FLEXIBLE OPERATION OF HYBRID AC/DC MICROGRIDS: FROM SMART METER ANALYTICS TO UNCERTAINTY MODELING

LIANG ZIPENG

PhD

The Hong Kong Polytechnic University

2026

The Hong Kong Polytechnic University

Department of Electrical and Electronic Engineering

Flexible Operation of Hybrid AC/DC Microgrids: From Smart Meter Analytics to Uncertainty Modeling

Liang Zipeng

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

September 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

__________________________ (Signed)

LIANG Zipeng (Name of student)

Abstract

The global energy system is undergoing a fundamental transformation driven by decarbonization, decentralization, and electrification. Declining costs of renewable energy sources (RES), rapid growth in electric vehicles (EVs), and increasing demands for flexibility and resilience are reshaping modern power networks. Microgrids, localized systems capable of operating in both grid-connected and islanded modes, have emerged as a key enabler of this transition. In particular, hybrid AC/DC microgrids (HMGs), which integrate AC and DC subsystems through bi-directional converters (BDCs), offer enhanced efficiency by reducing unnecessary power conversions while accommodating diverse energy resources and loads.

Despite their promise, the reliable and flexible operation of HMGs faces three interrelated challenges: non-convex BDC modeling, limited observability of load behavior, and significant renewable generation uncertainty. At the same time, the widespread deployment of smart meters enables appliance-level load analysis through non-intrusive load monitoring (NILM), which is essential for demand response and operational flexibility. However, existing NILM methods often struggle with low-frequency real-world data and fail to exploit contextual information. Moreover, prevailing uncertainty modeling techniques in microgrid optimization, such as convex hulls or polyhedral sets, tend to be overly conservative, sensitive to outliers, or computationally prohibitive. These limitations motivate the development of new data-driven models, learning frameworks, and uncertainty-aware optimization methods.

This thesis addresses these challenges through a unified framework that integrates smart meter analytics, converter modeling, and robust optimization. Five core contributions are made. First, data-driven convex BDC models are developed to approximate efficiency characteristics while pre-

serving convexity. Sufficient conditions are derived to guarantee non-simultaneous rectification and inversion, eliminating the need for binary variables and enabling scalable optimization. Case studies demonstrate substantial reductions in computational burden with minimal loss of accuracy.

Second, a heterogeneous multi-gate mixture-of-experts (HMMOE) framework is proposed for NILM. By combining convolutional, recurrent, and attention-based experts with adaptive task weighting, the method jointly improves appliance state detection and power estimation. Extensive experiments on public and real-world datasets confirm significant performance gains over state-of-the-art approaches.

Third, NILM is extended to a multimodal, dual-task architecture that incorporates contextual information such as weather and indoor conditions. An analytical task-weighting solution is derived, and the resulting appliance-level insights are embedded into a robust hybrid building microgrid model to evaluate demand response potential more accurately.

Fourth, novel renewable uncertainty sets are introduced, including pairwise convex hull, intersected convex hull, and probability-driven data-correlated formulations. These sets provide adjustable conservativeness, tighter feasible regions, and improved out-of-sample performance compared with traditional approaches.

Finally, tailored solution algorithms are developed for large-scale robust optimization involving renewable uncertainty, hydrogen storage, and EVs. Innovations include semi-convex relaxations, enhanced column-and-constraint generation, and quad-level decomposition with efficient uncertainty exploration.

By bridging appliance-level data analytics and system-level decision-making, this thesis provides both theoretical foundations and practical tools for enabling flexible, intelligent, and robust operation of renewable-penetrated hybrid microgrids.

List of Publications

A. International Publications

[1] **Z. Liang**, C. Y. Chung, Q. Wang, H. Chen, H. Yang, and C. Wu, "Fortifying renewable-dominant hybrid microgrids: A bi-directional converter based interconnection planning approach", *Engineering*, vol. 51, pp. 130–143, 2025.

[2] **Z. Liang**, C. Y. Chung, W. Zhang, Q. Wang, W. Lin, and C. Wang, "Enabling high-efficiency economic dispatch of hybrid AC/DC networked microgrids: Steady-state convex bi-directional converter models", *IEEE Trans. Smart Grid*, vol. 16, no. 1, pp. 45–61, 2025.

[3] **Z. Liang**, X. Yin, C. Y. Chung, S. K. Rayeem, X. Chen, and H. Yang, "Managing massive RES integration in hybrid microgrids: A data-driven quad-level approach with adjustable conservativeness", *IEEE Trans. Ind. Inform.*, vol. 21, no. 10, pp. 7698–7709, 2025.

[4] **Z. Liang**, C. Y. Chung, H. Yang, *et al.*, "A heterogeneous multiple-experts approach to low-frequency nonintrusive load monitoring", *IEEE Trans. Smart Grid*, pp. 1–18, 2025, Early Access.

[5] **Z. Liang**, C. Y. Chung, X. Yin, Q. Wang, and H. Yang, "Incorporating data-driven uncertainty set into hybrid microgrids with high renewable energy penetration and bi-directional converters", *IEEE Trans. Sustain. Energy*, pp. 1–19, 2025, Early Access.

[6] **Z. Liang**, C. Y. Chung, H. Dong, S. He, X. Yin, and J. Chen, "Energy management in microgrids: Sufficient conditions for convex bi-directional converter operation model", in *2024 IEEE PES 16th Asia-Pacific Power and Energy Engineering Conference*, Nanjing, 2024, 1–5 **(Best Paper Award)**.

[7] **Z. Liang**, S. K. Rayeem, C. Y. Chung, S. He, Y. Zhu, and X. Yin, “Energy scheduling of vpps with renewable generation: A hybrid data-driven convex hull approach”, in *2024 IEEE Conference on Energy Internet and Energy System Integration*, Shenyang, 2024, pp. 1–6.

B. Submitted International Publications

[1] **Z. Liang**, C. Y. Chung, X. Yin, Q. Wang, and J. Zhu, “Efficient energy management of hybrid microgrids via data-correlated uncertainty modeling and semi-convex relaxation”, *IEEE Trans. Ind. Inform.*, Under Review.

[2] **Z. Liang**, C. Y. Chung, X. Yin, X. Fu, C. Li, and J. Zhu, “Nilm-aided robust energy management of hybrid microgrids via dual-task learning”, *IEEE Trans. Smart Grid*, Under Review.

Acknowledgements

My heartfelt thanks go to everyone who helped, encouraged, and accompanied me during my three years of Ph.D. study at Hong Kong PolyU.

First and foremost, I wish to express my most profound and sincere gratitude to my supervisor, Professor C.Y. Chung. His invaluable guidance, encouragement, and support have been the cornerstones of my doctoral studies. I feel extremely fortunate to be his first Ph.D. student after his return to Hong Kong in 2022, and to have witnessed the remarkable growth of the research team, from just a few members to a vibrant group of dozens of researchers within a short span of two years. Under his mentorship, I had the privilege of participating in and completing his first research project on smart meter data analysis, a responsibility for which I am immensely grateful, and of collaborating with team members to prepare research proposals. These experiences have been immensely inspiring and rewarding, and I have learned greatly from his vision, leadership, and dedication.

I would like to extend my sincere appreciation to Dr. Qin Wang. His insights on valuable practical experiences in power system operation and planning greatly broadened my perspective and deepened my understanding of the field. I am sincerely thankful to all members of our research team for their friendship, collaboration, and help during my Ph.D. journey. In particular, I would like to acknowledge Dr. Junwei Liu, Dr. Jinpeng Tian, Dr. Zilin Li, Dr. Wenjie Zhang, Dr. Wei Lin, Dr. Chongyu Wang, Dr. Wei Qiu, Dr. Shibo Chen, Dr. Suhan Zhang, Dr. Yingping Cao, Dr. Haosen Yang, Dr. Hong Tan, Dr. Song Ke, Dr. Xin Yin, Dr. Yi Yu, Dr. Ziqiang Wang, Dr. Kui Zhang, Dr. Ruohan Guo, Dr. Dandan Peng, Mr. Shangyang He, Mr. Hanjiang Dong, Mr. Junkai Li, Ms. Nancy Li, and many other colleagues who have made my doctoral life intellectually stimulating and enjoyable.

My heartfelt thanks also go to my parents for their unconditional love and support, and to my girlfriend for her constant encouragement and understanding. They have always been my strongest source of strength and motivation, standing behind me in every stage of life. Without their sacrifices, encouragement, and belief in me, this journey would not have been possible.

I gratefully acknowledge the generous financial support from The Hong Kong Polytechnic University through the three-year Presidential PhD Fellowship Scheme.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Background

The global energy landscape is undergoing a fundamental transformation, shifting from a centralized, fossil-fuel-based model to a decentralized and decarbonized system. This change is driven by the need to combat climate change, the declining costs of renewable energy, and a growing demand for a more resilient power supply. Microgrids, localized electrical grids that can operate either autonomously or connected to the main utility grid, are central to this transition [1]. They are transforming the traditional one-way flow of electricity into a dynamic and intelligent bidirectional network. There are three primary types of microgrids, distinguished by their use of AC and DC: AC microgrids, DC microgrids, and hybrid AC/DC microgrids (HMGs).

1.1.1 AC Microgrids

AC microgrids offer significant advantages in terms of infrastructure compatibility and technological maturity [2]. Their seamless integration with existing AC power distribution systems eliminates the need for extensive infrastructure modifications, allowing utilities and consumers to leverage existing transformers, switchgear, distribution lines, and electrical devices. This compatibility significantly reduces implementation costs and deployment time while minimizing power supply disruptions during the transition phase. Additionally, the prevalence of AC systems ensures a larger trained

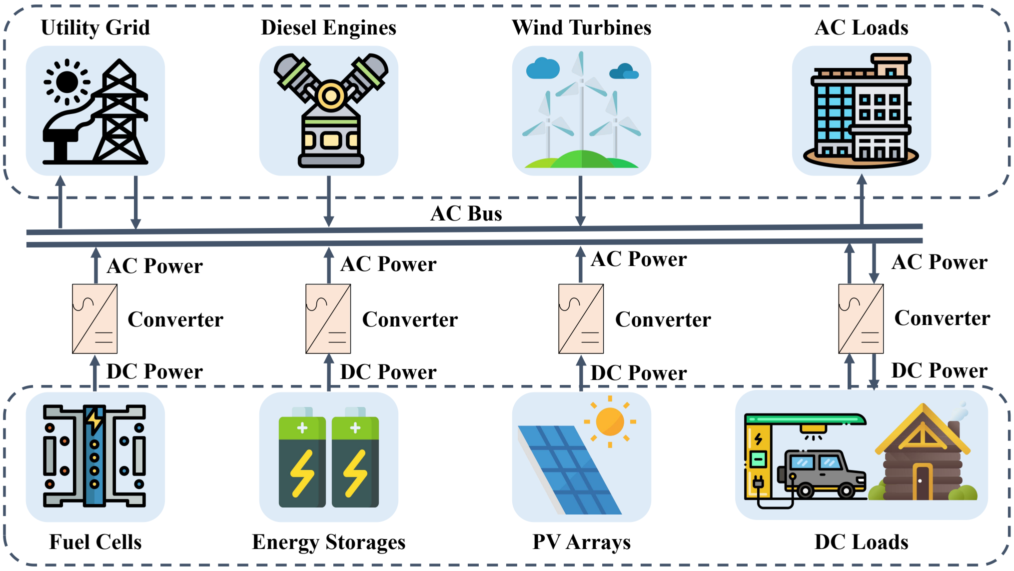


Figure 1.1: Typical AC microgrid structure

workforce for maintenance and repair operations, contributing to improved system serviceability and operational reliability [3].

The technological advantages of AC microgrids stem from over a century of electrical system development and standardization. The mature AC market provides a vast selection of distributed energy resources (DER) technologies, including renewable energy sources like wind turbines and hydroelectric generators that can be easily integrated due to inherent AC compatibility. Furthermore, AC microgrids benefit from well-established standards and protocols developed by organizations such as the IEEE, IEC, and NEC, ensuring regulatory compliance, equipment interoperability across different manufacturers, and adherence to proven safety and reliability guidelines [4]. These established frameworks enable diverse microgrid components, such as generators, inverters, energy storage systems, and load control devices, to operate as a cohesive system while maintaining high safety standards and operational dependability [5].

Despite their numerous benefits, AC microgrids face several operational challenges that can impact system efficiency and complexity. One primary disadvantage is the potential for higher energy losses during the conversion process when integrating DC sources such as solar panels [6]. Since AC microgrids operate on alternating current while many renewable energy systems generate direct current, power electronic devices like inverters are required to convert DC to AC power. This conversion process, along with associated components such as transformers for voltage level adjustments, introduces energy losses that can reduce the overall system efficiency and result in lower net energy output compared to purely DC-based systems. Although conversion technology efficiency has improved significantly over the years, these losses remain a consideration in system design and operation.

The complexity of managing power quality and protection presents another significant challenge for AC microgrids. The integration of diverse energy sources with varying output characteristics, including different voltage and frequency profiles, makes maintaining stable power quality across the microgrid particularly demanding [7]. Power quality management becomes increasingly complex as each energy source may exhibit unique operational behaviors that must be harmonized to ensure consistent voltage and frequency levels throughout the system. Additionally, protection coordination requires sophisticated control systems and specialized equipment to detect and respond to faults such as short circuits or voltage fluctuations in a timely manner. This complexity necessitates advanced control and monitoring systems, adding to the overall system design and operational requirements. However, ongoing advancements in power electronics and control technologies continue to address these challenges, improving conversion efficiency and simplifying power quality and protection management in AC microgrids [8].

1.1.2 DC Microgrids

DC microgrids operate exclusively on direct current, offering significant advantages for modern energy applications. DC microgrids offer superior efficiency and operational simplicity by eliminating unnecessary power conversion processes that plague traditional AC systems [9]. Since many DERs such as solar panels and battery storage systems inherently generate or store energy in DC form, DC microgrids can maintain this power in its native format from generation through consumption, avoiding the energy losses typically associated with DC-to-AC conversion equipment like inverters. This elimination of conversion steps not only significantly increases overall system efficiency but also results in a simpler and potentially more reliable design with fewer components and potential failure points. The reduced system complexity translates to easier maintenance, lower associated

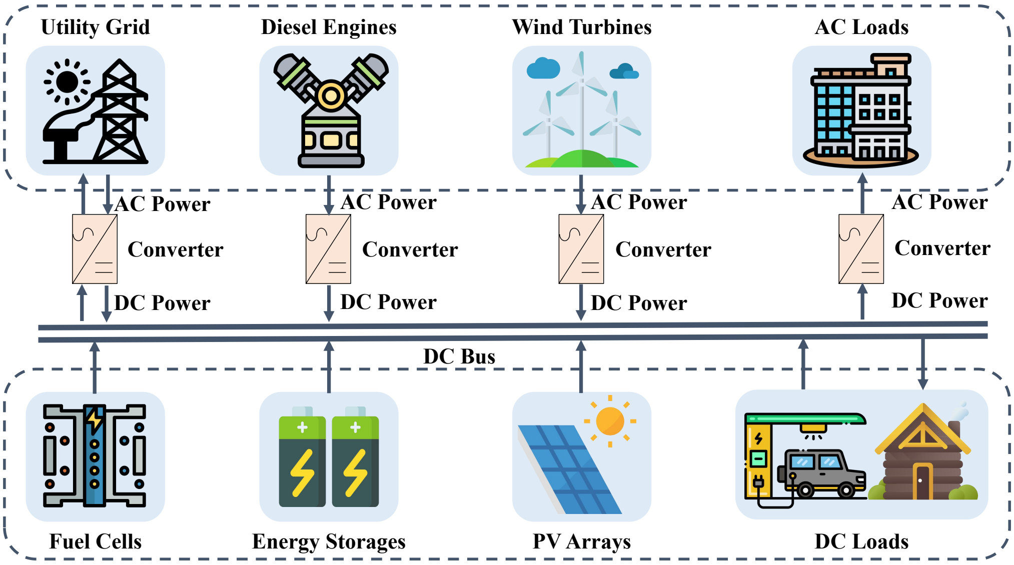


Figure 1.2: Typical DC microgrid structure

costs, and improved power quality by avoiding AC-specific issues such as phase imbalances and frequency variations that can compromise system stability [10].

The compatibility of DC microgrids with modern electrical infrastructure and emerging technologies provides additional operational advantages. Contemporary electronic devices including computers, LED lighting systems, and electric vehicles operate internally on DC power, requiring costly and inefficient AC-to-DC conversion when connected to traditional AC grids. DC microgrids can directly supply these devices with native DC power, optimizing energy utilization and eliminating conversion losses throughout the system. Furthermore, the seamless integration capabilities with renewable energy sources like photovoltaic solar panels and energy storage technologies such as lithium-ion batteries make DC microgrids particularly well-suited for localized applications including residential complexes, commercial buildings, and remote communities with high penetration of DC-based systems and renewable energy sources [11]. This alignment with modern energy demands and the increasing proliferation of DC-based technologies positions DC microgrids as an efficient and future-oriented solution for distributed energy management.

DC microgrids face significant market and infrastructure limitations that stem from over a century of AC power system dominance. The global electrical infrastructure has been predominantly designed around AC systems since the early developments by Tesla and Edison, resulting in a vast, well-established market for AC devices with extensive manufacturing and support infrastructure. In contrast, the DC device market remains relatively small and underdeveloped, leading to limited variety and availability of DC-compatible appliances and industrial equipment. This market disparity creates practical challenges including higher costs for DC equipment procurement, reduced availability of maintenance services, limited spare parts supply, and fewer options for system designers. The smaller market ecosystem makes sourcing appropriate DC equipment more difficult and potentially more expensive, which can significantly impact the economic viability of DC microgrid implementations [12].

The absence of widely adopted standards and protocols for DC distribution systems presents another critical disadvantage that complicates DC microgrid deployment and operation. While AC grids benefit from well-established standards developed over more than a century, ensuring equipment interoperability, safety, and reliability, DC microgrids lack similar comprehensive standardization frameworks [13]. This deficiency results in inconsistencies across voltage levels, connector types, and communication protocols, leading to compatibility issues between equipment from different manufacturers and complicating system design, installation, and future expansion efforts. Without universally accepted standards, each DC microgrid often requires custom engineering solutions, increasing both costs and system complexity while potentially compromising safety and reliability assurance. Overcoming these disadvantages will require coordinated efforts from industry stakeholders and regulatory bodies to develop comprehensive DC power standards and expand the market for DC-compatible devices, ultimately creating the standardized ecosystem necessary for widespread DC microgrid adoption [14].

1.1.3 Hybrid AC/DC Microgrids

HMGs incorporate both AC and DC power systems to leverage the advantages of both types of current [15]. They are designed to optimize the performance and integration of different DERs and loads that are native to AC or DC, using power electronic converters to manage the flow of energy between the two systems [16]. This architecture is emerging as a strategically vital solution for future energy systems.

