chapter 6 : Reinforcement Learning 92

6.1 Reinforcement Learning Framework 92

6.2 Markov Decision Process 93

6.3 Deep Reinforcement Learning 95

6.4 Reinforcement Learning Taxonomy 96

6.5 Summary 99

Chapter 7 : The DRL-Based Volt/Var Optimization Scheme 100

7.1 VVO Problem Formulation as MDP 100

7.2 Twin-Delayed Deep Deterministic Policy Gradient (TD3) RL Algorithm. 103

7.3 The Workflow of the Proposed TD3 Algorithm 106

7.4 DRL-based VVO results and discussions 109

7.4.1 Case study description 109

7.4.2 Numerical analysis – Learning setup 111

7.4.3 Numerical analysis – Learning performance 113

7.4.4 Numerical analysis – VVO performance 114

7.4.5 Numerical analysis – Computation time 116

7.5 Summary 116

Chapter 8 : Thesis Conclusion and Future Work 118

8.1 Summary 118

8.2 Contributions 120

8.3 Future Work 121

References 124

Appendix 154

LIST OF TABLES

Table 1-1. Comparison Between Passive and Active Distribution Networks [26].....8

Table 1-2. Comparison Between Different Volt/Var Technologies [43] 10

Table 2-1. Comparison of Objectives and Resources of Multi-timescale VVO Approaches 27

Table 2-2. Comparison of Related DRL-Based Approaches.....31

Table 5-1. Case Study A System Parameters.....65

Table 5-2. Case Study B System Parameters.....68

Table 5-3. Number of Control Actions of CBs and OLTC for C1 & C2 – Case Study A 75

Table 5-4. Number of Control Actions of CBs and OLTC for C1 & C2 – Case Study B 76

Table 5-5. Case Study A Performance Evaluation of Stage I – Power Loss 79

Table 5-6. Case Study B Performance Evaluation of Stage I – Power Loss 79

Table 5-7. Case Study A Performance Evaluation of Stage I – Max. & Min. Voltages 81

Table 5-8. Case Study B Performance Evaluation of Stage I – Max. & Min. Voltages 81

Table 5-9. Case Study A Maximum Utilization of DERs in the Day In % - Based on Information in Table 5-1 and Figures 5-1 and 5-2.....83

Table 5-10. Case Study B Maximum Utilization of DERs in the Day In % - Based on Information in Table 5-2, Figure 5-4 and Figure 5-5.....85

Table 5-11. Performance of Reduction In Power Loss -Stage I 87

Table 5-12. Performance of Voltage Limitation – Stage I.....87

Table 5-13. Performance of Voltage Deviation - Stage II 88

Table 7-1. Setting Parameters of TD3 112

Table 7-2. Statistical Analysis of Proposed and GA Algorithms 115

LIST OF FIGURES

Figure 1-1. Classification of distributed energy resources [11] 3

Figure 1-2. Structure of (a) passive distribution networks, (b) active distribution networks [34] 7

Figure 1-3. Types of machine learning algorithms [52] 12

Figure 1-4. Conceptual diagram of the proposed VVO algorithm 15

Figure 1-5. Thesis outline schematic 19

Figure 2-1. Road to Volt/Var optimization 21

Figure 2-2. Volt/var management market [118] 23

Figure 2-3. Categorization of VVO algorithms. 24

Figure 3-1. OLD of a distribution network 33

Figure 3-2. OLD of radial feeders. (a) OLD of radial primary feeder, (b) OLD of the radial secondary feeder 34

Figure 3-3. Classification of losses in a distribution network [52][143] 35

Figure 3-4. Structure of an OLTC [158] 38

Figure 3-5. Structure of a pole-mounted CB [157] 39

Figure 3-6. PV capability curve 41

Figure 3-7. EV charging classifications [134] 42

Figure 3-8. Operating modes of EV charging stations [161]-[163] 43

Figure 3-9. Main components of a BESS [166] 45

Figure 4-1. Pareto front for two objectives [131] 57

Figure 4-2. Flowchart of the NSGA-II algorithm [132] 59

Figure 4-3. Flowchart of IPM algorithm [194] 60

Figure 4-4. Main steps of the proposed two-timescale VVO algorithm 61

Figure 5-1. Case study A modified IEEE 33-bus system 65
Figure 5-2. Case study A input data (a) EV charging stations power demand, (b) forecasted demand load, and (c) forecasted DER data 66
Figure 5-3. Case study A Q limitation of DERs 67
Figure 5-4. Case study B modified IEEE 33-bus system..... 68
Figure 5-5. Case study B input data (a) EV charging stations power demand, (b) forecasted demand load, and (c) forecasted DER data 69
Figure 5-6. VVO studied cases 70
Figure 5-7. Flowchart of stage I; step 1 of the developed algorithm 72
Figure 5-8. Case study A Pareto front solutions of C1 and C2..... 76
Figure 5-9. Case study B Pareto front solutions of C1 and C2..... 76
Figure 5-10. Case study A power loss during different stages for C0 and C2-C4 78
Figure 5-11. Case study B power loss during different stages for C0 and C2-C4..... 79
Figure 5-12. Case study A maximum and minimum network voltages after stage I... 80
Figure 5-13. Case study B maximum and minimum network voltages after stage I... 81
Figure 5-14. Case study A maximum and minimum network voltages after stage II . 83
Figure 5-15. Case study A reactive output power of DERs of stage II 83
Figure 5-16. Case study B voltage deviation profiles of network before optimization and after stage II considering a) Vref= 1.02 p.u; b) Vref= 1.0 p.u; c) Vref= 0.98 p.u; and d) voltage deviation curves of the three considered Vrefs. 84
Figure 5-17. Case study B reactive output power of DERs of stage II..... 84
Figure 5-18. Case study A voltage profiles of buses 18, 22, and 29 after different stages 86
Figure 5-19. Case study B voltage profiles of buses 18, 22, and 29 after different stages

.....86
Figure 5-20. Real and reactive power mismatches of the developed algorithm.....88
Figure 6-1. Reinforcement learning framework [181].....93
Figure 6-2. Deep reinforcement learning framework [182].....95
Figure 6-3. Reinforcement learning taxonomy [145] 98
Figure 7-1. Two-timescale segregating for controlling the different VVO resources 101
Figure 7-2. Architecture of proposed TD3 algorithm.....106
Figure 7-3. Modified IEEE 33-bus system - testing RL 110
Figure 7-4. Neural network architecture a) critic architecture, b) actor architecture 112
Figure 7-5. Learning process of the proposed TD3 agent 113
Figure 7-6. Comparison of the proposed VVO TD3 based algorithm and GA – power loss 115
Figure 7-7. Comparison of the proposed VVO TD3 based algorithm and GA – voltage deviation.....116

NOMENCLATURE

nl	Number of branches.
${P^{st}}_{k,t}$	Active power at the start of the k_{th}branch at time t.
${P^{en}}_{k,t}$	Active power at the end of the k_{th}branch at time t.
${CB_m}^{actions}$	Number of control actions of the m_{th} CB.
${OLTC_n}^{actions}$	Number of control actions of the n_{th} OLTC.
M	Total number of CBs in the network.
N	Total number of OLTCs in the network.
V_{min}	Lower voltage limit according to the ANSI standard.
V_{max}	Higher voltage limit according to the ANSI standard.
$V_{i,t}$	Voltage of the i_{th} bus at time t.
I_{max}	Maximum current in lines.
$I_{i,t}$	Current in the i_{th} line at time t.
V	Number of PV plants in the network.
E	Number of EV charging stations in the network.
B	Number of BESS units in the network.
L	Number of the network lines.
R	Number of busses in the network.
${P_{i,t}}^{lineloss}$	Power loss in the i_{th} line at time t.
${P_{i,t}}^{load}$	Active power load at the i_{th} bus at time t.
${P_{x,t}}^{EV}$	Active power requirement of the x_{th} EV station at time t.
${P_t}^{Grid}$	Active power supply from the grid at time t.

$P_{x,t}^{PV}$	Active power generation of the x_{th} PV at time t.
$P_{x,t}^{BESS}$	Active power charged or discharged of the x_{th} BESS at time t.
$Q_{i,t}^{line}$	Reactive power in the i_{th} line at time t.
$Q_{i,t}^{load}$	Reactive power load at the i_{th} bus at time t.
Q_t^{Grid}	Reactive power supply from the grid at time t.
$Q_{x,t}^{EV}$	Reactive power absorbed or supplied from the x_{th} EV station at time t.
$Q_{x,t}^{PV}$	Reactive power absorbed or supplied from the x_{th} PV at time t.
$Q_{x,t}^{BESS}$	Reactive power absorbed or supplied from the x_{th} BESS at time t.
$Q_{x,t}^{Cap}$	Reactive power supplied by the x_{th} CB at time t.
$S_{f,t}$	kVA of the f_{th} line at time t.
$S_{f,t,max}$	Maximum thermal limit of the f_{th} line at time t.
PF_{min}	Minimum power factor of a node in the network.
$PF_{i,t}$	Power factor of the i_{th} node at time t.
PF_{max}	Maximum power factor of a node in the network.
$q_{s,i,t}$	Setting position of the i_{th} CB at time t.
Q_c	Unit size of an individual CB.
$Q_{Cap,i,t}$	Reactive output power of the i_{th} CB at time t.
$Q_{Cap,i,\max}$	Maximum rating of the i_{th} CB.
V_S	Primary substation voltage.
T_t	Tap position at time t.
V_{Tap}	Tap voltage adjustment.

$V_{0,t}$	Secondary output voltage of the OLTC at time t.
T_{min}	Minimum tap position of OLTC.
T_{max}	Maximum tap position of OLTC.
$T_{i,t}$	Tap position of the i_{th} OLTC at time t.
${S_{max}}^{PV}$	Rating of the PV inverter.
$P_{EV,i,t}$	Active power of the i_{th} EV at time t.
$P_{EV,max}$	Maximum charging limit of an EV.
${S_{max}}^{EV}$	Rating of the EV charging station inverter.
${S_{max}}^{BESS}$	Rating of the BESS inverter.
$E_{x,t}$	Energy level of the x_{th} BESS unit at time t.
${P_{x,t}}^{Charge}$	Active power charges the x_{th} BESS unit at time t.
${P_{x,t}}^{Discharge}$	Active power discharges the x_{th} BESS unit at time t.
η^{Charge}	Charging efficiency of the BESS unit.
$\eta^{Discharge}$	Discharging efficiency of the BESS unit.
$V_{p,t}$	Voltage of the pth bus at time t.
V_{ref}	Reference voltage.
ε	Epsilon value of the ε-constraint method.

CHAPTER 1 : INTRODUCTION

This chapter starts with highlighting the motivations behind the research work of the thesis. It then provides backgrounds on distributed energy resources, their impact on distribution networks, and the challenges and benefits associated with them. It also introduces active distribution networks and their emerging concerns, as well as volt/var optimization in active distribution networks. Finally, it states the research objectives and contributions of the thesis.

1.1 Motivations

The core function of utility operators is to fulfill customers' requirements by providing standardized electrical services while minimizing network losses. However, the increasing penetration of distributed energy resources (DERs) technologies of battery energy storage systems (BESSs), electric vehicles (EVs), and renewable energy sources (RESs) such as wind turbines (WTs) and photovoltaics (PVs), has increased the complexity of power network planning and operation. A primary concern due to such technologies' high usage is the undervoltage and overvoltage events [1]. DERs may allow electric power to flow back to the upstream transformer causing bus voltages to rise. This may damage customers' electrical appliances and may cause equipment malfunction and reduce energy efficiency. Furthermore, DERs output power's uncertainty may lead to voltage variations and sudden over-/under-generation that may drive the network to be out of synchronization [2]. Thus, there is a vital need for coordinating schemes to overcome these challenges.

To improve voltage profile, volt/var optimization (VVO), volt/var control (VVC), and conservation voltage reduction (CVR) are employed [3]. These algorithms consider distribution system switch-based assets that include capacitor banks (CBs), step voltage regulators (SVRs), and transformer on-load tap changers (OLTCs) to

optimize the network power flow. System data of a wide time-resolution is utilized to limit voltage variations and control injected reactive power to improve voltage profile, reduce system losses, and improve energy saving [4]-[8]. The time response of distribution system switch-based assets is, however, considered to be relatively slow. Moreover, due to the stochastic nature and fast fluctuations of DERs' generation, their employment to reduce the adverse impacts on a distribution network may not be effective [9]. On the other hand, the interface-inverters of DERs are power electronic devices that have faster time response. That is, they are capable of handling the fast-changing components of demand and generation. Therefore, the utilization of the reactive power of DERs to support the network will lead to improving network performance. Furthermore, their adoption will decrease the dependency on the distribution network switched-assets, reducing their wear-and-tear effect.

1.2 Distributed Energy Resources

In numerous aspects of generation, demand, and control, electricity grids have been transformed over the last few decades. In order to reduce greenhouse emissions, renewable generations gradually replaced fossil fuel generations. Moreover, with the continuation of electrification and digitization of devices and resources in people's daily life, advanced technologies such as electric vehicle charging and computing services are emerging. Due to these technological developments and environmental protections, DERs, such as BESSs, internal combustion engines, EVs, and distributed generations (DG) such as gas turbines, microturbines, PVs, fuel cells, and WTs, have emerged within the distribution system. Their capacities, however, are much smaller than conventional generation units. To meet the rising energy demand, DERs are installed near the loads in distribution networks. To offer enough power to new loads and new areas without expanding grids, the locations and capacities of DERs, which are

connected to the system, are optimized [10]. The DERs can be broadly classified into three categories: 1) distributed energy technologies that comprise conventional and non-conventional generators, 2) energy storage technologies of different energy storage systems, pumped storage and flywheels, and 3) e-transportation, including EVs, electric aircrafts, electric ships, and electric trains. The classification of DERs is presented in Figure 1-1.

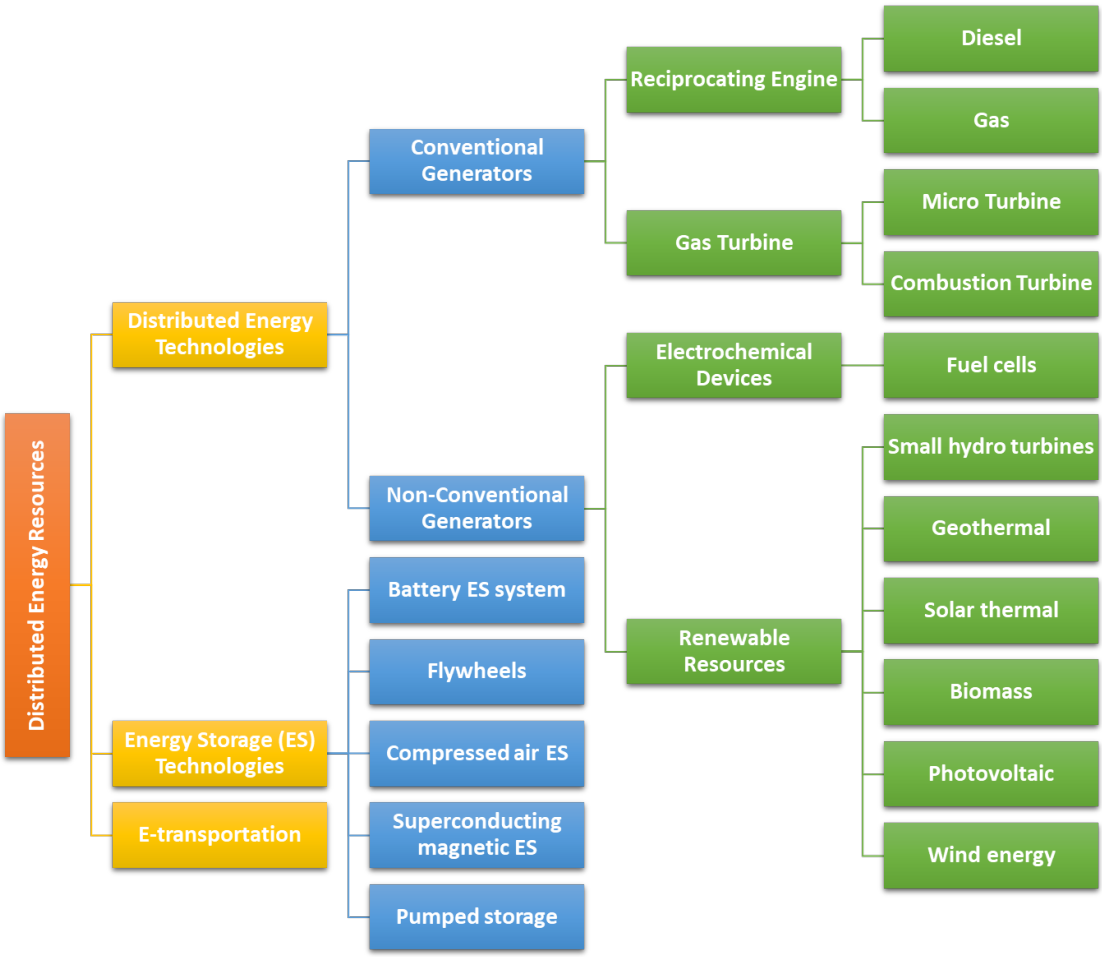


Figure 1-1. Classification of distributed energy resources [11]

According to the U.S. Energy Information Administration, EIA-861M, statistics and analysis report, the total residential PV generation capacity was 3,346MW in 2014 and has increased to 13,526MW in 2019 [12]. Furthermore, BESS's capacity in the U.S.

was only 862MW at the end of the year 2018 but is estimated to be over 2,500MW by 2023 [13]. Likewise, EV sales escalated from 118k in 2014 to 358k in 2018 to benefit from the advanced BESS technology [14]. The substantial amount of EVs and BESSs charging demand has increased stability issues in the distribution grid. Nonetheless, if properly coordinated together along with renewable resources, EVs will increase operational efficiency.

1.2.1 Interface-Inverter of DERs

Recently, with the proliferation of DERs, their associated inverters are widely applied to convert the DC power, which is generated by renewable resources, to AC power that can be injected into the system. For example, solar inverters convert the DC output of PV plants to AC, then feed the power into the network. Appropriate design and operation of inverters can improve the efficiency of power systems. Reference [15] proposed an inverter topology of single-stage operation considering battery charging capability to increase renewable sources' usage rate. In [16], the authors proposed connecting the PV inverter to the network using a phase-locked loop control system for more effective maximum power point tracking, which increased system efficiency. A multilevel transformer-less PV inverter is proposed in [17] to overcome the leakage current issue of the single-stage inverters. Further, results showed better system performance in terms of lower electromagnetic interference (EMI), lower total harmonic distortion (THD), and lower switching stress.

1.2.2 Challenges and Benefits Associated with DERs

When the number of DERs increases, the challenges of incorporating them into the distribution network increases as well. Both their bidirectional power flow and stochastic nature may lead to high voltage fluctuations [18]. The pressing challenge here is how to reduce and manage the negative effects of DERs, like high voltage

deviations. Keeping in mind that in conventional distribution networks, all voltage controlling devices have been principally designed to perform in the absence of DERs, and that voltage magnitudes are assumed to be decreasing along the distribution feeder from substations to customers. This assumption becomes no longer authentic after the integration of DERs. The bidirectional power flow led to bus voltages along the distribution feeder, violating these assumptions [19].

The impacts of DER decrease or increase comparatively depending on their location, size, and penetration level. These impacts may negatively affect the lifetime of the voltage controlling and reactive power compensating devices used along the distribution feeders, like CBs, SVRs, and OLTCs, due to the high DER output fluctuations. In addition, protection devices may be adversely affected since they are not designed to operate in such conditions. They can also increase the potential risk of overloading transformers due to, for example, adopting a large number of EVs and cause reverse power flow to substations due to an imbalance of demand and generation. The major significant challenges and impacts that have been addressed in literature concerning planning, controlling, and monitoring distribution networks with DERs include power flow, protection, fault current levels, thermal equipment ratings, and voltage magnitude levels issues [20].

On the other hand, the introduced benefits of DERs to electric utilities have been vividly increased. Most electric utilities discovered that the effective use of DERs may reduce the cost of injecting reactive power into distribution networks. Thus, electric utilities will be positively impacted by their efficient use. Several advantages have been realized from the employment of DERs in distribution networks, for instance, minimizing feeder losses, supporting the network voltage, reducing greenhouse gas emissions, and increasing the system reliability [21]–[23]. Furthermore, for economic

benefits, it is found that DERs can efficiently reduce both operation and maintenance costs and increase operating profit [24]. Optimal placement of DERs can also decrease the investment of network upgrading and expansion [25]. Moreover, DERs are capable of assisting the energy supply for loads of the distribution networks.

1.3 Active Distribution Networks

A conventional power distribution network is a component of a power delivery system that links the customers and transmission system. It has been operated initially and designed based on several fundamental assumptions that designers evaluate the distribution networks' efficiency and reliability. It is designed to transport power from the electric grid -high voltage side- to the end-users -low voltage side-. Further, it is designed based on low energy losses, currents and voltages within permissible ranges, minimum consumption, centralized generation, and unidirectional power flow [26]. Recently, due to the presence of high DER penetrations, the power flow in distribution networks is transformed from unidirectional power flow in passive distribution networks into bidirectional power flow in active distribution networks (ADNs), also known as the Smart Grids [19].

ADNs include protection and communication systems to achieve technical and economic benefits to systems by optimally managing the electrical energy. ADNs can flexibly tackle faults to enhance reliability and robustness [27]. Flexible resources can reduce energy outages, and energy dispatch strategies can be improved [28]. By controlling DG, storage devices, and active power, the operation of ADNs can be more efficient, and the operation cost can be reduced [29]. Flexible control methods and multiple components can help ADNs improve voltage stability and voltage profiles [30],[31]. The active power source can be controlled to reduce power loss, and the efficiency of ADNs would be improved [32]. New market mechanisms, such as peer-

to-peer trading, can be designed to benefit both consumers and prosumers [33]. Besides these advantages, ADNs also embody the benefits of all components which are applied in ADNs. The structure of passive and active distribution networks is shown in Figure 1-2.

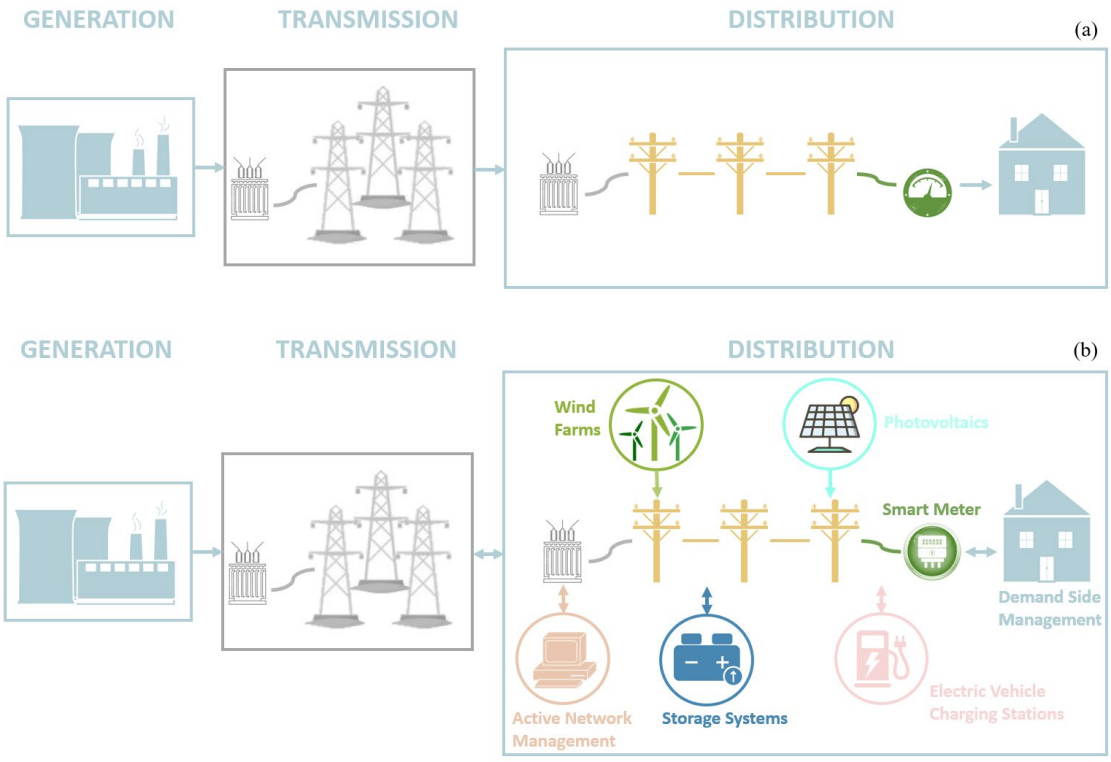


Figure 1-2. Structure of (a) passive distribution networks, (b) active distribution networks [34]

Overall, ADNs take advantage of advanced communication systems, sensors, measuring units, feedback control, and intelligent systems to improve the network's performance. Table 1-1 compares active and passive distribution networks based on several aspects. The benefits ADNs can contribute to are: enhancing capacity and efficiency of existing networks, improving power reliability and quality, avoiding construction of peak load power plants, optimizing facility utilization, improving

resilience to disruption, integrating DERs to the grid, enabling predictive maintenance, allowing self-healing responses to system faults and disturbances, reducing greenhouse gas emissions, facilitating the expanded deployment of RESs, automating maintenance and operation, and decreasing the need for inefficient generation during peak demand periods.

Table 1-1. Comparison Between Passive and Active Distribution Networks [26]

Feature	Passive Distribution Network	Active Distribution Network
Communication	One-way	Two-way
Metering	Mechanical	Digital
Generation	Centralized	Distributed
Restoration	Manual	Self-healing
Monitoring	Manual check	Remote check
Control	Limited	Pervasive

1.4 Volt/Var Optimization in Active Distribution Networks

The main aim of VVO employment in distribution networks is to manage the uncertainty of power injections and loads and maintain convenient network voltages by supplying or absorbing reactive power as required. VVO actions have been achieved in conventional distribution networks through voltage regulating and reactive power compensation devices such as CBs, OLTCs, and SVRs. These VVO devices are designed to operate based on assumptions like unidirectional power flow that adopts the fact that bus voltages decrease along the distribution network within the permissible range stated by the American National Standards Institute (ANSI).

The coordinated VVO approaches that utilize conventional reactive power and voltage control devices in distribution networks have been extensively studied and investigated by researchers to minimize network losses or energy consumption. For

instance, the settings of SVRs and CBs are configured to maintain network bus voltages within the permissible range under different load conditions in [35]. Another study demonstrated that determining optimal settings of OLTCs positions and CBs hours in advance reduces active power loss [36]. The system's voltage stability can be improved by controlling voltage-controlling equipment [37]. Different control methods in distribution networks are analyzed based on VVO strategies [38]. After the development of ADNs, uncoordinated VVO algorithms led to excessive use of OLTCs and CBs, causing a reduction in their life expectancy [27]. Therefore, VVO algorithms that reduce their usage are required to relieve the stress on conventional volt/var devices.

Besides, considering decreasing costs and increasing interests in energy sources substitutes other than fossil-fuel-based sources, usage of DERs in transmission and distribution systems has increased significantly in recent years [39]. The integration of DERs in distribution networks has increased conventional VVO devices' challenges to cope with the initially-set assumptions for distribution networks. With the increment of the level of penetrations and stochastic nature of DERs, conventional feeders are too slow to manage rapid variations in DGs' output power. Conventional voltage controlling systems, such as local static var sources, are expensive and cannot cope with rapid fluctuations of output powers of DERs [40]. Therefore, a VVO algorithm that coordinates between DERs and conventional volt/var devices is vital.

1.4.1 Volt/Var Optimization Utilizing DERs

With the rapid advancement in DER inverter technologies, electric utilities discovered that inverter-based DERs are alternative reactive sources and can contribute to solving the introduced issues in ADNs, such as the rapid variations of network voltages. Based on the fact that inverter-based DERs are power electronic devices, they

can provide the required reactive power in less than 50 milliseconds to manage rapid voltage variations, for instance, as in [41], [42] handling transient cloud passing to PVs circumstances. Incorporating DERs has decreased the reliance on conventional distribution network devices such as OLTCs, CBs, STATCOM, and static var compensators. Moreover, CBs can only provide support to the network by supplying reactive power. However, it cannot consume reactive power. On the contrary, DERs inverters can support the network more efficiently by injecting or absorbing reactive power with faster response times. Thus, more flexible reactive power support can be attained [40]. A comparison between different volt/var technologies is provided in Table 1-2.

Table 1-2. Comparison Between Different Volt/Var Technologies [43]

Technologies	Inverters	Capacitor Banks	On-load Tap Changers	STATCOM
Time Response	Rapid	Slow	Slow	Rapid
Output Capacity	Flexible	Discrete	Discrete	Flexible
Initial Cost	Low	Moderate	Low	High

Coordinated reactive power compensation, attained from conventional VVO devices along with inverter-based DERs, can enhance system reliability and efficiency and provide substantial economic benefits to electric utilities [44]. For instance, utilizing PV inverters to increase line capacities and reduce network losses [45]. An optimal VVO using the PV inverters' capability to inject and consume the reactive power to reduce energy consumption and network losses while sustaining network voltages within the acceptable range is proposed in [46].

Another objective of VVO is to maintain voltage deviation within an adequate

range by utilizing distribution network devices and DERs. OLTCs and PV inverters control bus network voltages for either global or local control along the distribution feeder [47]. Inverter-based PVs can provide reactive power compensation to supply voltage support when generation fluctuations occur [48]. VVO can also conserve energy by preserving network voltages within the allowed range [49]. By controlling voltage magnitudes based on ANSI standards, electric utilities can efficiently deliver energy [50]. As a result, they will conserve money through the minimization of total power losses in distribution feeders [51].

1.5 Reinforcement Learning

In literature, machine learning approaches are classified based on the expected learning tasks and the available data. Figure 1-3 demonstrates the three top categories of machine learning algorithms. Which include 1) supervised learning (immediate feedback), where the training dataset includes both the inputs and the desired outputs, and the model tries to learn the functional relationship between them [52], 2) unsupervised learning (no feedback) where no clear output from the training dataset is provided, and the model learns similarities between input data and makes actions accordingly, and 3) Reinforcement learning (delayed feedback) where no output is provided and the model tries to learn the input-output relationship from the data.

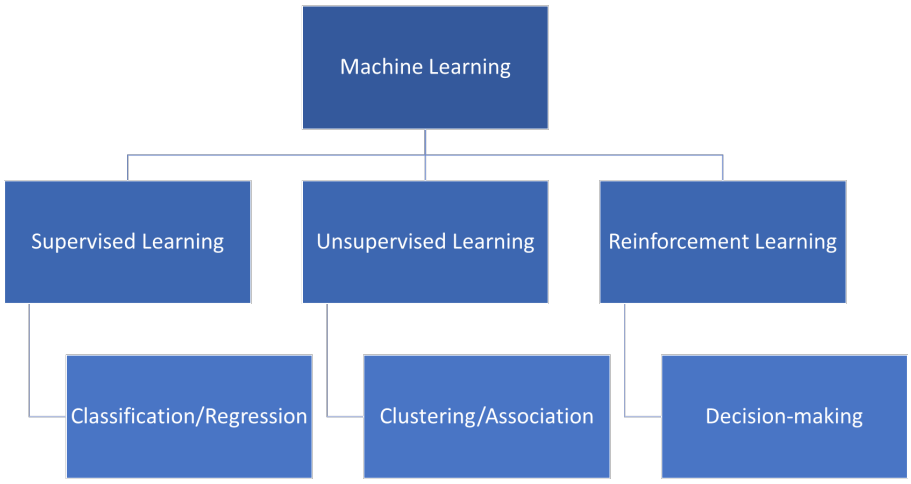


Figure 1-3. Types of machine learning algorithms [52]

Reinforcement learning, derived from neutral stimulus and response, is a machine learning approach capable of learning from scratch and relies on environmental interaction rather than complex mathematical models. Due to its success in addressing challenging sequential decision-making problems, it has become increasingly popular [52], [54]. Its combination with deep learning, called deep reinforcement learning (DRL), has achieved significant successes in games [55]–[57], robotics [58], [59], natural language processing [60], [61], finance and business management [62], [63]. Recently, few researches have reported the application of DRL in power systems and that it outperforms traditional approaches in overcoming the introduced challenges in smart grids. Traditional approaches mainly include convex optimization methods, programming methods, and heuristic methods.

Convex optimization methods are classical mathematical methods, such as the Lyapunov optimization algorithm [64]. The advantage of this method is that the mathematics are rigorous, and real-time management can be realized. However, this method relies on explicit objective functional expressions, difficult to abstract from many real-world optimization decision scenarios. Moreover, the Lyapunov condition

(required for the Lyapunov optimization algorithm) cannot be guaranteed in complicated, high-dimensional scenarios. While the programming methods, such as stochastic programming [65], [66], mixed integer programming [67], [68], and dynamic programming [69], [70], can handle diverse optimization problems. However, they tend to have high calculation costs and lack of realizing real-time decision-making in some scenarios. Also, programming algorithms rely on accurate predictions of load and renewable energy generation, which are difficult to achieve in real scenarios. Heuristic methods, such as ant colony optimization (ACO) [71], [72], PSO [73], [74], and GA [75], [76], are also able to handle diverse optimization problems. Especially for non-convex optimization problems, a heuristic method can achieve the local optimal solution with a certain probability, which helps solve large data scale and complicated scenarios. However, these methods are less robust and cannot be proven rigorously using mathematics.

Compared with convex optimization methods, the exact objective function is not necessary for DRL. In contrast, DRL uses a reward function to evaluate decision behavior. DRL can also handle higher dimensional data than convex optimization methods. Against the programming methods, DRL makes decisions according to the current state and makes real-time and online decisions. It can also handle uncertainty in the load and renewable energies profiles. In contrast to the heuristic methods, DRL is more robust with stable convergence results and is better suited for decision-making problems. Also, it has been proven that it can escape local optimal solutions since it performs stochastic optimization. Besides, the traditional approaches rely on the exact power system model, while DRL requires less accurate or no information about the system model.

Many researches have recently addressed DRL applications in power systems,

and most of them are published since 2018. These applications cover a wide range of decision, control, and optimization problems in the power systems, including energy management [77]-[80], demand response [81]-[84], electricity market [85]-[88], operational control [89]-[92], cybersecurity [93]-[95], economic dispatch [96], [97], system optimization [98], edge computing [99], energy routing [100] and many others.

1.6 Thesis Objectives

This work aims to develop a VVO algorithm that minimizes power loss and voltage deviation in an ADN. The algorithm will consider the volt/var impact of conventional devices of CBs and OLTCs along with DERs, including RESs, EV charging stations, and BESS. It will optimize conventional assets and DERs settings at multi-timescales to limit network voltage deviation, reduce system losses while considering the distribution transformer lifetime. The system operators' input on CBs and OLTC settings is also included in the optimization. Figure 1-4 describes the overall concept of the proposed VVO algorithm. The VVO controller comprises of two-timescales optimizers. The slow timescale optimizer sends control decisions to set the position of CBs and OLTCs in the ADN, based on demand and generation, to achieve power loss and number of position adjustments of CBs and OLTCs minimization simultaneously. While the fast timescale optimizer uses the settings of the CBs and OLTCs and sends control decisions to schedule the amount of reactive output power of DERs to achieve voltage variation minimization.

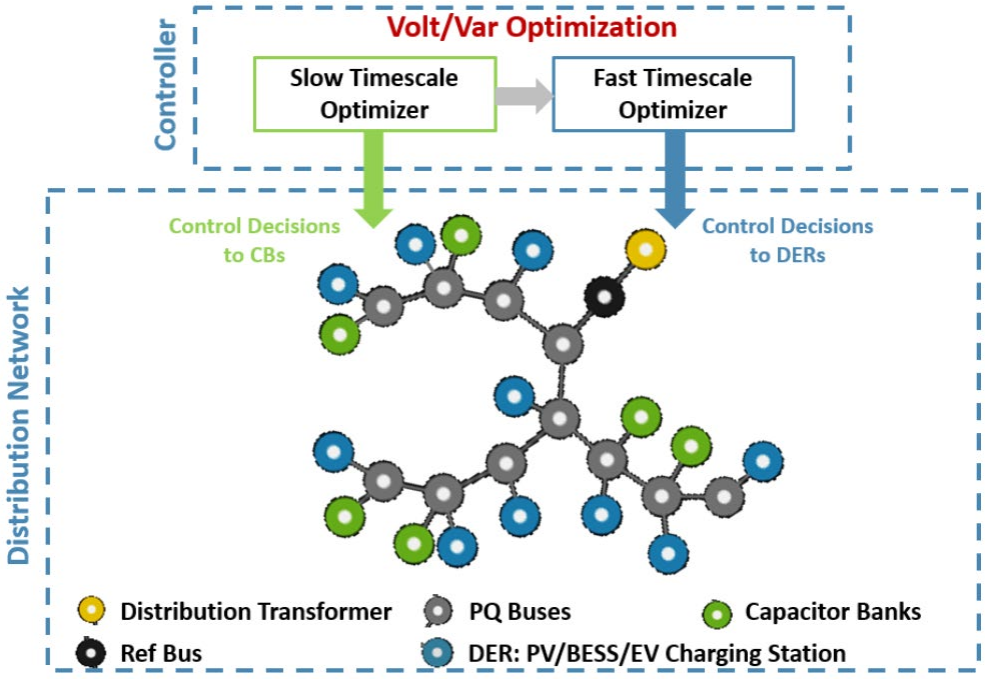


Figure 1-4. Conceptual diagram of the proposed VVO algorithm

The thesis objectives can be summarized as follows:

1. **Literature survey:** Conduct a literature survey on existing volt/var optimization algorithms focusing on devices and resources utilized, as well as timescale resolution considerations.
2. **VVO assets/resources survey:** Study the different volt/var control components and resources utilized in the VVO solving, including utility-owned assets and DERs.
3. **Comprehensive multi-objective and multi-timescale formulation:** To attain better performance of ADNs and overcome its rising concerns, the foremost issues should be considered in the VVO problem. Besides, to effectively achieve this aim, an optimization problem should be formulated to not improve one of the objectives at the expense of other objectives. Thus, we aim to develop a VVO algorithm that minimizes network power losses and the number of adjustments of utility-owned assets simultaneously and minimizes bus voltage deviations. In addition, to achieve

more efficient VVO performance and due to the uncertainty in DERs' generation, the VVO algorithm is preferred to have high time-resolution to manage fast-changing components of demand and generation. Hence, to match the time response of the utilized utility-owned assets and DERs, the VVO problem should be formulated with different timescales. This thesis aims to formulate the VVO algorithm with a multi-time scale resolution to match the utilized assets and resources accordingly.