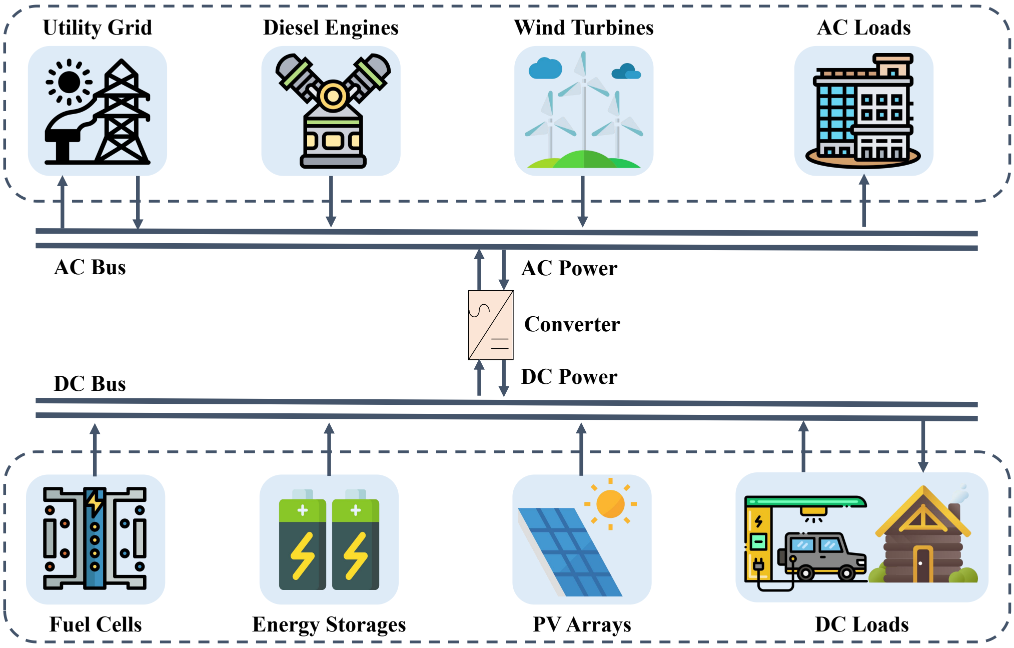


Figure 1.3: Typical AC/DC microgrid structure

HMGs offer superior flexibility, allowing for the direct and efficient integration of any type of distributed energy resource or load, whether it is AC- or DC-native. By creating parallel AC and DC buses, HMGs improve overall system efficiency by minimizing the need for power conversions [17]. For example, power from a DC solar panel can directly supply a DC electric vehicle charger, bypassing the conversion losses found in a pure AC system. This structure also enhances reliability, as a fault on one bus does not necessarily disrupt the other. However, the sophistication of HMGs leads to challenges. They require more complex control and management systems to handle power flow between the two buses and ensure stability. This complexity, along with the need for additional power electronics like bi-directional converters, results in higher initial costs compared to single-current microgrids [18]. HMGs provide exceptional flexibility and operational efficiency by accommodating diverse energy resources and load types within a single integrated system. This architecture allows for seamless integration of various DERs regardless of their native power format, enabling solar panels (DC) and wind turbines (AC) to coexist within the same grid without requiring standardization to a single current type. The system extends this flexibility to loads, supporting both traditional AC appliances and modern DC-based electronics and machinery, thereby acknowledging current AC infrastructure prevalence while embracing the growing availability of DC generation and consumption technologies. This design approach significantly improves system efficiency by enabling power to be distributed in its native form to appropriate loads, reducing unnecessary conversions and associated energy losses [19]. For example, DC power from solar panels can directly supply DC loads or charge battery storage systems without AC conversion, while AC power is converted only when specifically required by traditional loads or utility grid interactions.

The advanced control capabilities of hybrid microgrids enable sophisticated protection and isolation strategies that enhance system resilience and reliability [20]. These systems can intelligently manage both AC and DC networks, providing the ability to isolate different sections during faults or maintenance operations, allowing the DC network to continue operating independently during AC network issues and vice versa. This isolation capability proves especially valuable in critical applications requiring uninterrupted power supply and can be strategically employed during emergencies or peak demand periods. During grid outages, hybrid microgrids can implement islanding operations to isolate from the main grid while continuing to supply power to designated critical loads, demonstrating adaptive control that maintains operation even when portions of the microgrid are compromised. This comprehensive approach to modern energy management makes HMGs particularly suitable for complex energy systems that prioritize performance optimization while maintaining future-proof adaptability against evolving technology trends.

1.2 Literature Review

1.2.1 Bi-directional Converter Modeling in Hybrid AC/DC Microgrids

The integration of renewable energy sources into electric power systems is undergoing a remarkable transformation, facilitated by the utilization of emerging power electronics-based devices, such as BDCs [21]. These widely adopted BDCs have become integral to power communication in HMGs, profoundly influencing power communication schemes and generation strategies. Hence, it is crucial to accurately model the operation characteristics of BDCs in order to determine effective economic dispatch strategies for these MGs. The modeling of BDCs significantly impacts power communication schemes between two sub-MGs in HMGs and the generation strategy of generators within the

Table 1.1: Comparisons of different microgrids

Microgrids	**Advantages**	**Disadvantages**
AC Microgrids	• Compatibility with existing infrastructure • Wide selection of AC-based DER technologies and equipment • Established standards and protocols for operation	• Potential for higher losses in the conversion process for DC sources • Complexity in managing power quality and protection
DC Microgrids	• Higher efficiency for DC-based DERs due to elimination of conversion losses • Simpler and potentially more reliable due to fewer conversion steps • Better suited for integration with electronics, electric vehicles, and modern DC energy sources	• Smaller market for DC devices and less infrastructure than AC • Lack of widely adopted standards and protocols for DC distribution
Hybrid AC/DC Microgrids	• Flexibility in integrating various types of DERs and loads • Improved efficiency by reducing unnecessary conversions • Ability to isolate and protect different parts of the microgrid	• More complex control and management systems required • Higher initial costs due to additional equipment (converters, inverters)

two sub-MGs. Thus, accurately modeling BDC operation characteristics is of utmost importance in defining economic dispatch strategies of HMGs within the solution time requirements for practical applications.

A comprehensive model for the utilization of BDCs in HMG economic dispatch consists of two crucial elements: power conversion efficiency and power conversion direction. Regarding the first element, the power conversion efficiency model of BDCs can be broadly classified into static and dynamic conversion efficiency. More specifically, the static model encompasses both lossless conversion efficiency [22], [23] and scale-based conversion efficiency[24]. The lossless model simplifies the efficiency by assuming no energy loss during the conversion process and considers the power output of a BDC equal to the power input. However, this model overlooks the realistic energy loss during conversion. To address this unrealistic assumption, the scale-based model introduces a positive factor less than one to represent conversion loss proportional to the conversion power quantity. Although these static efficiency models can be easily implemented in HMG economic dispatch, they are often deemed overly simplistic for practical applications, as they disregard the conversion efficiency dynamics intrinsically linked with BDC utilization rates [25], [26]. This oversimplification overlooks the fact that the primary factors contributing to conversion loss in BDCs, namely conduction loss, magnetic loss, and switching loss, vary with the power magnitude of the BDCs.

In contrast to static models, dynamic models provide a more realistic representation of conversion loss. They can be classified into four types of efficiency models: full mathematical formulation based efficiency[27], high order polynomial based efficiency[28], [29], quadratic based efficiency[25], [30], and piece-wise linearization based efficiency[31]. Full mathematical formulation based efficiency accurately represents conversion loss using trigonometric function expressions, whereas high order polynomial based efficiency tries to express the BDC efficiency with high-order polynomials. However, incorporating these detailed loss terms in full mathematical formulation based and high order polynomial based models often results in a computationally challenging and occasionally unsolvable planning and operation model for HMGs. To mitigate this, Wei *et al.* reduced the polynomial order in high order polynomial based efficiency and employed quadratic functions of BDC utilization rate to express BDC efficiency [30]. Additionally, Zhao *et al.* proposed a piece-wise linearization method to linearize the full mathematical formulation based efficiency model [31], which could potentially be extended to other non-linear models such as high order polynomial based and quadratic based efficiency models. Although piece-wise linearization efficiency offers linear characteristics that accelerate the model solution process, it requires a large number of binary variables to ensure accurate linearization for non-linear efficiency models, leading to computational complexity. Although each of the four dynamic efficiency models has its strengths, none incorporates a convex approximation technique that avoids binary variables and forms a convex model. The existing non-convex efficiency models can get trapped in locally optimal solutions, and lead to computational difficulties, particularly for NMGs that include several BDCs in mixed AC/DC systems. Thus, there is a clear need to develop a convex model that accurately captures the dynamic power conversion efficiency of BDCs.

The other crucial component of the BDC model for hybrid AC/DC NMG economic dispatch is the conversion direction of BDCs, which determines whether the BDC operates in rectification mode or inversion mode [32]. It should be noted that a BDC cannot operate in both rectification mode and inversion mode simultaneously due to the distinct operations required in each mode [33]. In rectification mode, the power on the AC side of hybrid AC/DC NMGs is rectified to the power on the DC side, while in inversion mode, the DC power is inverted to AC power. Consequently, at any time, a BDC can only operate in one mode, not both [34]. To ensure this mutually exclusive operation, various measures have been implemented, including integer complementary constraints and non-linear complementary constraints[35]. The integer complementary constraints introduce binary variables to represent the operation state of rectification mode and inversion mode, ensuring that the BDC can only operate in one mode for each time slot[36]. In contrast, the non-linear complementary constraints set up a bi-linear constraint that ensures the product of rectification power and inversion power is zero for each BDC, preventing simultaneous rectification and inversion behavior[37]. Incorporating either of these measures into the operation models of hybrid AC/DC NMGs can yield a solution with non-simultaneously rectifying and inverting behaviors of BDCs. However, adding these constraints to operation models of HMGs results in a non-convex optimization problem. Such problems are computationally intensive and often lead to non-convergence of algorithms.

This issue is particularly acute in the widely used two-stage robust or stochastic operation models of hybrid AC/DC NMGs, which require dualizing the inner-level sub-model. According to the strong duality theorem, for a robust operation model, its inner-level sub-model must be a convex optimization problem to ensure zero duality gap between the primal and dual problems. Unfortunately, the current integer or non-linear complementary constraints do not ensure this zero duality gap, creating challenges for solving these models with typical C&CG algorithms[38] or Benders decomposition algorithms [39]. Although a nested C&CG algorithm has been proposed for robust models with mixed integer programming-based inner-level sub-models[40], [41], it is computationally burdensome and does not guarantee convergence due to the non-zero duality gap.

Although a variety of algorithms have been developed to tackle optimization models with non-convex complementary constraints, these non-convex models generally incur high computational costs and result in the non-convergence of algorithms. A straightforward solution is to remove these non-convex constraints to simplify the models into convex forms, thereby enhancing computational efficiency. However, omitting these constraints may result in a solution different from that of the original models incorporating non-convex constraints. Thus, there is a compelling need to identify the conditions that can ensure the non-simultaneous rectifying and inverting behaviors of BDCs. Under such conditions, simplified models that omit the non-convex complementary constraints can still yield results identical to those of the original non-convex models.

To the best of the authors' knowledge, no previous research has investigated the sufficient conditions to formulate a convex model of BDCs for the economic dispatch of hybrid AC/DC NMGs.

1.2.2 Smart Meter Analytics for Load Modeling

The continuing increase in global population and economic developments world-wide have significantly increased the energy consumption of residential and commercial sectors [17]. This issue is particularly prominent in high-growth regions such as the city of Hong Kong, where residential and commercial buildings account for over 94% of total electricity consumption [42]. Thus, efforts to encourage individuals to adopt more energy-efficient lifestyles are becoming increasingly crucial. One important means of assisting these efforts is to analyze the detailed electricity consumption characteristics of populations via load monitoring activities. Load monitoring has many other applications as well, including facilitating improvements in the accuracy of residential load forecasting, evaluating demand responses, and identifying unauthorized electric loads [43], [44]. Here, nonintrusive load monitoring (NILM) is one of the most widely used load monitoring methods owing to its ability to track the energy consumption of individual appliances in a household based solely on the data collected by a smart meter installed at the main electrical panel, without any direct physical connection with these individual appliances [45]. Currently, NILM approaches can be categorized primarily as combination optimization (CO), hidden Markov model (HMM), and deep learning model (DLM) approaches.

Approaches based on CO formulate the NILM process into an optimization problem by finding the best combination of individual appliance states that minimizes the difference between the actual aggregated power consumption and the total power consumption, which is derived by summing the predicted power values of all individual appliances. Various optimization models have been employed depending on the variables and objectives used in the CO process, including multi-objective programming [46], mixed-integer nonlinear programming [47], and mixed-integer linear programming [48], [49]. Most NILM-based optimization models represent nondeterministic polynomial (NP) hard problems for which finding an optimal solution is computationally challenging and the required computation time typically increases exponentially as the size of the optimization model increases. Although a number of algorithms have been developed to simplify the solution of these optimization problems, such as piece-wise linearization algorithms, branch-and-cutting algorithms, and heuristic algorithms [50], the resulting computational burden still limits the use of CO approaches to practical applications. Moreover, CO approaches are unable to model the time-series characteristics of power consumption, such as electricity consumption patterns and state transitions for appliances that have distinct operational states.

In contrast to the CO process, HMMs perform well with time-series data[51]. Moreover, several HMM variants have been proposed to enhance the ability of HMMs to model the power consumption patterns of appliances, such as factorial HMM[52], additive factorial HMM[53], hierarchical HMM[54], and hidden semi-Markov model[55]. However, the number of state configurations that must be incorporated into an HMM can expand rapidly with an increasing number of appliances, making the model complex and difficult to manage. In addition, currently available HMMs cannot account effectively for long-term dependencies in time-series data.

Compared to CO processes and HMMs, a DLM represents a network system serving as an expert learner that can often extract complex and nonlinear patterns in time-series data more accurately by learning the complex feature representations of the power consumption patterns of appliances from smart-meter data automatically. Currently existing DLM-based NILM expert approaches often adopt single-task deep learning techniques that focus solely on either conducting appliance state classification of on and off states or estimating power consumption via regression analysis, and include non-temporal and temporal DLMs. Non-temporal models focus on learning patterns from aggregated power consumption data without considering the sequential dependence of data over time. For example, García-Pérez *et al.* [56] designed a fully connected deep neural network (DNN) architecture to estimate the power consumption of each of the individual appliances in hospital buildings. Schirmer and Mporas [57] performed appliance recognition and energy disaggregation by extracting local features from high sampling frequency power consumption data using the convolutional and pooling operations in convolutional neural networks (CNNs). The lack of temporal awareness in non-temporal DLMs, such as DNNs [56] and CNNs [57], makes these models highly suitable for capturing high-level features such as appliance signatures, where appliance states can be accurately inferred from static, aggregated snapshots of appliance data. However, these models may face challenges in identifying the states of appliances whose behavior depends on sequential usage (e.g., the multiple cycles of a washing machine). This issue is addressed by temporal DLMs, making them highly suitable for NILM tasks involving time-dependent appliance states. For example, Yuan *et al.* [58] improved the load identification performance of DLMs by addressing limitations in the extraction of temporal dependencies in time-series data using gated recurrent unit (GRU) networks. Similarly, Hwang and Kang *et al.* [59] captured time-series characteristics for energy disaggregation using long short-term memory (LSTM) networks with feedback mechanisms. Kaselimi *et al.* [60] obtained enhanced characteristic extraction performance using bidirectional LSTM (BiLSTM) networks. Zhang *et al.* [61] captured long-range dependencies in time-series data by leveraging the

Table 1.2: Comparisons and summary of related works

Framework	Area	Heterogeneity Nature	Weighting Optimization
HMoE [66], [67]	NLP	By Size	×
AutoMoE [68], [69]	NLP	By Size & Placement	×
Our Work	NILM	By Architecture	✓

causal and dilated convolutions in a temporal convolutional network (TCN).

However, single-task DLM-based expert methods can obtain suboptimal characteristic extraction performance by neglecting the inherent relationships and complementary information between the appliance state classification and power consumption estimation tasks. For instance, accurate state identification can provide valuable constraints for power estimation, while precise power estimates can help validate state classifications. This has led to the development of integrated multitask frameworks that can jointly optimize these interdependent objectives. For example, Chen *et al.* [62] applied a single shared CNN expert for extracting the features of both tasks. Furthermore, Liu *et al.* [63] extended this shared framework using a TCN to capture time-series characteristics in the NILM process. Similarly, Dash *et al.* [64] extended the shared framework further by applying a CNN-BiLSTM-based architecture for estimating the power consumption and operational states of multiple appliances.

The concept of leveraging multiple experts within a single model has been formalized through the mixture-of-experts (MoE) architecture. While early multitask learning frameworks in NILM often relied on a single shared feature extractor, developments in the broader machine learning field have demonstrated the advantages of more flexible MoE designs [65]. In particular, the use of heterogeneous experts, where individual experts differ in model capacity or placement, has shown significant promise in domains such as natural language processing (NLP). For instance, some studies have introduced heterogeneity in expert sizes to better align model capacity with the complexity of input data [66], as exemplified by HMoE [67]. Others, such as AutoMoE [68], have employed neural architecture search to automate the construction of heterogeneous MoE frameworks [69].

These approaches collectively support the principle that architectural diversity among experts can enhance both performance and efficiency. However, the application of such heterogeneous expert architectures to the unique feature extraction challenges in low-frequency NILM remains relatively unexplored. As summarized in Table 1.2, existing NILM methods, including multitask variants, have predominantly employed homogeneous expert designs, which may be inadequate for capturing the wide variability present in aggregated power consumption data. Specifically, the existing multitask learning frameworks developed for NILM suffer from three primary deficiencies. 1) Existing approaches predominantly rely on a single shared network architecture to extract features for both appliance state classification and power consumption estimation tasks. However, these single architectures are limited in their ability to capture the diverse and complex characteristics of NILM data fully. This limitation becomes even more pronounced in real-world scenarios where the long sampling intervals on the order of 15 min typically employed by smart meters reduce the richness of appliance signatures, making it difficult for single-expert DLMs to achieve accurate disaggregation. 2) Existing NILM approaches tend to rely on homogeneous neural network architectures, such as CNN or LSTM architectures. Nonetheless, these different network architectures have very different strengths, where CNNs are adept at extracting local spatial and temporal features, such as the short-term, periodic patterns of some appliances, and LSTM networks excel at modeling sequential dependencies, such as the long-term dependencies of heating systems or refrigerators. Hence, neither architecture alone is sufficient to address the diverse characteristics of NILM data. 3) Current multitask learning techniques often rely on pre-determined or manually tuned fixed weights to balance the competing objectives of appliance state classification and power consumption estimation (i.e., regression) operations. However, the importance of each objective may vary in practice for different appliances, depending on the complexity of their power usage patterns. For example, an appliance with highly variable behavior may require greater emphasis on regression during training, whereas a simple binary on/off appliance may prioritize classification. Hence, the failure of static weight methods to adapt to these dynamic requirements can lead to suboptimal optimization for one or both tasks. Moreover, this lack of adaptability further limits the performance of multitask learning frameworks in NILM by preventing current methods from fully exploiting the interdependence between classification and regression tasks. Finally, though a considerable quantity of smart meter data is collected and produced everyday, how to use these data for flexible microgrid operation has not been studied yet.

1.2.3 Uncertainty Sets for Renewable Energy Modeling

The integration of RESs into electric power grids has become crucial to facilitate the global trend toward sustainable energy practices. The development of HMGs composed of ACsMs and DCsMs with power flow between the two systems facilitated via BDCs is pivotal in this transition [70]. In particular, such HMG infrastructure can accommodate both AC and DC RES outputs (e.g., from wind and solar energy resources) with maximum flexibility [17]. However, the energy management of these HMG systems suffers from significant challenges under high penetration of converter-interfaced RES units. One of the foremost challenges stems from the generation characteristics of RES units, which are highly dependent on weather conditions and thus difficult to predict accurately [71], [72]. The inherent uncertainty and intermittency of RES generation lead to frequent power fluctuations and mismatches between electricity supply and demand, particularly when deterministic energy management approaches are used. These fluctuations not only disrupt the balance of supply and demand but also exacerbate voltage stability issues within the microgrid, especially during periods of rapid changes in RES output. As the penetration of RES units increases, these challenges become more pronounced, requiring the application of uncertainty-based strategies to handle the stochastic nature of RES generation and ensure the robust and reliable operation of HMG systems.

Additionally, energy management is further complicated by the non-convex characteristics of BDCs. These converters play a critical role in facilitating energy exchange between different components of the microgrid, but their operation introduces unique modeling challenges. Specifically, non-convex constraints must be incorporated into HMG energy management models to ensure that a BDC cannot operate simultaneously in inverter mode and rectifier mode [73]. These non-convex constraints not only increase the complexity of the energy management models but also lead to convergence issues and a significant computational burden for the algorithms used to solve them.

Currently, uncertainty-based strategies can be broadly categorized into scenario-based stochastic energy management[17], chance-constrained energy management [74], and robust energy management based on uncertainty sets [75]. Here, robust energy management offers a number of advantages compared with other uncertainty optimization strategies. For example, robust optimization does not require precise knowledge of the probability distributions of RES outputs, as is the case with stochastic methods, and the optimal management strategies obtained are guaranteed to be robust against all possible fluctuations in RES outputs included within the constructed uncertainty set. Therefore, the construction of uncertainty sets directly affects the conservativeness of the obtained energy management strategies [76]. The most widely used types of US include BbUS, PUS, EUS, and CHUS, which offer various tradeoffs in terms of the conservativeness of the resulting solutions and their computational tractability depending on their shape and method of construction.

Valencia *et al.*[77] constructed a BbUS using the upper and lower bounds of RES generation outputs independently, and then Xiang *et al.*[78] reduced the extreme scenarios in the BbUS using Taguchi's orthogonal array testing method. Because the construction of a BbUS requires only the bounds of RES unit outputs, a BbUS presents an easily constructed linear and hyper-rectangular form [79], [80]. However, BbUSs tend to produce overly conservative strategies because they include the most severe RES generation scenarios, including those with vanishingly rare probabilities, such as where all RES units experience worst-case scenarios simultaneously. Attempts to reduce the conservativeness of robust optimization solutions without introducing excessive complexity include the development of distributionally robust optimization (DRO) methods. These methods obtain optimized planning solutions under the worst-case distributions within distribution ambiguity sets [81]. The distribution ambiguity sets employed include continuous distribution uncertainty sets based on the Wasserstein or Kullback–Leibler metric[82], and discrete distribution uncertainty sets[83]. However, continuous distribution uncertainty sets can lead to optimization problems that are computationally intensive, especially when the size of the problem increases. Moreover, while discrete methods generate a robust solution under the worst-case probability distribution of predicted RES outputs, they fail to yield a solution that is robust against the actual worst-case scenario of RES output. In comparison, a PUS presents a polyhedral shape because it is constructed under a constraint specifying that the total deviation of all RES outputs from their nominal values cannot exceed a predefined budget Γ established according to operator experience [84]. Studies have extended PUSs to manage RES uncertainty in multi-energy MGs [85] and ship-based HMGs [86]. Zheng *et al.* [41] proposed a novel robust operational model for energy storage systems, incorporating a non-anticipativity constraint and representing uncertainties through PUSs. This model was solved using a relaxed value function-based algorithm. The application of Γ ensures that PUSs produce solutions that are less conservative than BbUSs owing to the incorporation of more realistic RES output scenarios [76]. However, the determination of Γ is overly subjective. As a result, the PUS cannot always guarantee a robust solution against RES uncertainty, particularly when selecting a relatively low value of Γ (e.g., $\Gamma \leq 0.25$). Moreover, a key assumption employed in both a BbUS and PUS is that the outputs of all RES units are truly independent, and their variations are uniformly likely across any time interval. Nonetheless, inherently strong correlations have been observed between RES outputs within MGs [87], [88].

It is important to note that RESs in an HMG are typically located within a relatively small geographical area [89]. Consequently, they are subject to similar local weather conditions, seasonal patterns, and climatic variations. This proximity results in strong correlations among RES generation within an HMG, which can be more accurately modeled using correlations of RES outputs. In contrast, correlations among RES outputs can be captured explicitly by the ellipsoid shape of an EUS and the convex polytope shape of a CHUS [90]. El-Meligy *et al.* [91] proposed a data-driven approach to determine the appropriate parameters for the ellipsoid, formulating a minimal volume EUS that encompasses all measured RES output data. Accordingly, the resulting management solutions are generally less conservative than those produced by a BbUS or PUS[92]. In general, an EUS is defined as the set of all historical data that satisfy a quadratic form [93]. However, this introduces non-linear expressions in the EUS that make the solution of energy management problems more computationally intensive, particularly owing to the generation of quadratic programming problems[94].

This issue has been addressed by constructing a convex hull of historical RES data, which is the smallest convex polygon that contains all the data. Therefore, a CHUS is constructed using the outermost points in the dataset, and thereby represents underlying correlations in RES output uncertainties with a linear form[95], [96]. Chen *et al.*[97] developed a CHUS method to accurately depict the security region for integrated energy systems, while Zhong *et al.*[98] reduced the conservativeness of robust strategies in multi-energy MGs by extending the standard CHUS involving a single time interval to multi-interval CHUS. However, existing applications of CHUSs become increasingly computationally intensive as the number of RES units increases because the number of

CHUS vertices increases exponentially with increasing RES number. This can lead to an exponential increase in the number of constraints and a computationally intractable model. Qiu *et al.*[99] addressed this issue to some extent by approximating the CHUS with linear constraints. Compared with the BbUS, PUS, and EUS, a CHUS can minimize the size of the uncertainty set while encompassing all RES output data within a given set of data. Therefore, a CHUS can provide fully robust energy management strategies with the least conservativeness. Nonetheless, the size of the CHUS is currently determined solely by the outermost points of historical RES output data, which can lead to highly uneconomical strategies if these points are actually outliers obtained due to measurement error. While Zhou *et al.* [100] proposed a diamond-cut CHUS method to approximate the original CHUS with fewer vertices and enhance computational efficiency, the dimensionality of the CHUS space still expands exponentially as the number of RES units grows.

The BDC operational constraints applied in HMG energy management models include mixed integer constraints and nonlinear constraints. For example, Qiu *et al.*[84] devised mixed integer constraints by incorporating binary variables to represent the inverter and rectifier operation modes of BDCs, while Xu *et al.*[101] achieved the same outcome using a bilinear form without introducing additional binary variables. However, both mixed integer constraints and nonlinear constraints are non-convex. Consequently, robust optimization algorithms reliant on classical strong duality theory, such as Benders decomposition algorithms [102] and C&CG algorithms [103], which involve the iterative solution of upper and lower level problems with excellent efficiency, cannot be applied directly in the solution process because they cannot consistently achieve convergence. This issue has been mitigated by integrating non-convex complementary constraints into the upper-level model only, where strong duality theorem is not a requisite for the solution process, but excluding these constraints from the lower-level problem where satisfying this theorem is necessary [104], [105]. However, non-simultaneous inverter and rectifier operations cannot be ensured for BDCs in the lower-level model. Therefore, the final solution may not be physically realizable.

The above discussion demonstrated the presence of a number of issues in past research that have not been appropriately addressed for HMGs with BDCs and high RES penetration. **First**, while the above discussion clearly illustrated the advantages of the CHUS, existing research exhibits a lack of clear guidance for creating a computationally efficient uncertainty set that enables the accurate control of solution conservativeness while minimizing the negative impacts of possible outliers in the outermost points of RES datasets and reducing the computational load associated with the number of RES units. Furthermore, when the number of RES units exceeds 10, neither the Gift Wrapping algorithm [106] nor the QuickHull algorithm [107] is capable of computing CHUS with this quantity of RES units. The question of how to formulate a modified CHUS to enable computational tractability has not been addressed. **Second**, the application of non-convex BDC constraints in current research suffers from contrasting deficiencies, where the constraints are either rigorously incorporated into HMG energy management models [84], [101] or completely omitted [104], [105]. As discussed above, the former approach often results in non-convergence and a significant computational burden, while the latter approach cannot guarantee non-simultaneous inverter and rectifier operations in BDCs. In this regard, we note that the likelihood of violating the strong duality theorem has been demonstrated to decrease as the number of non-convex elements in a problem decreases[108]. However, to the best of our knowledge, no effort has been made to improve the convergence of solution algorithms reliant on strong duality theory when used in conjunction with non-convex BDC operational constraints by reducing the number of constraints down to a minimal set in HMG models. **Third**, the greater the number of vertices in the uncertainty set, the more likely the current column-and-constraint generation (C&CG) algorithms are to converge to a local optimum [109]. As mentioned before, the number of vertices in an uncertainty set, especially CHUS, grows exponentially with the number of RES units. Finding effective strategies to guide C&CG algorithms out of local optima remains an unresolved and significant challenge in the field. To encapsulate these points, a comparison is included in Table 1.3, which contrasts the features of our approach against those of existing methods.

1.3 Research Contributions

This thesis makes a series of original contributions that advance both the theoretical modeling and practical operation of HMGs with high penetration of renewable energy. These contributions span

Table 1.3: Comparison summary of existing studies

Ref.	BdC efficiency models	Convexity of BdC efficiency	Convexity of BdC direction	Uncertainty set type	Conservativeness controllability	Vertex controllability of uncertainty sets	Parallelizability of C&CG algorithms
[110], [111]	×	×	×	×	×	×	×
[112]	Static	✓	×	BcUS	✓	✓	×
[113]	Static	✓	×	×	×	×	×
[114]	Dynamic	×	×	×	×	×	×
[31]	Dynamic	×	×	×	×	×	×
[115]	Static	✓	×	BcUS	✓	✓	×
[116]	×	×	×	BcUS	✓	✓	×
[117]	×	×	×	BbUS	×	✓	×
[95]	×	×	×	CHUS	✓	×	×
[81]	×	×	×	×	✓	×	×
[82]	×	×	×	×	✓	×	×
[83]	×	×	×	×	✓	×	×

three major areas, i.e., converter modeling, smart meter analytics, and uncertainty-aware operation, and are summarized as follows:

1. **Development of Data-Driven Convex Models for BDCs**

   A key challenge in the operation of HMGs is that the efficiency characteristics and rectifying/inverting modes of BDCs introduce non-convexities into optimization models, often leading to computational intractability or convergence failures. This thesis proposes a convex approximation framework that integrates data-driven efficiency modeling with sufficient operational conditions to guarantee non-simultaneous rectifying/inverting behavior. The resulting convex models allow for accurate yet computationally efficient representations of BDCs, enabling scalable economic dispatch of large-scale hybrid microgrids. This is the first work to rigorously establish sufficient conditions for convex BDC modeling in hybrid networks.

2. **Heterogeneous Multi-Task Learning for Load Identification**

   To address the growing importance of demand-side flexibility, this thesis introduces a heterogeneous multi-gate mixture-of-experts (HMMOE) architecture for NILM. Unlike conventional single-task or homogeneous expert designs, the proposed framework integrates diverse neural network structures and dynamically allocates task-specific weights across multiple experts. This significantly improves both state classification accuracy and load estimation performance, particularly for low-frequency datasets that are common in real-world smart meter deployments. The approach enhances NILM's applicability for energy management in residential and commercial microgrids.

3. **Multimodal Multi-Task Load Disaggregation for Demand Response Evaluation**

   Building on the HMMOE framework, a two-task multimodal NILM approach is developed to incorporate additional contextual data such as temperature, humidity, and solar irradiance alongside smart meter signals. An analytical multi-objective optimization scheme is proposed to balance the competing goals of appliance state classification and load regression, yielding an optimal allocation of learning weights. By capturing both temporal and cross-modal dependencies, this contribution enables more robust demand response assessments, facilitating flexible scheduling of residential and commercial appliances in HMG operation.

4. **Advanced Uncertainty Set Modeling for Renewable Energy Sources**

   A series of novel uncertainty sets are proposed to overcome the limitations of conventional convex hull approaches, which often suffer from over-conservativeness, sensitivity to outliers, and high computational complexity. Specifically, the Intersected Convex Hull Uncertainty Set (ICHUS) and the Pair Convex Hull Uncertainty Set (PCHUS) are formulated to incorporate correlation structures and controllable conservativeness. Furthermore, a probability-driven data-correlated uncertainty set (PDCUS) is developed to leverage historical data distributions in constructing uncertainty bounds. These models strike a balance between robustness and cost-effectiveness, improving both the operational reliability and economic efficiency of renewable-penetrated HMGs.

5. **Customized Solution Algorithms for Robust Microgrid Operation**

   To solve the computationally demanding robust optimization problems arising from advanced uncertainty sets, the thesis designs tailored algorithmic frameworks, including: 1) A semi-convex relaxation method to handle complementarity constraints in hydrogen storage and electric vehicle scheduling. 2) A modified column-and-constraint generation (C&CG) algorithm embedded with nested decomposition strategies. 3) A quad-level optimization framework supported by tailored decomposition and initialization strategies to handle large-scale microgrids with multiple renewable clusters. These algorithms ensure tractability, convergence, and scalability, bridging the gap between theoretical models and real-world microgrid operation.

Taken together, the above contributions form a unified framework that links smart meter analytics with uncertainty-aware microgrid operation. The thesis not only advances modeling accuracy and computational efficiency but also demonstrates how data-driven techniques, advanced learning architectures, and robust optimization can be integrated to enhance the flexibility and robustness of future energy systems.

1.4 Thesis Overview

The thesis is structured in eight chapters, progressing logically from background and modeling foundations to data analytics, uncertainty set development, and advanced solution algorithms, before concluding with insights and future research directions. The layout is shown in Fig. 1.4, and summarized below:

1. **Chapter 1 – Introduction**

   This chapter introduces the motivation and background of hybrid AC/DC microgrids in the context of energy transition, outlines the challenges in converter modeling, smart meter analytics, and renewable uncertainty handling, and provides a comprehensive literature review. It concludes by highlighting the research gaps, stating the objectives, and summarizing the main contributions of the thesis.

2. **Chapter 2 – Data-Driven Convex Model for High-Efficiency Operation of Hybrid AC/DC Microgrids**

   This chapter focuses on the modeling of bi-directional converters. It develops data-driven convex efficiency approximations, establishes sufficient conditions for convexity, and proposes an iterative solution process. Case studies validate the proposed models under varying system configurations, renewable penetrations, and storage integrations, demonstrating both accuracy and scalability.

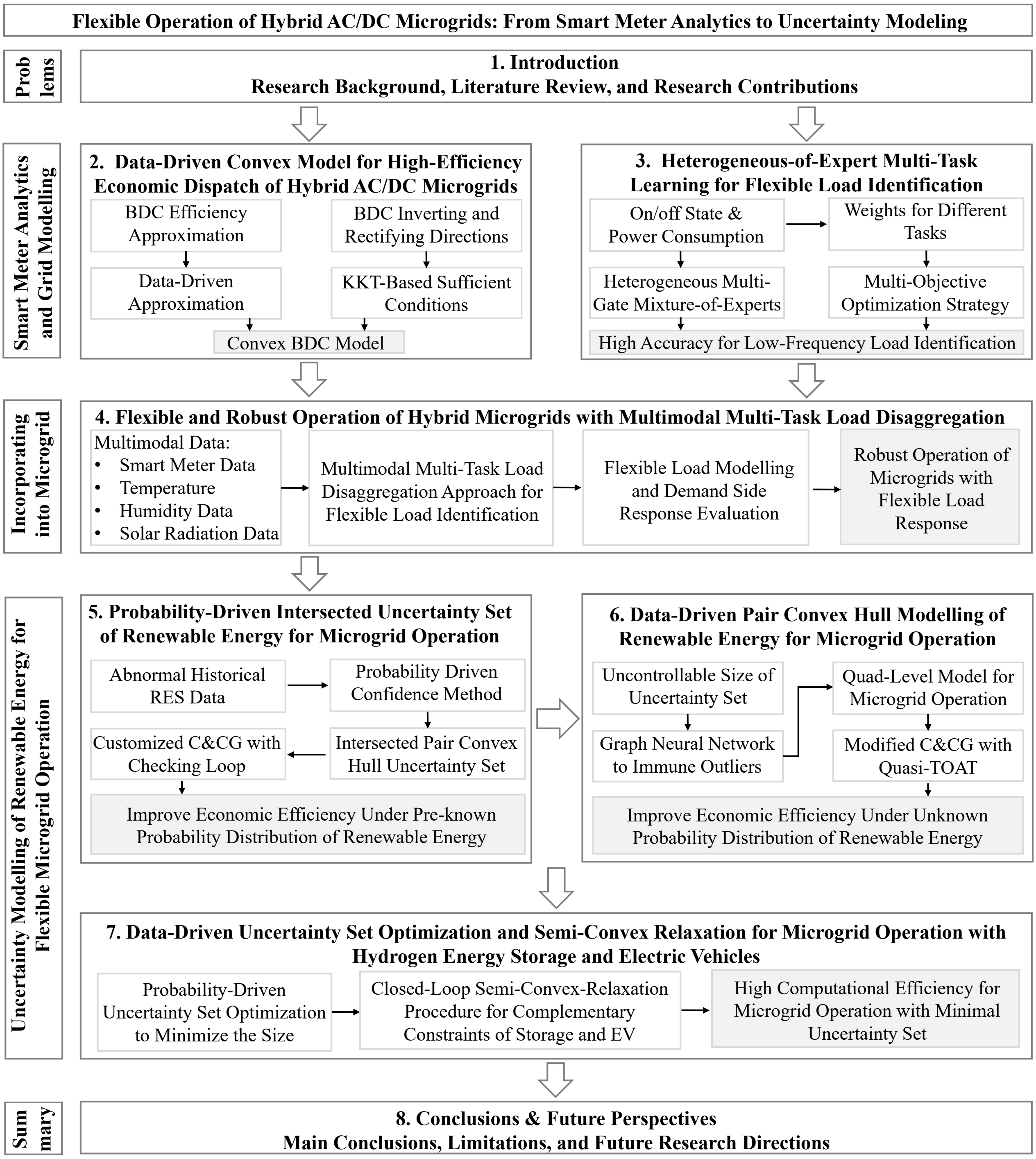


Figure 1.4: Overall framework of the thesis

3. **Chapter 3 – Heterogeneous Multi-Task Learning for Flexible Load Identification**

   Here, the thesis introduces the heterogeneous mixture-of-experts NILM framework, integrating multiple expert networks with a multi-gate mechanism. A dynamic multi-objective optimization strategy is proposed to adaptively balance classification and regression losses. Extensive experiments on public and real-world datasets showcase significant improvements over state-of-the-art NILM methods.