4. **Utility operator involvement:** Benefiting from utility operators' experience in network operation can add value and give a higher degree of flexibility in managing the distribution network. We aim to incorporate utility operators' preference range in the VVO problem concerning network power loss and the number of utility-owned assets' adjustments.
5. **Utility assets management independently:** The utility-owned assets can have different conditions and lifetime status, where the utility operator may require giving higher priority to the assets having severe conditions. In this thesis, we aim to optimize the number of adjustments of utility-owned assets, named: CBs and OLTCs, independently, depending on their conditions.
6. **DERs incorporation:** In order to handle the fast variations of demand and generation of DERs, devices/resources with fast time response should be considered. We aim to develop a VVO algorithm that integrates DERs that include PVs, BESSs, and EV charging stations in solving the VVO problem by utilizing their reactive output power. Further, carry out performance evaluation and examine the effect of the utilization of DERs, individually, on network power losses and bus voltage deviations.
7. **Deep RL (DRL)-based VVO approach:** To overcome the limitations of the

conventional optimization-based VVO algorithm, a real-time optimization approach that can handle DERs' uncertainty and does not require an accurate distribution network model to solve the VVO problem is pivotal. Thus, in this thesis, we propose a two-timescale deep RL-based VVO scheme for ADNs. First, understand preliminaries and basic terminologies of RL. Following, formulate the two-timescale VVO problem as a Markov decision process (MDP) and solve it using a suitable DRL approach.

8. **DRL and conventional optimization-based VVO comparison:** Evaluate and compare the performance of the conventional optimization and DRL-based VVO approaches in realizing power loss reduction and voltage deviation minimization of ADNs.

1.7 Thesis Contributions

The contributions of this thesis can be summarized as follows:

- A two-timescale VVO algorithm that incorporates EV charging stations, BESSs, and RESs, in addition to standard devices. Where, also the condition of the different distribution assets, namely, CBs and OLTCs, are considered independently.
- A VVO algorithm that engages the operator by considering its preference ranges of the planning objectives, namely, power loss and a number of adjustments of standard volt/var devices.
- Due to conventional optimization VVO scheme limitations, a DRL-based two-timescale VVO algorithm that jointly coordinates the different utilized VVO resources at the two timescales using a DRL approach, named twin-delayed deep deterministic policy gradient (TD3), is proposed.
- A model-free DRL approach that supports continuous action spaces since PVs, BESSs, and EV charging stations can continuously provide reactive power rather

than quantized levels as in CBs. In addition, the algorithm does not rely on a precise optimization model nor accurate and complete information of distribution networks.

1.8 Thesis Outline

The thesis content is organized into eight chapters, of which an introduction chapter (chapter 1), 6 main chapters (chapters 2-7), and a final concluding chapter (chapter 8). Chapter 1 introduces the challenges and provides background information related to research objectives, and chapter 10 summarizes the work carried out and provides future recommendations and research plans. The research chapters are divided into two VVO common chapters (chapters 2 and 3) and four VVO solving chapters (chapters 4-7). Chapter 2 presents the conducted literature review of VVO in active distribution networks, their history, and current status in the market. It also presents related RL-based VVO prior work; chapter 3 presents the volt/var optimization main features and resources. The research VVO solving chapters are further divided into two parts, each comprising 2 chapters with the following structure. Part I: Chapter 4 demonstrates the formulation of the developed volt/var optimization algorithm and employed optimization techniques; Chapter 5 presents the tested network with its parameters and assumptions of the examined case studies, the implemented code, and the simulated cases. It then discusses the solution to the volt/var optimization problem and its effectiveness. Part II: Chapter 6 introduces RL and DRL and provides their theory and fundamentals; chapter 7 formulates the DRL-based VVO problem and describes the workflow of the employed DRL algorithm. It then presents the solution and analysis of the two-timescale DRL-based VVO scheme. It also demonstrates a comparative study of DRL and conventional optimization-based VVO scheme performances. Figure 1-5 shows a schematic representation of the thesis outline.

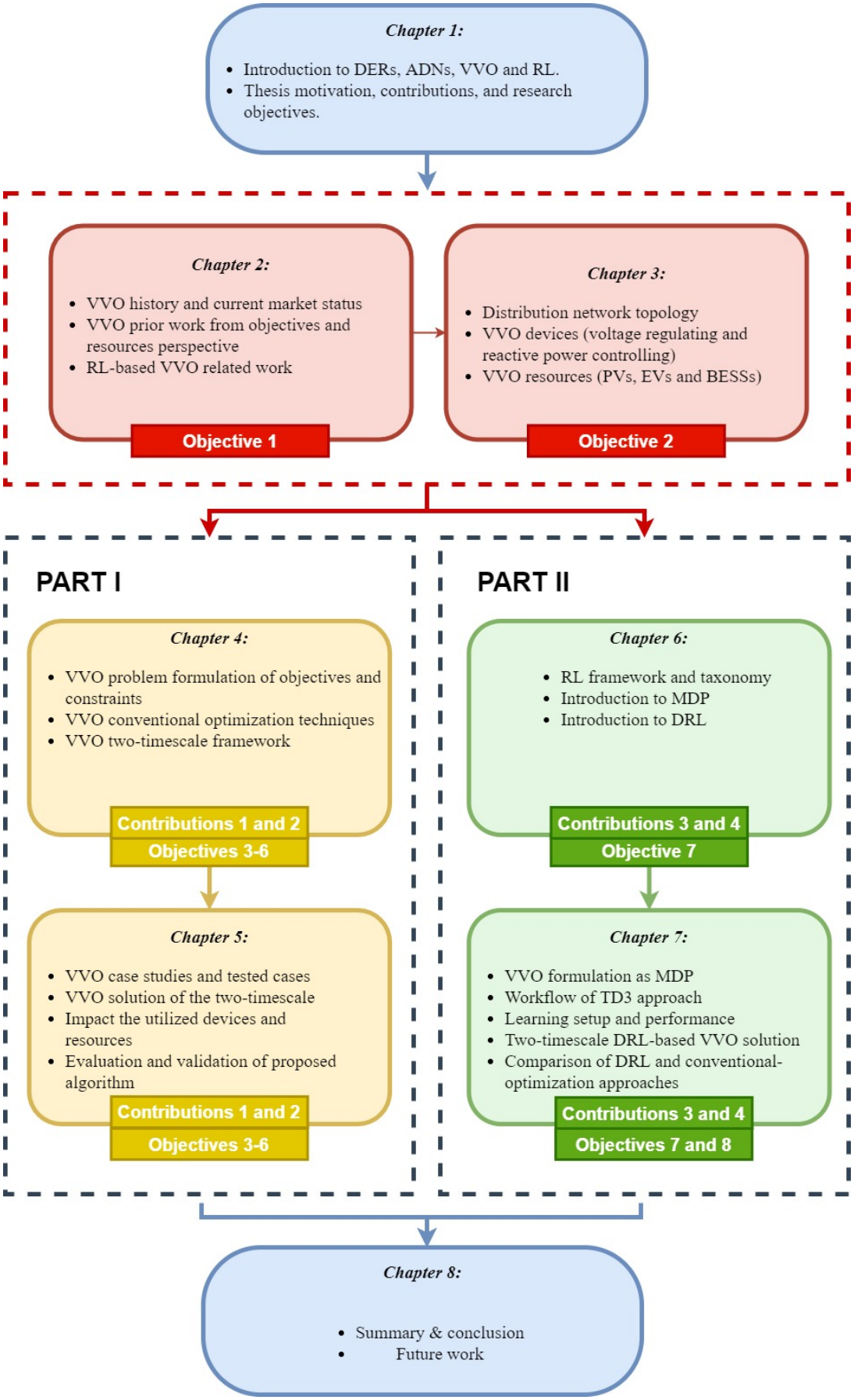


Figure 1-5. Thesis outline schematic

CHAPTER 2 : REVIEW OF RELATED WORK

This chapter presents a review of previous and related VVO approaches and their deployments throughout the years. It presents the volt/var management market's current status and a forecast for the coming years. It also presents the prior related VVO researches that employ RL algorithms.

2.1 Historical View on Volt/Var Optimization

Numerous studies have been dedicated to investigating and examining the impact of reactive power and voltage drop resources on distribution networks in the late 1800s, when they were first developed. Figure 2-1 shows the main studies of the road to the deployment of VVO. The first marked research related to the concept of volt/var control is [101], in 1932, performed basic local control. The study utilized series reactor behavior and transformer taps to study load ratio characteristics in local control circuits. Incomplete local measurements are used as inputs to find the settings of the utilized resources. Mostly, these measurements estimated the amount of reactive current flow in the distribution network. Accordingly, it was not reliable and lead to unsatisfactory performance, unacceptable power factor, and voltage levels. Moreover, the local control scheme lacked coordination on system-level and ineffective periodic maneuvers and trips of CBs. Generally, CBs and other resources' settings were not optimal, leading to an excess of reactive current network flow. Consequently, network power loss increased.

Most of the research work in the 1970s focused on static VAR optimization schemes [102]-[107]. The simple utilization of static capacitors in the distribution network's optimization was presented in [102]. In the 1980s, centralized control-based monitoring schemes were introduced. A basic communication design between VAR measurement devices at distribution feeder breakers was developed. The

communication to CBs was mostly one-way, and the VAR flow at feeder breakers was the only controllable parameter. The automatic tracking of CBs' settings did not exist, making the process of reconfiguration of distribution feeders to balance loads or isolate faults complicated. This control scheme did not consider other aspects that affect distribution networks' performance, like end-users loads and voltages. Subsequently, the control scheme was unreliable and inaccurate, which drove utilities to define VVO mechanisms.

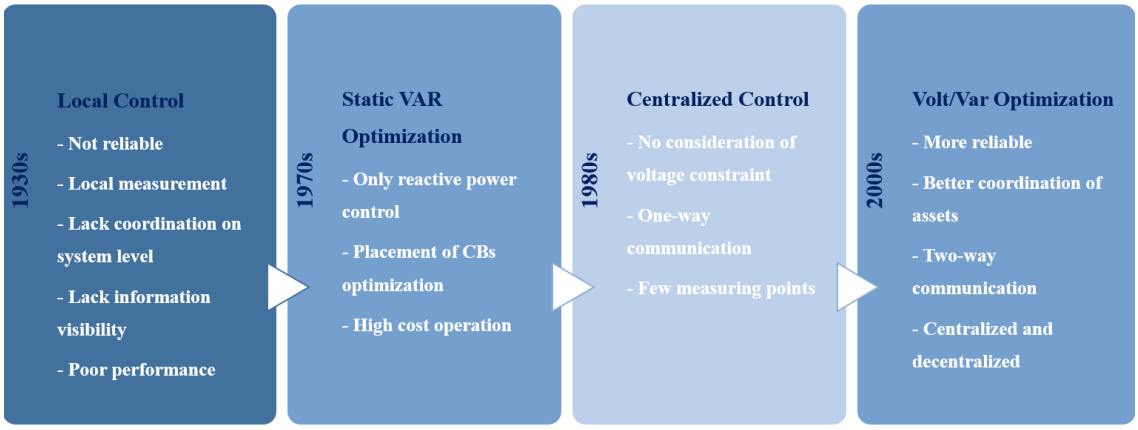


Figure 2-1. Road to Volt/Var optimization

Some of the heavily studied issues in the 1980s were the optimizations, aiming for power loss minimization, of the secondary volt/var control devices [108], and the allocation of reactive power sources [109]. In 1985, several volt/var schemes in distribution networks were published [110], [111]. It was the first use of the notation "volt/var" for reactive power and voltage optimization. Researchers studied the general volt/var control problem and then proposed design and control solutions.

In the 1990s, features like distribution management systems (DMS) and distribution automation were introduced in the distribution segment. The authors of [112] and [113] were the first researchers who studied the volt/var control concerns

related to DMS. Since the 2000s, desperate efforts are spent on solving the reactive power and voltage optimization problem for distribution networks [114]-[117]. However, with recent advancements in technology and distribution networks, more studies are required to overcome the introduced issues.

2.2 Volt/Var Management Market

The global volt/var management market is estimated to be worth $427 million in 2019 and is projected to reach $568 million by 2024, growing at a compound annual growth rate (CAGR) of 5.9% in the period of 2019-2024 [118]. The market is set to grow and focus on reducing transmission and distribution losses in power systems, optimizing power factor, and increasing electricity supply due to DGs' growing investment.

The distribution segment is driven by investments in smart grids and a need to reduce power outages and faults in distribution feeders. Upcoming plans for upgrading the existing infrastructure and new transmission and distribution projects in economies such as the US, China, India, the UK, France, Spain, Norway, Denmark, Belgium, and Ireland are expected to boost the volt/var management market growth from 2019 to 2024. Moreover, the utility and hardware segments are expected to grow at the fastest rate from 2019 to 2024 due to a need to reduce system-wide losses and minimize the distribution system and customer voltage variations. The increasing number of substation automation projects will drive the market in the coming years.

The leading players in the volt/var management market are Eaton (Ireland), Siemens (Germany), DC Systems (US), ABB (Switzerland), Open Systems International (US), GE (US), Advanced Control Systems (US), Varentec (US), Schneider Electric (France), Beckwith Electric (US), Landis+Gyr (Switzerland), S&C Electric Company (US), DVI(US) and Utilidata (US) [118].

In [118], the volt/var management market has been analyzed for five regions: Europe, North America, Asia Pacific, South America, and the Middle East & Africa. Figure 2-2 illustrates the market share for the years 2017-2024 for the aforementioned regions. Based on this study, North America is expected to be the largest market because of the number of investments in volt/var management solutions, increasing investments in RESs, and volt/var optimization projects in the utility sector to reduce losses.

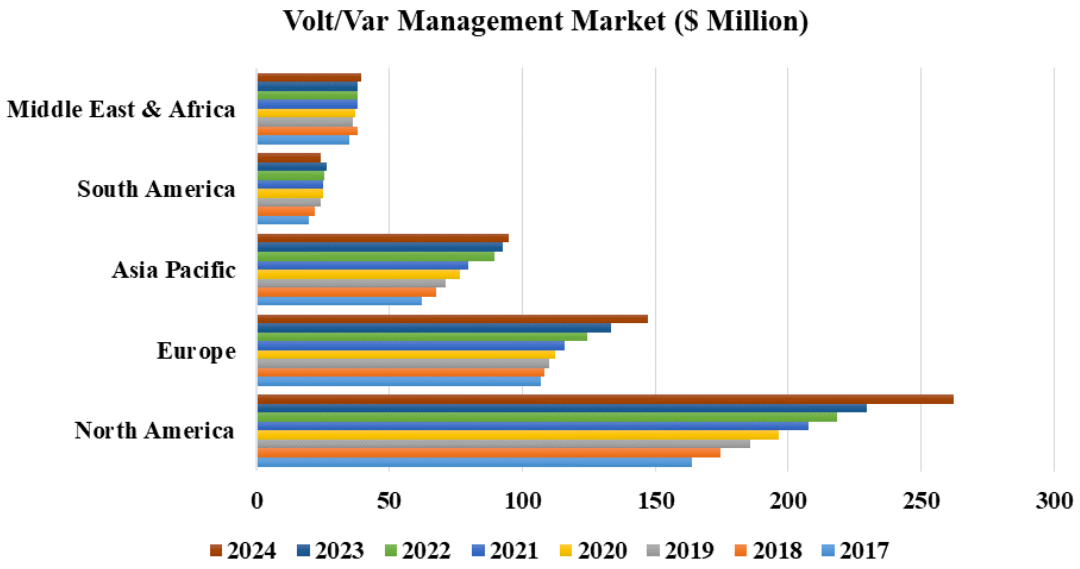


Figure 2-2. Volt/var management market [118]

2.3 Studies on Solving the Volt/Var Optimization Problem

VVO approaches can be categorized into several groups based on various aspects. In this thesis, the review of prior work on VVO algorithms is viewed, as depicted in Figure 2-3, based on two main features: the type of resources used in solving the VVO problem and the VVO problem formulation timescale.

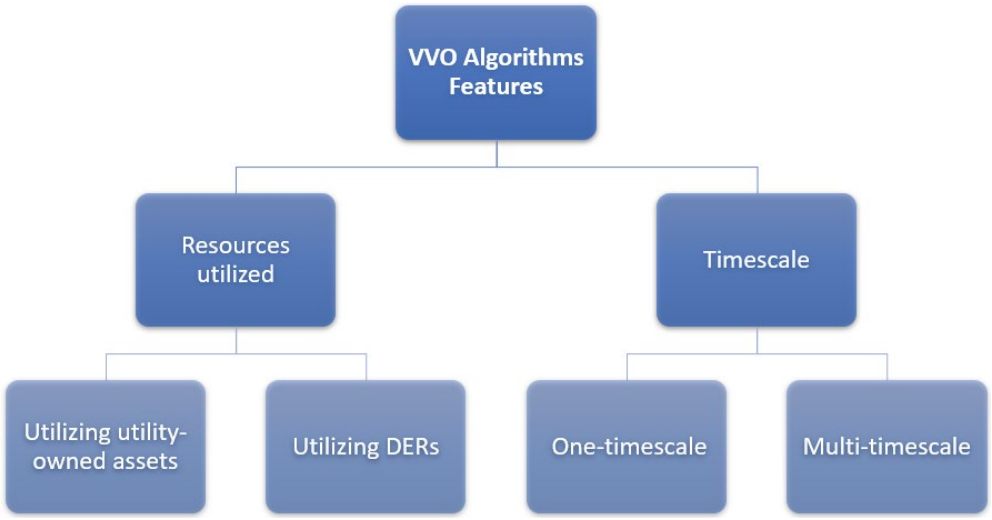


Figure 2-3. Categorization of VVO algorithms.

2.3.1 The Utilization of Utility-Owned Assets in VVO Algorithms

Several VVO methodologies utilizing utility-owned assets, CBs, OLTCs, and SVRs, have been proposed in the literature. Since CBs and OLTCs are discrete control devices, the single VVO problem is formulated as a mixed-integer nonlinear optimization problem (MINLP). Numerous algorithms employing MINLP have been proposed in literature [119]-[123]. Authors in [120] proposed a discretization penalty-based VVO algorithm using a nonlinear interior point method with limited control actions. While in [123], a prime-dual interior point (PDIP) method is proposed to solve a power dispatch problem. Ref. [124] proposed a static equivalent method for a reactive power flow model of an interconnected network.

In [5], a VVC approach using the PDIP technique with a penalty-based discretization algorithm to coordinate the reactive power output of CBs and OLTC settings is introduced. Ref. [6] proposed a model-free VVC scheme based on advanced metering infrastructure (AMI) data analytics to reduce power loss and maintain voltage levels. The approach is extended to integrate the CVR as an objective. The proposed method is based on adjusting voltage regulators setting points to switch On/Off CBs in

radial systems. Where [7] proposed a support vector regression-based model predictive control (MPC) VVO (trained and employed by the MPC in a closed-loop scheme to adjust the settings of CBs and OLTCs). Another model-free method is proposed in [10], where a CVR is implemented to control OLTCs via the distribution management system based on voltage feedback from AMI meter readings. Results showed that energy saving is achieved by optimizing off-peak voltage. To maintain voltage levels and achieve peak reduction by controlling SVRs, ref. [125] proposed a VVC that combined voltage alarms based on AMI with remote terminal unit measurements.

Due to multiple conflicting concerns that need to be considered, multi-objective VVO problems have been recently studied. In [126], the authors proposed an algorithm that aims to minimize the voltage stability index and reactive power loss. While in [127], the authors developed an algorithm to minimize voltage deviation and a weighted power loss. Ref. [128] proposed a multi-objective reactive power control algorithm to simultaneously minimize power loss of reactive compensation devices and minimize voltage deviation. Moreover, in multi-objective VVO problems, the number of adjustments of the utility-owned assets, CBs, and OLTCs, need to be considered. This is due to their frequent actions reduce their lifetime.

2.3.2 The Utilization of DERs VVO based Algorithms

Notwithstanding, the aforementioned approaches are based on adjusting the settings of SVRs, CBs, and OLTCs, which have considerably slow time responses that span several seconds to minutes [9], and the frequent operations reduce the lifetime of these assets as well as of the substation transformer. Besides, switched volt/var devices are costly, which adds to the overall system cost [129]. For an active distribution network (ADN) to be efficient, these devices' slow time response and the impact on system assets need complete coordination with DERs [129], representing alternative

means for reducing power losses and maintaining voltage levels. BESSs, for instance, are mainly introduced to support the fluctuations of other weather-dependent DERs like PVs and WTs [130],[131]. Moreover, EVs' batteries can also support grid volt/var related issues. Recent literature shows that DERs can provide fast responses in adjusting their reactive power via their grid-integrated inverters [132]-[136]. The number of DERs in an ADN is continuously increasing, and their response time is faster than the SVRs, CBs, and OLTCs [9]. Therefore, there is a vital need to coordinate among inverter-based, i.e., fast response, DERs, and the switch-based devices with the slow response to control voltage profile within the standard boundaries, reduce system losses, and enhance system asset lifetime.

In [137], a VVO approach combined with DER control is proposed for reactive power compensation and voltage control. The OLTCs and CBs, along with WT and PV resources, have been used to minimize power losses and voltage deviations. A model-free control of DERs for voltage regulation based on extreme seeking control is proposed in [138].

2.3.3 Multi-Timescale VVO based Algorithms

To efficiently solve the VVO problem and due to the different time responses of the different resources that can be utilized, multi-time scale VVO algorithms are emerging. In [9], a two-timescale approach is proposed to integrate PVs and BESSs using particle swarm optimization and GA. The work focused on minimizing the total system cost and minimizing power losses. Authors in [139] and [140] proposed a multi-time scale approach that coordinates settings of CBs, OLTCs, along with the output of RES to minimize power loss. In [141], in addition to the power loss minimization, the authors studied the effect of minimizing the number of setting adjustments of OLTCs, while utilizing the same resources as in [139]-[140] based on a two-stage algorithm. A

three-stage inverter-based VVC (TRI-VVC) framework is proposed, in [142], to minimize energy loss utilizing CBs and OLTCs, in the first stage and minimize voltage deviation utilizing PV inverters at different timescales in the second and third stages. The main differences between the proposed conventional optimization-based VVO approach in this thesis and multi-time scale approaches presented in literature are shown in Table 2-1.

Table 2-1. Comparison of Objectives and Resources of Multi-timescale VVO Approaches

Ref	Objectives of the Algorithm						Utilized resources (Q of Converters)		
	Voltage deviation	Engagement of operator	No. of control actions: CBs Independently		No. of control actions: OLTCs Independently		RES	EV	BESS
[9]	✓	✓	×		×		✓	×	×
[139],[140]	×	×	×		×		✓	×	×
[141]	×	×	×		✓		✓	×	×
[142]	✓	×	×		×		✓	×	×
Proposed	✓	✓	✓		✓		✓	✓	✓

2.4 Studies on Solving Volt/Var Optimization using Reinforcement Learning

The existing VVO algorithms deployed by electric utilities mainly incorporate their assets, namely, OLTCs, CBs, and VRs. However, as previously mentioned, their time response is considerably slow, from several seconds to minutes, and their frequent operation reduces their lifetime [9]. While DERs are equipped with smart inverters, which enable them to inject or absorb reactive power to/from the network with faster

time response [9]. Hence, DERs are utilized in solving the VVO problem for ADNs.

Further, most existing algorithms, to minimize power loss and/or bus voltage deviation, either employ day-ahead conventional optimization-based algorithms, such as PSO [9],[143], GA [3],[144], mixed-integer quadratic programming (MIQP) [145],[4], mixed-integer quadratically constrained programming (MIQCP) [146],[147], etc.. or real-time algorithms that require records of history data, such as model-predictive control (MPC) algorithm, etc. [7], [148]. Although such algorithms achieve promising performance, they are physical model-based control approaches that require the knowledge of a complete and accurate set of network models and parameters, such as the model of transmission lines, non-linear system components, and renewable energies. Nevertheless, this is not easily manageable for modern, large, and interconnected distribution networks with increasing complexity. Also, the aforementioned approaches highly depend on precise optimization models and have limited capabilities to manage the rapid and intermittent nature of generation and demand of ADNs. To overcome these issues and reduce their impact on VVO performance, RL approaches have emerged as one of the most promising tools to solve decision-making problems with uncertainties [60]. In this thesis, we focus on summarizing the most recent related researches that propose an RL-based approach.

The most common and simple RL algorithm is the *Q*-learning [54]. A dimensional Q-learning (DQL) approach is proposed in [149] for implementing reactive power optimization with discrete control variables. Where settings of utility-owned assets are configured to minimize power loss. The proposed algorithm employs the conventional *Q*-learning to explore the feasible area dimensionally to reduce the agent's memory. While, in [150], authors proposed a distributed Q-learning, model-free, multiagent-based RL algorithm for solving the optimal reactive power dispatch

problem. The proposed algorithm coordinates utility-owned assets along with PVs to minimize power loss. The authors then proposed a fully distributed multiple agent system based on sub-gradient RL in [151], where utility-owned assets are coordinated to minimize power loss and voltage deviations in the network. In [152], the authors proposed a batch RL algorithm to coordinate tap positions of OLTCs to regulate voltage levels in power distribution systems. The problem is solved with an action-value function and discrete action space based on a linearized power flow model.

However, as previously mentioned, in applications that involve high dimensional continuous action and/or state spaces, the traditional RL algorithms will suffer from "curse of dimensionality", which makes them inefficient in practical cases [153]. Subsequently, DRL approaches have emerged to overcome the dimensionality issue. DRL is the combination of deep learning (DL) and RL perceptions. DRL approaches employ neural networks (NNs) in their agent as value or policy function approximators, which increases the efficiency and applicability of the algorithm in real cases.

In [153], a model-free, soft actor-critic, off-policy DRL approach is proposed to solve the volt/var control (VVC) problem with discrete action space where utility-owned assets are coordinated to minimize the costs of power loss and devices' switching. While in [91], the authors proposed a two-timescale voltage regulation algorithm for distribution grids. On the fast timescale, the reactive output power of the DERs is scheduled to minimize bus voltage deviations using a conventional optimization algorithm. While, on the slower scale, CBs are coordinated to minimize voltage deviations using a DRL approach, named deep Q-network (DQN).

In [154], the authors proposed a DRL approach to solve the VVC problem, where settings of utility-owned assets are configured to minimize the costs of power

loss and device switching. Two policy gradient approaches are examined: constrained policy optimization and trust region policy optimization. The authors demonstrated that policy gradient methods could learn near-optimal solutions and determine control actions faster than conventional optimization methods. While, in [155], a DRL-based algorithm to coordinate multiple smart inverters for voltage regulation is proposed. The authors demonstrated that a well-trained DRL agent can schedule different smart inverters to maintain voltage levels within permissible ranges, minimize system losses, and achieve a reduction of PV production curtailment. The authors in [201] propose a multi-agent DQN based VVO approach with discrete action space. The proposed algorithm utilizes utility-owned assets along with DERs to minimize power loss.

Noticeably in the context of VVO, existing DRL approaches employ discrete spaces when representing action spaces. However, utilizing reactive output power of DERs in solving the VVO problem, and since their output is continuous, a continuous action space is preferred to precisely schedule their reactive output power. Discretizing the output of DERs could lead to higher DERs' generation curtailment due to their capacity limit. Further, the discretization will tremendously increase the action space size, leading to dimensionality issues. The main differences between the proposed and recent DRL approaches in literature are summarized in Table 2-2.

Table 2-2. Comparison of Related DRL-Based Approaches

Ref	Objectives		Resources Utilized		Characteristics				
	Power Loss	Voltage Deviation	Utility-Owned Assets	Inverter-Interface resources	Model-free	Off-policy	Actor-Critic	Multi-Timescale	Continuous action space
[153]	✓	×	✓	×	✓	✓	✓	×	×
[91]	×	✓	✓	✓	✓	×	×	✓	×
[154]	✓	×	✓	×	✓	×	×	×	×
[155]	×	✓	×	✓	✓	×	×	×	×
[201]	✓	×	✓	✓	✓	✓	×	×	×
Proposed	✓	✓	✓	✓	✓	✓	✓	✓	✓

2.5 Summary

In conclusion, most VVO studies utilized only utility-owned assets in the problem solving and did not consider the presence of DERs [5]-[7], [10], [119]-[128], and few research integrated existing DERs in distribution networks [137]-[142]. Moreover, most studies formulated the VVO problem as a one-timescale problem, which ineffective due to the stochastic nature of RESs in the ADNs. Thus, the existing gaps in the literature in regards to solving VVO include but not limited to the availability of VVO algorithm with comprehensive objective functions of the conflicting objectives, considering the presence of DERs, like PVs, BESSs, and EVs and utilizing them in the optimization problem, considering the conditions and lifetime of utility-owned assets, involving the utility operator in the decision-making process.

Further, to overcome raised concerns in ADNs more efficiently, DRL approaches have been proposed. Recent DRL VVO related studies adopted RL

algorithms that only support discrete action spaces. However, utilizing reactive output power of DERs in VVO problem solving, and since their output is continuous, a continuous action space is preferred to precisely schedule their reactive output power. Moreover, almost all recent DRL-based VVO studies [201], [91], and [153]-[155] formulated the VVO problem as a one-timescale problem, which is ineffective when utilizing both utility-owned assets and DERs. Hence, it will not manage the intermittent outputs of DERs in ADNs. Thus, the existing gaps in the literature in regard to solving VVO using a DRL approach include but are not limited to the availability of DRL-based VVO algorithm with comprehensive objective functions and multi-timescale resolutions, as well as, being able to handle continuous action spaces.

The first step towards formulating and solving the VVO problem is understanding the distribution network's topology and the possible utilized resources. The aforementioned points are presented in the next chapter.

CHAPTER 3 : VOLT/VAR OPTIMIZATION TOPOLOGY AND RESOURCES

One of the main parts of a power system is the distribution network. It links high power sources and customers. Understanding the topology of the distribution network aids in formulating and solving a VVO problem. In this chapter, the VVO distribution network's topology is emphasized, and the different VVO devices and resources utilized in solving the optimization problem are studied.

3.1 Distribution Network Structure

Generally, distribution networks consist of sub-transmission parts, distribution substations, distribution feeders, distribution transformers, and secondary feeders. Figure 3-1 shows the one-line diagram (OLD) of a distribution network.

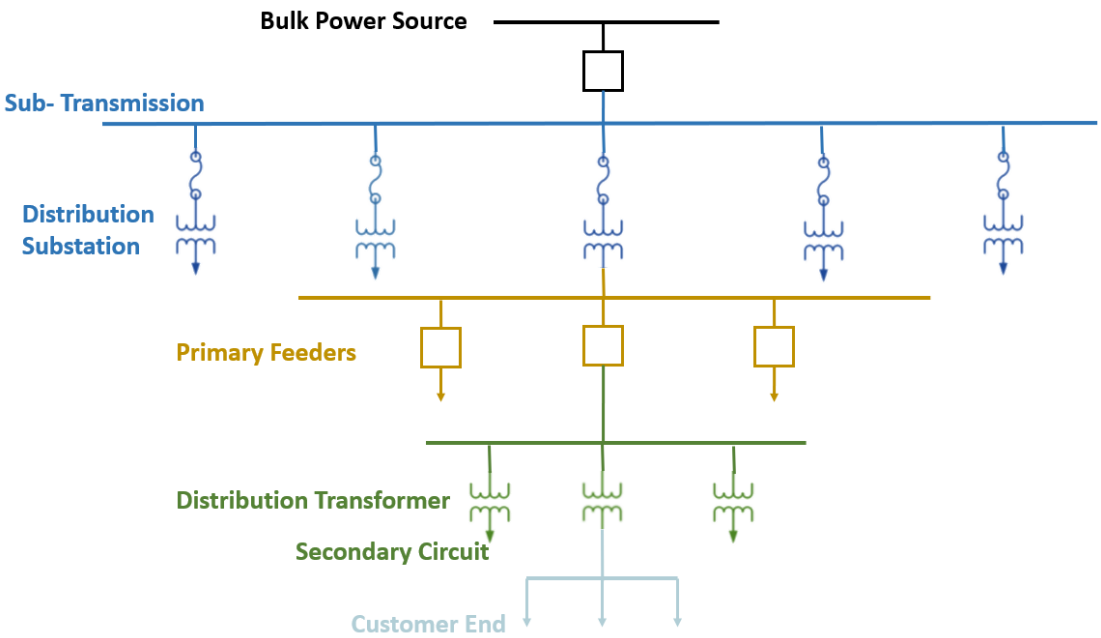


Figure 3-1. OLD of a distribution network

Typically, a distribution substation comprises a station bus, power transformer, breakers, voltage regulating devices like OLTCs and CBs, and switchgears. It is where the voltage level is reduced to primary feeder voltage. The distribution transformer then

reduces the voltage level further to secondary voltage. Finally, through service drops, the secondary circuits distribute power to customers.

The most common, low cost and simple topology of primary and secondary feeders is “radial”. A typical OLD of a radial primary and secondary feeder is shown in Figure 3-2(a) and Figure 3-2(b), respectively. Based on this fact, the focus of this thesis is on radial active distribution networks. Moreover, it should be emphasized that voltage levels at any point in a distribution network have to be in the acceptable ranges according to the ANSI C.84.1 standard [50].

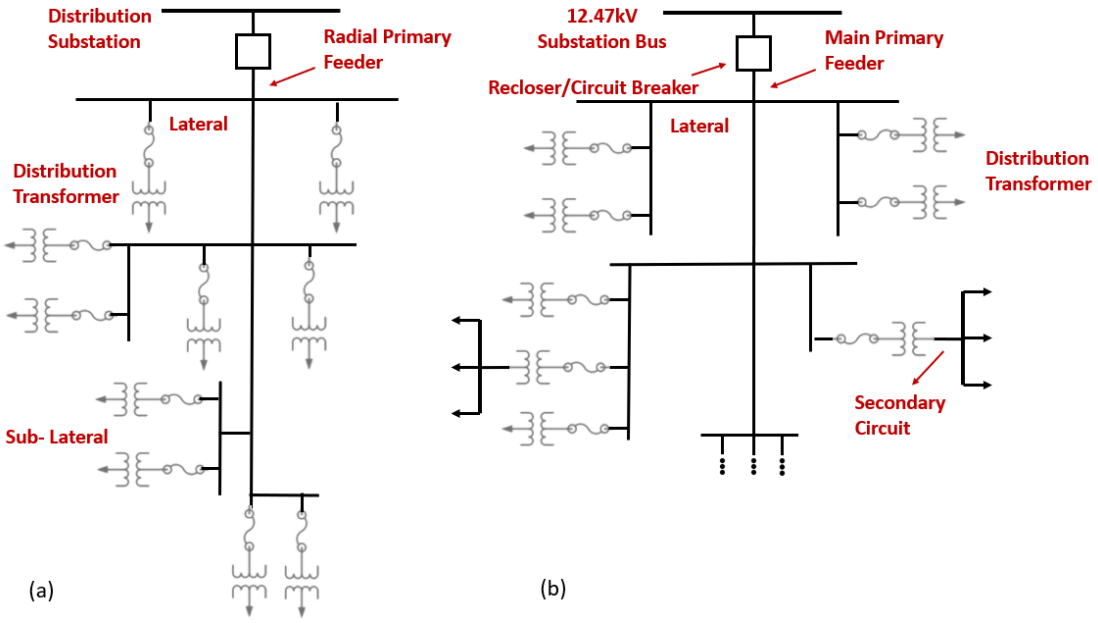


Figure 3-2. OLD of radial feeders. (a) OLD of radial primary feeder, (b) OLD of the radial secondary feeder

As to the control topology used, SCADA (supervisory control and data acquisition) is the commonly employed tool for monitoring and controlling stations via a master control center. Typically, SCADA comprises controlling and monitoring equipment, sensing and measuring schemes, and a two-way communication platform.

3.2 VVO Main Regulating Components

As previously mentioned, a key objective of VVO is to reduce losses of the distribution network. Generally, losses in a distribution network can be classified into two technical and non-technical losses. Figure 3-3 highlights the categorization of losses in a distribution network and a few examples for each type of loss. Most types of losses in a distribution network are technical losses since they depend on operational modes and network characteristics. Fixed technical losses do not vary with current changes, and it is found that they comprise approximately 25% to 30% of the technical losses [52]. On the other hand, since variable technical losses are proportional to the current squared, they comprise about 66% to 75% of the technical losses [143]. Since the technical losses contribute to higher losses, this thesis focuses on minimizing one of the variable technical losses, specifically, the active power loss in a distribution network.

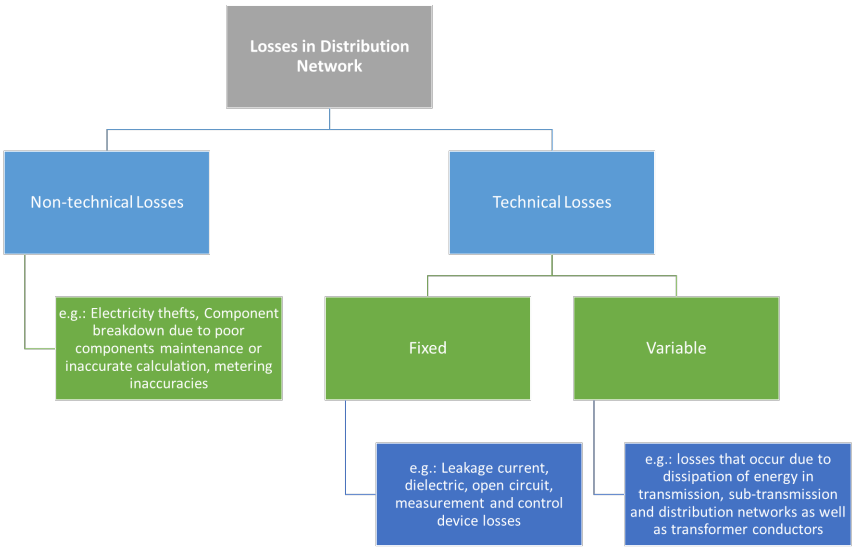


Figure 3-3. Classification of losses in a distribution network [52][143]

The main components and resources that can be utilized and integrated into the VVO algorithm are essential to be studied to achieve the loss minimization objective. As mentioned in previous chapters, voltage regulating components, such as VRs, are installed in the distribution feeders at specific nodes to regulate voltage levels and, such as OLTCs, are installed at the distribution transformers to adjust the downstream voltage level to the feeders based on load condition. Reactive power controlling components, like CBs, are installed at the distribution substation and/or distribution feeders to regulate reactive power flow. DERs existing at the network buses in the distribution network can also control reactive power flow by injecting/absorbing it. Thus, the utility-owned components and DERs are reviewed thoroughly in the following subsections.

3.2.1 Voltage Controlling Components

One way of adjusting the voltage level is by employing auto-transformers. It is considered a step-type regulating component. It is designed to regulate the line voltage between ±10% in 32 steps. Thus, each step can increase/decrease line voltage by 5/8%. The types of connection of the internal coils can be series or parallel. If coils are connected in series, the regulation range is ±10%, and if connected in parallel, the line voltage can be regulated up/down between ±5%.