4. **Chapter 4 – Flexible and Robust Operation of Hybrid Microgrids with Multimodal Multi-Task Load Disaggregation**

   Building on Chapter 3, this chapter extends NILM into a multimodal learning setting, incorporating environmental and contextual data to improve load disaggregation. It further embeds the multimodal NILM outputs into a robust energy management framework for building-level HMGs. Customized decomposition and relaxation methods are applied to solve the resulting optimization model, with case studies verifying performance gains in both load monitoring and energy scheduling.

5. **Chapter 5 – Probability-Driven Intersected Uncertainty Set Modeling of Renewable Energy Sources**

   This chapter shifts the focus to uncertainty-aware microgrid operation. A probability-driven intersected convex hull framework is proposed, enabling adjustable conservativeness and correlation modeling among renewable outputs. A customized C&CG algorithm with sufficient convexity conditions is developed to ensure tractability. Simulation results highlight improvements in renewable utilization and reduced operational costs compared with existing uncertainty sets.

6. **Chapter 6 – Data-Driven Pair Convex Hull Modeling of Renewable Energy Sources**

   Chapter 6 develops the pair convex hull uncertainty set (PCHUS), supported by a confidence-level approach to mitigate outliers and balance robustness with efficiency. A quad-level energy management model and a tailored decomposition algorithm are introduced, enabling large-scale deployment in renewable-dominant microgrids. Comparative studies evaluate both solution accuracy and computational performance.

7. **Chapter 7 – Data-Driven Uncertainty Set Optimization and Semi-Convex Relaxation for Microgrid Operation with Hydrogen Energy Storage and Electric Vehicles**

   This chapter extends the framework to incorporate hydrogen energy storage systems (HESS) and electric vehicles (EVs). A probability-driven data-correlated uncertainty set is proposed, while a semi-convex relaxation method is designed to address the complementarity constraints of HESS operation. The chapter further develops a hybrid algorithmic framework to efficiently solve the complex robust dispatch problem, verified by extensive simulations with varying EV penetration levels.

8. **Chapter 8 – Conclusions and Future Perspectives**

   The final chapter summarizes the key findings and contributions of the thesis, discusses the limitations of the current work, and outlines future research directions, including real-time microgrid control, integration of sector coupling (e.g., hydrogen and transportation systems), and further applications of machine learning in uncertainty-aware decision making.

Chapter 2

Data-Driven Convex Model for High-Efficiency Operation of Hybrid AC/DC Microgrids

The use of bi-directional converters (BDCs) is crucial for enhancing power exchange in hybrid AC/DC networked microgrids (NMGs). However, the dynamic nature of their conversion efficiency and the non-convex conversion direction expression of BDC models result in a highly non-convex programming problem, which leads to significant computational challenges. This paper proposes a least squares approximation approach to simplify the complex trigonometric function that characterizes the dynamic changes in BDC conversion efficiency with power. The simplified expression transforms the original non-convex relationship into a computationally efficient convex form. Subsequently, we systematically explore and validate the sufficient conditions to achieve non-simultaneous rectification and inversion behaviors of BDCs in a convex form. These explored conditions are further customized and extended to various practical application scenarios. Case studies are conducted on a hybrid AC/DC NMG with 66 nodes. The results demonstrate that our proposed least squares approximation method transforms the computationally intractable model into one that is solvable. Additionally, the explored conditions contribute to a substantial reduction in solution time by more than two orders of magnitude. These results verify the superiority of our proposed method, and showcase its applicability in practical scenarios.

2.1 Nomenclature

2.1.1 Acronyms

ACsM	AC sub-microgrid.
BDC	Bi-directional converter.
C&CG	Column-and-constraint generation.
DCsM	DC sub-microgrid.
DEs	Diesel engines.
FCs	Fuel cells.
KKT	Karush–Kuhn–Tucker.
LSA	Least squares approximation.
MGs	Microgrids.
NLP	Non-linear programming.
NMG	Networked microgrid.
PVs	Photovoltaic arrays.
WTs	Wind turbines.

2.1.2 Sets and Indices

$i, \mathcal{I}$	Index and set of nodes in ACsM.
$j, \mathcal{J}$	Index and set of nodes in DCsM.
$l, \mathcal{L}$	Index and set of BDCs connecting these sub-microgrids.
$t, \mathcal{T}$	Index and set of time slots.
$\pi(i)$	Set of all children nodes of node i in ACsM.
$\pi(j)$	Set of all children nodes of node j in DCsM.
$\delta(i)$	Set of all parent nodes of node i in ACsM.
$\delta(j)$	Set of all parent nodes of node j in DCsM.
$\Omega(i_1)$	Set of nodes in DCsM connected to node i_1 in ACsM.
$\Omega(j_1)$	Set of nodes in ACsM connected to node j_1 in DCsM.

2.1.3 Variables

$G^{\text{ac}}_{ii',t}$	Reactive power flow from i to i' during period t.
$H^{\text{ac}}_{ii',t}$	Active power flow from i to i' during period t.
$H^{\text{dc}}_{jj',t}$	Active power flow from j to j' during period t.
$P^{\text{ac}}_{i_1j_1,l,t}$	Received active power at node i_1 in ACsM from node j_1 in DCsM on BDC l during period t.
$P^{\text{acdc}}_{i_1j_1,l,t}$	Power flow on BDC l from ACsM to DCsM during period t.
$P^{\text{dc}}_{i_1j_1,l,t}$	Received active power at node j_1 in DCsM from node i_1 in ACsM on BDC l during period t.
$P^{\text{dcac}}_{i_1j_1,l,t}$	Power flow on BDC l from DCsM to ACsM during period t.
$P^{\text{de}}_{i,t}$	Power outputs of DEs in node i during period t.
$P^{\text{fc}}_{j,t}$	Power outputs of FCs in node j during period t.
$Q^{\text{de}}_{i,t}$	Reactive power of DEs at node i during period t.
$U^{\text{ac}}_{i,t}$	Voltage magnitude of node i during period t.
$U^{\text{dc}}_{j,t}$	Voltage magnitude of node j during period t.
$\rho^{\text{pv}}_{j,t}$	Curtailed PV power at node i during period t.
$\rho^{\text{w}}_{i,t}$	Curtailed wind power at node i during period t.

2.1.4 Parameters

$c_i^{\mathrm{de}}/c_j^{\mathrm{fc}}$	Fuel cost coefficients of DEs and FCs.
$c_i^{\mathrm{pv}}/c_j^{\mathrm{w}}$	Penalty cost coefficients of PVs and WTs.
$c_{i_1j_1,l}$	Loss penalty coefficient of BDC l.
$\underline{G}_{ii'}^{\mathrm{ac}}/\bar{G}_{ii'}^{\mathrm{ac}}$	Lower/upper reactive power flow on branch ii'.
$\underline{H}_{ii'}^{\mathrm{ac}}/\bar{H}_{ii'}^{\mathrm{ac}}$	Lower/upper active power flow on branch ii'.
$\underline{H}_{jj'}^{\mathrm{dc}}/\bar{H}_{jj'}^{\mathrm{dc}}$	Lower/upper active power flow on branch jj'.
$L_{i,t}^{\mathrm{ac}}/Q_{i,t}^{\mathrm{ac}}$	Active/reactive load at node i during period t.
$L_{j,t}^{\mathrm{dc}}$	Active load at node j during period t.
$\bar{P}_i^{\mathrm{de}}$	Active power capacity of DEs at node i.
$P_{i,t}^{\mathrm{w}}$	Predicted wind power at node i during period t.
$\bar{P}_j^{\mathrm{fc}}$	Active power capacity of FCs at node j.
$P_{j,t}^{\mathrm{pv}}$	Predicted PV power at node j during period t.
$\bar{P}_{i_1j_1,l}^{\mathrm{acdc}}$	Capacity of BDC l from ACsM to DCsM.
$\bar{P}_{i_1j_1,l}^{\mathrm{dcac}}$	Capacity of BDC l from DCsM to ACsM.
$\bar{Q}_i^{\mathrm{de}}$	Reactive power capacity of DEs at node i.
$r_{ii'}^{\mathrm{ac}}/x_{ii'}^{\mathrm{ac}}$	Resistance/reactance of branch ii' in ACsM.
$r_{jj'}^{\mathrm{dc}}$	Resistance/reactance of branch jj' in DCsM.
$\underline{U}_i^{\mathrm{ac}}/\bar{U}_i^{\mathrm{ac}}$	Lower/upper voltage magnitude of node i.
$\underline{U}_j^{\mathrm{dc}}/\bar{U}_j^{\mathrm{dc}}$	Lower/upper voltage magnitude of node j.
$h_{1/2/\ldots/7,l}$	Loss coefficients.
κ_l	Inductor loss coefficient of BDC l.
m_l	Modulation factor of BDC l.
$k_{\mathrm{E,on},l}/$ $k_{\mathrm{E,off},l}/$ $k_{\mathrm{D},l}$	Switch-on/switch-off/diode loss coefficients of BDC l.

2.2 Topology and Operation Model of Hybrid AC/DC NMGs

This section provides a comprehensive overview of the BDC-based operation model for hybrid AC/DC NMGs. We begin with an overview of the structure of a hybrid AC/DC NMG. Then, we outline the key assumptions necessary for the BDC-based operation model. Finally, we detail the BDC-based operation model of hybrid AC/DC NMGs.

2.2.1 Topology of Hybrid AC/DC NMGs

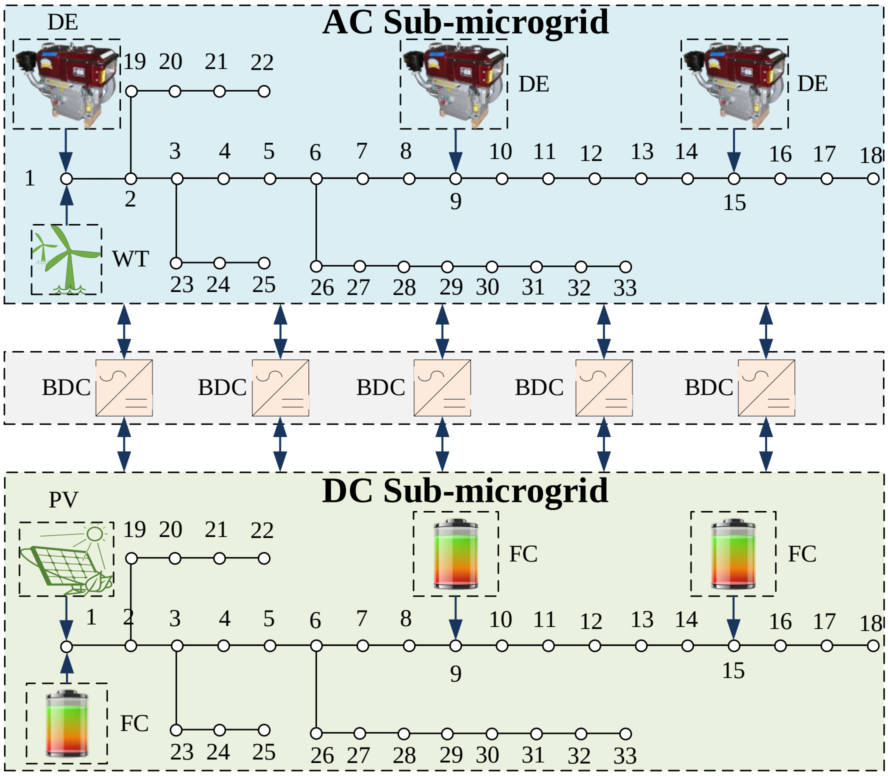


Figure 2.1: Topology of a hybrid AC/DC NMG

The typical topology of a hybrid AC/DC NMG is presented in Fig. 2.1. The entire MG comprises three primary components: an ACsM, a DCsM, and multiple BDCs. The ACsM encompasses DEs, WTs, AC loads, and an AC network, while the DCsM consists of FCs, PVs, DC loads, and a DC network [118]. Interconnection between these two sub-MGs is facilitated by several BDCs, which handle complex power conversion processes. It is noteworthy that, the power output from WTs and PVs is heavily influenced by meteorological conditions, which are inherently variable and difficult to predict. Conversely, DEs and FCs offer a more stable and controllable power supply, potentially mitigating the variability of WTs and PVs. Nonetheless, integrating diverse types of distributed generators, such as DEs and WTs, into a single node can introduce synchronization issues related to frequency and voltage [119]. These challenges necessitate the use of sophisticated power electronics to harmonize the frequency and voltage of the power produced by the various distributed generators to align with the grid's standards.

2.2.2 Assumptions

To properly formulate the BDC-based operation model of hybrid AC/DC NMGs, we make the following assumptions:

1) The MGs are assumed to work in an islanded mode, which can be easily modified for connection to the main power grid.

2) The model is tailored for short-term economic dispatch over time intervals ranging from minutes to hours. Consequently, it does not delve into the dynamic behaviors of BDCs, which are to be managed in scenarios involving real-time control.

3) The Linearized Distribution Flow (LinDistFlow) is applied to model the power flow equations in both AC and DC sub-microgrids.

4) This model only considers deterministic scenarios of power outputs from WTs and PVs, and can be easily extended to the stochastic or robust economic dispatch model with uncertain outputs of WTs and PVs.

This paper focuses on the short-term operation characteristics of BDCs in short-term economic dispatch issues, examining two particular characteristics: the non-simultaneous rectification and inversion behaviors, and the dynamic conversion efficiency. The conversion efficiency of BDCs dynamically varies with the power flow on BDC $P_{i_1 j_1,l,t}^{\mathrm{acdc/dcac}}$, which involves a complex combination of polynomial and trigonometric functions as described in equation (2.22). Moreover, the non-convex constraint arising from the non-simultaneous rectification and inversion behaviors of BDCs adds further complexity. Incorporating these exact but complex expressions directly into economic dispatch models for hybrid AC/DC NMGs can lead to challenging solutions, even when employing advanced commercial optimization solvers, such as GUROBI and CONOPT. To strike a balance between real-time control and economic dispatch, we propose an approximation method with high accuracy to convert the dynamic efficiency model into a convex form. This method accurately reflects the dynamic conversion efficiency of BDCs in economic dispatch schemes, while reducing computational complexity and enabling solvability of the economic dispatch models. Additionally, we explore sufficient conditions to relax the non-convex constraint related to the non-simultaneous rectifying and inverting behaviors of BDCs, resulting in a convex model that significantly expedites the solution process.

Although the economic dispatch model considers the dynamic conversion efficiency of BDCs in relation to energy efficiency, it does not account for the dynamic processes of BDCs for several reasons. Firstly, the time scale of converter dynamics (milliseconds to seconds) is significantly smaller than that of the typical short-term economic dispatch, which spans minutes to hours [35]. As a result, the system state and control command during the operation of the converter are well stabilized before the economic dispatch instructions take effect. Secondly, economic dispatch is primarily concerned with costs and overall system efficiency over a specific dispatch period, which are not directly impacted by BDC dynamics [120]. Lastly, dynamic characteristics of BDCs are highly non-linear and difficult to simplify into a convex form. Including these complex, non-convex characteristics would make it extremely challenging to find practical solutions for short-term economic dispatch models in microgrids.

2.2.3 BDC-Based Operation Model of Hybrid AC/DC NMGs

The operation model for islanded hybrid AC/DC NMGs incorporating BDCs is denoted as ***Model I***. The economic dispatch cost of such islanded hybrid AC/DC NMGs includes the fuel costs of DEs and FCs, as well as the conversion losses of BDCs resulting from the power conversion process [34], [115]. In contrast to conventional power generators, e.g., DEs and FCs, WTs and PVs do not have fuel or significant operational costs. Consequently, their operational costs are typically negligible.

Therefore, the objective function can be formulated as follows:

$$\min \sum\nolimits_{t\in\mathcal{T}} \left(\begin{array}{l} \sum_{i\in\mathcal{I}} c_i^{\mathrm{de}} P_{i,t}^{\mathrm{de}} + \sum_{j\in\mathcal{J}} c_j^{\mathrm{fc}} P_{j,t}^{\mathrm{fc}} \\ + \sum_{l\in\mathcal{L}} c_{i_1 j_1,l} \left(P_{i_1 j_1,l,t}^{\mathrm{acdc}} + P_{i_1 j_1,l,t}^{\mathrm{dcac}} \right) \end{array} \right) \tag{2.1}$$

Eq. (2.1) minimizes the total dispatch cost associated with hybrid AC/DC NMGs over the whole horizon $\mathcal{T}$. This is achieved by finding the optimal value of the total dispatch cost while ensuring that all operational constraints of the components within the hybrid AC/DC NMGs are satisfied.

Note that, hybrid AC/DC NMGs consist of three primary components: ACsM, DCsM and BDCs. Accordingly, the operation constraints of the hybrid AC/DC NMGs are categorized into three clusters: ACsM constraints, DCsM constraints, and BDCs constraints. More precisely, the operation constraints related to the ACsM involve those for DEs, WTs, and the AC network. Similarly, the operation constraints pertaining to the DCsM cover those for FCs, PVs, and the DC network. The Lagrange multipliers associated with the operation constraints in the hybrid AC/DC NMGs are defined after a colon.

1) Operation constraints for the ACsM

An ACsM operates as a localized AC grid and possesses the ability to operate independently of the central power grid [103]. It consists of a small-scale network of electricity users and AC-tied distributed energy resources, including DEs and WTs. DEs can provide adjustable levels of active and reactive power, within their operational limits, offering a backup power solution and compensating for periods when wind conditions are inadequate for WT power generation [121]. Conversely, there are occasions when the output from WTs must be regulated, particularly during instances of power overproduction relative to the current demand, to prevent surplus generation and ensure power balance within the ACsM. The constraints for the ACsM are detailed as follows.

$$0 \le P_{i,t}^{\mathrm{de}} \le \bar{P}_i^{\mathrm{de}}, \forall i,t : \underline{\lambda}_{i,t}^{\mathrm{de1}}, \bar{\lambda}_{i,t}^{\mathrm{de1}} \tag{2.2}$$

$$0 \le Q_{i,t}^{\mathrm{de}} \le \bar{Q}_i^{\mathrm{de}}, \forall i,t : \underline{\lambda}_{i,t}^{\mathrm{de2}}, \bar{\lambda}_{i,t}^{\mathrm{de2}} \tag{2.3}$$

$$0 \le \rho_{i,t}^{\mathrm{w}} \le P_{i,t}^{\mathrm{w}}, \forall i,t : \underline{\lambda}_{i,t}^{\mathrm{w}}, \bar{\lambda}_{i,t}^{\mathrm{w}} \tag{2.4}$$

$$U_{i',t}^{\mathrm{ac}}=U_{i,t}^{\mathrm{ac}}-\left(r_{ii'}^{\mathrm{ac}}H_{ii',t}^{\mathrm{ac}}+x_{ii'}^{\mathrm{ac}}G_{ii',t}^{\mathrm{ac}}\right)/U_{0}^{\mathrm{ac}},\forall i,i',t:\beta_{ii',t}^{\mathrm{ac1}} \tag{2.5}$$

$$\begin{aligned}&P_{i,t}^{\mathrm{de}}-L_{i,t}^{\mathrm{ac}}+\sum\nolimits_{j_1\in\Omega(i_1)}\sum\nolimits_{l\in\mathcal{L}}\left(P_{i_1j_1,l,t}^{\mathrm{ac}}-P_{i_1j_1,l,t}^{\mathrm{acdc}}\right)\\&=\sum\nolimits_{i'\in\pi(i)}H_{ii',t}^{\mathrm{ac}}-\sum\nolimits_{i''\in\delta(i)}H_{i''i,t}^{\mathrm{ac}}-\left(P_{i,t}^{\mathrm{w}}-\rho_{i,t}^{\mathrm{w}}\right),\forall i,t:\beta_{i,t}^{\mathrm{ac2}}\end{aligned} \tag{2.6}$$

$$Q_{i,t}^{\mathrm{de}}-Q_{i,t}^{\mathrm{ac}}=\sum\nolimits_{i'\in\pi(i)}G_{ii',t}^{\mathrm{ac}}-\sum\nolimits_{i''\in\delta(i)}G_{i''i,t}^{\mathrm{ac}},\forall i,t:\beta_{i,t}^{\mathrm{ac3}} \tag{2.7}$$

$$\underline{H}_{ii'}^{\mathrm{ac}}\leq H_{ii',t}^{\mathrm{ac}}\leq\bar{H}_{ii'}^{\mathrm{ac}},\forall i,i',t:\underline{\lambda}_{ii',t}^{\mathrm{ac1}},\bar{\lambda}_{ii',t}^{\mathrm{ac1}} \tag{2.8}$$

$$\underline{G}_{ii'}^{\mathrm{ac}}\leq G_{ii',t}^{\mathrm{ac}}\leq\bar{G}_{ii'}^{\mathrm{ac}},\forall i,i',t:\underline{\lambda}_{ii',t}^{\mathrm{ac2}},\bar{\lambda}_{ii',t}^{\mathrm{ac2}} \tag{2.9}$$

$$\underline{U}_{i}^{\mathrm{ac}}\leq U_{i,t}^{\mathrm{ac}}\leq\bar{U}_{i}^{\mathrm{ac}},\forall i,t:\underline{\lambda}_{i,t}^{\mathrm{ac3}},\bar{\lambda}_{i,t}^{\mathrm{ac3}} \tag{2.10}$$

For ACsM, constraints (2.2) and (2.3) state the active and reactive power capacity of DEs, and constraint (2.4) limits the available power curtailment of WTs. Constraints (2.5)-(2.7) define the branch power flow in ACsM using the LinDistflow model[122], which is commonly used in modeling MG power flow[123]. Constraints (2.8)-(2.10) respectively bound the branch active power, branch reactive power, and node voltage magnitude limits of the ACsM network.

2) Operation constraints for the DCsM

Similar to an ACsM, a DCsM operates on a localized DC network of electricity users with DC-connected distributed energy resources, such as FCs with controllable power outputs, and PVs with less uncontrollable power outputs [36]. However, the technical details and operation constraints of a DCsM are specifically tailored to the characteristics of DC power systems. Notably, DC systems do not involve reactive power due to the unidirectional flow of current and constant voltage. The constraints for the DCsM are outlined as follows.

$$0\leq P_{j,t}^{\mathrm{fc}}\leq\bar{P}_{j}^{\mathrm{fc}},\forall j,t:\underline{\lambda}_{j,t}^{\mathrm{fc}},\bar{\lambda}_{j,t}^{\mathrm{fc}} \tag{2.11}$$

$$0\leq\rho_{j,t}^{\mathrm{pv}}\leq P_{j,t}^{\mathrm{pv}},\forall j,t:\underline{\lambda}_{j,t}^{\mathrm{pv}},\bar{\lambda}_{j,t}^{\mathrm{pv}} \tag{2.12}$$

$$U_{j',t}^{\mathrm{dc}}=U_{j,t}^{\mathrm{dc}}-r_{jj'}^{\mathrm{dc}}H_{jj',t}^{\mathrm{dc}}/U_{0}^{\mathrm{dc}},\forall j,j',t:\beta_{jj',t}^{\mathrm{dc1}} \tag{2.13}$$

$$
\begin{aligned}
&P_{j,t}^{\mathrm{fc}} - L_{j,t}^{\mathrm{dc}} + \sum\nolimits_{i_1 \in \Omega(j_1)} \sum\nolimits_{l \in \mathcal{L}} \left(P_{i_1 j_1,l,t}^{\mathrm{dc}} - P_{i_1 j_1,l,t}^{\mathrm{dcac}} \right) \\
&= \sum\nolimits_{j' \in \pi(j)} H_{jj',t}^{\mathrm{dc}} - \sum\nolimits_{j'' \in \delta(j)} H_{j''j,t}^{\mathrm{dc}} - \left(P_{j,t}^{\mathrm{pv}} - \rho_{j,t}^{\mathrm{pv}} \right), \forall j,t : \beta_{j,t}^{\mathrm{dc2}}
\end{aligned}
\tag{2.14}
$$

$$
\underline{H}_{jj'}^{\mathrm{dc}} \leq H_{jj',t}^{\mathrm{dc}} \leq \bar{H}_{jj'}^{\mathrm{dc}}, \forall j, j', t : \underline{\lambda}_{jj',t}^{\mathrm{dc1}}, \bar{\lambda}_{jj',t}^{\mathrm{dc1}} \tag{2.15}
$$

$$
\underline{U}_{j}^{\mathrm{dc}} \leq U_{j,t}^{\mathrm{dc}} \leq \bar{U}_{j}^{\mathrm{dc}}, \forall j, t : \underline{\lambda}_{j,t}^{\mathrm{dc2}}, \bar{\lambda}_{j,t}^{\mathrm{dc2}} \tag{2.16}
$$

For DCsM, the active power capacity of FCs is limited by (2.11), and the available power curtailment of PVs is bounded by (2.12). Constraints (2.13) and (2.14) state the branch power flow model in DCsM. The branch active power flow and node voltage magnitude in DCsM are constrained by (2.15) and (2.16), respectively.

3) Operation constraints for BDCs

BDCs are power electronic devices designed to convert electrical power between AC and DC in both directions. To avert the risks of overheating and potential damage, these converters are engineered with a maximum power handling capacity limit, whether they function as rectifiers or inverters. Additionally, BDCs are designed to either convert AC to DC or DC to AC at any given moment[24], [31]. They cannot physically perform both operations at the same time because the power electronic switches are configured to facilitate the flow of current in one direction depending on the mode of operation. The corresponding constraints for BDCs are presented as follows.

$$
0 \leq P_{i_1 j_1,l,t}^{\mathrm{acdc}} \leq \bar{P}_{i_1 j_1,l}^{\mathrm{acdc}}, \forall l, t : \underline{\lambda}_{i_1 j_1,l,t}^{\mathrm{ad}}, \bar{\lambda}_{i_1 j_1,l,t}^{\mathrm{ad}} \tag{2.17}
$$

$$
0 \leq P_{i_1 j_1,l,t}^{\mathrm{dcac}} \leq \bar{P}_{i_1 j_1,l}^{\mathrm{dcac}}, \forall l, t : \underline{\lambda}_{i_1 j_1,l,t}^{\mathrm{da}}, \bar{\lambda}_{i_1 j_1,l,t}^{\mathrm{da}} \tag{2.18}
$$

$$
P_{i_1 j_1,l,t}^{\mathrm{acdc}} \cdot P_{i_1 j_1,l,t}^{\mathrm{dcac}} = 0, \forall l, t \tag{2.19}
$$

$$
P_{i_1 j_1,l,t}^{\mathrm{ac}} = P_{i_1 j_1,l,t}^{\mathrm{dcac}} \cdot \eta_{\mathrm{BDC},l}^{\mathrm{dcac}} (P_{i_1 j_1,l,t}^{\mathrm{dcac}}), \forall l, t \tag{2.20}
$$

$$
P_{i_1 j_1,l,t}^{\mathrm{dc}} = P_{i_1 j_1,l,t}^{\mathrm{acdc}} \cdot \eta_{\mathrm{BDC},l}^{\mathrm{acdc}} (P_{i_1 j_1,l,t}^{\mathrm{acdc}}), \forall l, t \tag{2.21}
$$

$$
\begin{aligned}
&\eta_{\mathrm{BDC},l}^{\mathrm{acdc/dcac}}\left(P_{i_1 j_1,l,t}^{\mathrm{acdc/dcac}}\right)=1-(\kappa_l P_{i_1 j_1,l,t}^{\mathrm{acdc/dcac}}+\\
&h_{1,l}(I_{l,t})^{k_{\mathrm{E,on},l}}+h_{2,l}(I_{l,t})^{k_{\mathrm{E,off},l}}+h_{3,l}(I_{l,t})^{k_{\mathrm{D},l}}\\
&+h_{4,l}I_{l,t}\left(\tfrac{1}{\pi}+\tfrac{m_l}{4}\cos\left(\theta_{l,t}\right)\right)+h_{5,l}(I_{l,t})^2\left(\tfrac{1}{8}+\tfrac{m_l}{3\pi}\cos\left(\theta_{l,t}\right)\right)\\
&+h_{6,l}I_{l,t}\left(\tfrac{1}{\pi}-\tfrac{m_l}{4}\cos\left(\theta_{l,t}\right)\right)+h_{7,l}(I_{l,t})^2\left(\tfrac{1}{8}-\tfrac{m_l}{3\pi}\cos\left(\theta_{l,t}\right)\right))
\end{aligned}
\tag{2.22}
$$

For BDCs, the allowable rectification and inverter power of each BDC is enforced by its rates in (2.17) and (2.18). A complementary constraint (2.19) is introduced to ensure that power does not flow in both directions simultaneously through each BDC. This constraint stipulates that if $P_{i_1 j_1,l,t}^{\mathrm{acdc}} \neq 0$, then $P_{i_1 j_1,l,t}^{\mathrm{dcac}} = 0$, and if $P_{i_1 j_1,l,t}^{\mathrm{dcac}} \neq 0$, then $P_{i_1 j_1,l,t}^{\mathrm{acdc}} = 0$. This is rooted in the physical limitations of BDCs, which are engineered to only permit power flow in a single direction at any given moment. Constraints (2.20) and (2.21) calculate the power after BDC conversion using dynamic BDC loss function $\eta_{\mathrm{BDC},l}^{\mathrm{acdc/dcac}}(\cdot)$ in (2.22), where $I_{l,t}=\sqrt{4[(P_{i_1 j_1,l,t}^{\mathrm{acdc/dcac}})^2+q_{l,t}^2]/\left[3\left(V_{i_1}^{\mathrm{AC}}\right)\right]^2}$ and $\theta_{l,t}=\bar{\theta}_l-\arctan\left(q_l, P_{i_1 j_1,l,t}^{\mathrm{acdc/dcac}}\right)$. Here, parameters $V_{i_1}^{\mathrm{AC}}$, $\bar{\theta}_l$ and q_l are the AC-side voltage magnitude of hybrid AC/DC NMGs, the angle coefficient between BDC and grid voltage, and the reactive power required to maintain BDC operation. The dynamic BDC efficiency in (2.22) is derived from hardware tests reported in [27], with the complete expression further detailed in [31]. The power conversion loss of BDCs is determined by various loss terms: 1) inductor loss in the first line; 2) IGBT turn-on loss, IGBT turn-off loss, and diode reverse recovery loss in the second line; 3) IGBT conduction loss in the third line; 4) diode conduction loss in the last line. Note that the conversion efficiency of BDCs can be influenced by various factors, including topology, voltage magnitude, work mode, and conversion power. This work focuses on the primary influencing factors, namely work mode and conversion power, while neglecting the impacts of less critical variables such as topology and voltage magnitude. Though the parameters in (2.22) may have different values in the rectification mode and the inversion mode, this does not affect the applicability of the methodology in this paper. Thus, we assume the same values for these parameters to simplify the expression.

Remark 1: The computational complexity of solving the operation model for the hybrid AC/DC NMG (2.1)-(2.22) arises primarily from the non-linear and non-convex characteristics of the BDC operation constraints. These include a complementary constraint (2.19) as well as conversion efficiency constraints (2.20)-(2.22). Specifically, constraint (2.19) incorporates a bi-linear term involving $P^{\text{dcac}}_{i_1 j_1,l,t}$ and $P^{\text{acdc}}_{i_1 j_1,l,t}$, while constraints (2.20)-(2.22) encompass both polynomial and trigonometric expressions. These two non-linear constraints render the model a non-convex optimization problem, significantly increasing the computational burden of the solving process. Additionally, hybrid AC/DC NMGs are generally integrated with highly penetrated RESs, which exhibit uncertain and intermittent power outputs [124], [125]. To accurately integrate these outputs into dispatch models, it is essential to employ uncertainty-oriented approaches such as scenario-based stochastic programming [126], [127]. To comprehensively represent the potential variability in RES power generation [128], these models must incorporate massive RES scenarios to reflect diverse weather patterns, diurnal fluctuations, and seasonal variations. As a result, the integration of extensive RES scenarios into the hybrid AC/DC NMG model, which already contains non-convex constraints (2.19)-(2.22), can create computational difficulties, potentially leading to solutions that are hard to obtain within reasonable timeframes or computational resources.

To address these challenges, Sections 2.3 and 2.4 introduce approaches designed to convexify these complex constraints, and finally develop a computationally efficient model for economic dispatch in BDC-based hybrid AC/DC NMGs.

2.3 BDC Efficiency Function Approximation

In this section, we use $\eta^{\text{dcac}}_{\text{BDC},l}(x)$ in (2.22) to demonstrate the proposed approximation method. A similar approach can be applied to $\eta^{\text{acdc}}_{\text{BDC},l}(x)$ in (2.22). To convexify such a highly non-convex function $\eta^{\text{dcac}}_{\text{BDC},l}(x)$, we aim to find an approximating function $\varphi_l(x)$ for $\eta^{\text{dcac}}_{\text{BDC},l}(x)$ by minimizing the integrated value of the approximation error over the interval $[0,1]$. The approximation problem can be formulated as follows:

$$\min \delta = \sum\nolimits_{l\in\mathcal{L}} \int_{[0,1]} \left[\varphi_l(x) - \eta^{\text{dcac}}_{\text{BDC},l}(x)\right]^2 dx \tag{2.23}$$

We note that choosing a linear function to approximate the efficiency in constraint (2.22) introduces bilinear terms into constraints (2.20) and (2.21), which continues to result in the non-convexity issue in the model. According to the definition of convex optimization, the functions in equality constraints must be affine. Thus, to address this and ensure model convexity, we adopt the following translated inverse proportionality function form for $\varphi_l(x)$ to approximate efficiency in constraint (2.22). This adjustment effectively transforms the model into a convex one.

$$\varphi_l(x) = o_{0,l}x^{-1} + o_{1,l}, \forall l \in \mathcal{L} \tag{2.24}$$

We note that the feasible region of each BDC l in models (2.23) and (2.24) is independent of each other. Thus, the entire model (2.23) and (2.24) can be separated into $|\mathcal{L}|$ small-scale optimization problems. For each BDC $\forall l \in \mathcal{L}$, we have

$$\min \delta_l = \int_{[0,1]} [\varphi_l(x) - \eta^{\text{dcac}}_{\text{BDC},l}(x)]^2 dx \tag{2.25}$$

$$\varphi_l(x) = o_{0,l}x^{-1} + o_{1,l} \tag{2.26}$$

To solve the model (2.25) and (2.26), we first take the partial derivative of the function (2.25) at $o_{0,l}$ and $o_{1,l}$, which gives us:

$$\partial\delta/\partial o_{0,l} = 2\int_{[0,1]} [\varphi_l(x) - \eta^{\text{dcac}}_{\text{BDC},l}(x)]x^{-1}dx = 0 \tag{2.27}$$

$$\partial\delta/\partial o_{1,l} = 2\int_{[0,1]} [\varphi_l(x) - \eta^{\text{dcac}}_{\text{BDC},l}(x)]x^{0}dx = 0 \tag{2.28}$$

Then, (2.27) and (2.28) can be reformulated as follows:

$$\int_{[0,1]} (o_{0,l}x^{-1} + o_{1,l})x^{-1}dx = \int_{[0,1]} \eta^{\text{dcac}}_{\text{BDC},l}(x)x^{-1}dx \tag{2.29}$$

$$\int_{[0,1]} (o_{0,l}x^{-1} + o_{1,l})x^{0}dx = \int_{[0,1]} \eta^{\text{dcac}}_{\text{BDC},l}(x)x^{0}dx \tag{2.30}$$

With the definition of the inner product of functions x^a and x^b, as well as x^a and $\eta^{\text{dcac}}_{\text{BDC},l}(x)$ on the interval [0,1] in (2.31) and (2.32), the model (2.29) and (2.30) can be written in matrix form (2.33)-(2.35).

$$\lfloor x^a, x^b \rfloor = \int_{[0,1]} x^a x^b dx \tag{2.31}$$

$$\lfloor x^a, \eta_{\text{BDC},l}^{\text{dcac}}(x) \rfloor = \int_{[0,1]} x^a \eta_{\text{BDC},l}^{\text{dcac}}(x) dx \tag{2.32}$$

$$\mathbf{E}\mathbf{o}_l = \varphi_l \tag{2.33}$$

$$\mathbf{E} = \begin{bmatrix} \lfloor x^{-1}, x^{-1} \rfloor & \lfloor x^{-1}, x^{0} \rfloor \\ \lfloor x^{0}, x^{-1} \rfloor & \lfloor x^{0}, x^{0} \rfloor \end{bmatrix}, \mathbf{o}_l = \begin{bmatrix} o_{0,l} \\ o_{1,l} \end{bmatrix} \tag{2.34}$$

$$\varphi_l = \left[\lfloor x^{-1}, \eta_{\text{BDC},l}^{\text{dcac}}(x) \rfloor \ \lfloor x^{0}, \eta_{\text{BDC},l}^{\text{dcac}}(x) \rfloor \right]^{\text{T}} \tag{2.35}$$

Since x^{-1} and 1 are linearly independent, we know that the Gram determinant value det $\mathbf{E} \neq 0$. Thus, there exists a unique solution $\mathbf{o}_l^*$ for Eq. (2.36), which can be computed as follows.

$$\mathbf{o}_l^* = \mathbf{E}^{-1}\varphi_l \tag{2.36}$$

After calculating the solution of Eq. (2.36), we can obtain the optimal value of $o_{0,l}^*$ and $o_{1,l}^*$. Using these values, we can rewrite (2.22) as follows.

$$\eta_{\text{BDC},l}^{\text{dcac}}(P_{i_1 j_1,l,t}^{\text{dcac}}) = o_{0,l}^*(P_{i_1 j_1,l,t}^{\text{dcac}})^{-1} + o_{1,l}^* \tag{2.37}$$

$$\eta_{\text{BDC},l}^{\text{acdc}}(P_{i_1 j_1,l,t}^{\text{acdc}}) = o_{0,l}^*(P_{i_1 j_1,l,t}^{\text{acdc}})^{-1} + o_{1,l}^* \tag{2.38}$$