A step-type voltage regulating device comprises a tap changing part and a control part fitted with a control system. The controller sends commands to the tap changing part to adjust the tap's position based on the load condition. In addition to voltage level regulation, the controller also regulates the Bandwidth (BW) based on potential and current transformers measurements. Generally, the controller settings are voltage set point (represents the output voltage desired), bandwidth (shows the difference between the measured and desired voltages), and time delay (shows the

difference between the voltage exceeding BW limit time and the tap changing time). The tap starts to change when the difference between the set and measured voltages is greater than 50% of the BW value. In conventional voltage regulating devices, the time delay is typically between 10-20 seconds [157].

Some grids employ a voltage regulating device named Auto-Booster. It is a single-phase device comprised of four tap steps, where each step has a 1.5%-2.5% regulation range. Therefore, it has 6%-10% of regulation [157].

The most common components used to regulate and maintain voltage levels within the acceptable ranges are the VRs and LTCs. VRs adjust the primary winding taps to regulate the secondary voltage. While tap changers are employed to regulate downstream voltage. Distribution transformers can be equipped with offload tap changer type on the primary side or OLTC on the secondary side. No-load tap changers are also a type of distribution transformer tap changer; however, they need to be disconnected from the circuit to adjust taps. This thesis utilizes OLTC in solving the VVO problem.

The structure of a high voltage rating OLTC is shown in Figure 3-4. It comprises a diverter switch and a tap selector. First, based on the network voltage needs, a tap is pre-selected offline. The diverter switch then adjusts the tap position online. Currently, the switching time of the diverter switch is between 40-60 ms, and the total operation time of the OLTC is 3-10 seconds [158]. Low voltage rating OLTC employs a selector switch mechanism.

According to the tap mechanism, the OLTCs can be classified as either resistive-/reactive-based oil types. Both types of classifications can be either OLTC based on switch selector mechanism or diverter switch with tap selector.

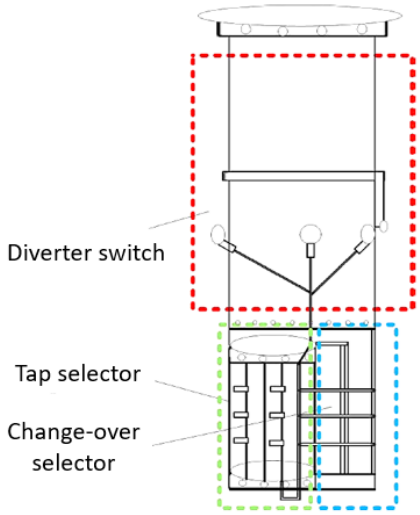


Figure 3-4. Structure of an OLTC [158]

3.2.2 Reactive Power Controlling Components

Shunt capacitors are the common reactive power controlling device utilized in distribution grids to regulate voltage by controlling the reactive power flow. In distribution networks, the size of a shunt capacitor ranges from 50 to 400 kVar, and with the banking feature employed, they can supply 300 to 1800 kVar [157]. CBs, generally, are installed in different locations on a distribution feeder as required. They can control the reactive power flow of a local load, group of loads, or a feeder branch. Consequently, they can improve voltage profiles, reduce feeder losses, and reduce maximum kVA demand. Typically, the type of CB mostly installed in distribution networks is a pole-mounted CB. Figure 3-5 shows the CB structure of the pole-mounted type.

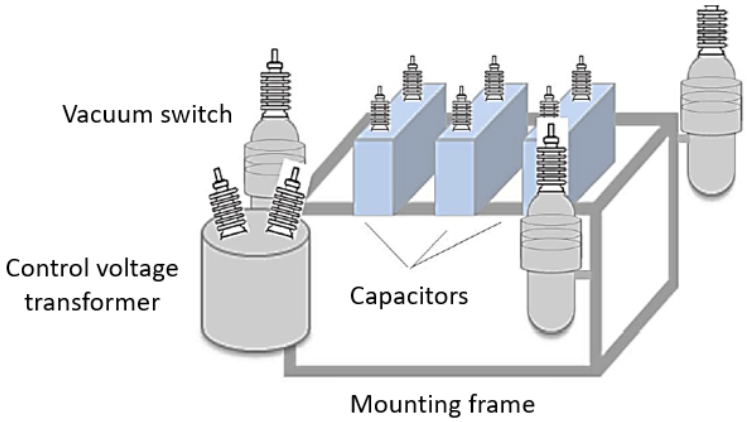


Figure 3-5. Structure of a pole-mounted CB [157]

Standard [159] gives the main specifications of a CB. They are designed to operate continuously in severe environmental circumstances. The tolerable average and annual temperatures are 46 and 65 ˚C for isolated and single row capacitors, respectively, and 40 and 25 ˚C for metal-enclosed or multiple row capacitors, respectively. Generally, CBs must operate uninterruptedly without exceeding 135% of their kVar, 180% of their RMS current, 110% of their peak voltage, and 120% of their RMS voltage. It is worth noting that the size of a CB is proportional to the frequency of the system.

In this thesis, switchable shunt CBs are utilized at different locations to regulate voltage by controlling the reactive power flow at different time intervals. Generally, CBs need to be equipped with adequate control systems to send commands to open/close kVar. The main control parameters of CBs are temperature since kVar increases with the increase of temperature, kVar control since kVar demand increases with an increase in load, current since it is proportional to kVar demand, the voltage to regulate voltage and switch time to be able to switch CBs in different load conditions.

3.2.3 Distributed Energy Resources

a. Photovoltaic

PV penetration has been increasing rapidly, aiming to reduce the dependence on fossil fuel resources and reduce conventional power plants' carbon emissions. However, its major concern is that its generation is dependent on weather conditions. Therefore, the output is very stochastic and variable. Although weather forecasts can be performed with acceptable accuracy and can anticipate PV generation accurately, sudden clouds can disrupt its output. The generation of a PV can increase/decrease by 15% of its capacity per minute with fast fluctuating cloud coverage [160]. Subsequently, the integration of PVs at the distribution level introduced significant challenges in operating and controlling the distribution network.

As much as technical limitations permit, utilities aim to increase PV penetration levels in the distribution network. Thus, the more accurately the concerns regarding PV integration are studied and evaluated, the higher the penetration level that can be considered into the network without risking technical and operational limitations. PVs are used for different operational aims. The utilization of PVs in distribution networks can assist in satisfying the high demand and, at the same time, present itself as an environmentally friendly alternative to fossil fuel resources. It has recently been proven that PV systems support the distribution network with real power and support the grid by supplying/absorbing reactive power. The amount of output reactive power of a PV depends on its real power generation at that specific time and the apparent rating power of its interface-inverter.

PVs can operate in two different modes, as demonstrated in Figure 3-6. Generally, during the daytime, they operate on the border between quadrants I and II, where $P>0$ and $Q=0$, which implies that they only inject active power. However, they can operate in a mode where they can inject active power while absorbing reactive power (quadrant II) or injecting reactive power (quadrant I). At nighttime, they can also

support the distribution network by operating at quadrant I or II, where P = 0 and Q > 0 or Q < 0, respectively.

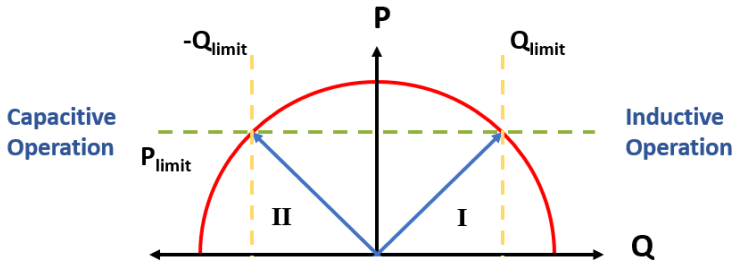


Figure 3-6. PV capability curve

b. Electric Vehicle Charging Stations

The introduction of advanced technology of EVs such as vehicle to grids (V2Gs) adds technical challenges to planning and controlling the distribution network. However, it also brings new opportunities as it can also be utilized to support the network. New studies showed that their advanced interface-inverters could provide reactive power control [132]-[136]. Utilizing such technologies with reactive power capabilities significantly affects several aspects of energy conservation, active and/or reactive power optimization, and control schemes for ADNs. Since they can be utilized in the V2G mode as a reliable reactive power source, the required reactive power and the capacity and number of operations of CBs and OLTCs in a distribution network will be affected. Thus, they can help minimize reactive and/or active power losses and conserving network energy consumption when employed.

Generally, EVs can be categorized based on their technology and their charging levels [134]. Figure 3-7 illustrates the classifications of EVs. High penetration of different types of EVs at different locations can lead to significant fluctuations in demand and load profile of a distribution network. Several researchers tried to study

and predict the trajectory of these fluctuations based on different EV level penetrations [161]-[163].

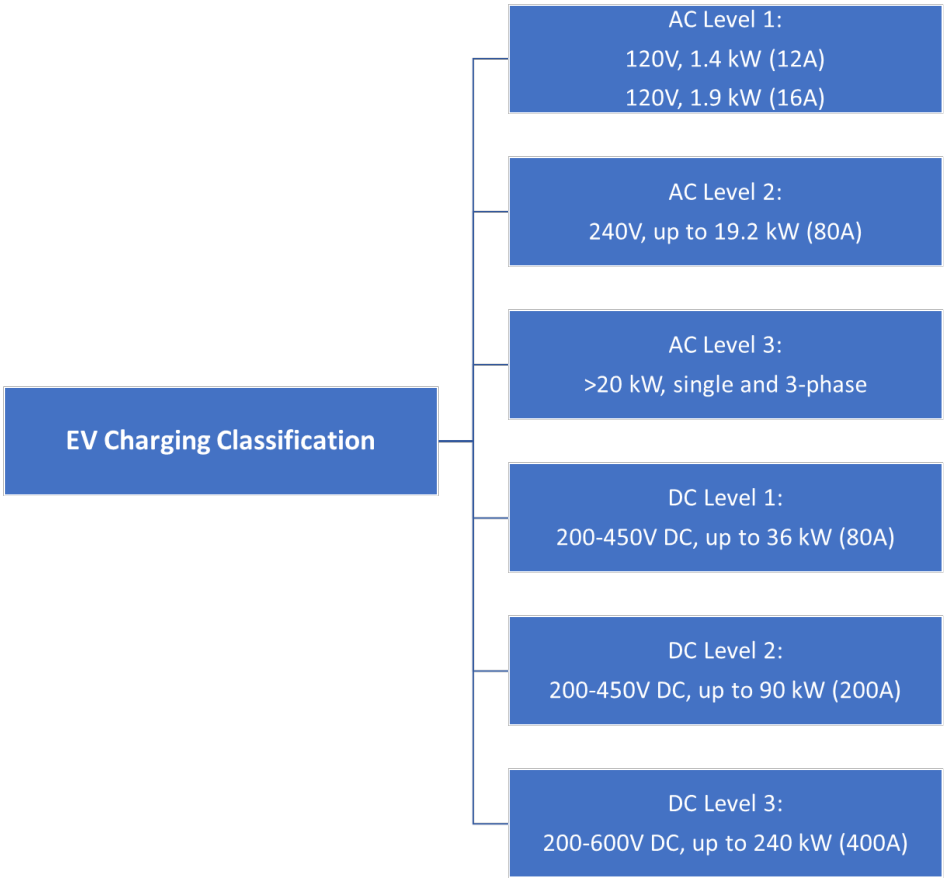


Figure 3-7. EV charging classifications [134]

When an EV supplies reactive power to the network, it affects the kW's overall value, load rise, and system overcompensation issues. It could increase the distribution feeder's capacity. Thus, reduce power loss and improve voltage profile provided that there is no overcompensation in reactive load. Also, it could lead to a rise in voltage. Consequently, there will be an increase in loads, which can be higher/lower than the reduction losses. Thus, the system's overall kW input might increase/decrease. The aforementioned points are covered within the power flow that runs in the VVO algorithm.

EV charging stations can adopt different modes of operation [161]-[163]. The operating modes of their AC/DC inverters are shown in Figure 3-8. Mostly, EVs operate on the boundary between the I and IV quadrants, where P>0 and Q=0. This implies that EVs mostly consume active power from the grid. However, it is possible for an EV to operate in a mode wherein it still absorbs active power for charging and, at the same time, supplies reactive power to the network, quadrant IV [132]. Consequently, EV charging stations can help improve voltage profile, minimize power loss, and increase feeder capacity.

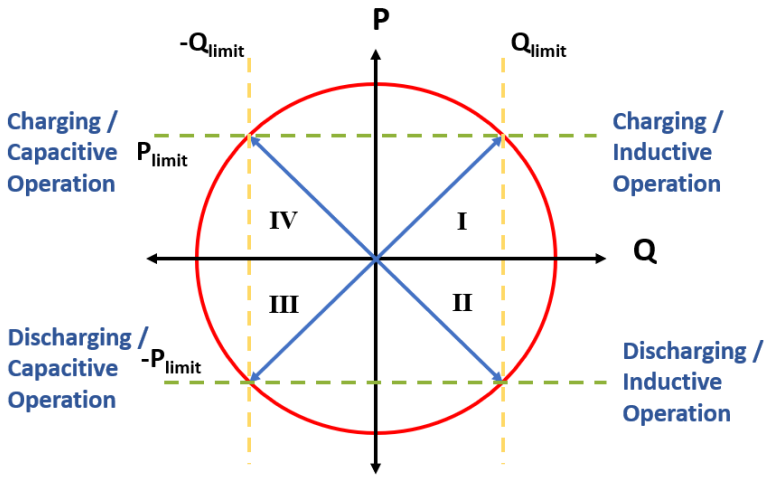


Figure 3-8. Operating modes of EV charging stations [161]-[163]

Generally, the bidirectional charger topology requires minimal adjustments to support the grid with reactive power [134]. Single-phase inverters' ability in injecting reactive power has been verified in [164],[165]. Based on the mentioned studies, it would be possible for the inverter to inject reactive power if the dc-link capacitor's voltage rate increased by at least 3%. It is also demonstrated that the capacitor dc link's current ripple rate is robust enough for inverters to function in capacitive mode. It is to be noted that inverter losses might inconsiderably increase in capacitive operation mode

compared to the charger's normal mode. However, the input inductor current and the EV battery are not affected [134]. An EV can supply reactive power to the grid through the capacitor dc-link without engaging the battery. Consequently, EVs' battery life will not be affected by reactive power support.

Moreover, research studies showed that EV charging stations could inject reactive power even if no EVs are connected to the charging station [164]. During charging mode, the quantity of reactive power supplied by an EV charging station to the network is constrained by the consumed active power and the charger power limit, as shown in Figure 3-8. Therefore, if an EV is consuming the maximum P from a charging station, the station will not be able to inject reactive power. However, the charger can be rated 10%-20% higher than the maximum real power consumed during charging for the EV charging station to support the grid at different operating conditions [134].

This thesis evaluates EV charging stations' utilization in supplying/absorbing reactive power to/from the grid on the set of VVO objectives. Further, the proposed VVO algorithm seeks to find the optimal reactive power amount injected/consumed for each EV charging station.

c. Battery Energy Storage Systems

BESSs can provide distribution networks several benefits in different aspects regarding the quality, control, and stability reliability at transformer/feeder levels in the network. Generally, BESSs are integrated to smooth the intermittent behavior of other DERs, like PVs, wind farms, and EVs. It also aids the integration of these DERs into the grid. BESS can provide voltage regulation by consuming/injecting active and/or reactive power through their inverters' four-quadrant modes. Utilizing BESS can decrease the installation and number of CBs and OLTCs, leading to cutting down costs

and enhancing utility-owned components' lifetime.

Typically, a BESS is employed in distribution networks to improve power quality, smoothing DER's stochastic outputs, reactive power support, frequency regulation, voltage regulation, peak shaving, etc. The main components of a BESS are a four-quadrant inverter located in the grid inverter panel, a battery management system (BMS) of mostly lithium-ion batteries, and control systems in the breaker and meter panel, such as energy management system (EMS) and thermal management system (TMS). Figure 3-9 demonstrates the main components of a BESS. As EV charging stations, BESSs and DGs' key difference is that it can exchange reactive and active powers through its four-quadrant inverters bi-directionally.

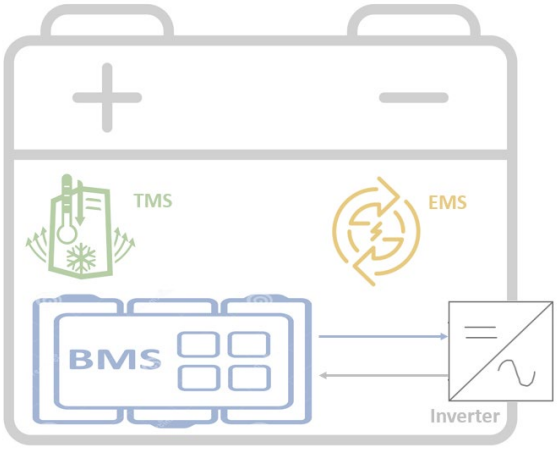


Figure 3-9. Main components of a BESS [166]

The principle operation of BESSs is simple. They store energy in their batteries, and based on the network's active and reactive demands, they supply/consume active and reactive powers. Since they can deliver power for a short time during discharging, they are considered backup power. Typically, BESS's discharging time is usually 1-3 hours at each operating cycle, and the charging time is less than two times the discharge time. Its discharging power is equal to the rated power, while the charging power is less

than or equal to the rated power. Thus, the discharge time can be obtained by dividing the rated energy and rated power [167].

In general, there are three operation modes for a BESS: charging, discharging, and standby. Its discharge could be active/reactive inductive, fully active, and active/reactive capacitive based on the four-quadrant diagram of Figure 3-8. VVO can guide BESSs to discharge during heavy load time since their main tasks are peak shaving. VVO will also benefit since the demand will decrease. Usually, BESSs charge during the light load period. During charging, they consume active power, which slightly increases the active power loss. However, since the power loss is low at the light load period, the BESS charging mode's impact on the system is insignificant.

In the standby mode, BESSs are typically disconnected from the network; however, they are on standby and available to be connected back when the network requires their support. Based on Figure 3-8, while in standby mode, BESS can inject/absorb reactive power ($Q > 0$ or $Q < 0$, $P = 0$). This is an interesting operation mode for BESS that could be created in the future. For example, a control command can be sent to a BESS in standby mode from the VVO to be integrated into the network and supply reactive power. However, this mode requires further economic and technical investigations. Overall, BESSs support the distribution networks in local control, feeder optimization, and energy management. Subsequently, benefit VVO objectives.

It is worth noting that the new IEEE standards specify that DERs can provide voltage regulation, subject to the system operator's approval, through adjusting reactive power output.

3.3 Summary

Based on the aforementioned study of components and resources in this chapter, from the VVO point of view, it can be concluded that DERs, like EVs and BESSs, are

favored when compared to CBs and OLTCs since they can supply active and reactive power. They are also preferred compared with DGs, like PVs, as their output is controlled and not stochastic. Therefore, in this thesis, the main volt/var voltage regulating and reactive power compensation components and resources utilized for performing the proposed VVO are CBs and OLTCs, along with PVs, EVs, and BESSs.

PART I: CONVENTIONAL OPTIMIZATION-BASED VOLT/VAR OPTIMIZATION SCHEME

CHAPTER 4 : THE TWO-TIMESCALE VOLT/VAR OPTIMIZATION MODEL

The formulation of the VVO's different objectives and constraints in the different timescales is the first step towards obtaining the solution. In this chapter, the VVO problem is formally defined, and the optimization technique used in solving it is studied. Further, the framework of the proposed VVO algorithm is emphasized.

4.1 Proposed VVO Problem Formulation

In an ADN with high penetration of DERs, the variations in the voltage profile have mainly slow- and fast-varying components. The slow components are primarily due to the changing nature of the demand. While the intermittent outputs of the DERs are responsible for the fast-varying components in the voltage profile. Hence having an objective to regulate the voltage profile based on VVO requires a solution that considers two timescales to solve the optimization problem under minimum system losses effectively. Furthermore, this solution should consider the system operator preference that optimizes the changes in settings of distribution system assets to enhance their lifetime. Thus, we view the optimization problem's solution in two stages, I and II, that match the two timescales. Due to the slow time response in demand change, the standard switch-based CBs and OLTCs are the primary resources utilized in stage I, and their response is optimized on h basis [139]. The objectives of this stage are to reduce system losses while minimizing the number of setting changes. This stage provides the means for the system operators to input their preference setting adjustments of CBs and OLTCs to preserve the assets' lifetime. The fast inverter-based DERs are the primary sources that are optimized in stage II. Their reactive power outputs are controlled and optimized based on an h/n timescale.

4.1.1 Stage I: h Time-Resolution Multi-Objective Formulation

In this first stage, power losses in lines are minimized while reducing the

number of setting adjustments, i.e., increasing/decreasing the setting positions of CBs and OLTCs over 24 hours. Here, the problem is presented as a multi-objective optimization with two conflicting objectives as formulated in (4-1) and (4-2), where the control variables are s_m, s_n, ${Q_x}^{PV}$, ${Q_x}^{EV}$ and ${Q_x}^{BESS}$.

minimize P_{loss}:

$$min \sum_{t=1}^{24} \sum_{k=1}^{nl} {P^{st}}_{k,t} - {P^{en}}_{k,t} \tag{4-1}$$

minimize $control\ actions$:

$$min \sum_{t=1}^{24} [\alpha\ .\ cap_{actions} + \beta\ .\ tap_{actions}] \tag{4-2}$$

The implemented procedure in minimizing control actions (voltage variation or reactive power injected) is emphasized using α and β weights presented in (4-2).

Furthermore, a set of comprehensive grid-related and resources-related constraints need to be considered:

- **Bus Voltage and Current Magnitudes' Constraints:**

The voltage of each node of the network has to be in the permissible range, and the current flowing in the network branches cannot exceed a specific value according to the ANSI [50] standard. Therefore, voltages and currents are constrained by (4-3) and (4-4), respectively.

$$V_{min} \leq \left|V_{i,t}\right| \leq V_{max} \tag{4-3}$$

$$\left|I_{i,t}\right| \leq I_{max} \tag{4-4}$$

- **Active and Reactive Power Balance Constraint:**

Based on the optimal power flow equations, the active and reactive powers at each node in the network must balance (4-5) and (4-6) illustrate the considered balance constraints of the active and reactive powers.

$$\sum_{i=1}^{L} {P_{i,t}}^{lineloss} + \sum_{i=1}^{R} {P_{i,t}}^{load} + \sum_{x=1}^{E} {P_{x,t}}^{EV} = \tag{4-5}$$

$$P_t^{Grid} + \sum_{x=1}^{V} P_{x,t}^{PV} \pm \sum_{x=1}^{B} P_{x,t}^{BESS}$$

$$\sum_{i=1}^{L} Q_{i,t}^{line} + \sum_{i=1}^{R} Q_{i,t}^{load} = Q_t^{Grid} \pm \sum_{x=1}^{E} Q_{x,t}^{EV} \pm \sum_{x=1}^{V} Q_{x,t}^{PV} \pm \sum_{x=1}^{B} Q_{x,t}^{BESS} + \sum_{x=1}^{M} Q_{x,t}^{Cap} \quad (4\text{-}6)$$

- **Feeder Thermal Limit Constraint:**

Due to the thermal limits of the network branches' material, there is a limitation for the maximum kVAs that can be transferred through them. It is important to note that exceeding the maximum value can damage the branches. The maximum kVA for a branch is constrained by (4-7).

$$S_{f,t} \leq S_{f,t,max} \quad (4\text{-}7)$$

- **System Power Factor (PF) Constraint:**

In a distribution network, the PF of a node in the network needs to be maintained within allowed limits by operators as given in (4-8).

$$PF_{min} \leq \left|PF_{i,t}\right| \leq PF_{max} \quad (4\text{-}8)$$

- **Capacitor Banks Constraints:**

For CB constraint, it is considered that the setting of an individual capacitor is $q_{s,t}$, therefore the output of a CB is obtained by (4-9). The CBs are considered discrete control variables, and their reactive output power is constrained by (4-10).

$$Q_{Cap,i,t} = q_{s,t} \cdot Q_c \quad (4\text{-}9)$$

$$0 \leq Q_{Cap,i,t} \leq Q_{Cap,i,\max} \quad (4\text{-}10)$$

- **On-Load Tap Changer Constraints:**

Transformer taps are considered as discrete variables where each transformer tap-step levels-up/-down voltage in a specific range. The output voltage and the tap position of an OLTC should satisfy (4-11) and (4-12), respectively.

$$V_{0,t} = V_S + V_{Tap} . T_t \quad (4\text{-}11)$$

$$T_{min} \leq T_{i,t} \leq T_{max} \tag{4-12}$$

- **Photovoltaic Constraint:**

The reactive power supplied or absorbed by a PV unit is constrained by (4-13).

$$\left|{Q_{x,t}}^{PV}\right| \leq \sqrt{\left({S_{max}}^{PV}\right)^2 - \left({P_{x,t}}^{PV}\right)^2} \tag{4-13}$$

- **Electric Vehicle Charging Stations Constraints:**

Due to physical constraints, the active power of an EV cannot exceed the charging limit [168] and should satisfy (4-14). Moreover, the reactive power an EV charging station can supply/absorb is constrained as given by (4-15).

$$0 \leq P_{EV,i,t} \leq P_{EV,max} \tag{4-14}$$

$$\left|{Q_{x,t}}^{EV}\right| \leq \sqrt{\left({S_{max}}^{EV}\right)^2 - \left({P_{x,t}}^{EV}\right)^2} \tag{4-15}$$

- **Battery Energy Storage System Constraints:**

The reactive power supplied or absorbed by a BESS unit is constrained by (4-16). In addition, an individual BESS unit should satisfy energy level constraints (4-17) -(4-20) [169].

$$\left|{Q_{x,t}}^{BESS}\right| \leq \sqrt{\left({S_{max}}^{BESS}\right)^2 - \left({P_{x,t}}^{BESS}\right)^2} \tag{4-16}$$

$$E_{x,t} + {P_{x,t}}^{Charge}\eta^{Charge}\Delta t + \frac{{P_{x,t}}^{Discharge}\Delta t}{\eta^{Discharge}} = E_{x,t+1} \tag{4-17}$$

$$0 \leq {P_{x,t}}^{Charge} \leq {P_{x,t}}^{Charge,max} \tag{4-18}$$

$${P_{x,t}}^{Discharge,max} \leq {P_{x,t}}^{Discharge} \leq 0 \tag{4-19}$$

$${E_x}^{min} \leq E_{i,t} \leq {E_x}^{max} \tag{4-20}$$

4.1.2 Stage II: h/n Time-Resolution Single Objective Formulation

The purpose of the second stage is to minimize further the voltage deviation of the network buses over 24 hours, considering h/n time-resolution. Thus, it is a single-

objective optimization problem as formulated in (4-21) where the control variables are ${Q_x}^{PV}$, ${Q_x}^{EV}$ and ${Q_x}^{BESS}$.

$$\min \sum_{t=1}^{93} \sum_{p=1}^{R} (V_{p,t} - V_{ref})^2 \quad (4\text{-}21)$$

The constraints of this stage are the same grid constraints given by (4-3) -(4-8), and the DERs constraints given by (4-13) -(4-20).

In summary, the optimization algorithm consists of two stages. Stage I, on an h basis, aims to simultaneously minimize power loss in lines by (4-1) and the number of control actions of CBs and OLTCs by (4-2) subject to the constraints (4-3) to (4-20). This stage limits voltage deviation within the standard boundaries and provides means for the system operator direct input in setting standard switched devices. Stage II aims to minimize the voltage deviation on an h/n basis by (4-21), subject to the constraints in (4-3) -(4-8) and (4-13) -(4-20), which results in further reductions in system losses.

4.2 Solutions for the VVO Problem

In solving VVO, the optimization solver is an essential aspect of achieving appropriate voltage regulation and minimum active power losses. Many methods have been explored in the literature [170]-[185]. Depending on parameters, variables, and system complexity, the VVO problem can be categorized as a:

- **Multi-objective Optimization:** based on the fact that it requires optimizing more than one objective simultaneously subject to defined constraints.
- **Mixed Integer Non-linear Programming (MINLP):** Since the VVO problem is required to handle non-linear systems as it aims to conserve energy consumption and minimize grid loss. Moreover, it is required to deal with integer variables such as CBs (on/off), VRs, and OLTCs (positions). Thus, the formulation of the VVO problem may have nonlinearities.

As previously mentioned, several optimization approaches can be used in VVO

problem solvers. Since it is challenging to examine all of such approaches, an effort is directed towards finding an effective technique for the proposed VVO algorithm with adequate reliability, accuracy, and convergence time. Heuristics and meta-heuristic approaches are considered to provide VVO problems with precise solutions most of the time since they do not force large or real assumptions on the problem [186]-[188]. Furthermore, they have a large enough search space for candidate solutions, which leads to precise optimizations. However, the computational time is affected.

One of the optimization techniques employed to solve VVO problems is Bender's decomposition technique [189]. Its working principle is based on decomposing the high-dimensional optimization problems with many constraints into smaller sub-problems having less constraints, i.e., master and slave. Bender's decomposition approach can be a valid candidate for the VVO problem since the VVO problem comprises disparate constraints and objectives. Also, it can be beneficial in large distribution networks.

A stochastic optimization technique, such as Particle Swarm Optimization (PSO), is a meta-heuristic approach widely used in the power system. It can provide distribution networks with fast solutions. The accuracy, simplicity of working principle, and speed encourage to employ it in problems related to distribution networks [170].

Another extensively employed approach in distribution network optimizations such as VVO is the meta-heuristic genetic algorithm (GA) approach [173]. It has been proven that it can solve optimization problems with adequate precision and convergence time. Determinations of mutation and crossover steps in GA significantly affect the accuracy of the optimization. Moreover, a realistic value of initial populations should be chosen since a large initial population increases the computation time.

In summary, each of the aforementioned optimization approaches has its

advantages and disadvantages based on the topology and condition of the distribution network and the operator requirements in terms of accuracy and convergence time. Bender's decomposition technique can solve the VVO problem with acceptable accuracy and speed. However, creating weak bender cuts can affect its performance. Furthermore, this approach's complexity drove employing it only in large-scale optimization problems where it is recommended to divide the problem into sub-problems [190]. On the other hand, PSO has several advantages, including simple implementation, insensitive to design variables scaling, a straightforward concept, dealing with few parameters, derivative-free, and having a robust global search algorithm. However, the major drawback of PSO is that it tends to converge fast to local optimums. GA is another approach to have an easy working concept. It defeats other approaches in distribution network applications and performs adequately since distribution network optimization problems can be easily defined as chromosomes and fitness functions.

Moreover, GA can be easily transferred to other levels of the system model. It can also solve the VVO problem optimally in adequate time. In this thesis, the GA technique is employed to solve the proposed VVO problem, and it is further detailed in the following subsection.

4.2.1 Genetic Algorithm Features and Components

GA is a heuristic optimization technique that is based on evolutionary principles [191]. To achieve the optimized solution, it goes through four main vital steps:

1) **Initialization:** An initial population is randomly generated, representing hundreds/thousands of possible search space solutions. Its value depends on the size and nature of the optimization problem.

2) **Selection:** a portion of the initial population is chosen to form the new generation. Typically, the solutions are selected based on the given fitness function. Thus, the formulation of the fitness function affects the quality of the solution.
3) **Genetic operator:** The second population is generated from the previously chosen set by adopting the crossover and mutation operators.
4) **Convergence criteria:** the condition that the GA process considers algorithm convergence (stop iterating). The criteria differ based on the tackled problem. Generally, when successive iterations do not generate improved results, the algorithm converges to the optimum.

To generate results, GA optimization performs decoding, evaluates fitness function, and runs reproduction, crossover, and mutation. After the population size is defined, the initial population is generated. Populations have to be encoded as binary numbers in order for chromosomes to be identified. It is essential to highlight that the length of chromosomes significantly affects GA optimization. Briefly, optimized results of GA depend on 11 parameters: encoding of chromosomes, fitness function, initial population, population size, elitism preserving rate, coding method, mutation operation and its rate, crossover operation and its rate, termination criterion, length of chromosomes, and parent selection operation.

The most critical parameters are reproduction, crossover, and mutation. GA is considered an encoded process and uses the load flow solving solution to simplify the defined objective and constraints functions. Thus, the complexity of heavy mathematical formulations is avoided. Although the initial population is randomly generated, the continuous evaluation of fitness function assures the optimum solution's achievement. It is safe to say that the local optima can be successfully avoided by adequately tuning the GA optimizer parameters. Since the VVO problem involves

optimizing more than one variable simultaneously, a multi-objective GA (MOGA) approach is employed.

4.2.2 Multi-objective Genetic Algorithm

Multi-objective optimization is the mechanism of optimizing more than one conflicting objective simultaneously. It does not have a single solution, as in a single optimization problem. However, it finds a set of non-dominated or Pareto optimal solutions demonstrated in a Pareto Front. A solution is defined as a non-dominated solution if none of the objective functions can be improved without sacrificing other objectives [192]. Figure 4-1 presents a Pareto Front for two objectives. The green solutions represent feasible solutions for the two objectives, yet not optimal, where yellow solutions are non-feasible solutions, and the blue solutions represent the non-dominated solutions.

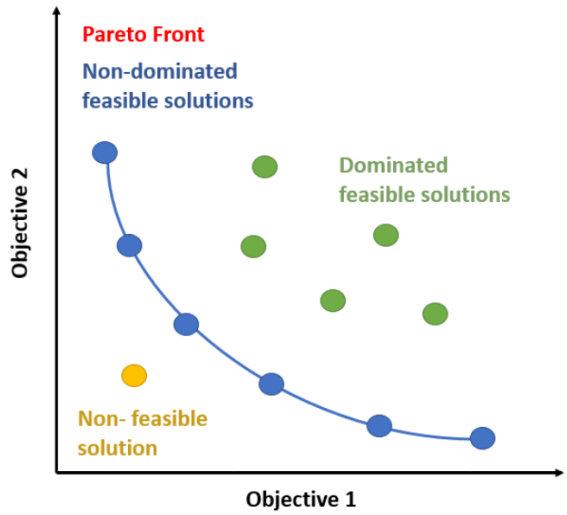


Figure 4-1. Pareto front for two objectives [192]

Elitist non-dominated sorting genetic algorithm (NSGA-II) is a development of the multi-objective GA approach. It is an evolutionary algorithm that is widely used in solving constrained multi-objective optimization problems. The NSGA-II algorithm

proposed in [192] enhances the convergence properties by introducing the idea of elitism to non-dominated sorting GA.

The sorting approach requires computing two parameters: 1) the domination count n_i, which represents the number of solutions that dominate solution i, 2) the domination set s_i, which represents the set of solutions that solution i dominates. The first non-dominated set of solutions is the solutions with $n_i = 0$. Then, from this set, each solution with $n_i = 0$ is visited and the domination count of each member k of its set s_i is reduced by one. Thus, solution i is removed from the domination set s_k. All members having $n_q = 0$ is saved in a separate list K, where K represents the second set of non-dominated optimal solutions. This process is then performed for each member of K to obtain the third non-dominated set. This procedure is repeated until all non-dominated sets of solutions are identified [193].

The computation of crowding distance requires the population of each objective to be sorted in ascending order first. With each objective being normalized, the total crowding distance is the summation of separate distance values for each objective. The crowded comparison operator $\succ_c$ ensures a uniform spread out of the set of optimal solutions. Assuming that every individual has two entities; rank and distance, the $\succ_c$ for two individuals m and n is defined as:

$$m \succ_c n \quad if$$

$$(m_{rank} > n_{rank}) \quad OR \quad (m_{rank} = n_{rank}\, AND\, m_{distance} > n_{distance})$$

Thus, a better solution than another, it has to have a better non-dominated front; higher rank. In the case of having the same rank, it requires better crowding distance.

The main steps of the NSGA-II approach are; first, as previously explained, an initial population is randomly generated based on the objective and constraints. The fitness function is then evaluated. Next, selecting individuals in the population is

performed based on ranking and crowding distance (an indication of an individual's closeness to its nearest neighbors). The crossover and mutation operators then run to generate the offspring (new population). The parent and children populations are combined, and based on the elitism and crowding distance, the individuals are selected [193]. If one of the convergence criteria is met, the algorithm stops and prints the Pareto Front with the non-dominated solutions. Otherwise, a new population is generated. A flowchart illustrating the NSGA-II algorithm is given in Figure 4-2.

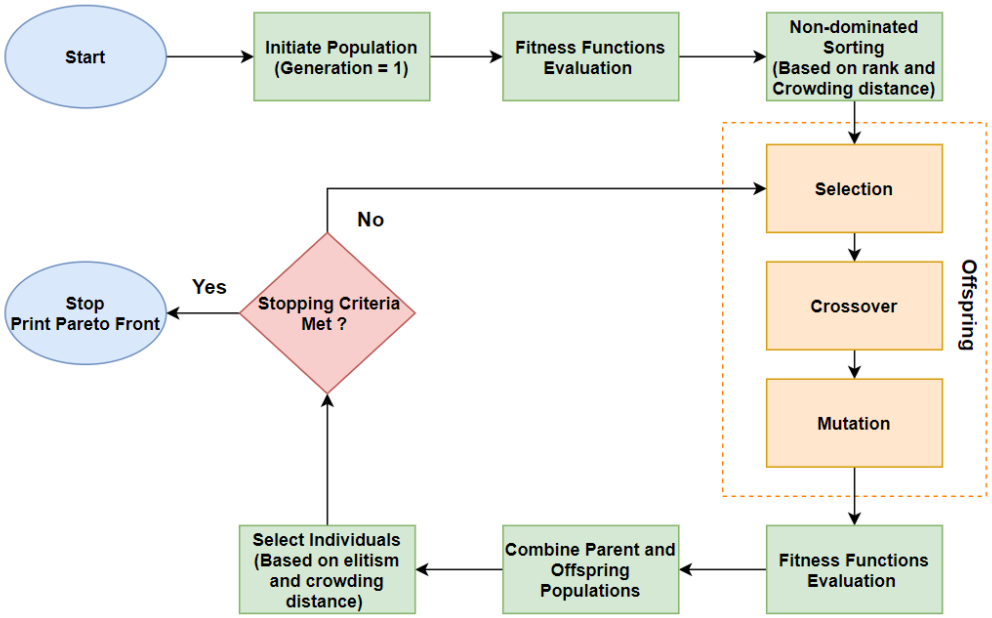


Figure 4-2. Flowchart of the NSGA-II algorithm [193]

4.2.3 Interior Point Method

One of the techniques used in solving large-scale nonlinear convex single optimization problems with many design variables is the interior point (barrier) method (IPM). It is known for its simplicity in mathematical modeling. It is based on the concept of iteratively approaching the optimal solution from the interior of the feasible set [194]. To force the optimal unconstrained value to be in the feasible space and avoid

the violation of constraints, the objective function is augmented by a barrier term.

The algorithm uses one of the two main types of steps at each iteration to solve the optimization problem: a direct step and a conjugate gradient (CG) step that uses a trust region. By default, the IPM takes the direct step first. If it cannot, it takes the CG step. When it does not take the direct step, the approximate problem is not locally convex near the current iterate. A flowchart illustrating the IPM algorithm employed is given in Figure 4-3.