Finally, *Model I* can be reformulated as *Model II* by replacing constraints (2.20)-(2.22) in *Model I* with constraints (2.37) and (2.38). ***Model II***:

$$\text{Objective :2.1} \tag{2.39}$$

$$s.t. 2.2 - 2.5, 2.7 - 2.10, 2.11 - 2.13, 2.15 - 2.19 \tag{2.40}$$

$$\begin{aligned} &P_{i,t}^{\text{de}} + \sum_{j_1 \in \Omega(i_1)} \sum_{l \in L} (o_{1,l}^* P_{i_1 j_1,l,t}^{\text{dcac}} + o_{0,l}^* \bar{P}_{i_1 j_1,l}^{\text{dcac}} - P_{i_1 j_1,l,t}^{\text{acdc}}) \\ &= \sum_{i' \in \pi(i)} H_{ii',t}^{\text{ac}} - \sum_{i'' \in \delta(i)} H_{i''i,t}^{\text{ac}} - (P_{i,t}^{\text{w}} - \rho_{i,t}^{\text{w}}) + L_{i,t}^{\text{ac}}, \forall i, t : \beta_{i,t}^{\text{ac2}} \end{aligned} \tag{2.41}$$

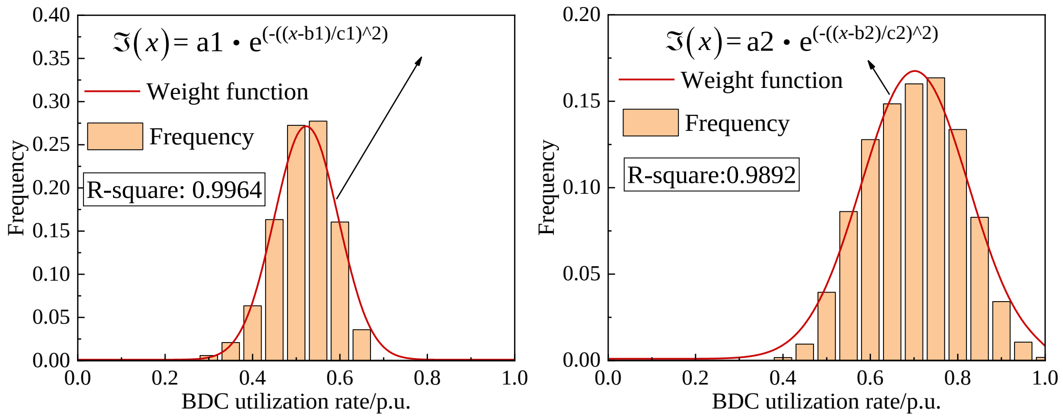


Figure 2.2: Determining $\Im\left(x\right)$ based on practical BDC operational data

$$\begin{aligned} & P^{\text{fc}}_{j,t}+\sum_{i_1\in\Omega(j_1)}\sum_{l\in L}(o^{*}_{1,l}P^{\text{acdc}}_{i_1 j_1,l,t}+o^{*}_{0,l}\bar{P}^{\text{acdc}}_{i_1 j_1,l}-P^{\text{dcac}}_{i_1 j_1,l,t}) \\ & =\sum_{j'\in\pi(j)}H^{\text{dc}}_{jj',t}-\sum_{j''\in\delta(j)}H^{\text{dc}}_{j''j,t}-(P^{\text{pv}}_{j,t}-\rho^{\text{pv}}_{j,t})+L^{\text{dc}}_{j,t},\forall j,t:\beta^{\text{dc2}}_{j,t} \end{aligned} \tag{2.42}$$

When the historical data on the BDC power flow is available, the approximation problem (2.23) can be extended as follows:

$$\min\delta=\sum\nolimits_{l\in\mathcal{L}}\int_{[0,1]}\Im\left(x\right)\left[\varphi_l(x)-\eta^{\text{dcac}}_{\text{BDC},l}(x)\right]^2dx \tag{2.43}$$

where $\Im\left(x\right)$ is a weight function that evaluates the contribution of each value of x to the approximation error. Here, we propose a data-driven approach to determine $\Im\left(x\right)$ using x values obtained from practical BDC operational data, and fitting data distribution for each BDC, as shown in Fig. 2.2.

Remark 2: Compared with *Model I*, *Model II* is a more simplified NLP problem. It only includes relatively simple non-linear terms in (2.19), while excluding both the polynomial and trigonometric functions. By using the proposed approximation method, the operation model of hybrid AC/DC NMGs can be transformed from computationally intractable to solvable. It is noteworthy that compared to existing linearization methods, such as piece-wise linearization or Taylor expansion-based linearization approaches [129], the proposed method does not introduce binary variables while minimizing the global approximation error. This allows the proposed method to well trade-off model complexity and approximation accuracy. To quantify these aspects, we measure model complexity by the computational time required to solve the model, and we assess approximation accuracy using both absolute and relative average errors. By minimizing the approximation error of the BDC expression, the resulting convex efficiency model consistently achieves the highest approximation accuracy while remaining computationally efficient.

2.4 Sufficient Conditions for Convex Model

We observe that, *Model II* remains a non-convex optimization problem because of the inclusion of the non-convex constraint (2.19). Consequently, we aim to explore the particular circumstances in which constraint (2.19) becomes inactive. Identifying these conditions enables us to formulate a convex model, denoted as *Model III*, by removing the non-convex constraint (2.19) from *Model II*. This removal simplifies the model and significantly accelerates the solving procedure, without compromising the optimality of the final solution.

Model III:

$$\text{Objective :}(2.1) \tag{2.44}$$

$$s.t.(2.2)-(2.5),(2.7)-(2.10),(2.11)-(2.13),(2.15)-(2.18),(2.41),(2.42) \tag{2.45}$$

For expression convenience, we first define

$$f^{\mathrm{ac}}\left(o_{1,l},\beta_{i_1,t}^{\mathrm{ac2}}\right)=\left(o_{1,l}^2-1\right)\sum\nolimits_{i_1\in i_1 j_1}\beta_{i_1,t}^{\mathrm{ac2}} \tag{2.46}$$

$$f^{\mathrm{dc}}\left(o_{1,l},\beta_{j_1,t}^{\mathrm{dc2}}\right)=\left(o_{1,l}^2-1\right)\sum\nolimits_{j_1\in i_1 j_1}\beta_{j_1,t}^{\mathrm{dc2}} \tag{2.47}$$

The KKT conditions are necessary for a solution to be considered optimal in an optimization problem. Specifically for *Model III*, which is a convex optimization problem, the KKT conditions fulfill a dual role. They are not just necessary but also sufficient for determining optimality. This implies that any optimal solution for *Model III* must satisfy all the KKT conditions. Next, we will establish the necessary conditions for the case when $P_{i_1 j_1,l,t}^{\mathrm{acdc}}P_{i_1 j_1,l,t}^{\mathrm{dcac}}>0,\forall l,t$, which means that constraint (2.19) is active. We then present three propositions that outline the specific conditions under which constraint (2.19) becomes non-active, all while utilizing the KKT conditions to guide our analysis.

Theorem 1. *Necessary conditions for* $P^{\text{acdc}}_{i_1j_1,l,t}P^{\text{dcac}}_{i_1j_1,l,t} > 0, \forall l, t$ *are* $f^{\text{ac}}\left(o_{1,l}, \beta^{\text{ac2}}_{i_1,t}\right) < 0$ *and* $f^{\text{dc}}\left(o_{1,l}, \beta^{\text{dc2}}_{j_1,t}\right)$ < 0.

Proof: Let ℓ be the Lagrangian function of *Model III*. Based on the stationarity condition of KKT conditions to *Model III*, we can obtain

$$\partial\ell/\partial P^{\text{acdc}}_{i_1j_1,l,t} = c_{i_1j_1,l} - \underline{\lambda}^{\text{ad}}_{i_1j_1,l,t} + \bar{\lambda}^{\text{ad}}_{i_1j_1,l,t} - \sum\nolimits_{i_1\in i_1j_1} \beta^{\text{ac2}}_{i_1,t} + \sum\nolimits_{j_1\in i_1j_1} o_{1,l}\beta^{\text{dc2}}_{j_1,t} = 0 \tag{2.48}$$

$$\partial\ell/\partial P^{\text{dcac}}_{i_1j_1,l,t} = c_{i_1j_1,l} - \underline{\lambda}^{\text{da}}_{i_1j_1,l,t} + \bar{\lambda}^{\text{da}}_{i_1j_1,l,t} - \sum\nolimits_{j_1\in i_1j_1} \beta^{\text{dc2}}_{j_1,t} + \sum\nolimits_{i_1\in i_1j_1} o_{1,l}\beta^{\text{ac2}}_{i_1,t} = 0 \tag{2.49}$$

Since $P^{\text{acdc}}_{i_1j_1,l,t}P^{\text{dcac}}_{i_1j_1,l,t} > 0$, the complementary slackness of KKT conditions yields $\underline{\lambda}^{\text{ad}}_{i_1j_1,l,t} = \underline{\lambda}^{\text{da}}_{i_1j_1,l,t} = 0$ and $\bar{\lambda}^{\text{ad}}_{i_1j_1,l,t}, \bar{\lambda}^{\text{da}}_{i_1j_1,l,t} \geq 0$. We can multiply Eq. (2.49) by a scalar of $o_{1,l}$, and then add it to Eq. (2.48) to replace $\sum_{j_1\in i_1j_1} o_{1,l}\beta^{\text{dc2}}_{j_1,t}$ in Eq. (2.48), resulting in:

$$\left(1+o_{1,l}\right)c_{i_1j_1,l} + \bar{\lambda}^{\text{ad}}_{i_1j_1,l,t} + o_1\bar{\lambda}^{\text{da}}_{i_1j_1,l,t} + \left(o^2_{1,l}-1\right)\sum\nolimits_{i_1\in i_1j_1} \beta^{\text{ac2}}_{i_1,t}=0 \tag{2.50}$$

Since $o_{1,l} > 0$, $c_{i_1j_1,l} > 0$, $\bar{\lambda}^{\text{ad}}_{i_1j_1,l,t} \geq 0$ and $\bar{\lambda}^{\text{da}}_{i_1j_1,l,t} \geq 0$, we obtain $f^{\text{ac}}\left(o_{1,l}, \beta^{\text{ac2}}_{i_1,t}\right) < 0$ based on Eq. (2.46).

By following a similar approach, we derive the following Eq. (2.51) and the condition $f^{\text{dc}}\left(o_{1,l}, \beta^{\text{dc2}}_{j_1,t}\right)$ < 0.

$$\left(1+o_{1,l}\right)c_{i_1j_1,l} + \bar{\lambda}^{\text{da}}_{i_1j_1,l,t} + o_1\bar{\lambda}^{\text{ad}}_{i_1j_1,l,t} + \left(o^2_{1,l}-1\right)\sum\nolimits_{j_1\in i_1j_1} \beta^{\text{dc2}}_{j_1,t}=0 \tag{2.51}$$

$f^{\text{ac}}\left(o_{1,l}, \beta^{\text{ac2}}_{i_1,t}\right) < 0$ and $f^{\text{dc}}\left(o_{1,l}, \beta^{\text{dc2}}_{j_1,t}\right) < 0$, as stated in the theorem, provide necessary conditions for the operation with $P^{\text{acdc}}_{i_1j_1,l,t}P^{\text{dcac}}_{i_1j_1,l,t} > 0$. ■

This theorem indicates that, if $f^{\text{ac}}\left(o_{1,l}, \beta^{\text{ac2}}_{i_1,t}\right) \geq 0$ or $f^{\text{dc}}\left(o_{1,l}, \beta^{\text{dc2}}_{j_1,t}\right) \geq 0$, the BDC will not simultaneously operate in inverting-rectifying states. In other words, the solution with simultaneously inverting-rectifying operations of BDCs is suboptimal. Leveraging this theorem, we explore the specific conditions that guarantee the BDC operates exclusively in either inverting or rectifying mode, never both. This ensures that the solution for *Model III* is not inferior to that of *Model II*, while greatly enhancing computational speed. These specific conditions are investigated in the following three propositions for hybrid AC/DC NMG economic dispatch based on the value of $o_{1,l}^2$.

Proposition 1. *For a given value of* $0 < o_{1,l} < 1, \forall l$, *a solution to Model III where* $P_{i_1 j_1,l,t}^{\mathrm{acdc}} P_{i_1 j_1,l,t}^{\mathrm{dcac}} > 0, \exists t$ *is suboptimal, if any of the following conditions is satisfied:*

1) $\exists t \in \mathcal{T}$, *such that* $0 \leq \rho_{i_1,t}^{\mathrm{w}} < P_{i_1,t}^{\mathrm{w}}$ *or* $0 \leq \rho_{j_1,t}^{\mathrm{pv}} < P_{j_1,t}^{\mathrm{pv}}$;

2) $\exists t \in \mathcal{T}$, *such that* $0 < P_{i_1,t}^{\mathrm{de}} \leq \bar{P}_{i_1}^{\mathrm{de}}$ *or* $0 < P_{j_1,t}^{\mathrm{fc}} \leq \bar{P}_{j_1}^{\mathrm{fc}}$;

3) $\exists t \in \mathcal{T}$, *such that* $\exists i' \in \pi(i_1)$, $H_{i_1 i',t}^{\mathrm{ac}} < H_{i_1 i'}^{\mathrm{ac,max}}$, $\beta_{i',t}^{\mathrm{ac2}} \leq 0, U_{i',t}^{\mathrm{ac}} + \left(r_{i_1 i'}^{\mathrm{ac}} H_{i_1 i',t}^{\mathrm{ac}} + x_{i_1 i'}^{\mathrm{ac}} G_{i_1 i',t}^{\mathrm{ac}}\right) / U_0^{\mathrm{ac}} < U_{i_1}^{\mathrm{ac,max}}$ *or* $\exists i \in \delta(i_1)$, $H_{i i_1}^{\mathrm{ac,min}} < H_{i i_1,t}^{\mathrm{ac}}$, $\beta_{i,t}^{\mathrm{ac2}} \leq 0$, $U_i^{\mathrm{ac,min}} < U_{i_1,t}^{\mathrm{ac}} + \left(r_{i i_1}^{\mathrm{ac}} H_{i i_1,t}^{\mathrm{ac}} + x_{i i_1}^{\mathrm{ac}} G_{i i_1,t}^{\mathrm{ac}}\right) / U_0^{\mathrm{ac}}$;

4) $\exists t \in \mathcal{T}$, *such that* $\exists j' \in \pi(j_1)$, $H_{j_1 j',t}^{\mathrm{dc}} < H_{j_1 j'}^{\mathrm{dc,max}}$, $\beta_{j',t}^{\mathrm{dc2}} \leq 0$, $U_{j',t}^{\mathrm{dc}} + r_{j_1 j'}^{\mathrm{dc}} H_{j_1 j',t}^{\mathrm{dc}} / U_0^{\mathrm{dc}} < U_{j_1}^{\mathrm{dc,max}}$, *or* $\exists j \in \delta(j_1)$, $H_{j j_1}^{\mathrm{dc,min}} < H_{j j_1,t}^{\mathrm{dc}}$, $\beta_{j,t}^{\mathrm{dc2}} \leq 0$, $U_j^{\mathrm{dc,min}} < U_{j_1,t}^{\mathrm{dc}} + r_{j j_1}^{\mathrm{dc}} H_{j j_1,t}^{\mathrm{dc}} / U_0^{\mathrm{dc}}$.

Proof: In the following, we will prove how each condition in the proposition leads to a contradiction in *Theorem 1* by taking Condition 1 as an example, while the proof of other Conditions can be referred to Appendix A.

Case 1.i): $\exists t \in \mathcal{T}, 0 \leq \rho_{i_1,t}^{\mathrm{w}} < P_{i_1,t}^{\mathrm{w}}$. Using the stationarity condition of KKT conditions for *Model III*, we have

$$\partial \ell / \partial \rho_{i,t}^{\mathrm{w}} = -\underline{\lambda}_{i,t}^{\mathrm{w}} + \bar{\lambda}_{i,t}^{\mathrm{w}} - \beta_{i,t}^{\mathrm{ac2}} = 0 \tag{2.52}$$

Applying the complementary slackness of KKT conditions to *Case 1.i)*, we obtain $\bar{\lambda}_{i_1,t}^{\mathrm{w}} = 0$ and $\underline{\lambda}_{i_1,t}^{\mathrm{w}} \geq 0$. Thus, Eq. (2.52) can be rewritten as $\beta_{i,t}^{\mathrm{ac2}} = -\underline{\lambda}_{i,t}^{\mathrm{w}} \leq 0$. Considering that $o_{1,l}^2 < 1$, we can derive $f^{\mathrm{ac}}\left(o_{1,l}, \beta_{i_1,t}^{\mathrm{ac2}}\right) \geq 0$. However, this contradicts the necessary condition of *Theorem 1*. Hence, we can conclude that a solution with $\exists t \in \mathcal{T}, 0 \leq \rho_{i_1,t}^{\mathrm{w}} < P_{i_1,t}^{\mathrm{w}}$ is suboptimal.

Case 1.ii): $\exists t \in \mathcal{T}, 0 \leq \rho_{j_1,t}^{\mathrm{pv}} < P_{j_1,t}^{\mathrm{pv}}$. By applying the stationarity condition of KKT conditions to *Model III*, we have

$$\partial \ell / \partial \rho_{j,t}^{\mathrm{pv}} = -\underline{\lambda}_{j,t}^{\mathrm{pv}} + \bar{\lambda}_{j,t}^{\mathrm{pv}} - \beta_{j,t}^{\mathrm{dc2}} = 0 \tag{2.53}$$

Applying the complementary slackness of KKT conditions to *Case 1.ii)*, we can derive $\bar{\lambda}_{j_1,t}^{\mathrm{pv}} = 0$ and $\underline{\lambda}_{j_1,t}^{\mathrm{pv}} \geq 0$. Substituting these values into Eq. (2.53), we obtain $\beta_{j_1,t}^{\mathrm{dc2}} = -\underline{\lambda}_{j_1,t}^{\mathrm{pv}}$. Thus, $f^{\mathrm{dc}}\left(o_{1,l}, \beta_{j_1,t}^{\mathrm{dc2}}\right) \geq 0$ with $o_{1,l}^2 < 1$, which contradicts the necessary condition of *Theorem 1*. Hence, a solution with $\exists t \in \mathcal{T}, 0 \leq \rho_{j_1,t}^{\mathrm{pv}} < P_{j_1,t}^{\mathrm{pv}}$ is suboptimal.

Case 2.i): $\exists t \in \mathcal{T}, 0 < P^{\mathrm{de}}_{i_1,t} \leq \bar{P}^{\mathrm{de}}_{i_1}$. By applying the stationarity condition of KKT conditions to *Model III*, we have

$$\partial \ell / \partial P^{\mathrm{de}}_{i,t} = c^{\mathrm{de}}_i - \underline{\lambda}^{\mathrm{del}}_{i,t} + \bar{\lambda}^{\mathrm{del}}_{i,t} + \beta^{\mathrm{ac2}}_{i,t} = 0 \tag{2.54}$$

With the complementary slackness of KKT conditions for *Case 2.i)*, we obtain $\underline{\lambda}^{\mathrm{de1}}_{i_1,t} = 0$ and $\bar{\lambda}^{\mathrm{de1}}_{i_1,t} \geq 0$. Substituting these values into Eq. (2.54), we obtain $\beta^{\mathrm{ac2}}_{i_1,t} = - c^{\mathrm{de}}_{i_1} - \bar{\lambda}^{\mathrm{de1}}_{i_1,t} < 0$. This implies that $f^{\mathrm{ac}}\left(o_{1,l}, \beta^{\mathrm{ac2}}_{i_1,t}\right) \geq 0$ when $o^2_{1,l} < 1$, contradicting the necessary condition of *Theorem 1*. Therefore, a solution with $\exists t \in \mathcal{T}, 0 < P^{\mathrm{de}}_{i_1,t} \leq \bar{P}^{\mathrm{de}}_{i_1}$ is not optimal.

Case 2.ii): $\exists t \in \mathcal{T}, 0 < P^{\mathrm{fc}}_{j_1,t} \leq \bar{P}^{\mathrm{fc}}_{j_1}$. Using the stationarity condition of KKT conditions for *Model III*, we have

$$\partial \ell / \partial P^{\mathrm{fc}}_{j,t} = c^{\mathrm{fc}}_j - \underline{\lambda}^{\mathrm{fcl}}_{j,t} + \bar{\lambda}^{\mathrm{fcl}}_{j,t} + \beta^{\mathrm{dc2}}_{j,t} = 0 \tag{2.55}$$

With the complementary slackness of KKT conditions for *Case 2.ii)*, we have $\underline{\lambda}^{\mathrm{fc1}}_{j_1,t} = 0$ and $\bar{\lambda}^{\mathrm{fc1}}_{j_1,t} \geq 0$. Consequently, Eq. (2.55) can be rewritten as $\beta^{\mathrm{dc2}}_{j_1,t} = - c^{\mathrm{fc}}_{j_1} - \bar{\lambda}^{\mathrm{fc1}}_{j_1,t} < 0$. Thus, $f^{\mathrm{dc}}\left(o_{1,l}, \beta^{\mathrm{dc2}}_{j_1,t}\right) \geq 0$ when $o^2_{1,l} < 1$. However, this contradicts the necessary condition of *Theorem 1*. A solution with $\exists t \in \mathcal{T}, 0 < P^{\mathrm{fc}}_{j_1,t} \leq \bar{P}^{\mathrm{fc}}_{j_1}$ is not optimal.

Case 3.i): $\exists t \in \mathcal{T}, \exists i' \in \pi(i_1), H^{\mathrm{ac}}_{i_1 i',t} < H^{\mathrm{ac,max}}_{i_1 i'}, \beta^{\mathrm{ac2}}_{i',t} \leq 0$ and $U^{\mathrm{ac}}_{i',t} + \left(r^{\mathrm{ac}}_{i_1 i'} H^{\mathrm{ac}}_{i_1 i',t} + x^{\mathrm{ac}}_{i_1 i'} G^{\mathrm{ac}}_{i_1 i',t}\right) / U^{\mathrm{ac}}_0 < U^{\mathrm{ac,max}}_{i_1}$. Using the stationarity condition of KKT conditions for *Model III*, we have

$$\partial \ell / \partial H^{\mathrm{ac}}_{ii',t} = -\beta^{\mathrm{ac1}}_{i,t} r^{\mathrm{ac}}_{ii'} / U^{\mathrm{ac}}_0 - \beta^{\mathrm{ac2}}_{i,t} + \beta^{\mathrm{ac2}}_{i',t} + \bar{\lambda}^{\mathrm{ac1}}_{ii',t} - \underline{\lambda}^{\mathrm{ac1}}_{ii',t} = 0 \tag{2.56}$$

$$\partial \ell / \partial U^{\mathrm{ac}}_{i,t} = \beta^{\mathrm{ac1}}_{i,t} + \bar{\lambda}^{\mathrm{ac3}}_{ii',t} - \underline{\lambda}^{\mathrm{ac3}}_{ii',t} = 0 \tag{2.57}$$

By introducing (2.57) into (2.56), we can obtain

$$\partial \ell / \partial H^{\mathrm{ac}}_{ii',t} = -\underline{\lambda}^{\mathrm{ac3}}_{i,t} r^{\mathrm{ac}}_{ii'} / U^{\mathrm{ac}}_0 + \bar{\lambda}^{\mathrm{ac3}}_{i,t} r^{\mathrm{ac}}_{ii'} / U^{\mathrm{ac}}_0 - \beta^{\mathrm{ac2}}_{i,t} + \beta^{\mathrm{ac2}}_{i',t} + \bar{\lambda}^{\mathrm{ac1}}_{ii',t} - \underline{\lambda}^{\mathrm{ac1}}_{ii',t} = 0 \tag{2.58}$$

Using the complementary slackness of KKT conditions for *Case 3.i)*, we derive $\bar{\lambda}^{\mathrm{ac1}}_{i_1 i',t} = 0$, $\underline{\lambda}^{\mathrm{acl}}_{i_1 i',t} \geq 0$, $\bar{\lambda}^{\mathrm{ac3}}_{i_1,t} = 0$ and $\underline{\lambda}^{\mathrm{ac3}}_{i_1,t} \geq 0$. Then, Eq. (2.58) can be rewritten as follows.

$$-\underline{\lambda}^{\mathrm{ac3}}_{i_1,t} r^{\mathrm{ac}}_{i_1 i'} / U^{\mathrm{ac}}_0 - \beta^{\mathrm{ac2}}_{i_1,t} + \beta^{\mathrm{ac2}}_{i',t} - \underline{\lambda}^{\mathrm{ac1}}_{i_1 i',t} = 0 \tag{2.59}$$

When $\beta^{\text{ac2}}_{i',t} \leq 0$, it implies that $\beta^{\text{ac2}}_{i_1,t} \leq 0$, which leads to $f^{\text{ac}}\left(o_{1,l}, \beta^{\text{ac2}}_{i_1,t}\right) \geq 0$ when $o^2_{1,l} < 1$. This contradicts the condition in *Theorem 1*. Thus, a solution with *Case 3.i)* is suboptimal.

Case 3.ii): $\exists t \in \mathcal{T}, \exists i \in \delta(i_1), H^{\text{ac,min}}_{ii_1} < H^{\text{ac}}_{ii_1,t}$, $U^{\text{ac,min}}_{i} < U^{\text{ac}}_{i_1,t} + \left(r^{\text{ac}}_{ii_1} H^{\text{ac}}_{ii_1,t} + x^{\text{ac}}_{ii_1} G^{\text{ac}}_{ii_1,t}\right)/U^{\text{ac}}_0$, and $\beta^{\text{ac2}}_{i,t} \leq 0$. Using the complementary slackness of KKT conditions for *Case 3.ii)*, we derive $\underline{\lambda}^{\text{ac1}}_{ii_1,t} = 0$, $\bar{\lambda}^{\text{ac1}}_{ii_1,t} \geq 0$, $\underline{\lambda}^{\text{ac3}}_{i,t} = 0$ and $\bar{\lambda}^{\text{ac3}}_{i,t} \geq 0$. Then, Eq. (2.58) can be rewritten as follows.

$$\bar{\lambda}^{\text{ac3}}_{i,t} r^{\text{ac}}_{ii_1}/U^{\text{ac}}_0 - \beta^{\text{ac2}}_{i,t} + \beta^{\text{ac2}}_{i_1,t} + \bar{\lambda}^{\text{ac1}}_{ii_1,t} = 0 \tag{2.60}$$

When $\beta^{\text{ac2}}_{i,t} \leq 0$, the inequality $\beta^{\text{ac2}}_{i_1,t} \leq 0$ also holds, which implies that $f^{\text{ac}}\left(o_{1,l}, \beta^{\text{ac2}}_{i_1,t}\right) \geq 0$ when $o^2_{1,l} < 1$. Finally, we obtain a contradiction with the condition in *Theorem 1*. Thus, a solution with *Case 3.ii)* is suboptimal.

3) Condition 4

Case 4.i): $\exists t \in \mathcal{T}, \exists j' \in \pi(j_1), H^{\text{dc}}_{j_1j',t} < H^{\text{dc,max}}_{j_1j'}$, $\beta^{\text{dc2}}_{j',t} \leq 0$ and $U^{\text{dc}}_{j',t} + r^{\text{dc}}_{j_1j'} H^{\text{dc}}_{j_1j',t}/U^{\text{dc}}_0 < U^{\text{dc,max}}_{j_1}$. Using the stationarity condition of KKT conditions for *Model III*, we have

$$\partial\ell/\partial H^{\text{dc}}_{jj',t} = -\beta^{\text{dc1}}_{j,t} r^{\text{dc}}_{jj'}/U^{\text{dc}}_0 - \beta^{\text{dc2}}_{j,t} + \beta^{\text{dc2}}_{j',t} + \bar{\lambda}^{\text{dc1}}_{jj',t} - \underline{\lambda}^{\text{dc1}}_{jj',t} = 0 \tag{2.61}$$

$$\partial\ell/\partial U^{\text{dc}}_{j,t} = \beta^{\text{dc1}}_{j,t} + \bar{\lambda}^{\text{dc2}}_{jj',t} - \underline{\lambda}^{\text{dc2}}_{jj',t} = 0 \tag{2.62}$$

By introducing (2.62) into (2.61), we obtain

$$\partial\ell/\partial H^{\text{dc}}_{jj',t} = -\underline{\lambda}^{\text{dc2}}_{j,t} r^{\text{dc}}_{jj'}/U^{\text{dc}}_0 + \bar{\lambda}^{\text{dc2}}_{j,t} r^{\text{dc}}_{jj'}/U^{\text{dc}}_0 - \beta^{\text{dc2}}_{j,t} + \beta^{\text{dc2}}_{j',t} + \bar{\lambda}^{\text{dc1}}_{jj',t} - \underline{\lambda}^{\text{dc1}}_{jj',t} = 0 \tag{2.63}$$

Applying the complementary slackness of KKT conditions to *Case 4.i)*, we derive $\bar{\lambda}^{\text{dc1}}_{j_1j',t} = 0$, $\underline{\lambda}^{\text{dc1}}_{j_1j',t} \geq 0$, $\bar{\lambda}^{\text{dc2}}_{j_1,t} = 0$ and $\underline{\lambda}^{\text{dc2}}_{j_1,t} \geq 0$. Then, Eq. (2.63) can be rewritten as follows.

$$-\underline{\lambda}^{\text{dc2}}_{j_1,t} r^{\text{dc}}_{j_1j'}/U^{\text{dc}}_0 - \beta^{\text{dc2}}_{j_1,t} + \beta^{\text{dc2}}_{j',t} - \underline{\lambda}^{\text{dc1}}_{j_1j',t} = 0 \tag{2.64}$$

With $\beta^{\text{dc2}}_{j',t} \leq 0$, we can deduce $\beta^{\text{dc2}}_{j_1,t} \leq 0$, yielding $f^{\text{dc}}\left(o_{1,l}, \beta^{\text{dc2}}_{i_1,t}\right) \geq 0$ if $o^2_{1,l} < 1$. This contradicts the condition in *Theorem 1*, which requires $P^{\text{acdc}}_{i_1j_1,l,t} P^{\text{dcac}}_{i_1j_1,l,t} > 0$ to obtain a feasible solution. Thus, a solution with *Case 4.i)* is suboptimal.

Case 4.ii): $\exists t \in \mathcal{T}, \exists j \in \delta(j_1), H_{jj_1}^{\text{dc,min}} < H_{jj_1,t}^{\text{dc}}$, $\beta_{j,t}^{\text{dc2}} \leq 0$ and $U_j^{\text{dc,min}} < U_{j_1,t}^{\text{dc}} + r_{jj_1}^{\text{dc}} H_{jj_1,t}^{\text{dc}}/U_0^{\text{dc}}$. Using the complementary slackness of KKT conditions for *Case 4.ii)*, we derive $\underline{\lambda}_{jj_1,t}^{\text{dc1}} = 0$, $\bar{\lambda}_{jj_1,t}^{\text{dc1}} \geq 0$, $\underline{\lambda}_{j,t}^{\text{dc2}} = 0$ and $\bar{\lambda}_{j,t}^{\text{dc2}} \geq 0$. Then, Eq. (2.63) can be rewritten as follows.

$$\bar{\lambda}_{j,t}^{\text{dc2}} r_{jj_1}^{\text{dc}}/U_0^{\text{dc}} - \beta_{j,t}^{\text{dc2}} + \beta_{j_1,t}^{\text{dc2}} + \bar{\lambda}_{jj_1,t}^{\text{dc1}} = 0 \tag{2.65}$$

When $\beta_{j,t}^{\text{dc2}} \leq 0$, it implies that $\beta_{j_1,t}^{\text{dc2}} \leq 0$. This means that $f^{\text{dc}}\left(o_{1,l}, \beta_{i_1,t}^{\text{dc2}}\right) \geq 0$ when $o_{1,l}^2 < 1$. However, this causes a contradiction with the condition in *Theorem 1*. Therefore, it can be concluded that a solution with *Case 4.ii)* is suboptimal.■

Overall, *Proposition 1* outlines the conditions to achieve the optimal solution with non-simultaneous inverting-rectifying operations of each BDC in the operation problem of hybrid AC/DC NMGs.

Specifically, Condition 1 states that the wind power connected with BDCs on the ACM side should not be completely cut off from hybrid AC/DC NMGs, or the photovoltaic power connected with BDCs on the DCM side should not be completely cut off from hybrid AC/DC NMGs. This condition is generally satisfied because cutting off all available wind power and photovoltaic power from hybrid AC/DC NMGs is not cost-effective.

Condition 2 indicates that the diesel engine unit connected with BDCs in the ACM side or the fuel cell unit linked with BDCs on the DCM side should operate above their minimum power outputs. This condition is generally met in practice since it is cost-effective for a controllable unit to run above its minimum power output.

Condition 3 defines the requirements for the children and parent nodes of the node i_1 connected to BDCs on the ACM side. Specifically, for the children nodes, the conditions include: a) the active power flow on the branch line $i_1 i'$ should be less than its upper power capacity, b) the Lagrangian multiplier corresponding to the active power balance Eq. (2.6) in the ACM side should not exceed zero, and c) the voltage magnitude at node i_1 should be less than its upper limit. For the parent node, the following conditions should be met: a) the active power flow on the branch line ii_1 must exceed its lower power limit, b) the Lagrangian multiplier corresponding to the active power balance Eq. (2.6) in the ACM side should not exceed zero, and c) the voltage magnitude at node i must be larger than its lower limit.

Condition 4 is similar to Condition 3 but applies to the children and parent nodes of the node j_1 connected to BDCs on the DCM side. For the children nodes, the conditions include: a) the active power flow on the branch line $j_1 i'$ should be less than its upper power capacity, b) the Lagrangian multiplier corresponding to the active power balance Eq. (2.14) in the DCM side should not exceed zero, and c) the voltage magnitude at node j_1 should be less than its upper limit. For the parent node, the following conditions should be met: a) the active power flow on the branch line jj_1 must exceed its lower power limit, b) the Lagrangian multiplier corresponding to the active power balance Eq. (2.14) in the DCM side should not exceed zero, and c) the voltage magnitude at node j must be larger than its lower limit.

Proposition 2. *As long as $o_{1,l} = 1, \forall l$, a solution to Model III where $P^{\mathrm{acdc}}_{i_1 j_1,l,t} P^{\mathrm{dcac}}_{i_1 j_1,l,t} > 0, \exists t$ is suboptimal, without any specific requirements on the operation conditions of hybrid AC/DC NMGs.*

Proof: When $o^2_{1,l} = 1$, the condition $f^{\mathrm{ac}}\left(o_{1,l}, \beta^{\mathrm{ac2}}_{i_1,t}\right) = f^{\mathrm{dc}}\left(o_{1,l}, \beta^{\mathrm{dc2}}_{j_1,t}\right) = 0$ always holds for any $\sum_{i_1 \in i_1 j_1} \beta^{\mathrm{ac2}}_{i_1,t} \in \mathbb{R}$, which does not satisfy the necessary conditions stated in *Theorem 1*. Thus, regardless of the value of $\sum_{i_1 \in i_1 j_1} \beta^{\mathrm{ac2}}_{i_1,t}$, a solution where a BDC operates in simultaneously inverting-rectifying states is consistently suboptimal if $o^2_{1,l} = 1$. ■

Proposition 3. *For a given value of $o_{1,l} > 1, \forall l$, a solution to Model III where $P^{\mathrm{acdc}}_{i_1 j_1,l,t} P^{\mathrm{dcac}}_{i_1 j_1,l,t} > 0, \exists t$ is suboptimal, if any of the following conditions is satisfied:*

1) $\exists t \in \mathcal{T}$, such that $0 < \rho^{\mathrm{w}}_{i_1,t} \leq P^{\mathrm{w}}_{i_1,t}$ or $0 < \rho^{\mathrm{pv}}_{j_1,t} \leq P^{\mathrm{pv}}_{j_1,t}$;

2) $\exists t \in \mathcal{T}$, such that $0 \leq P^{\mathrm{de}}_{i_1,t} < \bar{P}^{\mathrm{de}}_{i_1}, \underline{\lambda}^{\mathrm{de1}}_{i_1,t} \geq c^{\mathrm{de}}_{i_1}$ or $0 \leq P^{\mathrm{fc}}_{j_1,t} < \bar{P}^{\mathrm{fc}}_{j_1}, \underline{\lambda}^{\mathrm{fc1}}_{j_1,t} \geq c^{\mathrm{fc}}_{j_1}$;

3) $\exists t \in \mathcal{T}$, such that $\exists i' \in \pi\left(i_1\right)$, $H^{\mathrm{ac,min}}_{i_1 i'} < H^{\mathrm{ac}}_{i_1 i',t}, \beta^{\mathrm{ac2}}_{i',t} \geq 0$, $U^{\mathrm{ac,min}}_{i_1} < U^{\mathrm{ac}}_{i',t} + \left(r^{\mathrm{ac}}_{i_1 i'} H^{\mathrm{ac}}_{i_1 i',t} + x^{\mathrm{ac}}_{i_1 i'} G^{\mathrm{ac}}_{i_1 i',t}\right)/U^{\mathrm{ac}}_0$ or $\exists i \in \delta(i_1)$, $H^{\mathrm{ac}}_{ii_1,t} < H^{\mathrm{ac,max}}_{ii_1}$, $\beta^{\mathrm{ac2}}_{i,t} \geq 0$, $U^{\mathrm{ac}}_{i_1,t} + \left(r^{\mathrm{ac}}_{ii_1} H^{\mathrm{ac}}_{ii_1,t} + x^{\mathrm{ac}}_{ii_1} G^{\mathrm{ac}}_{ii_1,t}\right)/U^{\mathrm{ac}}_0 < U^{\mathrm{ac,max}}_i$;

4) $\exists t \in \mathcal{T}$, such that $\exists j' \in \pi(j_1)$, $H^{\mathrm{dc,min}}_{j_1 j'} < H^{\mathrm{dc}}_{j_1 j',t}$, $\beta^{\mathrm{dc2}}_{j',t} \geq 0$, $U^{\mathrm{dc,min}}_{j_1} < U^{\mathrm{dc}}_{j',t} + r^{\mathrm{dc}}_{j_1 j'} H^{\mathrm{dc}}_{j_1 j',t}/U^{\mathrm{dc}}_0$, or $\exists j \in \delta(j_1)$, $H^{\mathrm{dc}}_{jj_1,t} < H^{\mathrm{dc,max}}_{jj_1}$, $\beta^{\mathrm{dc2}}_{j,t} \geq 0$, $U^{\mathrm{dc}}_{j_1,t} + r^{\mathrm{dc}}_{jj_1} H^{\mathrm{dc}}_{jj_1,t}/U^{\mathrm{dc}}_0 < U^{\mathrm{dc,max}}_j$.