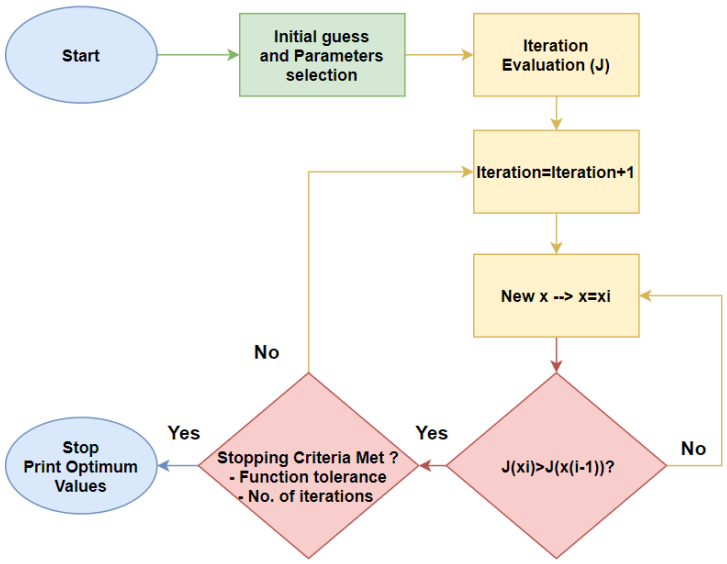


Figure 4-3. Flowchart of IPM algorithm [194]

4.3 Proposed VVO Algorithm Framework

In the proposed algorithm, stage I involves solving a nonlinear multi-objective optimization problem considering an h based time resolution. Stage II comprises solving a single nonlinear optimization problem based on an h/n time resolution. Figure 4-4 highlights the algorithm's main steps and the slow and fast stages' implementation procedures.

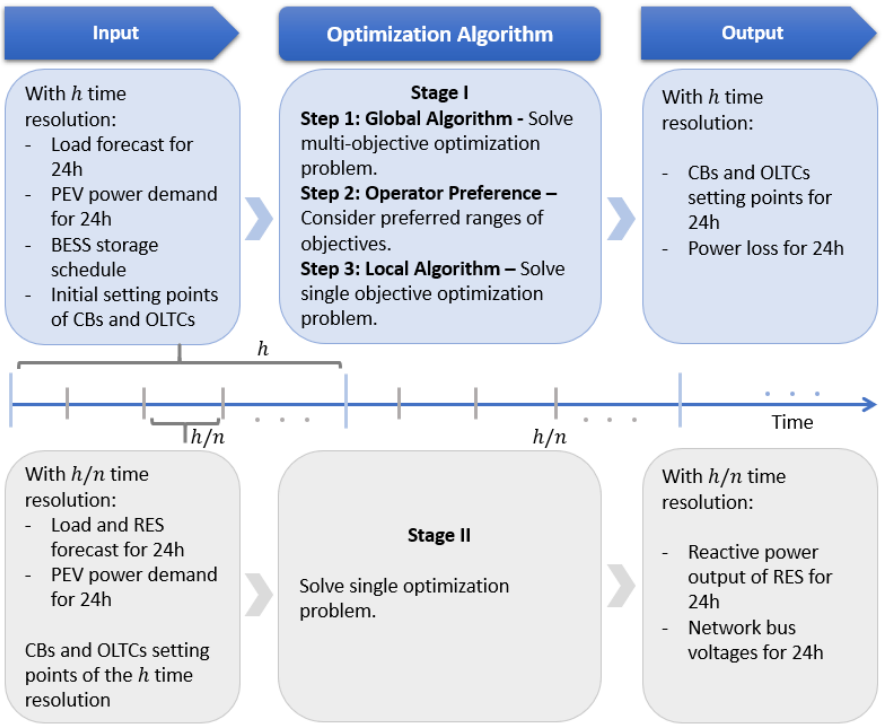


Figure 4-4. Main steps of the proposed two-timescale VVO algorithm

4.3.1 Stage I: Within the h Time-Resolution

The data required to solve stage I's optimization problem are the network topology, forecasted load demand, EV charging station demand, expected generation of DERs, and the BESSs charging/discharging schedules with an h time resolution. The outputs of this stage are the day-ahead power loss and the settings of the CBs and OLTCs with an h time resolution. These settings are kept constant and used in stage II. The nonlinear multi-objective problem in stage I is solved in three steps to ensure reaching an optimal solution considering the operator's preferences. These steps include a global search algorithm, system operators' engagement step, and a local search algorithm.

Step 1: Global search: A global search procedure is first used to explore the design space, assuming the settings of the CBs and OLTCs are not known. This paper uses a MOGA [195],[196] to coordinate the settings of the CBs and OLTCs based on h time

resolution and simultaneously solve (4-1) and (4-2) subject to (4-3)-(4-12). The solution to this step is the Pareto front, as it shows the boundary of the solution of the multi-objective problem in the objective space [197]. This boundary highlights all the non-dominated optimal solutions of the two objectives: power loss and the number of CBs and OLTCs adjustments.

Step 2: Operator preference: The operators' engagement in the VVO problem is included by considering their preferences for the range of power loss and the number of CBs and OLTCs adjustments. This gives the operators the capability of choosing the zone of solutions based on the network condition. The operator expertise and preference are reflected in single or multiple options extracted from the Pareto optimal solutions that resulted in step 1. These solutions are selected as candidate solutions for further optimization in step 3.

Step 3: Local search: The local search is used to achieve the optimal solution using an initial guess, the candidate solutions preferred by the operator in step 2. The multi-objective problem is converted to a single objective problem using the ε-constraint method. Where one of the objectives is selected, and the other objective acts as a restriction constraint. In this application, the power loss minimization in (4-1) is taken to be the single objective function to be solved, and the control actions objective function in (4-2) is transferred to the constraints of the optimization problem, as given in (4-22).

$$\sum_{t=1}^{24}[\alpha.cap_{actions} + \beta.tap_{actions}] \leq \varepsilon \qquad (4\text{-}22)$$

where ε ranges between the upper and lower bounds of the control actions objective function, defined by the operator, in an equidistant manner.

For solving a large-scale nonlinear optimization problem, the paper uses the interior point method (IPM) [198],[199] to further minimize the power loss in the

network by controlling the reactive power output of the DERs based on h time resolution.

4.3.2 Stage II: At the h/n Time Resolution

In this stage, the IPM [198],[199], is used to reduce the network voltage deviations by controlling the reactive output power of the DERs. The data required to solve the optimization problem of stage II is similar to stage I data yet in an h/n time resolution, in addition to the CBs and OLTCs settings obtained from stage I. Moreover, the key outputs of stage II are the network bus voltages and the reactive output power of each DER on an h/n time resolution.

4.4 Summary

The proposed VVO algorithm is formulated as a two-timescale algorithm to handle fast and slow varying demand components. The slow timescale, with h time resolution, is comprised of 3 steps. Step 1 aims to simultaneously optimize power loss and number of control actions as given in (4-1) and (4-2), respectively, subject to constraints (4-3) to (4-20). This step utilizes utility-owned assets using MOGA. In step 2, the optimization problem is transferred from multi-objective to single objective using the epsilon constraint method considering (4-1) as the objective function, which is then further optimized in step 3 utilizing DERs along with the utility-owned assets using IPM. While the fast time scale, with h/n, time resolution, aims to optimize voltage deviation as given in (4-21) using IPM and utilizing reactive output power of DERs subject to constraints (4-3) -(4-8) and (4-13) -(4-20).

CHAPTER 5 : THE TWO-TIMESCALE VOLT/VAR OPTIMIZATION SOLUTION

This chapter first demonstrates the tested network along with the parameter settings and the assumptions made of the two examined case studies. It presents the different studied scenarios in this thesis and guides to code implementation of the different timescales. Then, it examines the impact of the proposed algorithm in reducing the power losses and the voltage variations in the tested network, considering the different cases for both case studies. The proposed VVO algorithm is implemented on a gradually modified ADN. The system losses are monitored during the three steps of stage I. The enhancement in voltage profile variation results from the two stages of the VVO algorithm is evaluated.

5.1 Test System Description and Parameter Settings

IEEE 33-bus system is a radial distribution network widely used for testing in power system-related problems. It consists of 33 buses (1 slack bus and 32 PQ busses) and 32 transmission lines with a value of 12.66 kV rated voltage, load size of 3.715 MW and 2.3 MVar, one transformer with a maximum real power of 10 MW and maximum and minimum reactive power is 10 MVar and -10 MVar, respectively. The data of the standard IEEE 33-bus system is provided in Appendix A.

To test and validate the proposed conventional optimization-based VVO scheme, the algorithm is tested on two different case studies with a different number of resources, their locations and ratings.

5.1.1 Case Study A

The proposed algorithm is tested on a modified IEEE 33-bus system that includes CBs, OLTCs, and DERs, as in Figure 5-1. The parameter settings of the system are presented in Table 5-1. The following assumptions are made:

- 6 CBs are installed on buses 2, 7, 12, 20, 23, and 29.

- One OLTC is installed at the substation transformer located at bus 1 with ± 5% range and 20 tap positions.
- 4 PVs are located at buses 8, 13, 22, and 24, with an inverter rating of 1.5 MVA.
- BESSs are located at buses 16 and 30 with an inverter rating of 1 and 1.5 MVA, respectively.
- EV charging stations are at buses 9, 14, and 21, with an inverter rating of 0.3, 0.3, and 0.2 MVA, respectively.

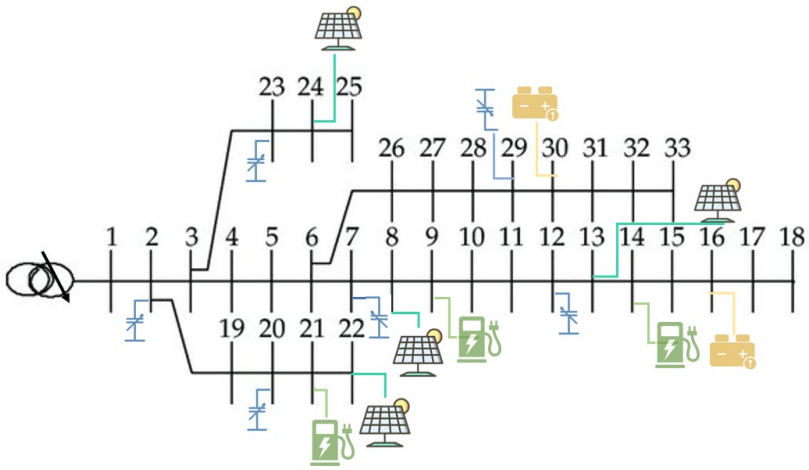


Figure 5-1. Case study A modified IEEE 33-bus system

Table 5-1. Case Study A System Parameters

Parameter	Value
Peak demand load	8.0 +2.6 MVA
Substation voltage (V_S)	1.04 p.u (12.67 kV)
$Q_{Cap,i,\max}$	300 kVar
V_{Tap}	0.05
T_{min}	-10
T_{max}	10
V_{min}	0.95 p.u
V_{max}	1.05 p.u
S_{max}^{PV}	1.5 MVA
$S_{max}^{BESS1,2}$	1 MVA, 1.5 MVA
$S_{max}^{EV1,2,3}$	0.3 MVA, 0.3 MVA, 0.2 MVA

The demand of the plug-in electric vehicle (PEV) depends on various aspects: location of the charging station, number of PEVs charging, the capacity of the battery, charging level, as well as the start of charging time and state of charge of the battery that varies based on different driving behaviors of PEVs. Using accurate data, EV patterns are generated. The varying demand for each EV charging station is shown in Figure 5-2(a). The residential day-ahead forecasted active and reactive power demand is shown in Figure 5-2(b). The reactive power is estimated by randomly varying the load power factor (PF) between 0.8 – 0.95 along the day. The load of PEVs charged at home is assumed as a part of the residential load. The forecasted PVs and BESSs generation profiles are estimated, as shown in Figure 5-2(c).

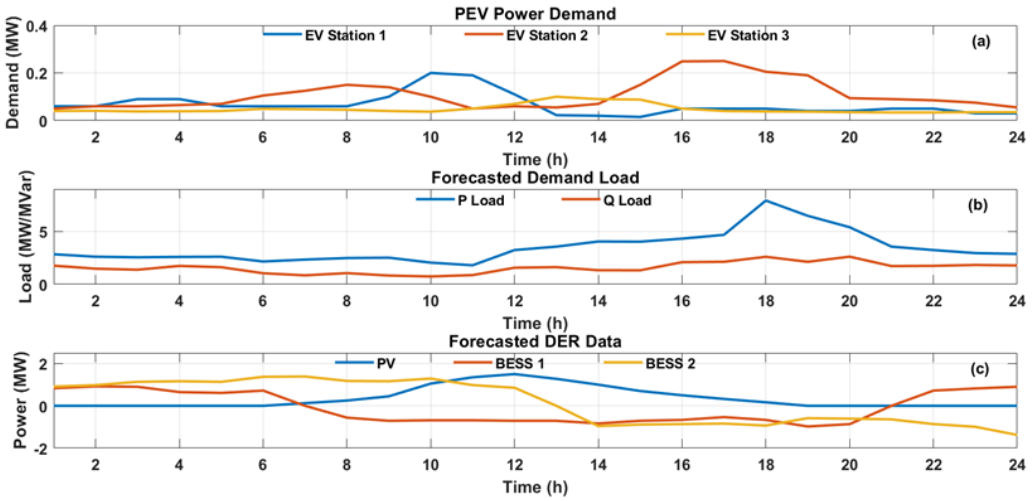


Figure 5-2. Case study A input data (a) EV charging stations power demand, (b) forecasted demand load, and (c) forecasted DER data

Furthermore, DERs inverters' capability and PQ operating range are assumed to be optimal; inverters' full capacity is utilized. Thus, the limit of injection/absorption of reactive power of PVs, EV charging stations, and BESSs is only constrained by (4-13),

(4-15), and (4-16), respectively. The limit of the absolute amount of reactive power that DERs' inverters can provide is shown in Figure 5-3.

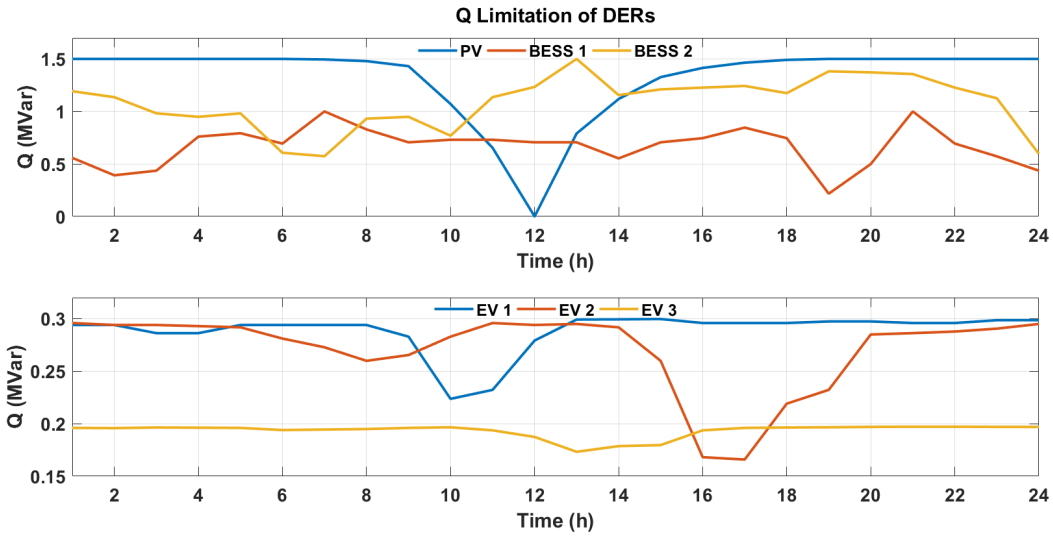


Figure 5-3. Case study A Q limitation of DERs

5.1.2 Case Study B

The VVO algorithm is evaluated on a modified IEEE 33-bus system that includes CBs, OLTCs, and DERs, as in Figure 5-4. The considered parameter settings of the system are presented in Table 5-2. The following assumptions are made:

- Three CBs are installed on buses 11, 20 and 29 with 1800kVar and 4 stages each
- One OLTC is installed at the substation transformer located at bus 1 with ± 5% range and 20 tap positions
- Four PVs are located at buses 8, 13, 22, and 24 with an inverter rating of 1.5 MVA.
- Two BESSs are located at buses 16 and 30 with an inverter rating of 1 MVA and 1.5 MVA, respectively.
- Three EV charging stations are at buses 9, 14, and 21 with an inverter rating of 0.5 MVA.

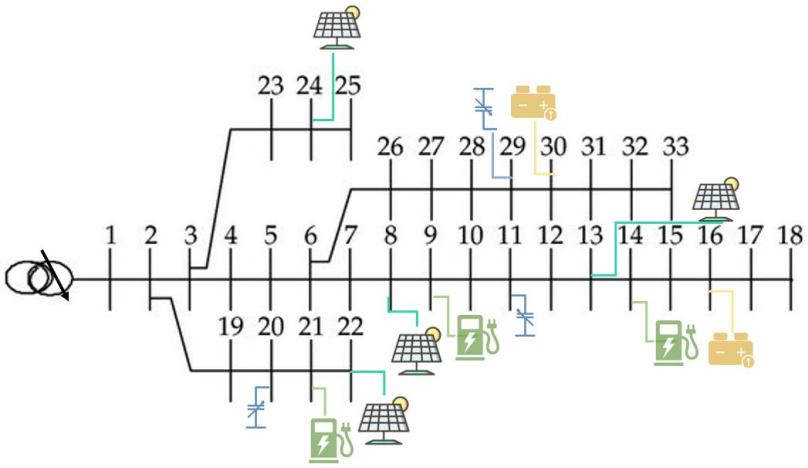


Figure 5-4. Case study B modified IEEE 33-bus system

Table 5-2. Case Study B System Parameters

Parameter	Value
Peak demand load	9.5 +j5.5 MVA
Substation voltage (V_S)	1.04 p.u (12.67 kV)
$Q_{Cap,i,\max}$	1800 kVar
V_{Tap}	0.05
T_{min}	-10
T_{max}	10
V_{min}	0.95 p.u
V_{max}	1.05 p.u
S_{max}^{PV}	1.5 MVA
$S_{max}^{BESS1,2}$	1 MVA, 1.5 MVA
$S_{max}^{EV1,2,3}$	0.5 MVA

The varying demand for each EV charging station is shown in Figure 5-5(a). The residential day-ahead forecasted active and reactive power demand is shown in Figure 5-5(b). Similar to case study A, the reactive power is estimated by randomly varying the load power factor (PF) between 0.8 – 0.95 along the day, and the load of PEVs charged at home is assumed as a part of the residential load. The forecasted PVs and BESSs generation profiles are estimated, as shown in Figure 5-5(c). The capability of DERs inverters and their PQ operating range is assumed to be optimal; inverters' full

capacity is utilized. Thus, the limit of injection/absorption of reactive power of PVs, EV charging stations, and BESSs is only constrained by (4-13), (4-15), and (4-16), respectively.

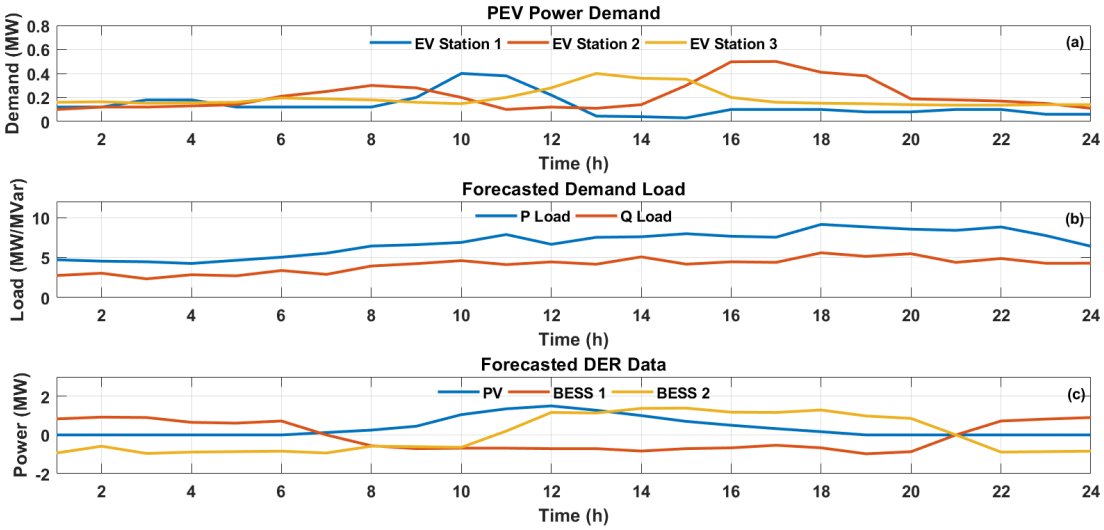


Figure 5-5. Case study B input data (a) EV charging stations power demand, (b) forecasted demand load, and (c) forecasted DER data

5.2 Simulated Cases

Four cases are simulated and compared with a base case. The proposed algorithm results in optimizing the power losses are compared in the four simulated cases, and case 4 is used to demonstrate the algorithm's effectiveness in optimizing the voltage deviation. Figure 5-6 highlights the simulated cases.

- ***The base case (C0)*** is considered a reference case in the study, without the VVO algorithm. The system is simulated in the presence of the DERs without CBs and OLTC.
- ***Case 1 (C1) and case 2 (C2)*** are compared in terms of the number of control actions of the switch-based settings. The system in C1 and C2 has inverter-based PVs only

and switch-based CBs and OLTC. A higher weight is assigned for CBs in C1 ($\alpha = 0.8$ and $\beta = 0.2$), and higher weight for the OLTC is considered in C2 ($\alpha = 0.2$ and $\beta = 0.8$).

- ***Case 3 (C3)*** considers a system with inverter-based BESSs and PVs only and switch-based CBs and OLTC. A higher weight is considered for the OLTC ($\alpha = 0.2$ and $\beta = 0.8$).
- ***Case 4 (C4)*** considers a system with inverter-based BESSs, EV charging stations and PVs, and switch-based CBs and OLTC ($\alpha = 0.2$ and $\beta = 0.8$). The proposed multi-objective algorithm is applied to the system considering three reference voltages

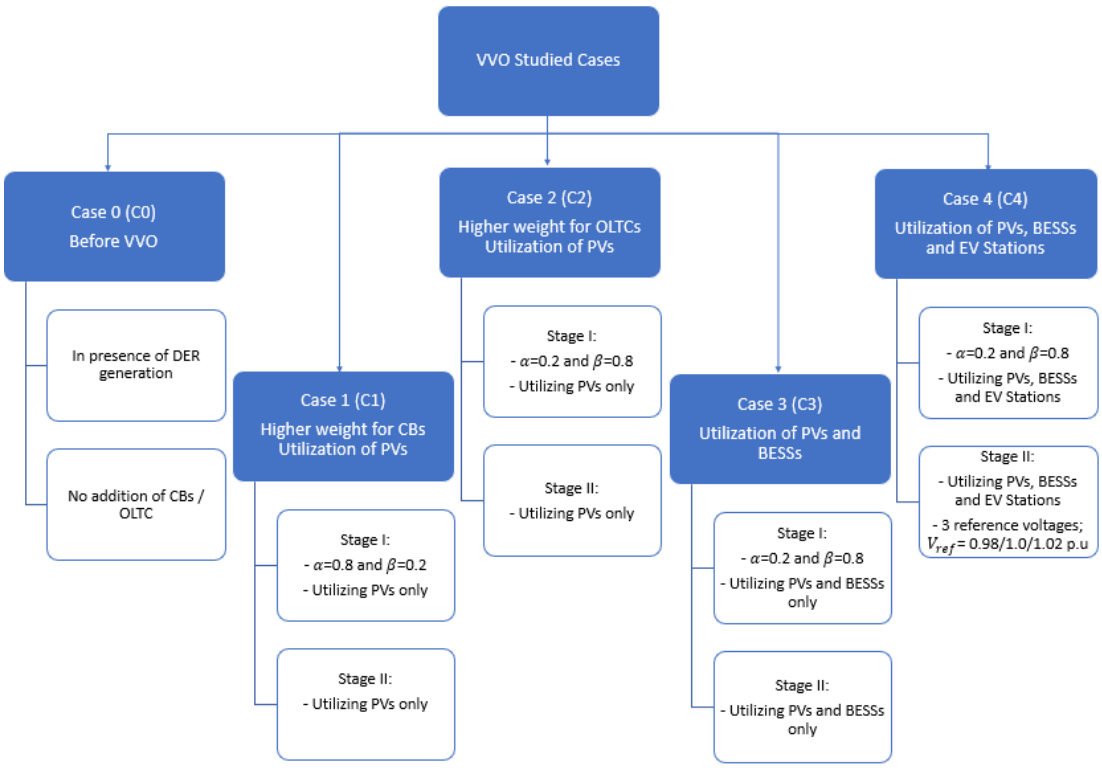


Figure 5-6. VVO studied cases

5.3 Code Implementation

This thesis considers the slow time resolution h to be an hour and n to be 4. Thus, stage I and stage II are solved based on hourly and 15-minute time resolutions, respectively. The VVO algorithm is implemented in MATLAB utilizing MATPOWER. MATPOWER is an open-source MATLAB package that incorporates simulation tools for solving optimal power flow and power flow problems based on buses, branches, and transformer matrices [200]. It is used to solve the power flow problem, with Newton-Raphson (NR) methodology, to obtain line losses and bus voltages.

5.3.1 Stage I: Hourly Basis

The objective function is to find the optimal dispatch of CBs and OLTCs in stage I; step 1, and reactive power of DERs in stage I; step 3 and stage II for the 24 hours for the given constraints. The tested network contains 6 CBs, and for each CB, the solution should contain 24 values representing the status of the CBs at each hour of the day. Thus, the total number of variables for CBs is 6*24=144, and for OLTCs is 1*24=24 since the tested network has one transformer. Subsequently, the total number of variables in stage I; step 1, is 168. The output is the optimization vector variable comprising of setting positions (integer) for each CB and OLTC at each hour of the day. The main steps of the developed algorithm are presented in Figure 5-7.

As previously mentioned, the evaluation of objective functions depends on the variables generated by the MOGA optimizer, and the fitness value of selected individuals requires decoding. For CBs, the value generated by MOGA can take integer values between 0 and 20, and for OLTC, since it is a 20-tap transformer, the value generated is between -10 and +10. These integer values are then decoded, and the load flow is solved for each individual of the population while respecting the defined constraints. Before running load flow every hour, the active and reactive demand and the DERs data are updated. The NR load flow runs for every hour of the day to develop

all possible power loss profiles. The MOGA optimizer determines the non-dominated solutions for power loss and the number of control actions based on the solved load flow.

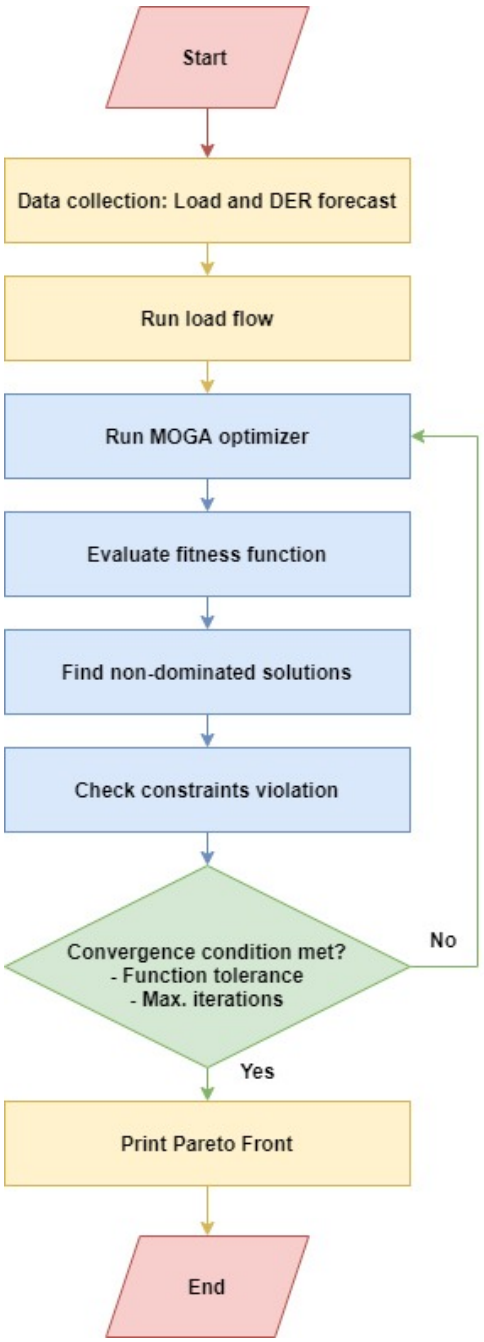


Figure 5-7. Flowchart of stage I; step 1 of the developed algorithm

The MOGA optimizer runs and generates several positions of OLTCs and settings of CBs. For every population generated, the power loss and number of control actions are calculated. It evaluates the objective functions and finds the non-dominated solutions that do not violate the given constraints. If the convergence criteria are met, it prints the optimized settings of CBs and OLTCs otherwise, it generates a new population and reevaluates the objective function until one of the convergence criteria is met.

In the MOGA optimization, the population size is set to 400 to avoid local minima and keep it under reproduction. The elite count is set to 20. For the selection function, the binary tournament approach is chosen. To provide the MOGA suitable time for convergence before termination, the generation is set to 600. Stall generation is set to 400 and function tolerance to 0.0001. The algorithm will terminate when the objective function value is less than the function tolerance.

In stage I, step 3, the objective is to reduce power loss further. Depending on the utilized DERs, the total number of variables is n*24. For instance, in C4, all the DERs are utilized, 4 PVs, 2 BESSs, and 3 EV charging stations along with the CBs and OLTC. Thus, the total number of variables is 168+ (9*24) =384. The output is the optimization variable comprising the amount of reactive power (continuous) supplied/absorbed by DERs each hour of the day. The main steps of the algorithm of stage I; step 3 are similar to stage I; step 1. However, instead of running the MOGA optimizer, the IPM optimizer runs to find optimal output reactive power for each DER at each hour. Based on the operator preference range, only the candidate solutions are further optimized to achieve a higher reduction in power loss. The CB and OLTC settings are fixed at this point and used in stage II.

In the IPM optimizer, the initial guess for all DERs is set to 0 MVar and for the CBs, and OLTC is the solution with the lowest number of control actions from the candidates' solutions. The maximum function evaluation is set to 5000, maximum iterations are set to 2000, and tolerance function and tolerance constraints are set to 0.0001.

5.3.2 Stage II: 15-min Basis

In stage II, the optimization problem is formulated with 15-min time resolution to minimize bus voltage deviations using the reactive output power of all existing DERs; therefore, the total number of variables is 9*93=837. The output is the optimization variable comprising the amount of reactive power (continuous) supplied/absorbed by DERs at each 15-min of the day. The main steps of the developed algorithm for this stage are similar to stage I; step 3. The CB and OLTC settings for every 4 timesteps (1 hour) are known from the previous step, and the initial guess for all DERs is set to 0 MVar. In this stage, the maximum function evaluation is set to 5000, maximum iterations are set to 3000, and tolerance function and tolerance constraints are set to 0.0001.

5.4 Proposed two-timescale results and discussions

5.4.1 Stage I: Hourly basis

In step 1, the MOGA optimization is implemented to solve the multi-objective optimization problem. The goal is to minimize power loss in lines and the number of control actions by controlling the setting positions of the CBs and OLTCs. Figure 5-8 and Figure 5-9 show the Pareto front from solving (4-1) and (4-2) for both case studies, respectively. The solution represents the relationship between the number of control actions in switch-based devices and system line losses. The global search algorithm is implemented in the two cases; C1 and C2. The defined weights in C1 for setting change of CBs α=0.8 and OLTC β=0.2 results in a minimum number of 18 and 23 control

actions for both case studies, respectively. This weight selection presents a significant difference compared with C2 system with selected weights of $\alpha = 0.2$ and β=0.8 and results in a minimum of 35 and 20 control actions for both case studies, respectively. Table 5-3 and Table 5-4 lists the number of control actions of each CB and OLTC for the two case studies, respectively.

System operators' experience is included in the VVO problem's solution by considering the operator preference range (OPR) extracted from the Pareto front. As indicated in Table 5-3, for case study A, the minimum number of control actions based on C1-OPR and C2-OPR are 18 and 35. While, for case study B, the minimum number of control actions based on C1-OPR and C2-OPR are 23 and 20, as highlighted in Table 5-4. The OPR is affected by the weights α and β assigned for CBs and OLTC. It can be seen that for case study A, the Pareto front of C1 shows optimal solutions having a lower number of control actions compared with the Pareto front of C2. In C1, the weight of the number of setting changes of CBs is higher than the weight of the OLTC, and the number of CBs is more significant than the OLTCs in the modified tested network. Thus, it leads to a lower total of control actions. However, for case study B, the Pareto front of C1 showed optimal solutions having approximately the same number of control actions compared with the Pareto front of C2. Hence, the difference between the total number of control actions in C1 and C2 depends on the number of CBs and OLCTs in the network. However, the same conclusion will be deduced; the type of asset, CB or OLTC, having higher weight will be more optimized than the other.

Table 5-3. Number of Control Actions of CBs and OLTC for C1 & C2 – Case Study A

Case #	CB1	CB2	CB3	CB4	CB5	CB6	Tap	Total
C1	2	4	1	1	1	1	8	18
C2	7	7	4	4	6	4	3	35

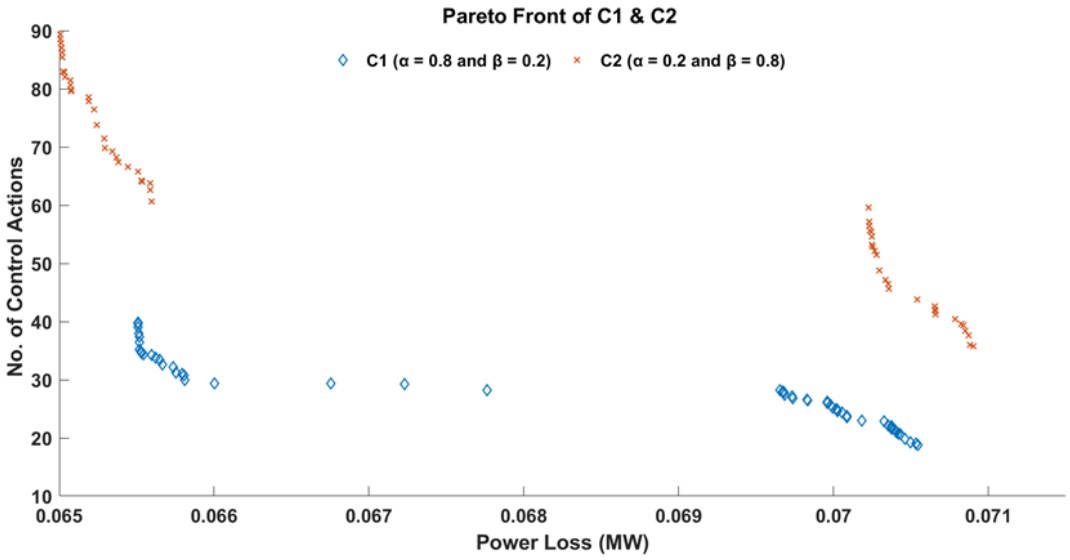


Figure 5-8. Case study A Pareto front solutions of C1 and C2

Table 5-4. Number of Control Actions of CBs and OLTC for C1 & C2 – Case Study B

Case #	CB1	CB2	CB3	Tap	Total
C1	3	2	2	16	23
C2	5	2	5	8	20

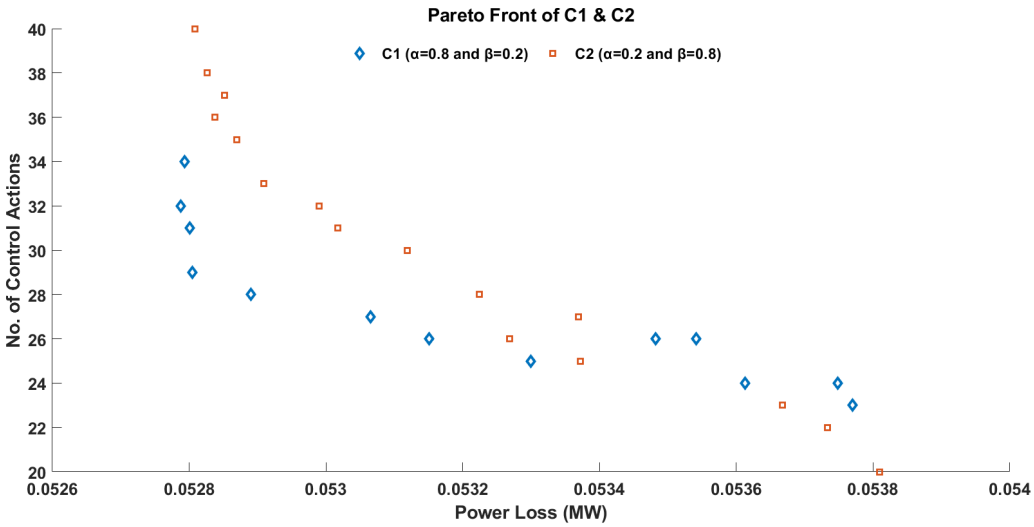


Figure 5-9. Case study B Pareto front solutions of C1 and C2

Furthermore, the selection of OPR depends on the operators' experience in

defining system assets condition and the power quality level expected by end-users. In general, the minimum number of control actions for C1 is lower than C2. A large number of CBs switching results in transient phenomena that affect distribution transformers and other system assets. However, many OLTC switching actions (increment or decrement) may degrade and damage the voltage regulating device, hence disabling the substation transformer. The total number of control actions of the OLTC is lower in C2 compared to C1. In this application, the importance of the lifetime of the OLTC, which may affect the whole system, is given high priority, and hence the Pareto front of C2 is used as the solution of step 1 in stage I and selected for implementation in the other cases. The system operator defined C2-OPR of 35-40 control actions and a range of 0.069-0.073 MW in power loss for case study A. While, for case study B, the system operator defined C2-OPR of 20-26 control actions and a range of 0.0533-0.0539 MW in power loss.

In step 3, the IPM optimization code is developed to minimize the power loss further, while the number of control actions is considered a constraint. The ε-values range from 35 to 40 and 20 to 26 for both case studies, respectively, and the optimal solution with the lowest number of control actions is taken as the initial guess. Figure 5-10 and Figure 5-11 show the day-ahead power loss curves estimated by Stage I of the proposed algorithm considering the modified system with the cases mentioned above for the two case studies. Figure 5-10(a) and Figure 5-11(a) show the power loss of C0 after step 1 of stage I of C2 and after step 3 of stage I of C2. The results of C2, C3, and C4 after step 3 are illustrated in Figure 5-10(b) and Figure 5-11(b). Table 5-5 and Table 5-6 highlight the day-ahead power loss values and the peak power loss and their percentage improvement for the different studied cases of both case studies, respectively.

For case study A, it can be noted from Figure 5-10 and Table 5-5 that the global algorithm reduces the day-ahead power loss to 0.07 MW (approximately 28%) with a peak of 0.36 MW and that the local algorithm controlling the reactive output power of PVs further reduces it to 0.065 MW (step 3 of C2) with a peak of 0.33 MW. The utilization of reactive power of BESSs and EV stations and the PVs reduces the power loss more by approximately 8% (C4) and the peak by 11%. While, for case study B, the day-ahead power loss is reduced to 0.05 MW (approximately 24%) with a peak of 0.12 MW after the global algorithm. While, the local algorithm utilizing the reactive output power of PVs further reduces the power loss to 0.04 MW with a peak of 0.11 MW. The utilization of reactive power of BESSs and EV stations along with the PVs reduces the power loss more by approximately 11% and the peak by 18%. Thus, the three steps of stage I achieve better results with all DERs in the network.

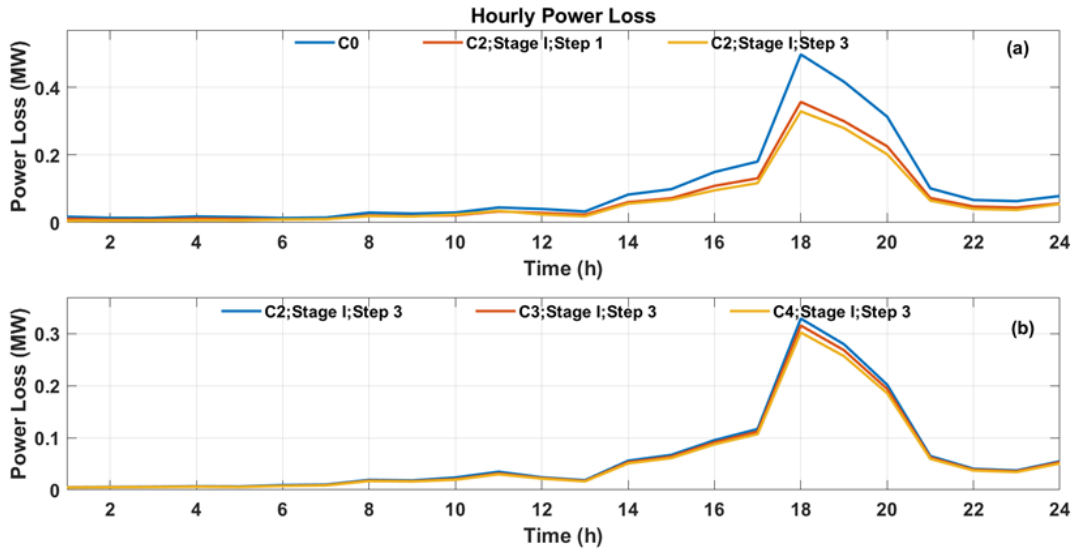


Figure 5-10. Case study A power loss during different stages for C0 and C2-C4

Table 5-5. Case Study A Performance Evaluation of Stage I – Power Loss

Case #	P_{loss} (MW)	% P_{loss} Improvement	P_{Peak} (MW)	% P_{Peak} Improvement
C0	0.098		0.498	
C2; Step 1	0.071	28.3	0.357	28.3
C2; Step 3	0.065	33.8	0.329	33.9
C3; Step 3	0.064	35.1	0.316	36.5
C4; Step 3	0.063	36.2	0.303	39.2

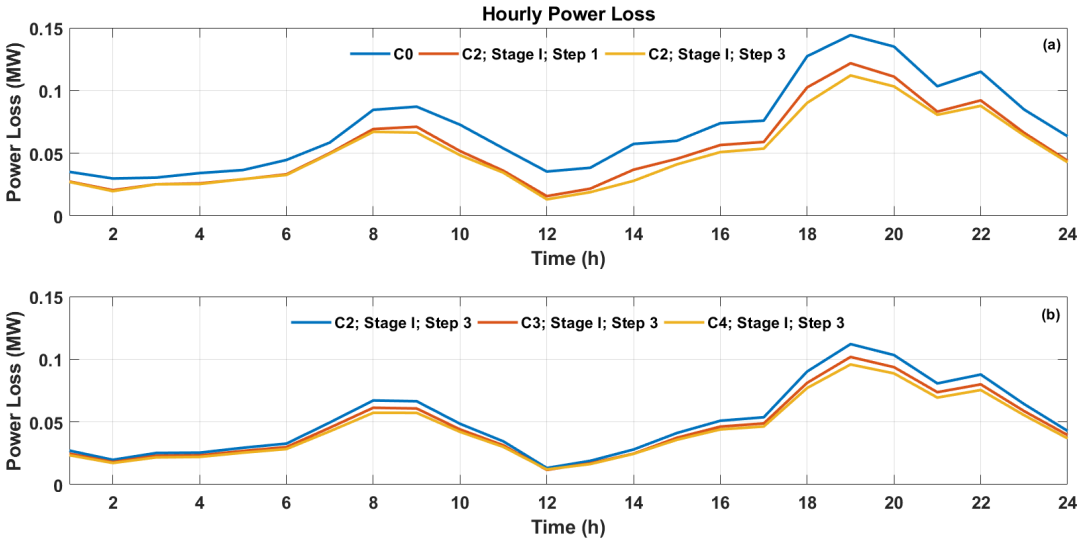


Figure 5-11. Case study B power loss during different stages for C0 and C2-C4

Table 5-6. Case Study B Performance Evaluation of Stage I – Power Loss

Case #	$\boldsymbol{P_{loss}}$ (MW)	% $\boldsymbol{P_{loss}}$ Improvement	$\boldsymbol{P_{Peak}}$ (MW)	% $\boldsymbol{P_{Peak}}$ Improvement
C0	0.070389		0.14422	
C2; Step 1	0.053809	23.6	0.12177	15.6
C2; Step 3	0.048726	30.7	0.11206	22.3
C3; Step 3	0.047119	33.1	0.10180	29.4
C4; Step 3	0.046040	34.6	0.09580	33.6

The impact of the three steps on the network bus voltages is also investigated. The maximum and minimum network voltages along the day are monitored and

evaluated during the different steps of stage I. Figure 5-12 and Figure 5-13 illustrates V_{max} and V_{min} of the network during different stage-I steps of the proposed algorithm and for different cases for both case studies. Figure 5-12(a) and Figure 5-13(a) illustrates the V_{max} and V_{min} of C0 and as extracted after step 1 and step 3 of C2 system. The voltage variations violate the maximum and minimum standard boundaries in C0. However, implementing the global and local algorithms impose bus voltage variations to be localized within the acceptable region for the C2 case. No under/overvoltage violations are detected during stage I, step 1. The V_{max} and V_{min} of the network voltages after step 3 of C2, C3, and C4, are shown in Figure 5-12(b) and Figure 5-13(b). The local search algorithm shows a reduction in the voltage variations. This reduction is more clearly detected in the minimum voltages, as shown in Figure 5-12(b) and Figure 5-13(b). Table 5-7 and Table 5-8 highlight the values of maximum and minimum voltages in the network of the two case studies for the different studied cases.

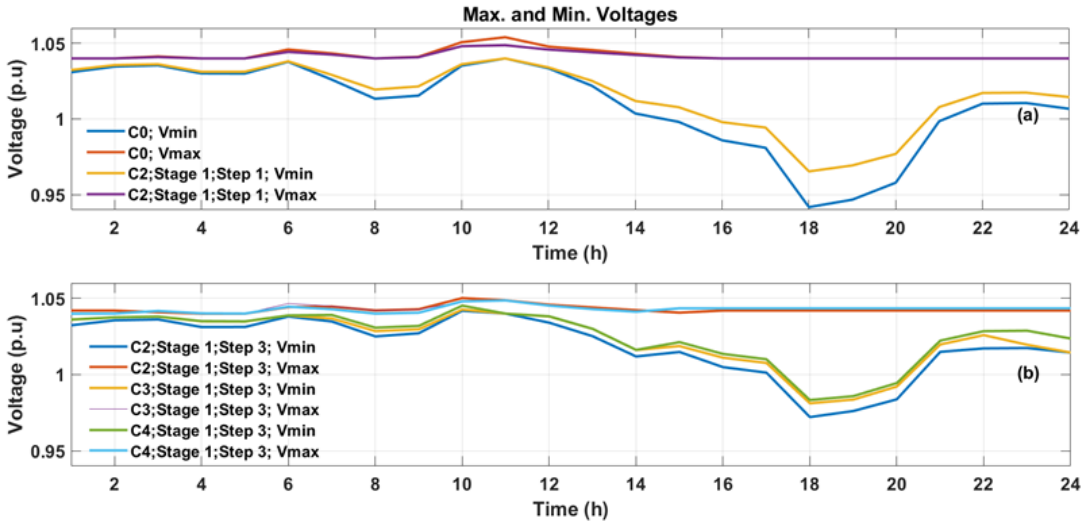


Figure 5-12. Case study A maximum and minimum network voltages after stage I

Table 5-7. Case Study A Performance Evaluation of Stage I – Max. & Min. Voltages

Case #	V_{max}(p.u)	V_{min}(p.u)
C0	1.054	0.9418
C2; Step 1	1.04873	0.96538
C2; Step 3	1.05000	0.97213
C3; Step 3	1.04911	0.98117
C4; Step 3	1.04892	0.98331

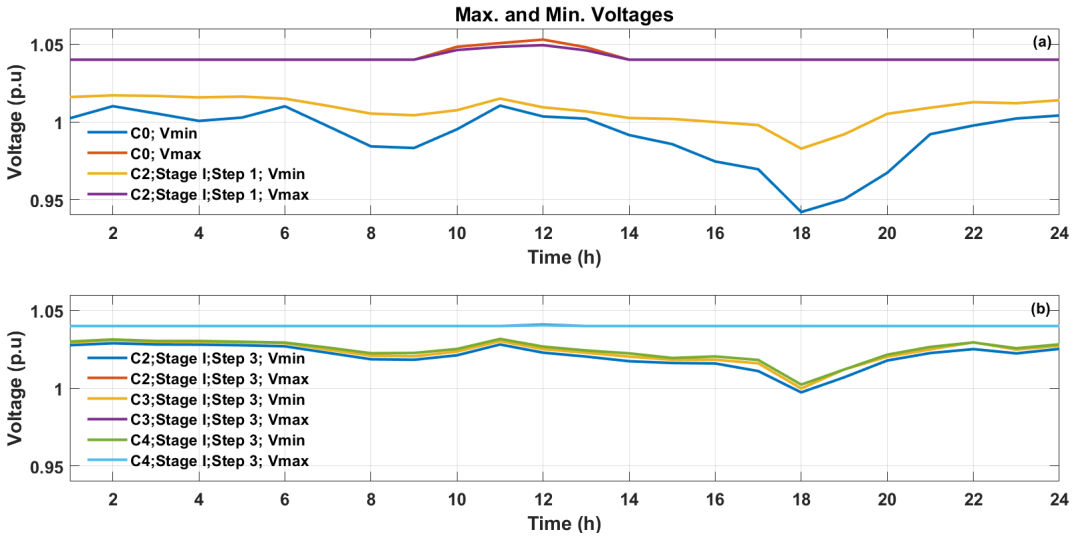


Figure 5-13. Case study B maximum and minimum network voltages after stage I

Table 5-8. Case Study B Performance Evaluation of Stage I – Max. & Min. Voltages

Case #	$\boldsymbol{V_{max}}$ (p.u)	$\boldsymbol{V_{min}}$ (p.u)
C0	1.0529	0.9420
C2; Step 1	1.0494	0.9828
C2; Step 3	1.0457	0.9973
C3; Step 3	1.0459	0.9998
C4; Step 3	1.0455	1.0000

5.4.2 Stage II: 15-minute basis

The IPM optimization solution is developed to schedule the reactive output power of the intended DERs in 15-minute time resolution. Based on the results of stage I, the goal of this stage is to minimize network bus voltage deviations with respect to a reference voltage considering C4. The maximum and minimum network voltages are

examined with different values of V_{ref}; 0.98 p.u, 1.0 p.u and 1.02 p.u for the two case studies. Figure 5-14 shows stage II results of the network's maximum and minimum voltage profiles along the day for case study A considering three reference voltages. It can be noted that as V_{ref} increases, the maximum deviation between V_{max} and V_{min} in the network voltages decreases. For all the system cases, the maximum and minimum voltages remain within the standard range for all three V_{ref} values. Figure 5-16(a), (b) and (c) shows results of voltage deviation of the network along the day after Stage II optimization compared with C0 voltage deviation, considering the three reference voltages, respectively. As can be seen, there is a significant reduction in voltage deviations utilizing the DERs. Figure 5-16(d) illustrates the voltage deviation curves of the three reference voltages. It is also noted that as V_{ref} increases, the voltage deviation decreases implying higher reduction of maximum deviation between V_{max} and V_{min} in the network voltages.

Stage II results of the scheduled reactive output power of the 4 PVs, 2 BESSs, and 2 EV stations are shown in Figure 5-15 and Figure 5-17 for the two case studies, respectively. The estimated reactive output power profile is illustrated for different reference voltages based on the demand and generation, DER's number, location, and inverter rating assumptions made in each case study. Table 5-9 and Table 5-10 presents the maximum utilization in the percentage of supplied/absorbed reactive power of each DER from its maximum capacity in the day for the two case studies. It can be noted that overall, as V_{ref} increases, the reactive power support from the DERs increase. It is essential to highlight that each DER's utilization is affected by other factors, like their location in the network and load demand and generation assumptions.

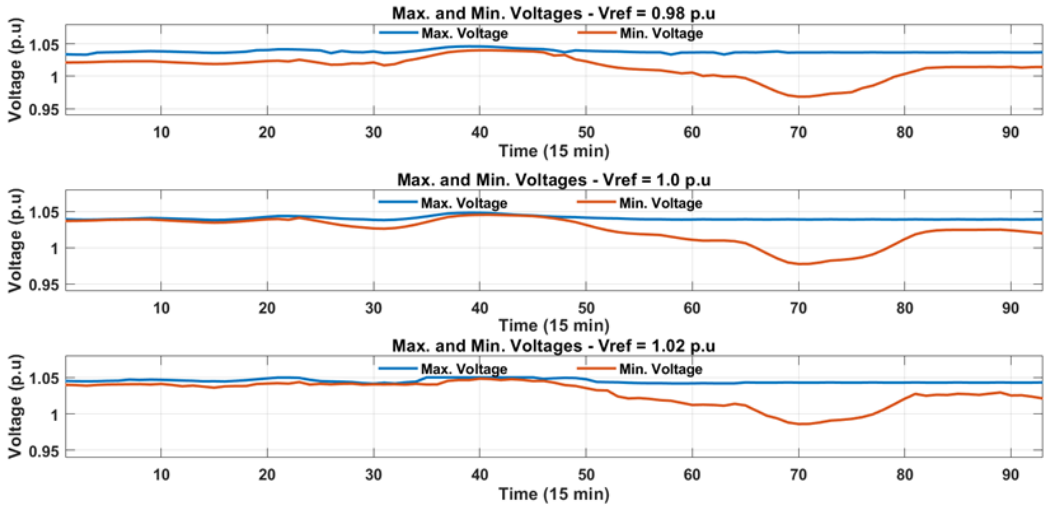


Figure 5-14. Case study A maximum and minimum network voltages after stage II

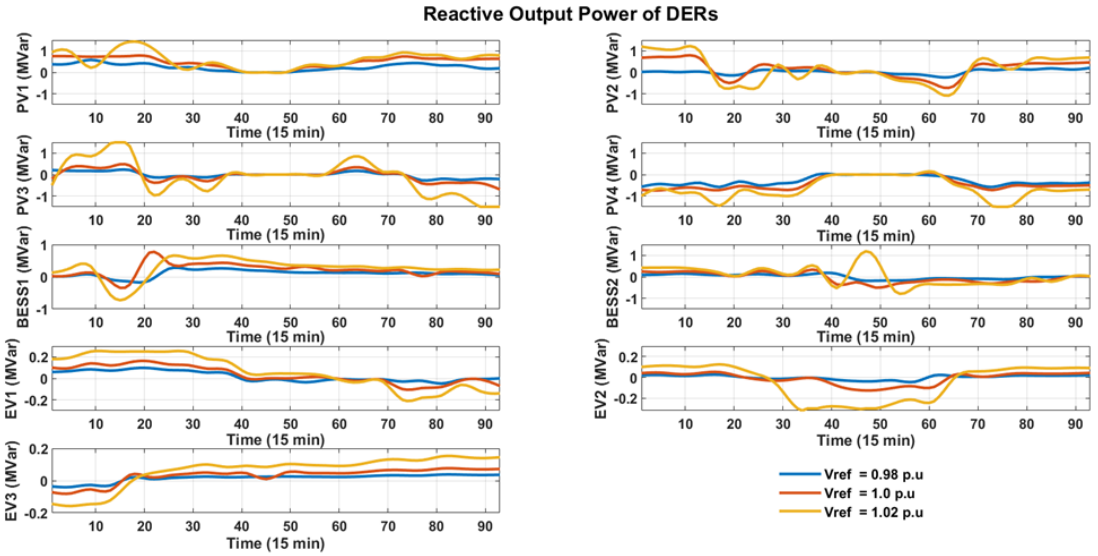


Figure 5-15. Case study A reactive output power of DERs of stage II

Table 5-9. Case Study A Maximum Utilization of DERs in the Day In % - Based on Information in Table 5-1 and Figures 5-1 and 5-2.

Vref	0.8 p.u	1.0 p.u	1.02 p.u
PV1	38.5	50.4	94.6
PV2	13.8	49.3	80.0
PV3	17.7	46.0	100.0
PV4	38.5	48.6	98.6
BESS1	25.9	64.8	75.0
BESS2	12.0	33.3	59.3
EV1	32.0	53.3	83.3

EV2	14.5	41.3	100.0
EV3	19.7	39.4	78.9

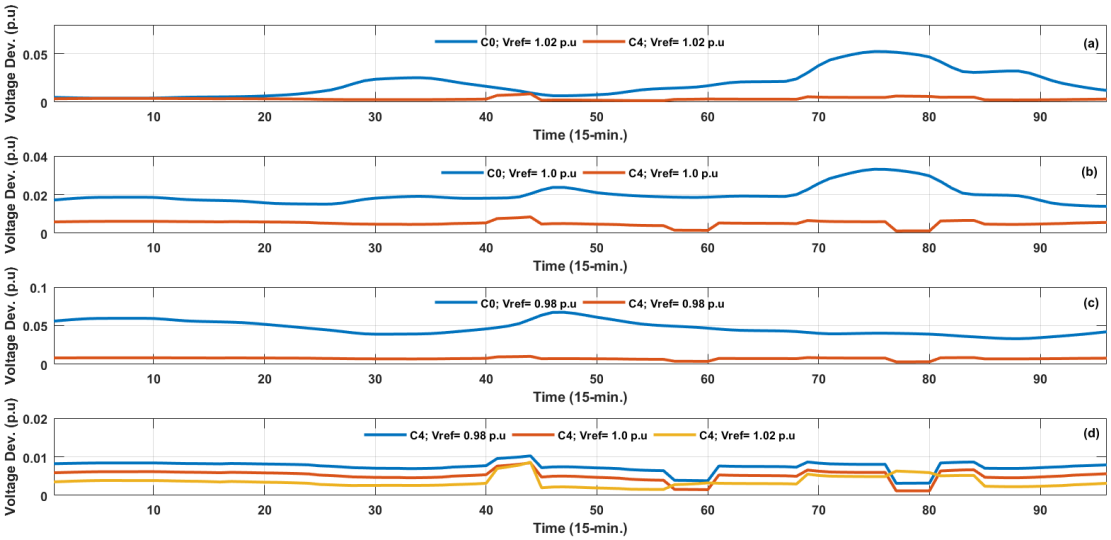


Figure 5-16. Case study B voltage deviation profiles of network before optimization and after stage II considering a) Vref= 1.02 p.u; b) Vref= 1.0 p.u; c) Vref= 0.98 p.u; and d) voltage deviation curves of the three considered Vrefs.

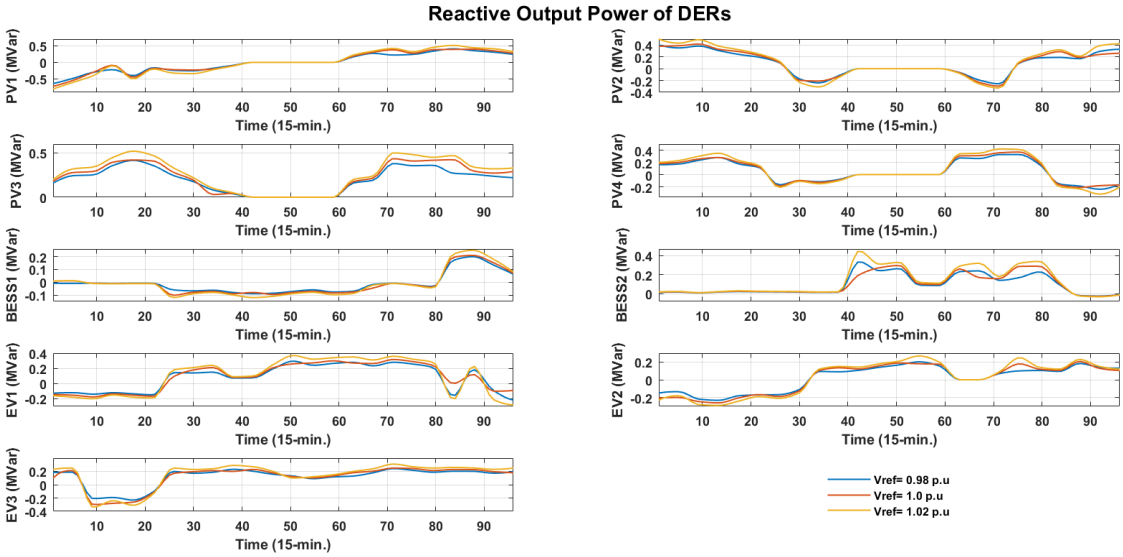


Figure 5-17. Case study B reactive output power of DERs of stage II

Table 5-10. Case Study B Maximum Utilization of DERs in the Day In % - Based on Information in Table 5-2, Figure 5-4 and Figure 5-5.

Vref	0.8 p.u	1.0 p.u	1.02 p.u
PV1	42.7	48	53.3
PV2	26.7	27.8	33.3
PV3	27.7	29	34.6
PV4	21.9	24.6	28.5
BESS1	20.1	21.4	25.2
BESS2	22.5	19.8	30
EV1	59.2	63.4	74.8
EV2	46.4	52.2	58
EV3	48.4	59.8	68.1

The system operator engagement in the VVO algorithm can also contribute to defining the potential buses for DERs investment. Figure 5-18 and Figure 5-19 show the voltage profile of three system buses with different reactive power resources for case studies A and B, respectively. Bus 18 is the farthest bus from the substation and has no var source. The base case minimum voltage is violated during the maximum load condition. However, the interaction with CBs, OLTC, and DERs brought the minimum voltage close to 1.0 p.u. Bus 22 is connected to a PV, and its voltage profile shows a nearly steady value at 1.04 p.u. Bus 29 is connected to a CB and shows a voltage variation within the standard boundaries.

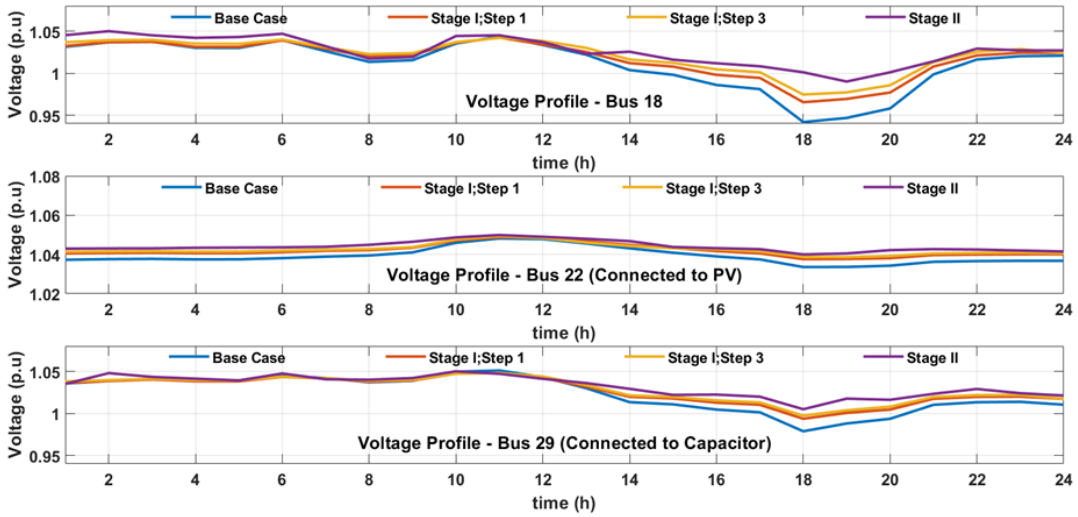


Figure 5-18. Case study A voltage profiles of buses 18, 22, and 29 after different stages

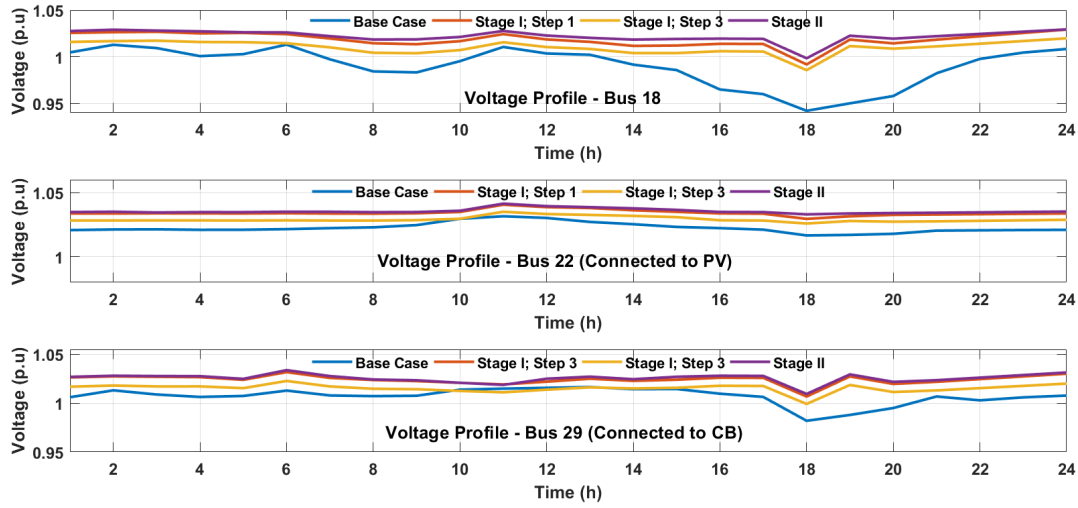


Figure 5-19. Case study B voltage profiles of buses 18, 22, and 29 after different stages

5.5 Evaluation and Validation

Considering case study A, to evaluate the reduction in losses and voltage limitation of stage I, two parameters are introduced; E_{PL} and E_{VL}[201]. Where, E_{PL} and E_{VL} represents the evaluation index of reduction in power loss and voltage limitation, obtained using (5-1) and (5-2), respectively.

$$E_{PL} = \frac{\Delta P>0}{24} \tag{5-1}$$

$$E_{VL} = \frac{Violation_{before}}{Satisfied_{After}} \quad (5\text{-}2)$$

Where $\Delta P > 0$ is the number of hours where the difference of power loss in C0 and C4 is greater than 0, $Violation_{before}$ is the number of hours in C0 where voltage violation existed, and $Satisfied_{After}$ is the number of hours in C4 where the voltage limit is satisfied. Table 5-11 and Table 5-12 illustrates the obtained performance indices based on the tested data. Table 5-11 shows that the proposed algorithm successfully reduced the power loss at all hours of the day with a maximum amount of 0.1954 MW. Moreover, from Table 5-12, it can be deduced that stage I of the proposed algorithm limited all bus voltages in the network to be within the permissible range.

Table 5-11. Performance of Reduction In Power Loss -Stage I

	$\boldsymbol{Avg.\Delta P}$	$\boldsymbol{Max.\Delta P}$	$\boldsymbol{\Delta P > 0}$
Power loss (MW)	0.0402	0.1954	100%

Table 5-12. Performance of Voltage Limitation – Stage I

	$\boldsymbol{Violation_{before}}$	$\boldsymbol{Satisfied_{After}}$	$\boldsymbol{E_{VL}}$
No. of hours	2	24	8.33%

Besides, to evaluate the minimization of voltage deviation of stage II based on 1.0 p.u reference voltage, $E_{VD,min}$ and $E_{VD,max}$ are introduced, representing the evaluation index of voltage deviation for the minimum and maximum voltage and obtained using (5-3) and (5-4), respectively.

$$E_{VD,min} = \frac{max.\Delta V_{min}}{V_{min,C2}} \quad (5\text{-}3)$$

$$E_{VD,max} = \frac{max.\Delta V_{max}}{V_{max,C2}} \quad (5\text{-}4)$$

Where $max.\Delta V_{min}$ is the difference between the minimum voltage in the network of C2 and C4, $V_{min,C2}$ is the minimum voltage in the network of C2, $max.\Delta V_{max}$ is the

difference between the maximum voltage in the network of C2 and C4 and $V_{max,C2}$ is the maximum voltage in the network of C2. Table 5-13 highlights the obtained indices. It can be deduced that the minimum voltages are more affected than the maximum voltages. This could be because, based on data used, there are higher minimum voltage violations and more hours in the day than the maximum voltages.

Table 5-13. Performance of Voltage Deviation - Stage II

	$Avg.\Delta V_{min}$	$max.\Delta V_{min}$	$E_{VD,min}$	$Avg.\Delta V_{max}$	$max.\Delta V_{max}$	$E_{VD,max}$
Voltage deviation (p.u)	0.0158	0.0392	4.2%	0.0063	0.0222	2.1%

The network's real and reactive power mismatches are assessed to validate the proposed algorithm's effectiveness and efficiency. Figure 5-20 shows the mismatch curves and the total generation (from grid and DERs), total demand (residential and EV stations load), and the obtained losses. It can be noted that the real and reactive power mismatch is zero, which proves the accuracy of the algorithm load flow solutions.

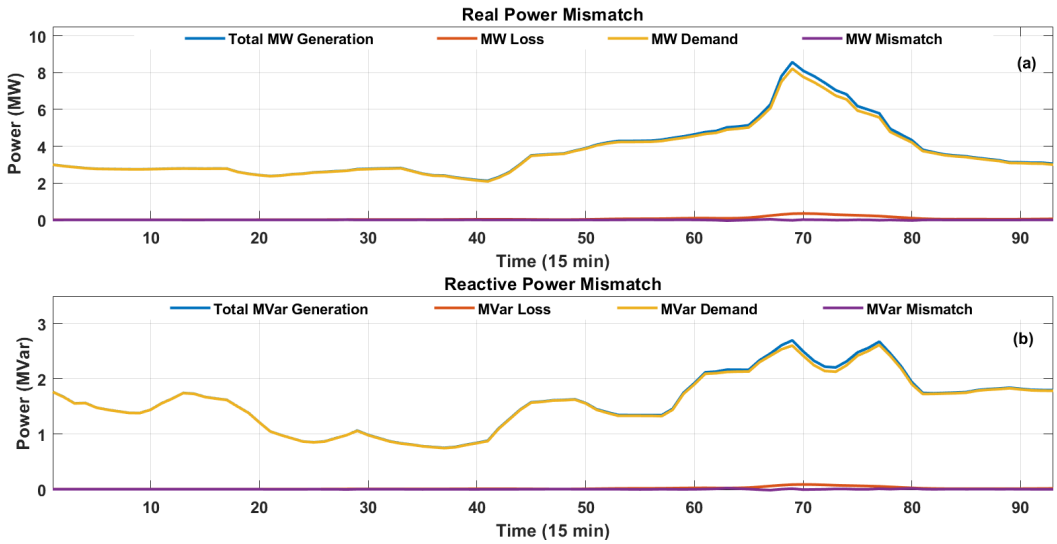


Figure 5-20. Real and reactive power mismatches of the developed algorithm

5.6 Summary

The proposed conventional optimization-based VVO algorithm is examined in two case studies. The algorithm is tested on a modified IEEE 33-bus system, where CBs, OLTCs, and DERs are added at different buses with different sizes in the two case studies. For both case studies, the algorithm is tested on 5 cases; without optimization, with importance given to CBs condition and PVs, with importance given to OLTCs condition and PVs, with importance given to OLTCs condition and PVs & BESSs, with importance given to OLTCs condition and PVs & BESSs & EVs. In stage I, MOGA and IPM approaches are used at step 1 and step 3, respectively, and in stage II, the IPM approach is employed.

The performance of the algorithm is examined in minimizing power loss, the number of control actions, and bus voltage deviations. Stage I results demonstrated the importance of weight factor α and β in considering the conditions of the utility-owned assets, CBs, and OLTCs. Moreover, the three-step methodology showed higher power loss reduction, by approximately 33% and 35% in case studies A and B, respectively, compared to if only the multi-objective step is considered (step 1 of stage I). Further, it is illustrated through cases 2 to 4 that the utilization of all existing DERs in the distribution network achieves higher power loss minimization, where a further 3% and 5% of power loss reduction is attained in case studies A and B, respectively. Also, stage I formulation restricted the bus voltages to be within the permissible range. In addition, stage II results showed a reduction in bus voltage deviations through scheduling the reactive output powers of DERs, where the difference between the maximum and minimum voltages in the network is observed to be minimized. The two examined case studies showed that overall, increasing the reference voltage demands a higher

contribution from the DERs. Additionally, statistical analyses are performed to evaluate and validate the proposed VVO algorithm. It is validated that the proposed algorithm successfully minimized power loss in lines and the number of utility-owned assets' control actions, simultaneously at stage I and bus voltage deviations at stage II.

5.7 Limitations

Even though the proposed conventional optimization-based VVO scheme demonstrated good performance and successful minimization of power loss and voltage deviation, several weakness points were realized in the optimization methodology. First, the optimization performed is day-ahead, which is based on forecasted data. Hence, not suitable for real-world applications. Therefore, a real-time optimization technique is required to achieve more reliable and accurate results. Moreover, DERs are stochastic in nature. To achieve more accurate solutions, their uncertainty and intermittent aspects have to be considered in solving the VVO problem. Thus, it is crucial for the VVO scheme to be capable of managing uncertainties of DERs' intermittent outputs, as well as active and reactive demands.

Further, developing an optimization technique that is adaptive and can handle changes in distribution network parameters due to aging and other factors like ambient temperature is vital. Also, the optimization is model-based. Thus, the performance of the VVO scheme is highly dependent on the model. In other words, if the model is not precise, the optimization results will not be realistic. For instance, the model dynamics have to be represented with high precision by knowing all the environment information, which limits its application on large-scale and real-life systems. This calls for a more efficient and intelligent solution, such as machine learning approaches, especially reinforcement learning (RL) [202]- [204].

PART II: REINFORCEMENT LEARNING-BASED VOLT/VAR OPTIMIZATION SCHEME

CHAPTER 6 : REINFORCEMENT LEARNING

As has been previously highlighted, to overcome the shortcomings of the developed conventional optimization-based VVO approach, a DRL-based VVO scheme is proposed to more efficiently mitigate the challenges introduced in ADNs. Firstly, this chapter starts with describing the RL framework and its preliminaries. It then introduces DRL and provides a brief description of its fundamentals.

6.1 Reinforcement Learning Framework

Reinforcement Learning [90] is an area of machine learning motivated by behaviorist psychology, which examines artificial agents' performance decision-making in an environment to reach a specific objective. The decision-making element in an RL problem is the agent, and everything the agent interacts with is called the environment. Specifically, the agent has to take actions sequentially to control a dynamic system, the environment. The environment is described by its dynamics, states, and function that define the state's evolution based on the agent's actions. The RL agent-environment interaction is done following discrete time steps, i.e., i.e., $\boldsymbol{t} = \boldsymbol{0}, \boldsymbol{1}, \boldsymbol{2}, \ldots$. At each timestep, the agent receives the current environment state $\boldsymbol{s_t} \in \boldsymbol{S}$ where $\boldsymbol{S}$ is the set of states and based on it selects an action $\boldsymbol{a_t} \in \boldsymbol{\mathcal{A}}$ where $\boldsymbol{\mathcal{A}}$ is the set of actions. After executing the action chosen by the agent, the environment moves to a new state $\boldsymbol{s_{t+1}}$ and the agent receives a scalar value, termed reward $\boldsymbol{r_{t+1}}$ that evaluates the correctness of its chosen action and consequently how far it is from the objective. Figure 6-1 displays the framework of RL and agent-environment interaction.

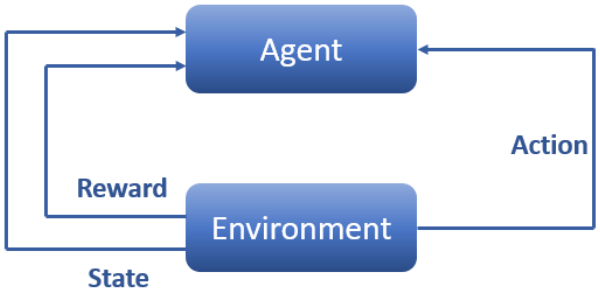


Figure 6-1. Reinforcement learning framework [90]

To achieve the defined objective, the agent has to learn a strategy to select the actions based on its interaction with the environment. This is known as a policy. Also, the states of a system define the required information that aids in predicting the evolution to the next state of the environment given an executed action. Thus, an RL agent's goal is to learn a policy and, based on it, select the proper actions, given an observed current state of the environment, in such a way that over time the expected sum of obtained rewards is maximized. The associated states, actions, and rewards differ from an RL problem to another; however, the framework remains the same. The continuous agent-environment interaction process makes decision-making behavior a step-by-step process. Most reinforcement learning tasks can be decomposed into sequences between initial and terminal states of agent-environment interactions, where a sub-sequence is called an episode. Upon reaching a terminal state, the environment is reset to the initial state.

6.2 Markov Decision Process

A preliminary to better understand RL's concept is first to comprehend its mathematical foundation, Markov Decision Process (MDP). An RL problem is formulated as an MDP. MDPs are defined by a 5-tuple $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{R}}, \boldsymbol{\gamma})$, where $\boldsymbol{\mathcal{S}}$ is the set of states, $\forall \boldsymbol{s} \in \boldsymbol{\mathcal{S}}$; $\boldsymbol{\mathcal{A}}$ is the set of actions, $\forall \boldsymbol{a} \in \boldsymbol{\mathcal{A}}$; $\boldsymbol{\mathcal{P}} : \boldsymbol{\mathcal{S}} \times \boldsymbol{\mathcal{A}} \times \boldsymbol{\mathcal{S}} \rightarrow [\mathbf{0}, \mathbf{1}]$ is the Markovian transition model which represents the probability that an action $\boldsymbol{a}$ in state $\boldsymbol{s}$

at time $\boldsymbol{t}$ is chosen will transition the system at time $\boldsymbol{t+1}$ to state $\boldsymbol{s'}$, that is $\boldsymbol{p_a(s, s') = p(s_{t+1} = s' \mid s_t = s, a_t = a)}$; $\boldsymbol{\mathcal{R} : \mathcal{S} \text{ x } \mathcal{A} \text{ x } \mathcal{S} \rightarrow \mathbb{R}}$ is the reward function, where $\boldsymbol{\mathcal{R}_a(s, s')}$ represents the agent's immediate reward after executing an action that transitioned the environment states from $\boldsymbol{s}$ to $\boldsymbol{s'}$; $\boldsymbol{\gamma}$ is the discount factor, where $\boldsymbol{\gamma}$=[$\boldsymbol{0}$,$\boldsymbol{1}$) denotes the balance between the future and current rewards. An essential property of MDPs is the Markov property [91], which states that the state transitions depend only on the recent action and state of the system and are independent of the prior environment actions and states, that is $\boldsymbol{p(s_{t+1} = s', r_{t+1} = r \mid s_t, a_t)}$ for all $\boldsymbol{s', r, s_t}$ and $\boldsymbol{a_t}$.

Let $\boldsymbol{\bar{\mathcal{R}} : \mathcal{S} \text{ x } \mathcal{A} \rightarrow \mathbb{R}}$ indicate the expected reward of a state-action pair $\boldsymbol{(s, a)}$, accordingly: $\boldsymbol{\bar{\mathcal{R}}(s, a) = \mathbb{E}[\mathcal{R} \mid s, a] = \sum_{s' \in \mathcal{S}} \mathcal{R}(s, a, s') p(s' \mid s, a)}$, where $\mathbb{E}[\cdot]$ represents the expectation operator. Then, the total discounted reward from time instant $\boldsymbol{t}$ and forward, known as the return, becomes $\boldsymbol{G_t = \sum_{t'=t}^{\infty} \gamma^{t'-t} R_{t'}}$.

A deterministic policy $\boldsymbol{\pi}$ is defined as the mapping from $\boldsymbol{\mathcal{S}}$ to $\boldsymbol{\mathcal{A}}$; $\boldsymbol{a = \pi(s), \forall s \in \mathcal{S}, \forall a \in \mathcal{A}}$. Then, the action-value function, which indicates the expected return when taking action $\boldsymbol{a}$ in state $\boldsymbol{s}$ following the policy $\boldsymbol{\pi}$, can be defined as $\boldsymbol{Q^{\pi}(s, a) = \mathbb{E}[G_t \mid \mathcal{S}_t = s, \mathcal{A}_t = a; \pi]}$. The action-value function indicates, for a given policy $\boldsymbol{\pi}$, the effectiveness of the state-action pair $\boldsymbol{(s, a)}$. The optimal action-value function then becomes the maximum action-value function overall policies; $\boldsymbol{Q^*(s, a) = max_{\pi} Q^{\pi}(s, a)}$. Let $\boldsymbol{Q^*(\cdot,\cdot)}$ express the optimal action-value function, the optimal policy is $\boldsymbol{\pi^* = arg\, max_a Q^*(s, a)}$. Subsequently, $\boldsymbol{Q^*(s, a)}$ satisfy the Bellman optimality equation in (6-1) [205].

$$\boldsymbol{Q^*(s, a) = \bar{\mathcal{R}}(s, a) + \gamma \sum_{s' \in \mathcal{S}} p(s' \mid s, a)\, max_{a' \in \mathcal{A}} Q^*(s', a')} \tag{6-1}$$

Hence, if (6-1) is solved, the optimal policy $\boldsymbol{\pi}^*$ can be easily obtained. When both states and actions are discrete, the action-value function can be depicted in the form of a table encompassing all possible pairs of $(\boldsymbol{s}, \boldsymbol{a}) \in \boldsymbol{\mathcal{S}}$ x $\boldsymbol{\mathcal{A}}$. In known transition probabilities $\boldsymbol{\mathcal{P}}$, the MDP can be solved using value iteration and policy iteration algorithms [205]. However, in the real world, the transition probabilities $\boldsymbol{\mathcal{P}}$ and the rewards $\boldsymbol{\mathcal{R}}$ are unknown. Therefore, the MDP can be solved using RL algorithms. In other words, RL represents an extension and generalization over MDPs.

6.3 Deep Reinforcement Learning

DRL combines the perceived role of deep learning with the decision-making feature of RL. It is an artificial intelligence (AI) approach closer to human thinking and is regarded as real AI. Figure 6-2 illustrates the framework of DRL. Deep learning gets the target observation information from the environment and provides the state information in the current environment. The reinforcement learning then maps the current state to the corresponding action and evaluates values based on the expected return [87], [91].

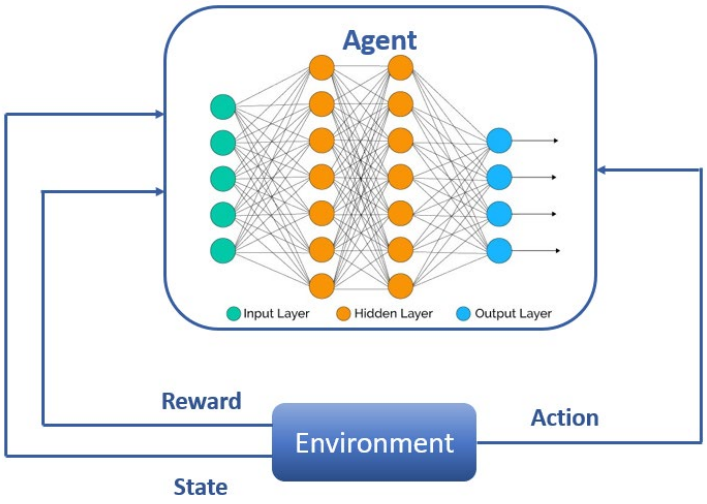


Figure 6-2. Deep reinforcement learning framework [91]

The journey from RL to DRL has gone through a long development process. In classical tabular RL, e.g., *Q*-learning approach, for small-scale state and action spaces, the representation of approximate value functions as arrays or tables is efficient. In this case, the methods can often find the exact optimal value functions and the optimal policies [54]. However, conventional RL approaches suffer from a difficult design issue when they come to real-world implementations. In problems comprising large-scale continuous state and/or action spaces, traditional RL algorithms suffer from 'curse of dimensionality'. Using deep neural networks (DNNs) compact low-dimensional depictions of high-dimensional inputs can be achieved to overcome the curse of dimensionality [80], where the approximate value functions are represented as a parameterized functional form with a weight vector instead of a table. Based on the capability to learn levels of abstractions from data, DRL can handle complex tasks with less previous knowledge [54], [60].

6.4 Reinforcement Learning Taxonomy

There are several ways of categorizing RL algorithms. RL algorithms have two major classes: whether it relies on an environment model during training; model-free or model-based. Model-based RL approaches first learn the characteristics of the environment, $\boldsymbol{p}(\boldsymbol{s_{t+1}} \mid \boldsymbol{s_t}, \boldsymbol{a_t})$ and $\boldsymbol{r}(\boldsymbol{s_t}, \boldsymbol{a_t})$ if not known already, then use them to find the optimal policy. The RL agent searches in the state-action space and selects the proper action based on the state of the environment it receives. However, learning the environment's actual model is typically time-consuming, has large computational complexity, and usually a challenging task in some cases. Also, an accurate model of the environment might not lead to better results [87]. On the other hand, a model-free algorithm does not rely on learning the environment's model, and it aims to learn the optimal policy through the agent-environment interactions and maximize the rewards

received. They are less computationally complex compared to model-based approaches. Moreover, in real-world cases, it is often easier to learn policies than a model. There are three types of model-free approaches: value-based, policy-based, and actor-critic.

The value-based approaches, also known as action-value approaches, rely on optimizing a specific action-value function dependent on the environment's states without explicitly representing policies. Optimizing the action-value function in an iterative scheme results in an improved policy that the agent can follow. The value function for state $\boldsymbol{s}$ when following policy $\boldsymbol{\pi}$ is $\boldsymbol{V^{\pi}(s)}$ as defined in (6-2), and the state-action function that indicates the effect of taking a particular action on the environment is $\boldsymbol{Q^{\pi}(s,a)}$ as defined in (6-3). The action is selected based on the evaluation of the state-action pairs ($\boldsymbol{s_t, a_t}$). This approach is known as Q-learning.

$$\boldsymbol{V^{\pi}(s)} = \mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{\boldsymbol{k=0}}^{\infty} \boldsymbol{\gamma^k \mathcal{R}_{t+k+1}} \middle| \boldsymbol{s_t = s}\right] \tag{6-2}$$

$$\boldsymbol{Q^{\pi}(s,a)} \leftarrow \boldsymbol{(1-\alpha)Q^{\pi}(s,a) + \alpha(\mathcal{R}\,(s,a) + \gamma\, max_{a'} Q^{\pi}(s',a'))} \tag{6-3}$$

Where $\boldsymbol{k}$ is the number of steps in an episode, and $\boldsymbol{\alpha}$ is the learning rate, which controls the agent's adeptness in the dynamic environment.

In contrast to value-based approaches, policy-based approaches explicitly represent a policy by its weights, independent of a value function. The main focus of this approach is to find the best parameters $\boldsymbol{\theta}$ of the policy, as defined in (6-4).

$$\boldsymbol{\pi_{\theta}(s,a)} = \mathbb{p}[\boldsymbol{a|s,\theta}] \tag{6-4}$$

This approach has the advantage of dealing with high-dimensional or continuous action and/or state spaces. They also show the ability to learn stochastic policies. However, they exhibit tendencies to converge to local optima. Also, assessing policies can be inefficient and has higher variances [202].

The third model-free approach is actor-critic. This approach maintains two sets

of parameters distinctly, separating the policy (named actor) and the action-value function (named critic). The actor's role is to update and maintain an action selection policy, while the critic's role is to estimate value functions associated with the actor's policy. It calculates the expected reward of being at state s and following policy π. In other words, the critic evaluates the effectiveness of policy π_θ for selected parameters θ. This approach has the advantage of requiring minimal computational effort when selecting actions. This is since the policy is separated. Thus, there is no need to search to select an action at each step of the episode. Figure 6-3 shows the taxonomy of the popular RL algorithms.

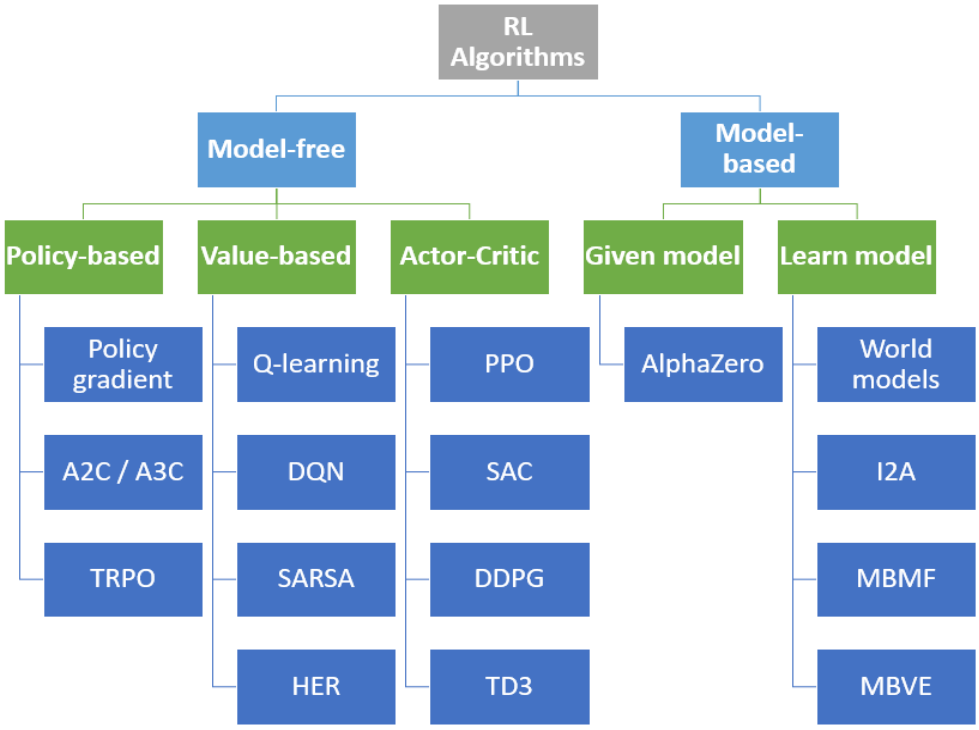


Figure 6-3. Reinforcement learning taxonomy [54]

Another feature of an RL algorithm is being on-policy or off-policy. An on-policy approach requires the policy π to be learned from experience sampled from it. This means that the target policy and the policy used for action selection are the same.

In this case, the exploration needs to be built into the policy, determining policy improvement speed. The experience of on-policy approaches is in the form of (s, a, r, s', a'). In contrast, the off-policy approaches attempt to learn the target policy from a behavior policy that the agent follows for action selection. In other words, it works with 2 different policies; the behavior policy is different from the target policy. This is considered an advantage since it can utilize one policy for learning and the other for exploration. Thus, the agent can improve the target policy while exploring. The experience of off-policy approaches is in the form of (s, a, r, s').

6.5 Summary

To overcome raised concerns in ADNs more efficiently, RL approaches have been proposed. RL is a machine learning approach that learns from scratch based on agent-environment interactions. The main RL elements include environment states, which describe the status of the environment, the action selected by the agent, and the reward, which can be considered an evaluation of the action chosen. In applications that involve high dimensional continuous state and/or action spaces, the traditional RL algorithms suffer from the "curse of dimensionality", which makes them inefficient in practical cases. Subsequently, DRL approaches have emerged to overcome this issue. DRL combines artificial neural networks with the RL concept, where the agent employs NNs as value or policy function approximators. This increases the efficiency and applicability of the algorithm in real cases. Hence, the only difference between Q-learning and DQN is the agent's brain.

The first step towards solving the DRL-based VVO problem is formulating it as an MDP by defining the main elements that include state and action spaces, as well as, formulating the reward function. The aforementioned points are presented in the next chapter.

CHAPTER 7 : THE DRL-BASED VOLT/VAR OPTIMIZATION SCHEME

Firstly, in this chapter, the two-timescale VVO problem is formulated as an MDP to be solved using an RL algorithm. The principles of the proposed RL approach are depicted, and its workflow is described. Also, this chapter provides a description of the tested network and illustrates the learning setup of the proposed TD3 algorithm. It then demonstrates the learning performance of the TD3 agent. Finally, it presents the comparison between the conventional optimization and TD3 based VVO approaches in minimizing power loss and bus voltage deviations for ADNs.

7.1 VVO Problem Formulation as MDP

For simplicity, the DRL-based VVO algorithm is designed to minimize the active power loss in lines at the slow scale and minimize the bus voltage deviation in the faster scale while satisfying certain constraints based on the utilized VVO resources. The considered VVO resources are CBs, OLTCs, and DERs, like PVs, BESSs, and EV charging stations. A DRL-based VVO formulation based on a two-timescale is proposed. Based on the fact that the time response of utility-owned assets is larger than DERs [9], the CBs and the OLTCs are utilized at a slow timescale, τ, and the DERs are coordinated at a faster timescale, t. That way, the utility-owned assets will handle the slow demand and generation variations, and the DERs will manage the fast demand and generation variations, such as DERs' intermittent outputs. Figure 7-1 illustrates the two-timescale segregating the control of the different VVO resources to achieve the defined objectives.

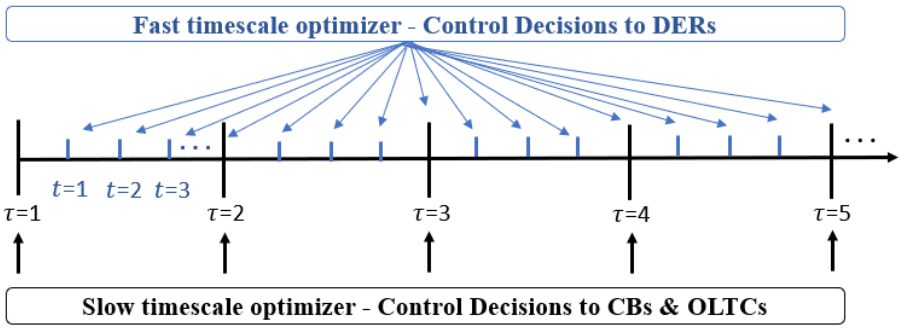


Figure 7-1. Two-timescale segregating for controlling the different VVO resources

Accordingly, at the slow timescale, the CBs and OLTCs are utilized to minimize the power loss in lines as formulated in (4-1) while satisfying grid constraints (4-3)-(4-8), CB constraints (4-9), (4-10), as well as, the OLTC constraints in (4-11), (4-12). Whereas the fast timescale aims to minimize bus voltage deviation, as formulated in (4-21), by coordinating the reactive output power of DERs, while satisfying grid constraints in (4-3)-(4-8) and DERs constraints in (4-13)- (4-20).

In formulating the VVO DRL-based problem as an MDP, the controller is denoted as the agent that decides on the utilized VVO resources' actions. The agent interacts with the environment, the distribution network, in a sequence of discrete-time steps to maximize the rewards. In the context of the proposed VVO DRL based algorithm, the states, actions, and rewards are defined as follows:

1) State Space

The information required by the agent to make decisions are set as the state of the VVO problem. Accordingly, the state space is defined to be $\boldsymbol{S} = [\boldsymbol{P}\ \ \boldsymbol{Q}\ \ \boldsymbol{PV}\ \ \boldsymbol{B}\ \ \boldsymbol{EV}]$. Where $\boldsymbol{P}$ and $\boldsymbol{Q}$ are the total real and reactive power injections, $\boldsymbol{PV}$ is the active power generation by the PV, $\boldsymbol{B}$ is the active power supply/absorbed by the BESS, and $\boldsymbol{EV}$ is the active power demand by the EV charging stations. Thus, at time step $\boldsymbol{t}$, the state vector is $\boldsymbol{s_t} = [\boldsymbol{p_t}\ \ \boldsymbol{q_t}\ \ \boldsymbol{pv_t}\ \ \boldsymbol{b_t}\ \ \boldsymbol{ev_t}]$, which encompasses continuous variables. Note

that the state vector's size will depend on the number of DERs in the distribution network.

2) Action Space

The utilized VVO resources are utility-owned assets and DERs. Thus, the action space is defined to be $\boldsymbol{\mathcal{A}} = [\boldsymbol{CB\ \ OLTC\ \ QPV\ \ QB\ \ QEV}]$. Where $\boldsymbol{CB}$ and $\boldsymbol{OLTC}$ are the setting positions of the CB and the OLTC, and $\boldsymbol{QPV}$, $\boldsymbol{QB}$, and $\boldsymbol{QEV}$ are the reactive output power via smart inverters of the PV, BESS, and EV charging stations. Thus, at time step $\boldsymbol{t}$, the action vector is $\boldsymbol{a_t} = [\boldsymbol{cb_t\ \ oltc_t\ \ qpv_t\ \ qb_t\ \ qev_t}]$, which encompasses discrete and continuous variables. Note, the action vector's size will depend on the number of VVO resources considered to contribute to achieving the objectives.

3) Reward function

The reward function can be formulated following two broad approaches: shaped reward and sparse reward [206]. The sparse reward approach gives the agent a positive reward if no constraints are violated and zero rewards otherwise. Even though the sparse rewards are simpler in designing, they do not encourage the RL agent to learn and may require more training to converge. Thus, to motivate the RL agent and achieve the two timescales' objective, the shaped reward formulation is followed.

The objective at the slow timescale is to minimize the total power loss in lines, thus, to give the agent a positive reward based on the power loss calculated from the load flow at each timestep, the positive reward, r_p, is as formulated in (7-1). It encourages the agent to minimize power loss since as power loss decreases, the agent will receive higher rewards. Furthermore, to penalize the agent if any of the constraints are violated, a negative value is subtracted from the positive reward. The minimum and maximum voltage constraints are adopted through, r_{vmin}, (7-2) and, r_{vmax}, (7-3), where

the penalty's value depends on how far the minimum and maximum voltages from the permissible limit. Hence, if constraints are less violated, a lower value will be subtracted from the positive value, and the agent will receive higher rewards. Accordingly, the total reward received at the slow timescale τ is given in (7-4).

$$\boldsymbol{r_p = N(1 - p_t)} \tag{7-1}$$

$$\boldsymbol{r_{vmin}} = \begin{cases} \boldsymbol{-M(0.95 - V_{min})} & \boldsymbol{if\ V_{min} < 0.95} \\ \boldsymbol{0} & \boldsymbol{otherwise} \end{cases} \tag{7-2}$$

$$\boldsymbol{r_{vmax}} = \begin{cases} \boldsymbol{-M(V_{max} - 1.05)} & \boldsymbol{if\ V_{max} > 1.05} \\ \boldsymbol{0} & \boldsymbol{otherwise} \end{cases} \tag{7-3}$$

$$\boldsymbol{R_\tau = r_p + r_{vmin} + r_{vmax}} \tag{7-4}$$

Where N and M are constants to scale up the positive and negative terms.

Similarly, the fast timescale's total reward function (7-7) comprises a positive reward that imitates the voltage deviation objective, as formulated in (7-5), and negative terms that mimic each constraint. The additional constraint in the fast timescale is the DERs' reactive output power restriction, r_{qDER}, which is comprehended by (7-6). The r_{qDER} term is employed for each DER in the distribution network.

$$\boldsymbol{r_{VD} = F(1 - VD)} \tag{7-5}$$

$$\boldsymbol{r_{qDER}} = \begin{cases} \boldsymbol{-G[|Q| - \sqrt{S^2 - PV^2}]} & \boldsymbol{if\ |Q| > \sqrt{S^2 - PV^2}} \\ \boldsymbol{0} & \boldsymbol{otherwise} \end{cases} \tag{7-6}$$

$$\boldsymbol{R_t = r_{VD} + r_{vmin} + r_{vmax} + r_{qDER}} \tag{7-7}$$

Where F and G are constants to scale up the positive and negative terms and VD is the summation of the difference of network bus voltages from the reference voltage computed by $(V_t - V_{ref})^2$.

7.2 Twin-Delayed Deep Deterministic Policy Gradient (TD3) RL Algorithm

As previously mentioned, if all transition probabilities are known, $\boldsymbol{Q^*(s, a)}$ can be easily derived, using (6-1), and subsequently the optimal policy $\boldsymbol{\pi^*}$. However, in

practical distribution systems, obtaining those transition probabilities is unfeasible. This calls for approaches that do not need the data of transition probabilities, instead estimates them. RL can support model-free approaches, where the RL agent can learn $\boldsymbol{\pi}^*$ by approximating $\boldsymbol{Q}^*(\boldsymbol{s}, \boldsymbol{a})$ without requiring an accurate model of the environment. *Q*-learning is the widely known and simple model-free RL approach [207]. It stores and updates the *Q*-values using a lookup table. However, *Q*-learning only supports discrete state and action spaces [208]. While, in the context of the VVO problem, both state and action spaces are continuous. Also, discretizing them will introduce a dimensionality issue [209]. This calls for RL model-free approaches that can support continuous action and state spaces, such as deep deterministic policy gradient (DDPG) and twin-delayed deep deterministic policy gradient (TD3).

The TD3 algorithm is an off-policy, model-free, actor-critic RL approach. It is built on the DDPG algorithm, which utilizes deep NNs instead of a table as in Q-learning. The NN acts as a function estimator, resulting in an improved generalization for high dimensional continuous state and action spaces. However, if the utilized RL approach overestimates the *Q*-values, the NN training will be unstable. Also, the estimation errors accumulate over time, leading to suboptimal policy updates or cause divergence. The TD3 algorithm addresses this issue by employing two critic networks that separate action selection updates and Q-values updates [210], which reduces the overestimation bias of the value function. Thus, the TD3 algorithm gives better performance and enhanced training speed than other DRL algorithms [210]. Indeed, TD3 applies several techniques to DDPG to overcome the overestimation issue.

The first technique is the double Q-learning [210], where the TD3 learns two *Q*-functions instead of one. Thus, it utilizes two critic networks and an independent actor-network. It simultaneously updates the two critics $Q_{\emptyset 1}$ and $Q_{\emptyset 2}$ using the recursive

Bellman equation in (7-8).

$$\boldsymbol{Q^{\pi}(s_t, a_t) = r(s_t, a_t) + \gamma\, Q^{\pi}(s_{t+1}, \pi(s_{t+1}))} \tag{7-8}$$

The mean-squared Bellman error function is used, in approximating the optimal *Q*-function, to signify how closely $Q_{\emptyset 1}$ and $Q_{\emptyset 2}$ satisfy the Bellman equation as in (7-9) and (7-10). It considers the smaller *Q*-value to develop the targets in the Bellman error loss functions, hence, reduces the *Q*-values overestimation bias issue.

$$\boldsymbol{L(\theta_i, R) = \mathbb{E}_{(s,a,r,\grave{s})\sim R}\left(Q_{\theta_i}(s,a) - y(r,\grave{s})\right)^2, i = 1,2} \tag{7-9}$$

$$\boldsymbol{y(r,\grave{s}) = r(s,a) + \gamma \min_{i=1,2} Q_{\theta_i}\left(\grave{s}, \pi_{\varphi'}(\grave{s}) + \epsilon\right)} \tag{7-10}$$

Where $Q_{\emptyset 1}$ and $Q_{\emptyset 2}$ are the value functions for two current networks under the same state and action.

The second technique is the target policy smoothing [210], where a small Gaussian noise component, ϵ, is added to the target action. This practice avoids over-fitting on the narrow spikes of *Q*-values. Subsequently, policy optimization is achieved by training the actor-network to select the action that maximizes the expected *Q*-function by $\boldsymbol{max_{\varphi} = \mathbb{E}_{s\sim R}[Q_1(s, \pi_{\varphi}(s))]}$. Using the gradient ascent of the expected return $\boldsymbol{\nabla_{\theta} J(\theta)}$, the actor network's parameters are updated only with respect to actor parameters [206], as provided in (7-11).

$$\boldsymbol{\nabla_{\varphi} J(\varphi) \approx \mathbb{E}_{s\sim R}[\nabla_a Q_{\theta_1}(s,a)|_{a=\pi(s)} \nabla_{\varphi} \pi_{\varphi}(s)]} \tag{7-11}$$

Delayed policy update [210] is the third technique, where the TD3 updates the actor-network less frequently than the critic networks. Hence, the model does not update the policy unless its value functions are updated sufficiently. This allows the network to be more stable before it is used to update the policy network. This practice will result in value estimates with lower variance, which improves NN learning stability and leads to a better policy. Further, the training stability is enhanced by adopting the

soft target update strategy [208], where the target networks' parameters are updated to trail the changes in the critic and actor networks at a slow pace by some portion τ as in (7-12) and (7-13).

$$\boldsymbol{\theta_i}' \leftarrow \boldsymbol{\tau\theta_i} + (\mathbf{1} - \boldsymbol{\tau})\boldsymbol{\theta_i}', \quad \boldsymbol{i} = \mathbf{1}, \mathbf{2} \tag{7-12}$$

$$\boldsymbol{\varphi}' \leftarrow \boldsymbol{\sigma\varphi} + (\mathbf{1} - \boldsymbol{\sigma})\boldsymbol{\varphi}' \tag{7-13}$$

7.3 The Workflow of the Proposed TD3 Algorithm

This section describes the workflow of the proposed TD3 algorithm to learn the optimal action generation policy. The algorithm's architecture is given in Figure 7-2, and the pseudocode is provided in Algorithm 1.

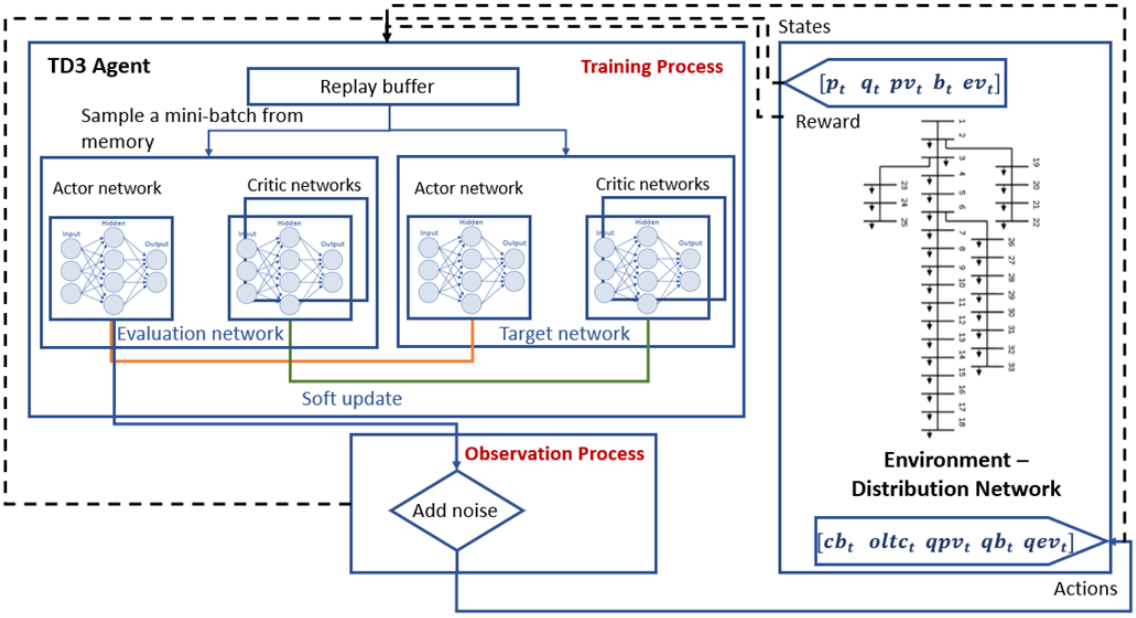


Figure 7-2. Architecture of proposed TD3 algorithm

Algorithm 1 TD3 learning process

Randomly initialize two critic networks Q_{θ_i} and actor-network π_{φ} with random weights $\theta_i (i = 1, 2)$ and φ
Initialize target networks $\theta'_i \rightarrow \theta_i (i = 1, 2)$ and $\varphi' \rightarrow \varphi$
Initialize relay buffer M
For **episode =1, E** do
  Obtain initial state s_t
    For **$t = 1$, R** do
      Generate an action a_t [$a_t = \pi(s_t) + \varepsilon(\varepsilon \sim N(0, \sigma))$]
      Execute action a_t
      Obtain the reward r_t and observe the next state s_{t+1}
      Store transition <s_t, a_t, r_t, s_{t+1}> in M
      Sample mini-batch of N transitions from M
      Compute
        $\tilde{a} = \pi_{\varphi'}(s_{t+1}) + clip\ (N(0, \delta'), -c, c)$.
        $y_t = r_t + \gamma \min(Q_{\theta_1}(s_{t+1}, \tilde{a}), Q_{\theta_1}(s_{t+1}, \tilde{a}))$
      Update $\theta_i (i = 1, 2)$ by minimizing the loss:
        $Loss(\theta_i) = \frac{1}{N}\sum_{t=1}^{N}\left(y_t - Q_{\theta_i}(s_t, a_t)\right)^2, (i = 1, 2)$
      If **t mod k** then
      Update θ_π using the sampled gradient:
      $\nabla_{\varphi} J(\varphi) \approx \frac{1}{N}\sum_{j=1}^{N} \nabla_a Q_{\theta_i}(s_t, a_t)|_{a=\pi_{\varphi}(s_j)} \nabla_{\varphi} \pi_{\varphi}(s_j)$
      Update the target networks:
      $\theta_1' \leftarrow \sigma\theta_1 + (1 - \sigma)\theta_1'$ and $\theta_2' \leftarrow \sigma\theta_2 + (1 - \sigma)\theta_2'$
      $\varphi' \leftarrow \sigma\varphi + (1 - \sigma)\varphi'$
    End if
  End for
End for

The architecture of the algorithm consists of two processes, the observation and the training processes. In the observation process, a replay buffer, M, having a fixed size cache, is first initialized. Transitions of <s_t, a_t, r_t, s_{t+1}> are sampled from the agent-environment interactions, based on the exploration policy, and stored in the replay buffer. Each time slot represents a single step. Hence, in the VVO context, a transition represents VVO results at time slot t based on the timescale. A randomly initialized actor-network with exploration noise is first used to take the actions – settings of CB and OLTC/ reactive output power of DERs-, $a_t = \pi(s_t) + \varepsilon(\varepsilon \sim N(0, \sigma))$. s_t is the TD3 agent's observation of the real and reactive power of the

distribution network at the beginning of time slot $\boldsymbol{t}$, as passed from time slot $\boldsymbol{t-1}$. The TD3 agent uses the $\boldsymbol{a_t}$ to examine the effect of different actions for the time slot $\boldsymbol{t}$. When time slot $\boldsymbol{t}$ is finished, the TD3 agent receives the immediate reward, $\boldsymbol{r_t}$, which resembles the computed active power loss and voltage deviation of this time slot. Meanwhile, the agent realizes the statistics of time slot $\boldsymbol{t}$ for the next state $\boldsymbol{s_{t+1}}$. These steps are repeated until the transitions of one day are collected, stored in the replay buffer.

The training process is comprised of an evaluation network and a target network. The evaluation network includes one actor-network and two critic networks. Where the evaluation actor-network represents the policy model that generates the set points of the CBs, OLTC, and the DERs, and the two evaluation critic networks generate the Q-values given state $\boldsymbol{s_t}$ and action $\boldsymbol{a_t}$. The target network comprises one target actor-network and two target critic networks, which helps update the evaluation critic networks. The network parameters of the actor and critic networks in the evaluation network are updated based on the training and using soft target updates, the parameters of the networks in the target network are copied from the networks of the evaluation networks.

Algorithm 1 illustrates the details of the TD3 algorithm. An $\boldsymbol{N}$ transitions are sampled from the replay buffer to form a mini-batch. Training the NN using mini-batches assures that the sampled transitions are identically and independently distributed, enabling efficient optimization of network parameters. The mini-batch is then fed to the networks for updating their parameters, where $\boldsymbol{s_t}$ is fed to the evaluation actor-network, $\boldsymbol{\pi_\varphi(s_t)}$, to generate the action, where $\boldsymbol{\varphi}$ denotes the parameters of the evaluation actor-network. $\boldsymbol{\pi_\varphi(s_t)}$ is input to the evaluation critic networks along with $\boldsymbol{a_t}$ to compute the Q-values, $\boldsymbol{Q_{\theta_1}(s_t, \pi_\varphi(s_t))}$, and $\boldsymbol{Q_{\theta_2}(s_t, \pi_\varphi(s_t))}$, respectively.

Where $\boldsymbol{\theta_1}$, and $\boldsymbol{\theta_2}$ represents the parameters of the evaluation critic networks, respectively. Using the deterministic policy gradient, as shown in line 19 of Algorithm 1, the evaluation actor-network parameters are updated. The evaluation critic networks' parameters are updated by minimizing the network's loss functions based on the TD error, as shown in line 16 in Algorithm 1. The two target critic networks use the $\boldsymbol{r_t}$ of the transition sample and the action $\widetilde{\boldsymbol{a}}$ from the target actor-network to compute the Q-values of both critic networks, $\boldsymbol{Q_{\theta_1}(s_{t+1}, \widetilde{a})}$ and $\boldsymbol{Q_{\theta_2}(s_{t+1}, \widetilde{a})}$, respectively. Subsequently, the target network parameters are copied from the corresponding networks of the evaluation network using a formula provided in lines 21 and 22 in Algorithm 1.

When the network parameters are updated, the TD3 agent reenters the observation process, collects new transition samples in the replay buffer, and repeats the training process's procedure. After training $\boldsymbol{E}$ steps, the optimal action generation policy for controlling the VVO resources is obtained.

7.4 DRL-based VVO results and discussions

7.4.1 Case study description

The proposed two-time scale VVO TD3 based algorithm is tested on a modified IEEE 33-bus system, as illustrated in Figure 7-3. Two CBs are installed at buses 12 and 29, each with a rating of 1.8 MVar and four switching stages. Two PV units are installed at buses 8 and 24 with an inverter rating of 0.6 MVA. A BESS unit and an EV charging station are installed at buses 16 and 21, with an inverter rating of 0.5 MVA, respectively. Also, one OLTC is installed at bus 1 with maximum, T_{max}, and minimum, T_{min}, setting positions of 10 and -10, and V_{tap}=0.005. The considered minimum, V_{min}, and maximum, V_{max}, voltage limits are 0.95 p.u and 1.05 p.u, respectively, and substation voltage, V_s, 1.0 p.u.

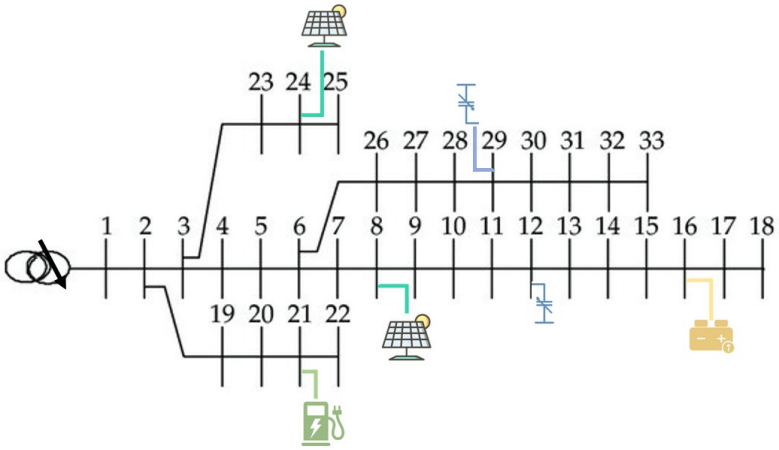


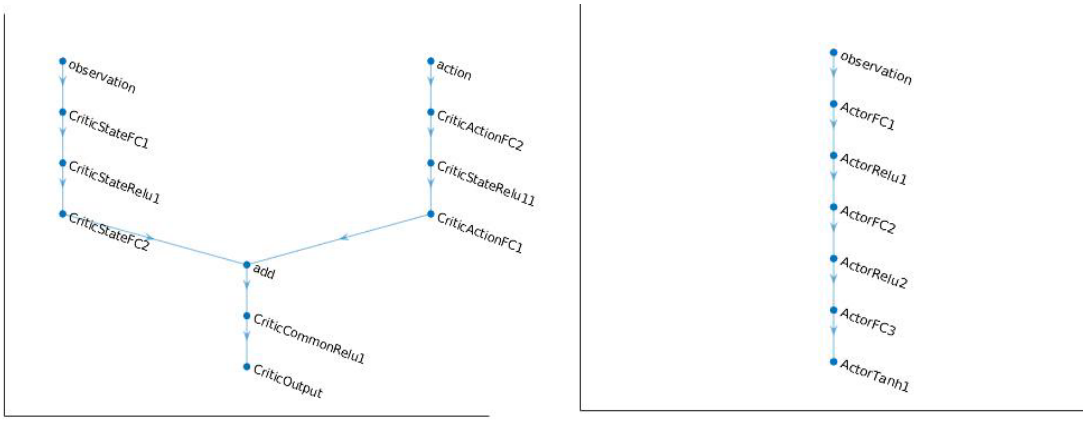
Figure 7-3. Modified IEEE 33-bus system - testing RL

The modified IEEE 33-bus system's real power injections are constructed based on historical data from [211]. The dataset is composed of electricity demand profiles of randomly selected 200 households from the 2009 RECS dataset of the United States' Midwest region. The residential power consumption profiles have been generated with modeling approaches in [212], [213] and validated using metering data of one year with 10-minute time resolution. The reactive power injections are estimated by randomly varying the load power factor (PF) between 0.8 – 0.95 along the day. The aggregated data are scaled up such that the maximum real and reactive demand over the year for the modified IEEE 33-bus system is 6.7 + 2.7 MVA. The selection of these ratings is chosen such that the voltage magnitudes at some time instants fall outside the permissible range. Also, an ideal PV real power generation profiles are assumed, and an adequate charging and discharging profile of the BESS unit is generated. In addition, the profile of the EV charging station of 10-minute time resolution is constructed from 348 EVs associated with the 200 households that compromise charging infrastructure of Level 1 (1920 W) and Level 2 (6600 W) from [211].

One-week data is randomly sampled out to validate and test the proposed TD3 algorithm, and the remaining data are used as the training dataset.

7.4.2 Numerical analysis – Learning setup

The adopted NN for the TD3 agent consists of one actor-network and two critic networks, each having two fully connected hidden layers with 400 and 300 nodes, respectively. For all hidden layers, the rectified linear (ReLU) activation units are employed. For the critic networks, the input layers' size is the state and action vectors' size. Since 4 DERs are utilized – 2 PVs, 1 BESS, and 1 EV charging station- the size of the state vector becomes [6,1], and the utilized VVO resources are 2 CBs, 1 OLTC, 2 PVs, 1 BESS, and 1 EV charging station. Accordingly, the size of the action vector becomes [7,1]. The output of each critic network is the *Q*-value, which is a single linear unit. For the actor-network, the output and input layers' size are the state and action vectors' size, respectively. A hyperbolic tangent (tanh) activation unit is adopted in the actor-network after the output. The created neural network architecture is provided in Figure 7-4. Adam optimizer [214] is employed to optimize the critic and actor networks' parameters with a learning rate of 0.001. The discount factor, minibatch size, experience buffer length, and the target smooth factor are chosen to be 0.99, 256, 100000, 0.005, respectively. The target policy is smoothed by adding Gaussian noise of (0, 0.55) to the target network's chosen action and clipped to (-0.5, 0.5). For the policy's exploration model, an uncorrelated additive Gaussian noise of (0, 0.55) with a 0.1 standard deviation and zero mean is added to the action space. Table 7-1 summarizes the TD3 parameters of the adopted TD3 algorithm.



(a) (b)

Figure 7-4. Neural network architecture a) critic architecture, b) actor architecture

Table 7-1. Setting Parameters of TD3

Parameter	Value
Size of hidden layers	400, 300
Activation function of hidden layers	ReLU
Minibatch size	256
Experience buffer length	100000
Target smooth factor	0.005
Learning rate	0.001
Discount factor	0.99
Maximum and minimum value of exploration	1, 0.1
Epsilon decay rate	0.005
Gradient decay factor	0.99
Squared gradient decay factor	0.999
Exploration model variance decay rate	0.0001
Target policy smooth model variance decay rate	0.0001
Scaling constants of reward functions: N, M, F, G	25, 35, 65, 45

We consider the slow time resolution, τ, to be an hour and the fast timescale, t, to be 10-minute. Thus, the proposed two-timescale DRL-based VVO is solved based on hourly and 10-minute resolutions, solving (7-4) and (7-7), respectively. MATPOWER MATLAB package is utilized to solve the exact load flow at each time step. Thus, the power loss and voltage deviation calculations are based on the Newton-

Raphson method.

7.4.3 Numerical analysis – Learning performance

The learning performance of the proposed TD3 RL algorithm is examined. Figure 7-5 demonstrates the episode reward -blue curve- and the average rewards of successive 40 episodes -red curve- of the learning process. Figure 7-5 shows that the agent was violating constraints for the first few episodes, which resulted in a small reward value. This is due to limited positive learning experiences in the earlier learning phase and that the action policy is not yet optimized. In other words, at the beginning of the training process, the action policy is incapable of satisfying VVO constraints and minimizing power loss and voltage deviation at the same time.

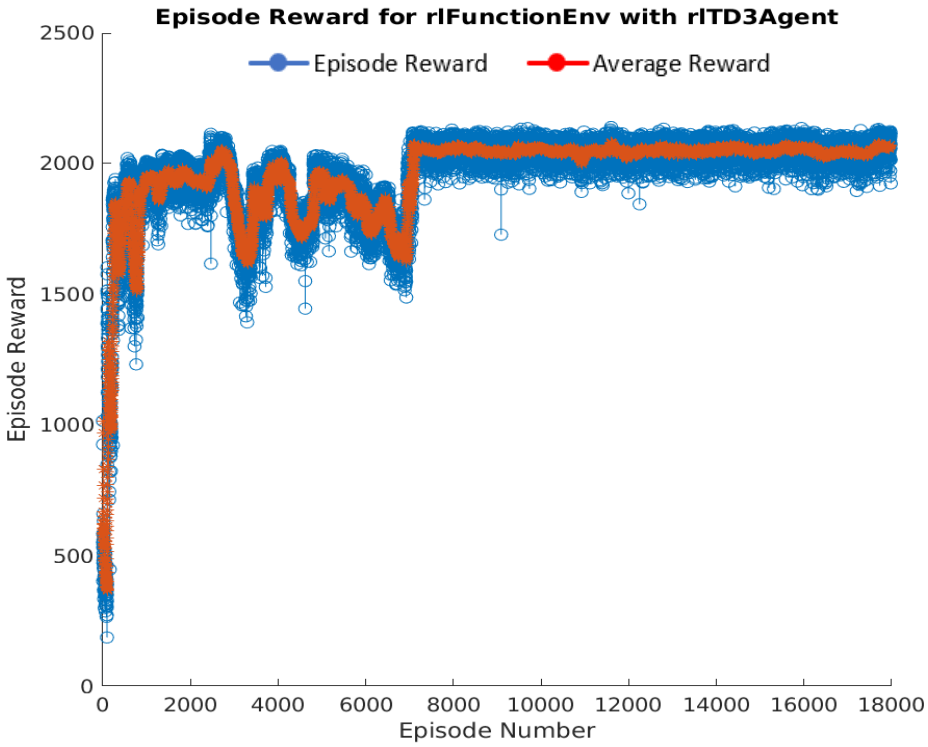


Figure 7-5. Learning process of the proposed TD3 agent

However, during the training process, the TD3 agent gradually obtains higher rewards most of the time. Nevertheless, it shows that the obtained rewards are not

stable, implying that the agent is still violating constraints; however, close to the constraints' limits. Also, the average reward curve shows continuous increments. Approximately at episode 8000, the TD3 agent exhibited a stable performance indicating that it optimized the action policy successfully. Thus, it can be concluded that the proposed TD3 algorithm enabled the distribution network to self-learn, controlling the different VVO resources for the different objectives.

7.4.4 Numerical analysis – VVO performance

The trained TD3 agent is then used to solve the two-timescale VVO problem. To illustrate the proposed VVO TD3 based algorithm's effectiveness, it is compared with one of the traditional optimization algorithms, GA. Both algorithms' performance is analyzed and compared in terms of power loss reduction at the slow timescale and voltage deviation at the fast timescale based on the defined objectives of the VVO problem.

GA is developed to solve the same VVO problem. The considered function tolerance, constraint tolerance, elite count, population size, maximum generations, and maximum stall generations are 1e-8, 1e-8, 50, 500, 20000, and 100, respectively.

Figure 7-6 illustrates the hourly power loss curves of one day using the proposed and GA algorithms. As can be seen, the proposed TD3 algorithm shows better minimization of power loss in the distribution network, which resulted in less total power loss for the day. Table 7-2 illustrates the total power loss and the peak loss of the day of both algorithms.

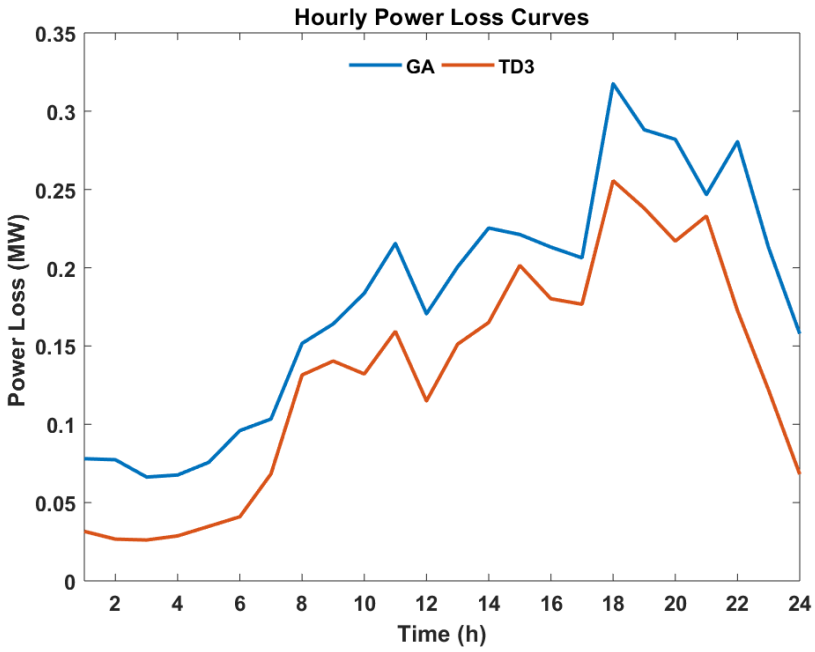


Figure 7-6. Comparison of the proposed VVO TD3 based algorithm and GA – power loss

Figure 7-7 shows the voltage deviation profiles, considering 1.0 p.u as the reference voltage, of the proposed and GA algorithms. Figure 7-7 shows that the proposed TD3 algorithm outperforms the GA and achieves less voltage deviation along the day. Table III demonstrates the maximum voltage deviation of the network buses along the day of both algorithms. Thus, it can be deduced from Figure 7-6, Figure 7-7, and Table 7-2 that the trained TD3 agent demonstrated effective control of the different utilized VVO resources, which resulted in better VVO performance.

Table 7-2. Statistical Analysis of Proposed and GA Algorithms

Algorithm	P_{loss} (MW)	P_{peak} (MW)	VD_{max} (p.u)
GA	0.1792	0.3176	0.0752
Proposed – TD3	0.1298	0.2556	0.0451

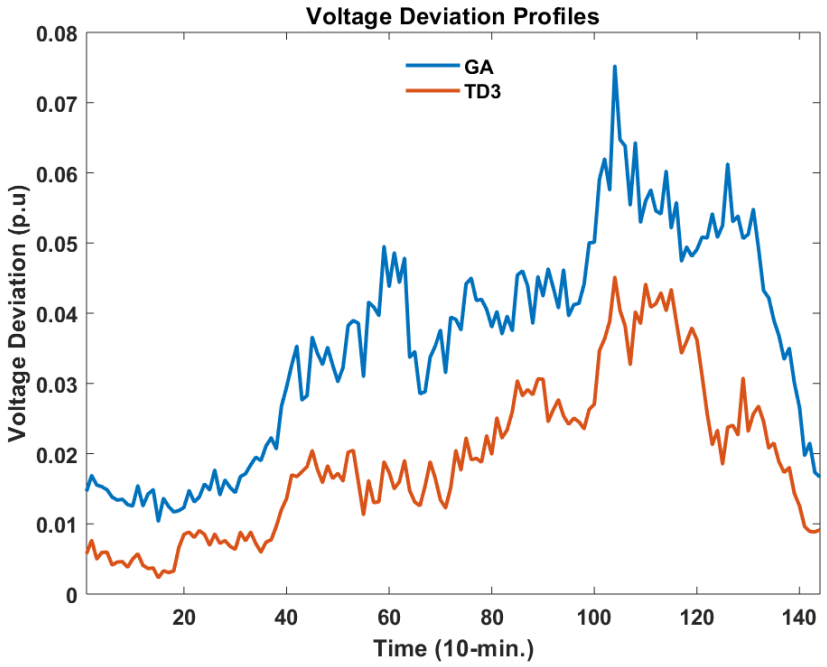


Figure 7-7. Comparison of the proposed VVO TD3 based algorithm and GA – voltage deviation

7.4.5 Numerical analysis – Computation time

Numerical analyses are carried out to examine the proposed two-timescale TD3 algorithm's computational efficiency for controlling the VVO resources. The algorithm training and testing are performed on a GPU of a machine equipped with a 36 GB and Intel(R) i9-9980XE 3.00 GHz CPU and 24576 MB GPU memory. The training phase's executive time is approximately 70 hours, while the execution and testing took 0.2 seconds. This figure is considered to be promising for real-time implementation required in power systems. On the other hand, the computation time of GA is approximately 4 hrs. Furthermore, the proposed TD3 algorithm is competitive at managing problems with high dimensional continuous action and state spaces in active distribution networks.

7.5 Summary

When formulating the RL problem as an MDP, the environment is the

distribution network, and the RL agent learns via interacting with it. The defined state space consists of network variables required by the agent to take actions: demand and generating figures. The action space includes the reactive control assets/resources of the network, including CBs, OLTCs, and DERs. The third element is the reward function, where it consists of positive and negative terms. The positive terms imitate each timescale's target objective function, and the negative terms denote violating constraints. The formulated total reward function for the slow and fast timescales are (7-4) and (7-7), respectively. The formulated MDP is then solved using the proposed TD3 algorithm, an off-policy, model-free, actor-critic approach.

The TD3 agent is then trained on a modified IEEE 33-bus system that includes CBs, OLTCs, and DERs, using one-year real-world data. The proposed TD3 agent's NN architecture consists of one actor-network and two critic networks, each having 2 ReLU hidden layers with 400, 300 nodes, respectively. Based on the utilized VVO resources, the state and action vector sizes are [6,1] and [7,1], respectively. The combination of hyperparameters that best suited the proposed VVO algorithm is defined. From the episode reward curve, it is observed that at the beginning, the agent violated constraints, which lead to low reward values. However, as training progressed, the reward is maximized, indicating better performance in minimizing voltage deviation and power loss. The performance of the TD3-based VVO approach compared with the conventional optimization-based VVO approach demonstrated improved realization of VVO objectives in both timescales.

CHAPTER 8 : THESIS CONCLUSION AND FUTURE WORK

This chapter gives a brief on the work conducted in this thesis regarding the VVO problem formulation and key results and analysis. It also provides probable future recommendation aspects for improving and enhancing VVO algorithms' performance for ADNs with high penetration of DERs.

8.1 Summary

Although the integration of DERs of RESs, BESSs, and EVs aided in network planning, operation, and maintenance, their increasing penetration led to the transitioning of the distribution network from passive to active distribution networks and introduced technical challenges. However, recently, it has been proven that DERs can support the network by supplying/absorbing reactive power via their interface-inverters. Thus, existing DERs in the distribution network can help overcome the challenges in control problems that include VVO. Thus, this thesis has focused on coordinating conventional volt/var devices and inverter-based resources to solve the VVO problem for active distribution networks with high penetration of DERs. The summary of the two proposed approaches in this thesis is as follows:

Part I- A non-linear multi-objective two-timescale VVO algorithm is proposed. The proposed methodology coordinates existing conventional switching-based CBs and OLTCs with inverter-based resources such as PVs, BESSs, and EV charging stations to improve voltage profile and reduce losses. A two-stage algorithm with different time resolutions is implemented. Stage I is formulated as a three-step methodology, where the first step aims to minimize active power loss in transmission lines and the number of adjustments of conventional volt/var assets simultaneously and limit voltage variation within the standard boundary. In this stage, the asset lifetime is considered in step 2 by incorporating the system operator preference for setting switching-based CBs

and OLTCs. Coordinating both standard and inverter-based DERs volt/var devices is achieved in this stage by optimizing the reactive power of inverter-based DERs given a slow time resolution that matches standard volt/var switching-based devices to further minimize active power loss in step 3. The second stage of the multi-objective optimization algorithm is formulated in a higher-time resolution to provide an additional decrease in network bus voltages deviations and system losses. This operation enhancement is achieved by defining the optimal scheduling of the reactive output power in a higher time resolution that matches the inverter-based DERs.

The proposed methodology is developed and evaluated through simulation of various cases on a modified IEEE 33-bus system. The time resolution for the proposed algorithm is selected to match the switching-based and inverter-based devices. One hour and 15-minute time resolutions are applied in this application. Results show that voltage deviation and active power loss are better minimized when optimally coordinating and utilizing all existing standard and DERs resources. The optimizing algorithm reduces system losses by 36% and the peak power by 39%, as well as it limits the voltage deviation within the standard boundary. The system operator experience in setting standard volt/var devices is considered, and thus enhancing the lifetime of distribution assets, CBs, and OLTCs.

Part II- In the second part of the thesis, a model-free two-timescale DRL-based algorithm is proposed to solve the VVO problem for ADNs. The proposed algorithm coordinates different VVO resources at different timescales to take into consideration their response time. The slow timescale -one hour- minimizes active power loss in lines using CBs and OLTCs, while the faster scale -10 min.- aims to minimize bus voltage deviations by scheduling the reactive output power of PVs, BESSs, and EV charging stations. The VVO problem is modeled as MDP and solved using a DRL approach that

supports continuous action spaces. The numerical analysis of the proposed TD3 algorithm is conducted on a modified IEEE 33-bus system. The TD3 agent is trained with real-world data, and its performance is analyzed and compared with a conventional optimization approach. Eventually, the trained TD3 agent can effectively control the different VVO resources when facing other demand and generation circumstances. Results and analyses demonstrate promising results of the proposed DRL approach for VVO.

8.2 Contributions

In literature, researchers focused on solving various consequent aspects of high penetration of DERs in ADNs like over/under voltage violations, sudden over/under generations, and reactive power control and dispatch. Tackling such challenges increased the contribution of conventional volt/var assets, like CBs and OLTCs, which affects their lifetime and conditions. In this thesis, the number of control actions of the conventional volt/var assets is considered one of the objectives to be minimized. Also, the different assets' condition is considered by assigning different weight factors for them where the greatly affected assets can be given higher priority to be less utilized.

Moreover, in this thesis, in addition to the utilization of conventional volt/var assets, different DERs like PVs, BESSs, and EV charging stations are coordinated along with the conventional assets to minimize power loss and voltage deviation. Further, to ensure the achievement of the optimal solution of power loss reduction, the formulation of its stage in the proposed algorithm minimizes power loss twice through a 3-step methodology. The proposed algorithm also considers the utility operator preference range of the network power loss and conventional volt/var assets' control actions. It also considers the time response of the utilized resources, conventional volt/var assets and DERs, by formulating the VVO algorithm to be a two-timescale algorithm where

it utilizes the resources having faster time response at a higher time resolution.

In addition to the conventional optimization-based VVO scheme, a two-timescale VVO algorithm using a DRL approach is proposed. The approach coordinates the utility-owned assets and inverter-interface resources, aiming to minimize active power loss and bus voltage deviations on the slow and fast timescales, respectively. The proposed DRL algorithm, named twin-delayed deep deterministic policy gradient (TD3), supports continuous action spaces. Thus, it can effectively manage the reactive output powers of the DERs. Further, the proposed approach is a model-free RL algorithm that does not depend on a specific optimization model. It is essential to highlight that the DRL algorithm is a real-time approach, unlike the conventional optimization algorithm, where a day-ahead optimization is performed.

8.3 Future Work

Based on the work performed in this thesis, to propose a VVO scheme that further increase the efficiency, accuracy and reliability of ADNs, several future work considerations related to VVO problem formulation and optimization methodology can be recommended as follows:

- VVO problem formulation related:
  - **Additional objectives:** expand the VVO problem to encompass other operational and security objectives and constraints. An objective of conservative voltage reduction (CVR) can be added to enhance network operation further.
  - **Weight factors selection:** in the proposed algorithm, the selection of weight factors of the conventional volt/var assets, alpha and beta, in step 1 of stage I is done manually by the utility operators. Their selection can be done in a smart way by having an algorithm that detects the criticality and conditions of the

assets.

- **Meshed distribution network:** the proposed conventional optimization and DRL-based VVO algorithms in this thesis have been examined on a radial distribution network. As a future research point, the proposed algorithms' performance can be investigated for other distribution network topologies, such as meshed networks.

- Optimization methodology related:
  - **Multi-agent:** in the proposed DRL algorithm, a single agent is adopted for controlling VVO assets/ resources of both the slow and fast timescales. The development of a multi-agent DRL algorithm can enhance the training performance and accelerate training time. One agent can be responsible for coordinating utility-owned assets on the slow timescale, and another agent schedules the reactive output power of DERs on the faster timescale.
  - **Scalability:** the proposed conventional optimization and DRL- based VVO algorithms have been tested only on the IEEE 33-bus system. Further studies could be carried out to investigate their performance and computational time for larger-scale networks.
  - **Computational requirements:** another crucial research line in DRL is the requirements of computing resources. In ADNs, as the number of utility-owned assets and DERs increases, the size of the state and action vectors will exponentially increase. Hence, training time will increase. A topic that could be further studied is the range of computational requirements for high dimensional data that will be adequate for ADNs.
  - **Training acceleration:** another research point could be to investigate possible ways, other than computing resources, to reduce the training time of the DRL

algorithm. For example, the approach of MDP formulation; selection of states, actions, and reward formulation.

REFERENCES

[1] C. L. Masters, “Voltage Rise: The Big Issue When Connecting Embedded Generation to Long 11 kV Overhead Lines,” *Power Eng. J.* vol. 16, no. 1, pp. 5–12, Feb. 2002.

[2] P. Moseley and J. Garche, *Electrochemical Energy Storage for Renewable Sources and Grid Balancing*. Amsterdam: Elsevier, 2015.

[3] T. Helmer and J. Sottnik, “Volt/VAR Control Options and How to Leverage AMI Data,” 03-Sep-2019. [Online]. Available: https://www.power-grid.com/2012/08/01/volt-var-control-options-and-how-to-leverage-ami-data/.

[4] M. Chamana and B. H. Chowdhury, "Optimal voltage regulation of distribution networks with cascaded voltage regulators in the presence of high PV penetration," *in IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1427-1436, July 2018, Available: 10.1109/TSTE.2017.2788869.

[5] A. Mohapatra, P. R. Bijwe and B. K. Panigrahi, "An Efficient Hybrid Approach for Volt/Var Control in Distribution Systems," in *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1780-1788, Aug. 2014, doi: 10.1109/TPWRD.2014.2306845.

[6] P. Bagheri and W. Xu, "Model-free volt-var control based on measurement data analytics", *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1471-1482, 2019.

[7] E. Pourjafari and M. Reformat, "A support vector regression based model predictive control for volt-var optimization of distribution systems," in *IEEE Access*, vol. 7, pp. 93352-93363, 2019.

[8] T. Ding, Q. Yang, Y. Yang, C. Li, Z. Bie, and F. Blaabjerg, "A Data-Driven Stochastic Reactive Power Optimization Considering Uncertainties in Active

Distribution Networks and Decomposition Method," in *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4994-5004, Sept. 2018, doi: 10.1109/TSG.2017.2677481.

[9] D. Jin, H. Chiang, and P. Li, "Two-timescale multi-objective coordinate volt/var optimization for active distribution networks", *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4418-4428, 2019. Available: 10.1109/tpwrs.2019.2914923.

[10] P. S. Georgilakis and N. D. Hatziargyriou, "Optimal Distributed Generation Placement in Power Distribution Networks: Models, Methods, and Future Research," in *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3420-3428, Aug. 2013, doi: 10.1109/TPWRS.2012.2237043.

[11] R. Lin, Y. Liu, Y. Man and J. Ren, "Towards a sustainable distributed energy system in China: decision-making for strategies and policy implications", *Energy, Sustainability and Society*, vol. 9, no. 1, 2019. Available: 10.1186/s13705-019-0237-9.

[12] "Form EIA-861M (formerly EIA-826) detailed data", *Eia.gov*, 2020. [Online]. Available: https://www.eia.gov/electricity/data/eia861m.

[13] "Most utility-scale batteries in the United States are made of lithium-ion" *Eia.gov*, 2020. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=4181 3.

[14] "Sales of plug-in electric vehicles in the U.S. 2018 | Statista", *Statista*, 2020. [Online]. Available: https://www.statista.com/statistics/416672/united-states-projected-plug-inelectric-vehicles-sales/.

[15] Z. Huang, S. Wong and C. K. Tse, "Design of a Single-Stage Inductive-Power-Transfer Converter for Efficient EV Battery Charging," in *IEEE Transactions on*

Vehicular Technology, vol. 66, no. 7, pp. 5808-5821, July 2017, doi: 10.1109/TVT.2016.2631596.

[16] Z. Zhou, "Design and Implementation of Digital Phase Locked Loop for Single-Phase Grid-Tied PV Inverters", *Electric Power Components and Systems*, vol. 46, no. 14-15, pp. 1662-1671, 2018. Available: 10.1080/15325008.2018.1511640.

[17] X. Zhu, H. Wang, W. Zhang, H. Wang, X. Deng and X. Yue, "A novel single-phase five-level transformer-less photovoltaic (PV) inverter," in CES Transactions on Electrical Machines and Systems, vol. 4, no. 4, pp. 329-338, Dec. 2020, doi: 10.30941/CESTEMS.2020.00040.

[18] A. Safavizadeh, G. R. Yousefi and H. R. Karshenas, "Voltage Variation Mitigation Using Reactive Power Management of Distributed Energy Resources in a Smart Distribution System," in *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1907-1915, March 2019, doi: 10.1109/TSG.2017.2781690.

[19] B. Kroposki et al., "Autonomous Energy Grids: Controlling the Future Grid With Large Amounts of Distributed Energy Resources," in *IEEE Power and Energy Magazine*, vol. 18, no. 6, pp. 37-46, Nov.-Dec. 2020, doi: 10.1109/MPE.2020.3014540.

[20] J. PeçasLopesa, J. Mutalec, P. Djapicc, N. Jenkinsc and N. Hatziargyrioub, "Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities", *Electric Power Systems Research*, vol. 77, no. 9, pp. 1189-1203, 2007.

[21] W. El-Khattam and M. A Salama, "Distributed generation technologies, definitions and benefits", *Electric Power Systems Research*, vol. 71, no. 2, pp. 119-128, 2004.

[22] P. Chiradeja and R. Ramakumar, "An approach to quantify the technical benefits

of distributed generation," in *IEEE Transactions on Energy Conversion*, vol. 19, no. 4, pp. 764-773, Dec. 2004, Available: 10.1109/TEC.2004.827704.

[23] M. A. Kashem and G. Ledwich, "Multiple distributed generators for distribution feeder voltage support," in *IEEE Transactions on Energy Conversion*, vol. 20, no. 3, pp. 676–684, Sept 2005.

[24] P. Benalcázar, J. Lara and M. Samper, "Distributed Photovoltaic Generation in Ecuador: Economic Analysis and Incentives Mechanisms," in *IEEE Latin America Transactions*, vol. 18, no. 03, pp. 564-572, March 2020, doi: 10.1109/TLA.2020.9082728.

[25] W. El-Khattam, K. Bhattacharya, Y. Hegazy, and M. M. A. Salama, "Optimal investment planning for distributed generation in a competitive electricity market," *IEEE Transactions on Power Systems,* vol. 19, no. 3, pp. 1674-1684, 2004.

[26] W. H. Kersting, Distribution System Modeling and Analysis. Boca Raton, NY: CRC Press, 2012.

[27] Y. Tan, Y. Cao, Y. Li, K. Y. Lee, L. Jiang and S. Li, "Optimal Day-Ahead Operation Considering Power Quality for Active Distribution Networks," in *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 425-436, April 2017, doi: 10.1109/TASE.2016.2629477.

[28] Y. Du, J. Wu, S. Li, C. Long and S. Onori, "Coordinated Energy Dispatch of Autonomous Microgrids With Distributed MPC Optimization," in *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5289-5298, Sept. 2019, doi: 10.1109/TII.2019.2899885.

[29] S. Wang, Z. Y. Dong, C. Chen, H. Fan and F. Luo, "Expansion Planning of Active Distribution Networks With Multiple Distributed Energy Resources and EV Sharing System," in *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 602-611,

Jan. 2020, doi: 10.1109/TSG.2019.2926572.

[30] Y. Guo, Q. Wu, H. Gao, X. Chen, J. Østergaard and H. Xin, "MPC-Based Coordinated Voltage Regulation for Distribution Networks With Distributed Generation and Energy Storage System," in *IEEE Transactions on Sustainable Energy*, vol. 10, no. 4, pp. 1731-1739, Oct. 2019, doi: 10.1109/TSTE.2018.2869932.

[31] P. Kou, D. Liang, R. Gao, Y. Liu and L. Gao, "Decentralized Model Predictive Control of Hybrid Distribution Transformers for Voltage Regulation in Active Distribution Networks," in *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2189-2200, Oct. 2020, doi: 10.1109/TSTE.2019.2952171.

[32] A. Gabash and P. Li, "Active-Reactive Optimal Power Flow in Distribution Networks With Embedded Generation and Battery Storage," in *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2026-2035, Nov. 2012, doi: 10.1109/TPWRS.2012.2187315.

[33] H. Liu, J. Li, S. Ge, X. He, F. Li and C. Gu, "Distributed Day-Ahead Peer-to-Peer Trading for Multi-Microgrid Systems in Active Distribution Networks," in *IEEE Access*, vol. 8, pp. 66961-66976, 2020, doi: 10.1109/ACCESS.2020.2983645.

[34] A. Ehsan and Q. Yang, "State-of-the-art techniques for modelling of uncertainties in active distribution network planning: A review", *Applied Energy*, vol. 239, pp. 1509-1523, 2019. Available: 10.1016/j.apenergy.2019.01.211.

[35] M. Chamana and B. H. Chowdhury, "Optimal Voltage Regulation of Distribution Networks With Cascaded Voltage Regulators in the Presence of High PV Penetration," in *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1427-1436, July 2018, doi: 10.1109/TSTE.2017.2788869.

[36] R.-Hsun Liang and Y.-S. Wang, "Fuzzy-based reactive power and voltage control

in a distribution system," in *IEEE Transactions on Power Delivery*, vol. 18, no. 2, pp. 610-618, April 2003, Available: 10.1109/TPWRD.2003.809740.

[37] Z. Tang, D. J. Hill and T. Liu, "Fast Distributed Reactive Power Control for Voltage Regulation in Distribution Networks," in *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 802-805, Jan. 2019, doi: 10.1109/TPWRD.2018.2868158.

[38] Q. Li, Y. Zhang, T. Ji, X. Lin and Z. Cai, "Volt/Var Control for Power Grids With Connections of Large-Scale Wind Farms: A Review," in *IEEE Access*, vol. 6, pp. 26675-26692, 2018, doi: 10.1109/ACCESS.2018.2832175.

[39] R. A. Walling, R. Saint, R. C. Dugan, J. Burke and L. A. Kojovic, "Summary of Distributed Resources Impact on Power Delivery Systems," in *IEEE Transactions on Power Delivery*, vol. 23, no. 3, pp. 1636-1644, July 2008, Available: 10.1109/TPWRD.2007.909115.

[40] C. Han et al., "STATCOM Impact Study on the Integration of a Large Wind Farm into a Weak Loop Power System," in *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 226-233, March 2008, doi: 10.1109/TEC.2006.888031.

[41] N. Kumar, I. Hussain, B. Singh and B. K. Panigrahi, "Rapid MPPT for Uniformly and Partial Shaded PV System by Using JayaDE Algorithm in Highly Fluctuating Atmospheric Conditions," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2406-2416, Oct. 2017, doi: 10.1109/TII.2017.2700327.

[42] E. Bianconi et al., "A fast current-based MPPT technique employing sliding mode control", *IEEE Transactions on Industrial Electronics*, vol. 60, no. 3, pp. 1168-1178, 2013. Available: 10.1109/tie.2012.2190253.

[43] J. Liu, Y. Xu, Z. Y. Dong and K. P. Wong, "Retirement-Driven Dynamic VAR Planning for Voltage Stability Enhancement of Power Systems With High-Level

Wind Power," in *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2282-2291, March 2018, doi: 10.1109/TPWRS.2017.2732441.

[44] R. G. Wandhare and V. Agarwal, "Reactive power capacity enhancement of a PV-grid System to increase PV penetration level in smart grid scenario," in *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1845-1854, July 2014, Available: 10.1109/TSG.2014.2298532.

[45] P. Jahangiri and D. C. Aliprantis, "Distributed Volt/VAr Control by PV Inverters," in *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3429-3439, Aug. 2013, doi: 10.1109/TPWRS.2013.2256375.

[46] T. Ding, C. Li, Y. Yang, J. Jiang, Z. Bie and F. Blaabjerg, "A Two-Stage Robust Optimization for Centralized-Optimal Dispatch of Photovoltaic Inverters in Active Distribution Networks," in *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 744-754, April 2017, doi: 10.1109/TSTE.2016.2605926.

[47] K. Turitsyn, P. Sulc, S. Backhaus and M. Chertkov, "Options for control of reactive power by distributed photovoltaic generators," in *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1063-1073, June 2011, Available: 10.1109/JPROC.2011.2116750.

[48] Y. Wang, T. Zhao, C. Ju, Y. Xu and P. Wang, "Two-Level Distributed Volt/Var Control Using Aggregated PV Inverters in Distribution Networks," in *IEEE Transactions on Power Delivery*, vol. 35, no. 4, pp. 1844-1855, Aug. 2020, doi: 10.1109/TPWRD.2019.2955506.

[49] V. B. Pamshetti, S. Singh and S. P. Singh, "Combined Impact of Network Reconfiguration and Volt-VAR Control Devices on Energy Savings in the Presence of Distributed Generation," in *IEEE Systems Journal*, vol. 14, no. 1, pp. 995-1006, March 2020, doi: 10.1109/JSYST.2019.2928139.

[50] The American National Standards Institute ANSI, American national standard for

electric power-systems and equipment voltage ratings (60) hertz, ANSI Std. C84.1, 1995-1996.

[51] U.S. Department of Energy, "Evaluations of conservation voltage reduction (CVR) on a national level", Pacific Northwest National Laboratory (PNNL), 2010.

[52] I. Ramalla, and N. Namburi, "Analytical Review of Loss Reduction Techniques in Indian Power Distribution Sector-Techno Managerial Approach", *International Journal of Research in Engineering and Technology*, vol. 03, no. 24, pp. 30-34, 2014. Available: 10.15623/ijret.2014.0324006.

[53] S. R. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed., Cambridge: MIT Press, 2018.

[54] V. Francois-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, “An introduction to deep reinforcement learning,” Foundations and Trends in Machine Learning, vol. 11, no. 3–4, pp. 219–354, Dec. 2018.

[55] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. J. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. T. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, “Mastering the game of Go without human knowledge, ” Nature, vol. 550, no. 7676, pp. 354–359, Oct. 2017.

[56] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, “Human-level control through deep reinforcement learning, ” Nature, vol. 518, no. 7540, pp. 529–533, Jan. 2015.

[57] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K.

Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of Go with deep neural networks and tree search” Nature, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[58] O. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. L. Weng, and W. Zaremba, “Learning dexterous in-hand manipulation” arXiv: 1808.00177 (2018).

[59] A. A. Rusu, M. Vecer´ık, T. Rothorl, N. Heess, R. Pascanu, and R. ¨ Hadsell, “Sim-to-real robot learning from pixels with progressive nets,” in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017.

[60] W. Y. Wang, J. W. Li, and X. D. He. (2018). Deep reinforcement learning for NLP. [Online]. Available: https://www.aclweb.org/anthology/ P18-5007.pdf.

[61] B. McCann, N. S. Keskar, C. M. Xiong, and R. Socher, “The natural language decathlon: multitask learning as question answering,” arXiv: 1806.08730 (2018).

[62] Y. Deng, F. Bao, Y. Y. Kong, Z. Q. Ren, and Q. H. Dai, “Deep direct reinforcement learning for financial signal representation and trading” in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 653–664, Mar. 2017.

[63] Z. H. Hu, Y. T. Liang, J. Zhang, Z. Li, and Y. Liu, “Inference aided reinforcement learning for incentive mechanism design in crowdsourcing,” in *Advances in Neural Information Processing Systems*, 2018.

[64] W. B. Shi, N. Li, C. C. Chu, and R. Gadh, “Real-time energy management in microgrids,” in *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 228–238, Jan. 2017.

[65] M. Zachar and P. Daoutidis, “Microgrid/macrogrid energy exchange: a novel

market structure and stochastic scheduling," in *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 178–189, Jan. 2017.

[66] C. Ordoudis, P. Pinson, and J. M. Morales, "An integrated market for electricity and natural gas systems with stochastic power producers," in *European Journal of Operational Research*, vol. 272, no. 2, pp. 642– 654, Jan. 2019.

[67] L. Zephyr and C. L. Anderson, "Stochastic dynamic programming approach to managing power system uncertainty with distributed storage" in *Computational Management Science*, vol. 15, no. 1, pp. 87–110, Jan. 2018.

[68] J. L. Duchaud, G. Notton, C. Darras, and C. Voyant, "Power ramp-rate control algorithm with optimal State of Charge reference via Dynamic Programming" *Energy*, vol. 149, pp. 709–717, Apr. 2018.

[69] H. T. Nguyen, L. B. Le, and Z. Y. Wang, "A bidding strategy for virtual power plants with the intraday demand response exchange market using the stochastic programming, " in *IEEE Transactions on Industry Applications*, vol. 54, no. 4, pp. 3044–3055, Aug. 2018.

[70] A. I. Mahmutogullari, S. Ahmed, O. Cavus, and M. S. Akturk, "The value of multi-stage stochastic programming in risk-averse unit commitment under uncertainty," arXiv:1808.00999, (2018).

[71] Megantoro, Prisma, F. D. Wijaya, and E. Firmansyah, "Analyze and optimization of genetic algorithm implemented on maximum power point tracking technique for PV system," in *Proceedings of 2017 International Seminar on Application for Technology of Information and Communication*, 2017.

[72] I. E. S. Naidu, K. R. Sudha, and A. C. Sekhar, "Dynamic stability margin evaluation of multi-machine power systems using genetic algorithm, " in *International Proceedings on Advances in Soft Computing, Intelligent Systems*

and Applications, M. S. Reddy, K. Viswanath, and S. P. K. M, Eds. Singapore: Springer, 2018.

[73] G. N. Nguyen, K. Jagatheesan, A. S. Ashour, B. Anand, and N. Dey, "Ant colony optimization based load frequency control of multiarea interconnected thermal power system with governor dead-band nonlinearity, " in *Smart Trends in Systems, Security and Sustainability*, X. S. Yang, A. K. Nagar, and A. Joshi, Eds. Singapore: Springer, 2018.

[74] R. Srikakulapu and V. U, "Optimized design of collector topology for offshore wind farm based on ant colony optimization with multiple travelling salesman problem," in *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 6, pp. 1181–1192, Nov. 2018.

[75] H. Li, D. Yang, W. Z. Su, J. H. Lu, X. H. Yu, "An overall distribution ¨ particle swarm optimization MPPT algorithm for photovoltaic system under partial shading, " in *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 265–275, Jan. 2019.

[76] H. J. Gu, R. F. Yan, and T. K. Saha, "Minimum synchronous inertia requirement of renewable power systems," in *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1533–1543, Mar. 2018.

[77] Z. Q. Wan, H. P. Li, and H. B. He, "Residential energy management with deep reinforcement learning," in *Proceedings of 2018 International Joint Conference on Neural Networks*, 2018.

[78] B. V. Mbuwir, M. Kaffash, and G. Deconinck, "Battery scheduling in a residential multi-carrier energy system using reinforcement learning," in *Proceedings of 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, 2018.

[79] J. D. Wu, H. W. He, J. K. Peng, Y. C. Li, and Z. J. Li, “Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus,” *Applied Energy*, vol. 222, pp. 799–811, Jul. 2018.

[80] Z. Q. Wan, H. P. Li, H. B. He, and D. Prokhorov, “Model-free real-time EV charging scheduling based on deep reinforcement learning,” in *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5246–5257, Sep. 2019.

[81] R. Z. Lu and S. H. Hong, “Incentive-based demand response for smart grid with reinforcement learning and deep neural network,” *Applied Energy*, vol. 236, pp. 937–949, Feb. 2019.

[82] X. Y. Zhang, “A data-driven approach for coordinating air conditioning units in buildings during demand response events,” Ph. D. dissertation, Department, Virginia Tech, Virginia, 2019.

[83] J. Hao, “Multi-agent reinforcement learning embedded game for the optimization of building energy control and power system planning,” arXiv: 1901.07333v1 (2019).

[84] T. Remani, E. A. Jasmin, and T. P. Imthias Ahamed, “Residential load scheduling with renewable generation in the smart grid: a reinforcement learning approach,” in *IEEE Systems Journal*, vol. 13, no. 3, pp. 3283– 3294, Sep. 2019.

[85] T. Chen and W. C. Su, “Indirect customer-to-customer energy trading with reinforcement learning,” in *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4338–4348, Jul. 2018.

[86] H. W. Wang, T. W. Huang, X. F. Liao, H. Abu-Rub, and G. Chen, “Reinforcement learning for constrained energy trading games with incomplete information, ” in *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3404–3416, Oct. 2017.

[87] H. C. Xu, X. Li, X. Y. Zhang, and J. B. Zhang, “Arbitrage of energy storage in electricity markets with deep reinforcement learning,” arXiv: 1904.12232 (2019).

[88] L. Xiao, X. Y. Xiao, C. H. Dai, M. Pengy, L. C. Wang, and H. V. Poor, “Reinforcement learning-based energy trading for microgrids,” arXiv: 1801.06285 (2018).

[89] B. J. Claessens, P. Vrancx, and F. Ruelens, “Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control,” in *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3259–3269, Jul. 2018.

[90] R. Rocchetta, L. Bellani, M. Compare, E. Zio, and E. Patelli, “A reinforcement learning framework for optimal operation and maintenance of power grids,” *Applied Energy*, vol. 241, pp. 291–301, May 2019.

[91] Q. Yang, G. Wang, A. Sadeghi, G. Giannakis and J. Sun, "Two-Timescale Voltage Control in Distribution Grids Using Deep Reinforcement Learning", in *IEEE Transactions on Smart Grid*, pp. 1-1, 2019. Available: 10.1109/tsg.2019.2951769.

[92] R. S. Diao, Z. W. Wang, D. Shi, Q. Y. Chang, J. J. Duan, and X. H. Zhang, “Autonomous voltage control for grid operation using deep reinforcement learning,” arXiv: 1904.10597 (2019).

[93] Z. Ni and S. Paul, “A multistage game in smart grid security: a reinforcement learning solution,” in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2684–2695, Sep. (2019).

[94] S. Paul and Z. Ni, “A study of linear programming and reinforcement learning for one-shot game in smart grid security,” in *Proceedings of 2018 International Joint Conference on Neural Networks*, 2018.

[95] Y. Chen, S. W. Huang, F. Liu, Z. S. Wang, and X. W. Sun, “Evaluation of reinforcement learning-based false data injection attack to automatic voltage control, ” in *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2158–2169, Mar. 2018.

[96] W. R. Liu, P. Zhuang, H. Liang, J. Peng, and Z. W. Huang, “Distributed economic dispatch in Microgrids based on cooperative reinforcement learning,” in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2192–2203, Jun. 2018.

[97] C. J. Han, B. Yang, T. Bao, T. Yu, and X. S. Zhang, “Bacteria foraging reinforcement learning for risk-based economic dispatch via knowledge transfer,” *Energies*, vol. 10, no. 5, pp. 638, May 2017.

[98] T. Hirata, D. B. Malla, K. Sakamoto, K. Yamaguchi, Y. Okada, and T. Sogabe, “Smart grid optimization by deep reinforcement learning over discrete and continuous action space, ” in *Bulletin of Networking, Computing, Systems, and Software*, vol. 8, no. 1, pp. 19–22, Jan. 2019.

[99] S. Munir, S. F. Abedin, N. H. Tran, and C. S. Hong, “When edge computing meets Microgrid: a deep reinforcement learning approach,” in *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7360–7374, Oct. 2019.

[100] D. L. Wang, Q. Y. Sun, Y. Y. Li, and X. R. Liu, “Optimal energy routing design in energy internet with multiple energy routing centers using artificial neural network-based reinforcement learning method, ” *Applied Sciences*, vol. 9, no. 3, pp. 520, Feb. 2019.

[101] L. Blume, "Characteristics of Load Ratio Control Circuits For Changing Transformer Ratio Under Load", *Transactions of the American Institute of Electrical Engineers*, vol. 51, no. 4, pp. 952-956, 1932. Available: 10.1109/t-

aiee.1932.5056195.

[102] H. Dura, "Optimum Number, Location, and Size of Shunt Capacitors in Radial Distribution Feeders A Dynamic Programming Approach," in *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-87, no. 9, pp. 1769-1774, Sept. 1968, doi: 10.1109/TPAS.1968.291982.

[103] S. Narita and M. A. A. Hammam, "A Computational Algorithm for Real-Time Control of System Voltage and Reactive Power Part I - Problem Formulation", *IEEE Transactions on Power Apparatus and Systems*, vol. 90, no. 6, pp. 2495-2501, 1971. Available: 10.1109/tpas.1971.292861.

[104] R. Billinton and S. Sachdeva, "Real and Reactive Power Optimization by Suboptimum Techniques", *IEEE Transactions on Power Apparatus and Systems*, vol. 92, no. 3, pp. 950-956, 1973. Available: 10.1109/tpas.1973.293661.

[105] A. Kishore and E. Hill, "Static Optimization of Reactive Power Sources by use of Sensitivity Parameters", *IEEE Transactions on Power Apparatus and Systems*, vol. 90, no. 3, pp. 1166-1173, 1971. Available: 10.1109/tpas.1971.292881.

[106] R. Shoults and Mo-Shing Chen, "Reactive power control by least squares minimization", *IEEE Transactions on Power Apparatus and Systems*, vol. 95, no. 1, pp. 325-334, 1976. Available: 10.1109/t-pas.1976.32109.

[107] S. Narita and M. Hammam, "A Computational Algorithm for Real-Time Control of System Voltage and Reactive Power Part II - Algorithm of Optimization", *IEEE Transactions on Power Apparatus and Systems*, vol. 90, no. 6, pp. 2502-2508, 1971. Available: 10.1109/tpas.1971.292862.

[108] M. Ilic-Spong, J. Christensen and K. Eichorn, "Secondary voltage control using pilot point information", *IEEE Transactions on Power Systems*, vol. 3, no. 2, pp. 660-668, 1988. Available: 10.1109/59.192920.

[109] S. Rama Iyer, K. Ramachandran and S. Hariharan, "New technique for optimal reactive-power allocation for loss minimisation in power systems", *IEE Proceedings C Generation, Transmission and Distribution*, vol. 130, no. 4, p. 178, 1983. Available: 10.1049/ip-c.1983.0030.

[110] S. Civanlar and J. Grainger, "Volt/Var Control on Distribution Systems with Lateral Branches Using Switched Capacitors and Voltage Regulators, Part I: The Solution Method", *IEEE Power Engineering Review*, vol. 5, no. 11, pp. 53-53, 1985. Available: 10.1109/mper.1985.5528388.

[111] J. Grainger and S. Civanlar, "Volt/Var Control on Distribution Systems with Lateral Branches Using Switched Capacitors and Voltage Regulators, Part II: The Overall Problem", *IEEE Power Engineering Review*, vol. 5, no. 11, pp. 52-52, 1985. Available: 10.1109/mper.1985.5528387.

[112] I. Roytelman, B. Wee and R. Lugtu, "Volt/var control algorithm for modern distribution management system", *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1454-1460, 1995. Available: 10.1109/59.466504.

[113] I. Roytelman, and S. M. Shahidehpour, “Practical aspects of distribution automation in normal and emergency conditions,” *IEEE Transactions on Power Delivery*, vol. 8, no. 4, pp. 2002-2008, Oct. 1993.

[114] G. Bakare, G. Krost, G. Venayagamoorthy and U. Aliyu, "Comparative Application of Differential Evolution and Particle Swarm Techniques to Reactive Power and Voltage Control", in *International Conference on Intelligent Systems Applications to Power Systems*, 2007.

[115] A. Vaccaro, G. Velotto and A. Zobaa, "A Decentralized and Cooperative Architecture for Optimal Voltage Regulation in Smart Grids", *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4593-4602, 2011. Available:

10.1109/tie.2011.2143374.

[116] B. de Souza and A. de Almeida, "Multiobjective Optimization and Fuzzy Logic Applied to Planning of the Volt/Var Problem in Distributions Systems", *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1274-1281, 2010. Available: 10.1109/tpwrs.2010.2042734.

[117] S. Auchariyamet and S. Sirisumrannukul, "Volt/VAr control in distribution systems by fuzzy multiobjective and particle swarm," in *International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2009, doi: 10.1109/ECTICON.2009.5137000.

[118] "Volt/Var Management market global forecast to 2024", marketsandmarkets, 2019.

[119] R. A. Jabr and I. Dˇzafi´c, "Sensitivity-based discrete coordinate-descent for volt/var control in distribution networks," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4670–4678, 2016.

[120] M. B. Liu, C. A. Canizares, and W. Huang, "Reactive power and voltage control in distribution systems with limited switching operations," *IEEE Transactions on Power Systems* , vol. 24, no. 2, pp. 889–899, 2009.

[121] L. Zhang, W. Tang, J. Liang, P. Cong, and Y. Cai, "Coordinated day ahead reactive power dispatch in distribution network based on real power forecast errors," *IEEE Transactions on Power Systems* , vol. 31, no. 3, pp. 2472–2480, 2016.

[122] Z. Yang, A. Bose, H. Zhong, N. Zhang, Q. Xia, and C. Kang, "Optimal reactive power dispatch with accurately modeled discrete control devices: A successive linear approximation approach," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2435–2444, 2017.

[123] Q. Nguyen, H. V. Padullaparti, K. Lao, S. Santoso, X. Ke, and N. Samaan, “Exact optimal power dispatch in unbalanced distribution systems with high PV penetration,” *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 718–728, 2019.

[124] J. Yu, W. Dai, W. Li, X. Liu, and J. Liu, “Optimal reactive power flow of interconnected power system based on static equivalent method using border PMU measurements,” *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 421–429, Jan. 2018.

[125] I. Roytelman and J. Medina, "Volt/VAR control and conservation voltage reduction as a function of advanced DMS," *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Minneapolis, MN, 2016, pp. 1-4.

[126] M. A. Medina, C. A. Coello Coello, and J. M. Ramirez, “Reactive power handling by a multi-objective teaching learning optimizer based on decomposition,” *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3629–3637, Nov. 2013.

[127] Z. Wang, H. Chen, J. Wang, and M. Begovic, “Inverter-less hybrid voltage/var control for distribution circuits with photovoltaic generators,” *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2718–2728, Nov. 2014.

[128] P. Dong, L. Xu, Y. Lin and M. Liu, “Multi-objective coordinated control of reactive compensation devices among multiple substations,” *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 2395–2403, May 2018.

[129] J.O. Petinrin and M. Shaaban, "Impact of renewable generation on voltage control in distribution systems," *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 770–783, 2016.

[130] B. Cheng and W. Powell, “Co-optimizing battery storage for the frequency regulation and energy arbitrage using multi-scale dynamic programming,” *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1997–2005, May 2018.

[131] D. Q. Hung, N. Mithulananthan, and R. Bansal, “Integration of PV and BES units in commercial distribution systems considering energy loss and voltage stability,” *Applied Energy*, vol. 113, pp. 1162–1170, Jan. 2014.

[132] L. H. Macedo, J. F. Franco, M. J. Rider, and R. Romero, "Optimal operation of distribution networks considering energy storage devices," in *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2825-2836, Nov. 2015, Available: 10.1109/TSG.2015.2419134.

[133] R. A. Jabr, "Linear decision rules for control of reactive power by distributed photovoltaic generators," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2165-2174, March 2018, Available: 10.1109/TPWRS.2017.2734694.

[134] K. Muttaqi, M. Islam, and D. Sutanto, "Future power distribution grids: integration of renewable energy, energy Storage, electric Vehicles, superconductor, and magnetic bus", *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 2, pp. 1-5, 2019. Available: 10.1109/tasc.2019.2895528.

[135] M. Kraiczy, T. Stetz and M. Braun, "Parallel operation of transformers with on load tap changer and photovoltaic systems with reactive power control," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6419-6428, Nov. 2018, Available: 10.1109/TSG.2017.2712633.

[136] J. Wang, G. Bharati, S. Paudyal, O. Ceylan, B. Bhattarai, and K. Myers, "Coordinated electric vehicle charging with reactive power support to distribution

grids", *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 54-63, 2019. Available: 10.1109/tii.2018.2829710.

[137] O. Gandhi, W. Zhang, C. Rodriguez-Gallegos, M. Bieri, T. Reindl, and D. Srinivasan, "Analytical approach to reactive power dispatch and energy arbitrage in distribution systems with DERs", *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6522-6533, 2018. Available: 10.1109/tpwrs.2018.2829527.

[138] H. Nazaripouya, H. Pota, C. Chu, and R. Gadh, "Real-time model-free coordination of active and reactive powers of distributed energy resources to improve voltage regulation in distribution systems", *IEEE Transactions on Sustainable Energy*, pp. 1-1, 2019. Available: 10.1109/tste.2019.2928824.

[139] Y. Xu, Z. Dong, R. Zhang, and D. Hill, "Multi-timescale coordinated voltage/var control of high renewable-penetrated distribution systems", *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4398-4408, 2017. Available: 10.1109/tpwrs.2017.2669343.

[140] R. Jha, A. Dubey, C. Liu, and K. Schneider, "Bi-level Volt-VAR optimization to coordinate smart inverters with voltage control devices", *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1801-1813, 2019. Available: 10.1109/tpwrs.2018.2890613.

[141] F. Nazir, B. Pal, and R. Jabr, "A two-stage chance constrained volt/var control scheme for active distribution networks with nodal power uncertainties", *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 314-325, 2019. Available: 10.1109/tpwrs.2018.2859759.

[142] C. Zhang, Y. Xu, Z. Dong and J. Ravishankar, "Three-stage robust inverter-based voltage/var control for distribution networks with high-level PV," in IEEE

Transactions on Smart Grid, vol. 10, no. 1, pp. 782-793, Jan. 2019, Available: 10.1109/TSG.2017.2752234.

[143] R. Anilkumar, G. Devriese, and A. K. Srivastava, "Voltage and Reactive Power Control to Maximize the Energy Savings in Power Distribution System With Wind Energy," in *IEEE Transactions on Industry Applications*, vol. 54, no. 1, pp. 656-664, Jan.-Feb. 2018, doi: 10.1109/TIA.2017.2740850.

[144] B. A. de Souza and A. M. F. de Almeida, "Multiobjective Optimization and Fuzzy Logic Applied to Planning of the Volt/Var Problem in Distributions Systems," in *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1274-1281, Aug. 2010, doi: 10.1109/TPWRS.2010.2042734.

[145] W. Sheng, K. Liu, S. Cheng, X. Meng, and W. Dai, “A trust region sqp method for coordinated voltage control in smart distribution grid,” in *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 381–391, 2016.

[146] F. Capitanescu, I. Bilibin, and E. Romero Ramos, “A comprehensive centralized approach for voltage constraints management in active distribution grid,” in *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 933–942, 2014.

[147] H. Ahmadi, J. R. Mart´ı, and H. W. Dommel, “A framework for volt-var optimization in distribution systems,” in *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1473–1483, 2015.

[148] M. Falahi, K. Butler-Purry, and M. Ehsani, “Dynamic reactive power control of islanded microgrids,” in *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3649–3657, Nov. 2013.

[149] X. Y. Shang, M. S. Li, T. Y. Ji, L. L. Zhang, and Q. H. Wu, "Discrete reactive power optimization considering safety margin by dimensional Q-learning," *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, Bangkok, 2015,

pp. 1-5.

[150] Y. Xu, W. Zhang, W. Liu, and F. Ferrese, "Multiagent-Based Reinforcement Learning for Optimal Reactive Power Dispatch", in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1742-1751, 2012. Available: 10.1109/tsmcc.2012.2218596.

[151] W. Zhang, W. Liu, X. Wang, L. Liu, and F. Ferrese, "Distributed Multiple Agent System Based Online Optimal Reactive Power Control for Smart Grids", in *IEEE Transactions on Smart Grid*, vol. 5, no. 5, pp. 2421-2431, 2014. Available: 10.1109/tsg.2014.2327478.

[152] H. Xu, A. Dominguez-Garcia and P. Sauer, "Optimal Tap Setting of Voltage Regulation Transformers Using Batch Reinforcement Learning", in *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1990-2001, 2020. Available: 10.1109/tpwrs.2019.2948132.

[153] W. Wang, N. Yu, Y. Gao, and J. Shi, "Safe Off-policy Deep Reinforcement Learning Algorithm for Volt-VAR Control in Power Distribution Systems", in *IEEE Transactions on Smart Grid*, pp. 1-1, 2019. Available: 10.1109/tsg.2019.2962625.

[154] W. Wang, N. Yu, J. Shi, and Y. Gao, "Volt-VAR Control in Power Distribution Systems with Deep Reinforcement Learning", in *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, 2019.

[155] C. Li and C. Jin, "Coordination of PV Smart Inverters Using Deep Reinforcement Learning for Grid Voltage Regulation", in *IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2019.

[156] T&TEC, "Incentive mechanisms for managing transmission and distribution

losses, regulated industries commission", 2005.

[157] T. Gonen, *Electric Power Distribution System Engineering*, 2nd ed. CRC Press, Taylor & Francis Group, 2007.

[158] D. Dohnal, "On-load tap-changers for power transformers", *Reinhausen.com*, 2013. [Online]. Available: https://www.reinhausen.com/PortalData/1/Resources/tc/research_development/vacuum_technology/PB252_en_Power_Transformers.pdf.

[159] "IEEE standard for shunt power capacitors," in *IEEE Std 18-1992* , pp.1-31, 25 Feb. 1993, doi: 10.1109/IEEESTD.1993.114400.

[160] National Renewable Energy Laboratory (NREL), "Review of PREPA technical requirements for interconnecting wind and solar generation", 2013.

[161] S. Shao, T. Zhang, M. Ipattanasomporn, and S. Rahman, "Impact of TOU rates on distribution load shapes in a smart grid with PHEV penetration," *IEEE PES Transmission and Distribution Conference and Exposition,* 2010.

[162] S. Acha, T. C. Green and N. Shah, "Effects of optimised plug-in hybrid vehicle charging strategies on electric distribution network losses," *IEEE PES T&D 2010*, 2010, pp. 1-6, Available: 10.1109/TDC.2010.5484397.

[163] K. Clement, H. Haesen, and J. Driesen, "Coordinated charging of multiple plug-in hybrid electric vehicles in residential distribution grids," *IEEE PES Power Systems Conference and Exposition*, 2009.

[164] L. J. Borle, "Zero average current error control methods for bidirectional AC-DC converters", Curtin University of Technology, Perth, Australia, 1999.

[165] M. C. Kisacikoglu, B. Ozpineci, L. M. Tolbert, and F. Wang, "Single-phase inverter design for V2G reactive power compensation," *26th Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, Fort Worth, TX, USA,

2011.

[166] H. Hesse et al., "Ageing and efficiency aware battery dispatch for arbitrage markets using mixed integer linear programming", *Energies*, vol. 12, no. 6, p. 999, 2019. Available: 10.3390/en12060999.

[167] Cooperative.com. 2020. [online] Available at: https://www.cooperative.com/programsservices/bts/Documents/Reports/Battery-Energy-Storage-Overview-Report-Update-April-2019.pdf.

[168] D. A. Atanackovic, and V. Dabic, "Deployment of real-time state estimator and load flow in BC Hydro DMS - challenges and opportunities," *IEEE Power and Energy Society General Meeting (PES)*, Vancouver, BC, Canada, 2013.

[169] R. Zafar, J. Ravishankar, J. E. Fletcher and H. R. Pota, "Multi-Timescale Model Predictive Control of Battery Energy Storage System Using Conic Relaxation in Smart Distribution Grids," in *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7152-7161, Nov. 2018, doi: 10.1109/TPWRS.2018.2847400.

[170] T. Niknam, B. B. Firouzi, and A. Ostadi, "A new fuzzy adaptive particle swarm optimization for daily Volt/Var control in distribution networks considering distributed generators," *Applied Energy*, vol. 87, no. 6, pp. 1919-1928, 2010.

[171] S. Auchariyamet and S. Sirisumrannukul, "Optimal daily coordination of volt/VAr control devices in distribution systems with distributed generators," *45th International Universities Power Engineering Conference UPEC2010*, Cardiff, Wales, 2010, pp. 1-6.

[172] R.-h. Liang and Y.-s. Wang, "Fuzzy-based reactive power and voltage control in a distribution system," *IEEE Transactions on Power Delivery*, vol. 18, no. 2, pp. 610-618, April 2003.

[173] S. Jashfar, M. M. Hosseini-Biyouki, and S. Esmaeili, "A stochastic programming

to Volt/VAR total harmonic distortion control in distribution networks including wind turbines," *Electric Power Components and Systems*, no. 7, pp. 733-746, 2015.

[174] A. Padilha-Feltrin, D. Quijano Rodezno, and J Sanches Mantovani, "Volt-VAR multiobjective optimization to peak-load relief and energy efficiency in distribution Networks," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 618-626, April 2015.

[175] M. Manbachi, H. Farhangi, A. Palizban, and S. Arzanpour, " A novel Volt-VAR optimization engine for smart distribution networks utilizing vehicle to grid dispatch," *International Journal of Electrical Power and Energy Systems*, vol. 74, pp. 238-251, 2016.

[176] T. Niknam, M. Zare, and J Aghaei, "Scenario-based multiobjective volt/var control in distribution networks including renewable energy sources," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 2004-2019, Oct. 2012.

[177] A. R. Malekpour and T. Niknam, "A probabilistic multi-objective daily Volt/Var control at distribution networks including renewable energy sources," *Energy*, vol. 36, no. 5, pp. 3477-3488, 2011.

[178] D. F. Pires, A. Martins, and C. Antunes, "A multi objective model for VAR planning in radial distribution networks based on tabu search," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1089-1094, May 2005.

[179] R.-H. Liang, Y.-K. Chen, and Y.-T. Chen, "Volt/Var control in a distribution system by a fuzzy optimization approach," *International Journal of Electrical Power & Energy Systems*, no. 2, pp. 278-287, 2011.

[180] A. R. Malekpour, S. Tabatabaei, and T. Niknam, "Probabilistic approach to multi-objective VoltNar control of distribution system considering hybrid fuel

cell and wind energy sources using Improved Shuffled Frog Leaping Algorithm," *Renewable Energy*, vol. 39, no. I, pp. 228-240, 2012.

[181] M. Zare and T. Niknam, "A new multi-objective for environmental and economic management of Volt/Var Control considering renewable energy resources," *Energy*, vol. 55, pp. 236-252, 2013.

[182] T. Niknam, "A new HBMO algorithm for multi objective daily Volt/Var control in distribution systems considering Distributed Generators," *Applied Energy*, vol. 88, no. 3, pp. 778-788, 2011.

[183] M. Zare, T. Niknam, R. Azizipanah-Abarghooee, and B. Amiri, "Multiobjective probabilistic reactive power and voltage control with wind site correlations," *Energy*, vol. 66, pp. 810-822, 2014.

[184] T. Niknam, "A new approach based on ant colony optimization for daily VoltNar control in distribution networks considering distributed generators," *Energy Conversion and Management*, vol. 49, no. 12, pp. 3417-3424, 2008.

[185] Y. Mohamed Shuaib, M. Surya Kalavathi, and C. Christober Asir Rajan, "Optimal capacitor placement in radial distribution system using Gravitational Search Algorithm," *International Journal of Electrical Power and Energy Systems*, vol. 64, pp. 384-397, 2015.

[186] J. Radosavljević, *Metaheuristic optimization in power engineering*. London: The Institution of Engineering & Technology, 2018.

[187] T. S. Medeiros and N. Kagan, "Bio-inspired metaheuristics applied to Volt/VAr control optimization problem in smart grid context," *2016 17th International Conference on Harmonics and Quality of Power (ICHQP)*, Belo Horizonte, 2016, pp. 295-300, doi: 10.1109/ICHQP.2016.7783328.

[188] A. Mohapatra, P. Bijwe and B. Panigrahi, "An efficient hybrid approach for

Volt/Var control in distribution systems", *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1780-1788, 2014. Available: 10.1109/tpwrd.2014.2306845.

[189] D. Jakus, J. Vasilj and P. Sarajcev, "Voltage control in MV distribution networks through coordinated control of tap changers and renewable energy sources," *2015 IEEE Eindhoven PowerTech, Eindhoven*, 2015, pp. 1-6, doi: 10.1109/PTC.2015.7232391.

[190] R. Rahmaniani, T. Crainic, M. Gendreau and W. Rei, "The Benders decomposition algorithm: A literature review", *European Journal of Operational Research*, vol. 259, no. 3, pp. 801-817, 2017. Available: 10.1016/j.ejor.2016.12.005.

[191] V. Miranda, J. V. Ranito, L. V. Proenca, "Genetic algorithm in optimal multistage distribution network planning" *IEEE Transactions on Power System*, vol. 9, no. 4, pp.1891-1898, 1996.

[192] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, April 2002, Available: 10.1109/4235.996017.

[193] Y. Yusoff, M. Ngadiman and A. Zain, "Overview of NSGA-II for optimizing machining process parameters", *Procedia Engineering*, vol. 15, pp. 3978-3983, 2011. Available: 10.1016/j.proeng.2011.08.745.

[194] Uppsala University, "Parallel Global Optimization of ABB's metal process models using Matlab", 2014.

[195] M. Jafari, T. O. Olowu, and A. I. Sarwat, "Optimal smart inverters Volt-VAR curve selection with a multi-objective Volt-VAR optimization using evolutionary algorithm approach," *2018 North American Power Symposium (NAPS)*, Fargo, ND, 2018, pp. 1-6.

[196] B. A. de Souza and A. M. F. de Almeida, "Multiobjective Optimization and Fuzzy Logic Applied to Planning of the Volt/Var Problem in Distributions Systems," in *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1274-1281, Aug. 2010, doi: 10.1109/TPWRS.2010.2042734.

[197] C. Coello Coello, D. A. Van Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer US, 2014.

[198] R. Ramos de Souza, A. Roberto Balbo, L. Nepomuceno, E. Cassia Baptista, E. Martins Soler, and R. Bento Nogueira Pinheiro, "A primal-dual interior/exterior point method, with combined directions and quadratic test in reactive optimal power flow problems," *IEEE Latin America Transactions*, vol. 15, no. 8, pp. 1413-1421, 2017.

[199] Z. Jing, L. Yanhua, Z. Dongying, and D. Yue, "Multi-time-scale coordinative and complementary reactive power and voltage control strategy for wind farms cluster," *2018 International Conference on Power System Technology (POWERCON)*, Guangzhou, 2018, pp. 1927-1934.

[200] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, “MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2 2011.

[201] Y. Zhang, X. Wang, J. Wang and Y. Zhang, "Deep Reinforcement Learning Based Volt-VAR Optimization in Smart Distribution Systems," in *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 361-371, Jan. 2021, doi: 10.1109/TSG.2020.3010130.

[202] M. J. Han, R. May, X. X. Zhang, X. R. Wang, S. Pan, D. Yan, Y. Jin, and L. G. Xu, “A review of reinforcement learning methodologies for controlling occupant

comfort in buildings, ”Sustainable Cities and Society, vol. 51, pp. 101748, Nov. 2019.

[203] M. Ding and X. Q. Yin, “A review on multi-agent technology in microgrid control,” Electronics Science Technology and Application, vol. 5, no. 1, pp. 1–13, 2018.

[204] L. F. Cheng and T. Yu, “A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems, ” International Journal of Energy Research, vol. 43, no. 6, pp. 1928–1973, May 2019.

[205] C. J. Watkins and P. Dayan, “Q-learning,” Mach. Learn., vol. 8, no. 3-4, pp. 279–292, 1992.

[206] "Deep Reinforcement Learning Doesn't Work Yet", Alexirpan.com, 2018. [Online]. Available: https://www.alexirpan.com/2018/02/14/rl-hard.html.

[207] T. P. Lillicrap et al., “Continuous control with deep reinforcement learning,” 2015, arXiv:1509.02971.

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

[209] S. Fujimoto, H. van Hoof, and D. Meger, “Addressing function approximation error in actor-critic methods,” 2018, arXiv:1802.09477.

[210] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, “Deterministic policy gradient algorithms,” in *Proceeding International Conference Machine Learning (ICML)*, 2014, pp. 1–9.

[211] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand", *Nature Energy*, vol. 3, no. 3, pp. 193-201, 2018.

Available: 10.1038/s41560-017-0074-z.

[212] M. Muratori, M. C. Roberts, R. Sioshansi, V. Marano, and G. Rizzoni, "A highly resolved modeling technique to simulate residential power demand," *Applied Energy*, vol. 107, no. 0, pp. 465 - 473, 2013.

[213] M. Muratori, V. Marano, R. Sioshansi, and G. Rizzoni, "Energy consumption of residential HVAC systems: a simple physically-based model," in *2012 IEEE Power and Energy Society General Meeting*, San Diego, CA, USA: IEEE, 22-26 July 2012.

[214] D. Kingma and J. Lei Ba, "Adam: A Method for Stochastic Optimization", 2014.

APPENDIX

Appendix A: Data of standard IEEE 33-bus system.

```
%% bus data
%   bus_i   type    Pd  Qd  Gs  Bs  area    Vm  Va  baseKV  zone    Vmax
Vmin
bus = [  %% (Pd and Qd are specified in kW & kVAr here, converted to MW &
MVAr below)
    1   3   0   0   0   0   1   1   0   12.66   1   1   1;
    2   1   100 60  0   0   1   1   0   12.66   1   1.1 0.9;
    3   1   90  40  0   0   1   1   0   12.66   1   1.1 0.9;
    4   1   120 80  0   0   1   1   0   12.66   1   1.1 0.9;
    5   1   60  30  0   0   1   1   0   12.66   1   1.1 0.9;
    6   1   60  20  0   0   1   1   0   12.66   1   1.1 0.9;
    7   1   200 100 0   0   1   1   0   12.66   1   1.1 0.9;
    8   1   200 100 0   0   1   1   0   12.66   1   1.1 0.9;
    9   1   60  20  0   0   1   1   0   12.66   1   1.1 0.9;
    10  1   60  20  0   0   1   1   0   12.66   1   1.1 0.9;
    11  1   45  30  0   0   1   1   0   12.66   1   1.1 0.9;
    12  1   60  35  0   0   1   1   0   12.66   1   1.1 0.9;
    13  1   60  35  0   0   1   1   0   12.66   1   1.1 0.9;
    14  1   120 80  0   0   1   1   0   12.66   1   1.1 0.9;
    15  1   60  10  0   0   1   1   0   12.66   1   1.1 0.9;
    16  1   60  20  0   0   1   1   0   12.66   1   1.1 0.9;
    17  1   60  20  0   0   1   1   0   12.66   1   1.1 0.9;
    18  1   90  40  0   0   1   1   0   12.66   1   1.1 0.9;
    19  1   90  40  0   0   1   1   0   12.66   1   1.1 0.9;
    20  1   90  40  0   0   1   1   0   12.66   1   1.1 0.9;
    21  1   90  40  0   0   1   1   0   12.66   1   1.1 0.9;
    22  1   90  40  0   0   1   1   0   12.66   1   1.1 0.9;
    23  1   90  50  0   0   1   1   0   12.66   1   1.1 0.9;
    24  1   420 200 0   0   1   1   0   12.66   1   1.1 0.9;
    25  1   420 200 0   0   1   1   0   12.66   1   1.1 0.9;
    26  1   60  25  0   0   1   1   0   12.66   1   1.1 0.9;
    27  1   60  25  0   0   1   1   0   12.66   1   1.1 0.9;
    28  1   60  20  0   0   1   1   0   12.66   1   1.1 0.9;
    29  1   120 70  0   0   1   1   0   12.66   1   1.1 0.9;
    30  1   200 600 0   0   1   1   0   12.66   1   1.1 0.9;
    31  1   150 70  0   0   1   1   0   12.66   1   1.1 0.9;
    32  1   210 100 0   0   1   1   0   12.66   1   1.1 0.9;
    33  1   60  40  0   0   1   1   0   12.66   1   1.1 0.9;
];

%% generator data
%   bus Pg  Qg  Qmax    Qmin    Vg  mBase   status  Pmax    Pmin    Pc1 Pc2
Qc1min  Qc1max  Qc2min  Qc2max  ramp_agc    ramp_10 ramp_30 ramp_q  apf
gen = [
    1   0   0   10  -10 1   100 1   10  0   0   0   0   0   0   0   0   0
0   0   0;
];

%% branch data
%   fbus    tbus    r   x   b   rateA   rateB   rateC   ratio   angle
status  angmin  angmax
branch = [  %% (r and x specified in ohms here, converted to p.u. below)
    1   2   0.0922  0.0470  0   0   0   0   0   0   1   -360    360;
    2   3   0.4930  0.2511  0   0   0   0   0   0   1   -360    360;
    3   4   0.3660  0.1864  0   0   0   0   0   0   1   -360    360;
    4   5   0.3811  0.1941  0   0   0   0   0   0   1   -360    360;
    5   6   0.8190  0.7070  0   0   0   0   0   0   1   -360    360;
    6   7   0.1872  0.6188  0   0   0   0   0   0   1   -360    360;
    7   8   0.7114  0.2351  0   0   0   0   0   0   1   -360    360;
    8   9   1.0300  0.7400  0   0   0   0   0   0   1   -360    360;
    9   10  1.0440  0.7400  0   0   0   0   0   0   1   -360    360;
    10  11  0.1966  0.0650  0   0   0   0   0   0   1   -360    360;
    11  12  0.3744  0.1238  0   0   0   0   0   0   1   -360    360;
```

```
    12  13  1.4680  1.1550  0   0   0   0   0   0   1   -360    360;
    13  14  0.5416  0.7129  0   0   0   0   0   0   1   -360    360;
    14  15  0.5910  0.5260  0   0   0   0   0   0   1   -360    360;
    15  16  0.7463  0.5450  0   0   0   0   0   0   1   -360    360;
    16  17  1.2890  1.7210  0   0   0   0   0   0   1   -360    360;
    17  18  0.7320  0.5740  0   0   0   0   0   0   1   -360    360;
    2   19  0.1640  0.1565  0   0   0   0   0   0   1   -360    360;
    19  20  1.5042  1.3554  0   0   0   0   0   0   1   -360    360;
    20  21  0.4095  0.4784  0   0   0   0   0   0   1   -360    360;
    21  22  0.7089  0.9373  0   0   0   0   0   0   1   -360    360;
    3   23  0.4512  0.3083  0   0   0   0   0   0   1   -360    360;
    23  24  0.8980  0.7091  0   0   0   0   0   0   1   -360    360;
    24  25  0.8960  0.7011  0   0   0   0   0   0   1   -360    360;
    6   26  0.2030  0.1034  0   0   0   0   0   0   1   -360    360;
    26  27  0.2842  0.1447  0   0   0   0   0   0   1   -360    360;
    27  28  1.0590  0.9337  0   0   0   0   0   0   1   -360    360;
    28  29  0.8042  0.7006  0   0   0   0   0   0   1   -360    360;
    29  30  0.5075  0.2585  0   0   0   0   0   0   1   -360    360;
    30  31  0.9744  0.9630  0   0   0   0   0   0   1   -360    360;
    31  32  0.3105  0.3619  0   0   0   0   0   0   1   -360    360;
    32  33  0.3410  0.5302  0   0   0   0   0   0   1   -360    360;
];
```

Appendix B: Publications.

In the period of this thesis research, several publications have been prepared, which can be listed as follows:

Publications under review:

1) Journal paper titled “Performance Assessment of Two-timescale Multi-objective Volt/Var Optimization Scheme Considering EV Charging Stations, BESSs, and RESs in Active Distribution Networks” submitted to Electric Power System Research (EPSR) on 26th November 2020.

Publications in preparation:

2) Journal paper titled “Model-Free Deep Reinforcement Learning in Two-Timescale Volt/Var Optimization with Continuous Action Spaces for Active Distribution Networks” to be submitted to IEEE transactions on Smart Grid.
3) Journal paper titled “Implementation of Deep Reinforcement Learning for Power System Applications on MATLAB Using Reinforcement Learning Toolbox”
4) Conference paper titled “Performance Analysis of Deep Reinforcement Learning Schemes for Power System Applications”