Proof:

1) Condition 1

Case 1.i): $\exists t \in \mathcal{T}, 0 < \rho^{\mathrm{w}}_{i_1,t} \leq P^{\mathrm{w}}_{i_1,t}$. Based on Condition 1, we deduce that $\underline{\lambda}^{\mathrm{w}}_{i_1,t} = 0$ and $\bar{\lambda}^{\mathrm{w}}_{i_1,t} \geq 0$ by using the complementary slackness of KKT conditions for *Model III*. Eq. (2.52) can be rephrased as $\beta_{i_1,t}^{\mathrm{ac2}} = \bar{\lambda}_{i_1,t}^{\mathrm{w}} \geq 0$, and thus $f^{\mathrm{ac}}\left(o_{1,l}, \beta_{i_1,t}^{\mathrm{ac2}}\right) \geq 0$ with $o_{1,l}^2 > 1$. This finding contradicts the conclusions in *Theorem 1*. Thus, a solution with $\exists t \in \mathcal{T}, 0 < \rho_{i_1,t}^{\mathrm{w}} \leq P_{i_1,t}^{\mathrm{w}}$ is suboptimal.

Case 1.ii): $\exists t \in \mathcal{T}, 0 < \rho_{j_1,t}^{\mathrm{pv}} \leq P_{j_1,t}^{\mathrm{pv}}$. Based on *Case 1.ii)*, we derive $\underline{\lambda}_{j_1,t}^{\mathrm{pv}} = 0$ and $\bar{\lambda}_{j_1,t}^{\mathrm{pv}} \geq 0$ by applying the complementary slackness of KKT conditions to *Model III*. Eq. (2.53) can be expressed as $\beta_{j_1,t}^{\mathrm{dc2}} = \bar{\lambda}_{j_1,t}^{\mathrm{pv}} \geq 0$, implying $f^{\mathrm{dc}}\left(o_{1,l}, \beta_{j_1,t}^{\mathrm{dc2}}\right) \geq 0$ with $o_{1,l}^2 > 1$. We obtain a contradiction in *Theorem 1*. That is, a solution with $\exists t \in \mathcal{T}, 0 < \rho_{j_1,t}^{\mathrm{pv}} \leq P_{j_1,t}^{\mathrm{pv}}$ is suboptimal.

2) Condition 2

Case 2.i): $\exists t \in \mathcal{T}, 0 \leq P_{i_1,t}^{\mathrm{de}} < \bar{P}_{i_1}^{\mathrm{de}}$ and $\underline{\lambda}_{i_1,t}^{\mathrm{de1}} \geq c_{i_1}^{\mathrm{de}}$. Based on $\exists t \in \mathcal{T}, 0 \leq P_{i_1,t}^{\mathrm{de}} < \bar{P}_{i_1}^{\mathrm{de}}$ in *Case 2.i)*, we derive $\bar{\lambda}_{i_1,t}^{\mathrm{de1}} = 0$ by using the complementary slackness of KKT conditions to *Model III*. Thus, Eq. (2.54) can be rewritten as $\beta_{i_1,t}^{\mathrm{ac2}} = \underline{\lambda}_{i_1,t}^{\mathrm{de1}} - c_{i_1}^{\mathrm{de}}$. Since $\underline{\lambda}_{i_1,t}^{\mathrm{de1}} \geq c_{i_1}^{\mathrm{de}}$, we derive $\beta_{i_1,t}^{\mathrm{ac2}} \geq 0$, yielding $f^{\mathrm{ac}}\left(o_{1,l}, \beta_{i_1,t}^{\mathrm{ac2}}\right) \geq 0$ with $o_{1,l}^2 > 1$. Thus, we conclude a contradiction in *Theorem 1*, indicating that a solution with $\exists t \in \mathcal{T}, 0 \leq P_{i_1,t}^{\mathrm{de}} < \bar{P}_{i_1}^{\mathrm{de}}$ and $\underline{\lambda}_{i_1,t}^{\mathrm{de1}} > c_{i_1}^{\mathrm{de}}$ is suboptimal.

Case 2.ii): $\exists t \in \mathcal{T}, 0 \leq P_{j_1,t}^{\mathrm{fc}} < \bar{P}_{j_1}^{\mathrm{fc}}$ and $\underline{\lambda}_{j_1,t}^{\mathrm{fc1}} \geq c_{j_1}^{\mathrm{fc}}$. Based on $\exists t \in \mathcal{T}, 0 \leq P_{j_1,t}^{\mathrm{fc}} < \bar{P}_{j_1}^{\mathrm{fc}}$ in *Case 2.ii)*, we derive $\bar{\lambda}_{j_1,t}^{\mathrm{fc1}} = 0$ by applying the complementary slackness of KKT conditions to *Model III*. Thus, Eq. (2.55) can be rewritten as $\beta_{j_1,t}^{\mathrm{dc2}} = \underline{\lambda}_{j_1,t}^{\mathrm{fc1}} - c_{j_1}^{\mathrm{fc}}$. Since $\underline{\lambda}_{j_1,t}^{\mathrm{fc1}} \geq c_{j_1}^{\mathrm{fc}}$, we derive $\beta_{j_1,t}^{\mathrm{dc2}} \geq 0$, and this implies that $f^{\mathrm{dc}}\left(o_{1,l}, \beta_{j_1,t}^{\mathrm{dc2}}\right) \geq 0$ with $o_{1,l}^2 > 1$.

3) Condition 3

Case 3.i): $\exists t \in \mathcal{T}, \exists i' \in \pi(i_1), H_{i_1 i'}^{\mathrm{ac,min}} < H_{i_1 i',t}^{\mathrm{ac}}, U_{i_1}^{\mathrm{ac,min}} < U_{i',t}^{\mathrm{ac}} + \left(r_{i_1 i'}^{\mathrm{ac}} H_{i_1 i',t}^{\mathrm{ac}} + x_{i_1 i'}^{\mathrm{ac}} G_{i_1 i',t}^{\mathrm{ac}}\right) / U_0^{\mathrm{ac}}$, and $\beta_{i',t}^{\mathrm{ac2}} \geq 0$. Based on the complementary slackness of KKT conditions to *Model III*, we obtain $\underline{\lambda}_{i_1 i',t}^{\mathrm{ac1}} = 0$, $\underline{\lambda}_{i_1,t}^{\mathrm{ac3}} = 0$, $\bar{\lambda}_{i_1 i',t}^{\mathrm{ac1}} \geq 0$ and $\bar{\lambda}_{i_1,t}^{\mathrm{ac3}} \geq 0$. Thus, Eq. (2.58) can be rewritten as:

$$\bar{\lambda}_{i_1,t}^{\mathrm{ac3}} r_{i_1 i'}^{\mathrm{ac}} / U_0^{\mathrm{ac}} - \beta_{i_1,t}^{\mathrm{ac2}} + \beta_{i',t}^{\mathrm{ac2}} + \bar{\lambda}_{i_1 i',t}^{\mathrm{ac1}} = 0 \tag{2.66}$$

This implies that $\beta_{i_1,t}^{\mathrm{ac2}} \geq 0$ when $\beta_{i',t}^{\mathrm{ac2}} \geq 0$. Finally, we can conclude that $f^{\mathrm{ac}}\left(o_{1,l}, \beta_{i_1,t}^{\mathrm{ac2}}\right) \geq 0$ with $o_{1,l}^2 > 1$, which presents a contradiction in *Theorem 1*. Thus, a solution with *Case 3.i)* in Condition 3 is suboptimal.

Case 3.ii): $\exists t \in \mathcal{T}, \exists i \in \delta(i_1), H_{ii_1,t}^{\mathrm{ac}} < H_{ii_1}^{\mathrm{ac,max}}, \beta_{i,t}^{\mathrm{ac2}} \geq 0$, and $U_{i_1,t}^{\mathrm{ac}} + \left(r_{ii_1}^{\mathrm{ac}} H_{ii_1,t}^{\mathrm{ac}} + x_{ii_1}^{\mathrm{ac}} G_{ii_1,t}^{\mathrm{ac}}\right) / U_0^{\mathrm{ac}} < U_i^{\mathrm{ac,max}}$. Based on the complementary slackness of KKT conditions to *Model III*, we obtain $\bar{\lambda}_{ii_1,t}^{\mathrm{ac1}} = 0$,

$\bar{\lambda}_{i,t}^{\text{ac3}} = 0$, $\underline{\lambda}_{ii_1,t}^{\text{ac1}} \geq 0$ and $\underline{\lambda}_{i,t}^{\text{ac3}} \geq 0$. Thus, Eq. (2.58) can be rewritten as:

$$-\underline{\lambda}_{i,t}^{\text{ac3}} r_{ii_1}^{\text{ac}} / U_0^{\text{ac}} - \beta_{i,t}^{\text{ac2}} + \beta_{i_1,t}^{\text{ac2}} - \lambda_{ii_1,t}^{\text{ac1}} = 0 \tag{2.67}$$

If $\beta_{i,t}^{\text{ac2}} \geq 0$, then we have $\beta_{i_1,t}^{\text{ac2}} \geq 0$. This contradicts *Theorem 1*, since the function $f^{\text{ac}}\left(o_{1,l}, \beta_{i_1,t}^{\text{ac2}}\right) \geq 0$ for $o_{1,l}^2 > 1$. Thus, a solution with *Case 3.ii)* in Condition 3 is suboptimal.

4) Condition 4

Case 4.i): $\exists t \in \mathcal{T}, \exists j' \in \pi(j_1)$, $H_{j_1 j'}^{\text{dc,min}} < H_{j_1 j',t}^{\text{dc}}$, $\beta_{j',t}^{\text{dc2}} \geq 0$ and $U_{j_1}^{\text{dc,min}} < U_{j',t}^{\text{dc}} + r_{j_1 j'}^{\text{dc}} H_{j_1 j',t}^{\text{dc}} / U_0^{\text{dc}}$. Based on the complementary slackness of KKT conditions to *Model III*, we can deduce that $\underline{\lambda}_{j_1 j',t}^{\text{dc1}} = 0$, $\underline{\lambda}_{j_1,t}^{\text{dc2}} = 0$, $\bar{\lambda}_{j_1 j',t}^{\text{dc1}} \geq 0$ and $\bar{\lambda}_{j_1,t}^{\text{dc2}} \geq 0$. Then, Eq. (2.63) can be rewritten as:

$$\bar{\lambda}_{j_1,t}^{\text{dc2}} r_{jj'}^{\text{dc}} / U_0^{\text{dc}} - \beta_{j_1,t}^{\text{dc2}} + \beta_{j',t}^{\text{dc2}} + \bar{\lambda}_{j_1 j',t}^{\text{dc1}} = 0 \tag{2.68}$$

Thus, when considering *Case 4.i)* in Condition 4, we find that it leads to a suboptimal solution for *Model III*, as it results in contradictory inequalities $\beta_{j_1,t}^{\text{dc2}} \geq 0$ and $f^{\text{dc}}\left(o_{1,l}, \beta_{j_1,t}^{\text{dc2}}\right) \geq 0$ when $\beta_{j',t}^{\text{dc2}} \geq 0$.

Case 4.ii): $\exists t \in \mathcal{T}, \exists j \in \delta(j_1)$, $H_{jj_1,t}^{\text{dc}} < H_{jj_1}^{\text{dc,max}}$, $\beta_{j,t}^{\text{dc2}} \geq 0$ and $U_{j_1,t}^{\text{dc}} + r_{jj_1}^{\text{dc}} H_{jj_1,t}^{\text{dc}} / U_0^{\text{dc}} < U_j^{\text{dc,max}}$. Based on the complementary slackness of KKT conditions to *Model III*, we obtain $\bar{\lambda}_{jj_1,t}^{\text{dc1}} = 0$, $\bar{\lambda}_{j,t}^{\text{dc2}} = 0$, $\underline{\lambda}_{jj_1,t}^{\text{dc1}} \geq 0$ and $\underline{\lambda}_{j,t}^{\text{dc2}} \geq 0$. Therefore, Eq. (2.63) can be rewritten as:

$$-\underline{\lambda}_{j,t}^{\text{dc2}} r_{jj'}^{\text{dc}} / U_0^{\text{dc}} - \beta_{j,t}^{\text{dc2}} + \beta_{j_1,t}^{\text{dc2}} - \underline{\lambda}_{jj_1,t}^{\text{dc1}} = 0 \tag{2.69}$$

Thus, we have $\beta_{j_1,t}^{\text{dc2}} \geq 0$ and $f^{\text{dc}}\left(o_{1,l}, \beta_{j_1,t}^{\text{dc2}}\right) \geq 0$ when $\beta_{j,t}^{\text{dc2}} \geq 0$, However, this contradicts the conditions stated in *Theorem 1*. Thus, we can conclude that *Case 4.ii)* in Condition 4 results in a suboptimal solution to *Model III*. ■

By explicitly stating the sufficient conditions for relaxing the non-convex constraint (2.19), we offer clear guidance for formulating a proper convex economic dispatch model for hybrid AC/DC NMGs. Specifically, *Proposition 1* establishes that if $0 < o_{1,l} < 1, \forall l$, satisfying any one of the derived conditions provides a theoretical guarantee that the relaxed convex model yields a physically feasible solution (non-simultaneous rectification/inversion). These conditions correspond to typical efficient operating regions. *Proposition 2* states that if $o_{1,l} = 1, \forall l$, then the non-convex constraint (2.19) can be fully relaxed with no effect on the final solutions. *Proposition 3* indicates that if $o_{1,l} > 1, \forall l$, it is likely impermissible to relax the non-convex constraint (2.19). the relaxation is theoretically invalid under standard conditions. However, the model remains universally applicable regardless of these conditions through the iterative verification process described in Section 2.5.5. This process acts as a rigorous fallback mechanism: in the rare instances where these sufficient conditions are not met, the algorithm detects the exact constraint violation and selectively re-enforces the non-convex constraints, ensuring a valid solution is always obtained.

2.5 Condition Discussions and Generalizations

In this section, we modify and extend the sufficient conditions for the operation of hybrid AC/DC NMGs in various practical circumstances.

2.5.1 Circumstance I: Hybrid AC/DC NMGs Operate in Night Mode

In the circumstance where hybrid AC/DC NMGs operate in night mode, the PV generation is null during night periods due to the absence of solar resources. This renders $\underline{\lambda}^{\mathrm{pv}}_{j,t} > 0$ and $\bar{\lambda}^{\mathrm{pv}}_{j,t} > 0$ in constraint (2.12) during these operational periods. Thus, constraint (2.53) should be rewritten as follows.

$$\beta^{\mathrm{dc2}}_{j,t} = \underline{\lambda}^{\mathrm{pv}}_{j,t} - \bar{\lambda}^{\mathrm{pv}}_{j,t} \tag{2.70}$$

In this case, it is not possible to determine whether the value of $f^{\mathrm{ac}}\left(o_{1,l}, \beta^{\mathrm{ac2}}_{i_1,t}\right)$ in *Theorem 1* is positive or negative, regardless of the value of $o_{1,l}$. As a result, *Case 1.ii)* in both *Proposition 1* and *Proposition 2* does not guarantee the zero value of the simultaneous rectifying and inverting power. Thus, when hybrid AC/DC NMGs operate in night mode, it is necessary to satisfy at least one of the other conditions, aside from *Case 1.ii)*, to ensure the convexity of *Model III*.

2.5.2 Circumstance II: $o_{1,l}$ Values of BDCs in Different Intervals

The values of $o_{1,l}$ of different BDCs connected to the same branch may differ significantly, leading to a notable impact on the conditions outlined in Section 2.4. Hence, it is essential to examine these conditions based on whether the $o_{1,l}$ values of BDCs on the same branch belong to the same interval, i.e., $\forall l, 0 < o_{1,l} < 1$ or $\forall l, o_{1,l} = 1$ or $\forall l, o_{1,l} > 1$. If the $o_{1,l}$ values of all BDCs on the same branch do not fall within the same interval, identifying the sufficient conditions for *Model III* requires seeking the intersection of different conditions with the corresponding propositions. Therefore, we propose the following two corollaries.

Corollary 1. *For the case that* $\exists l_1 \in \mathcal{L}, 0 < o_{1,l_1} < 1$ *and* $\exists l_2 \in \mathcal{L}, o_{1,l_2} > 1$*, a solution to Model III where* $P^{\text{acdc}}_{i_1 j_1, l_1, t} P^{\text{dcac}}_{i_1 j_1, l_1, t} + P^{\text{acdc}}_{i_1 j_1, l_2, t} P^{\text{dcac}}_{i_1 j_1, l_2, t} > 0, \exists t$ *is suboptimal, if any of the following conditions is satisfied:*

1) $\exists t \in \mathcal{T}$*, such that* $0 < \rho^{\text{w}}_{i_1,t} < P^{\text{w}}_{i_1,t}$ *or* $0 < \rho^{\text{pv}}_{j_1,t} < P^{\text{pv}}_{j_1,t}$*;*

2) $\exists t \in \mathcal{T}$*, such that* $0 < P^{\text{de}}_{i_1,t} < \bar{P}^{\text{de}}_{i_1}, \underline{\lambda}^{\text{de1}}_{i_1,t} \geq c^{\text{de}}_{i_1}$ *or* $0 < P^{\text{fc}}_{j_1,t} < \bar{P}^{\text{fc}}_{j_1}, \underline{\lambda}^{\text{fc1}}_{j_1,t} \geq c^{\text{fc}}_{j_1}$*;*

3) $\exists t \in \mathcal{T}$*, such that* $\exists i' \in \pi(i_1)$*,* $H^{\text{ac,min}}_{i_1 i'} < H^{\text{ac}}_{i_1 i',t} < H^{\text{ac,max}}_{i_1 i'}$*,* $\beta^{\text{ac2}}_{i',t} = 0$*,* $U^{\text{ac,min}}_{i_1} < U^{\text{ac}}_{i',t} + \left(r^{\text{ac}}_{i_1 i'} H^{\text{ac}}_{i_1 i',t} + x^{\text{ac}}_{i_1 i'} G^{\text{ac}}_{i_1 i',t}\right)/U^{\text{ac}}_0 < U^{\text{ac,max}}_{i_1}$ *or* $\exists i \in \delta(i_1)$*,* $H^{\text{ac,min}}_{i i_1} < H^{\text{ac}}_{i i_1,t} < H^{\text{ac,max}}_{i i_1}$*,* $\beta^{\text{ac2}}_{i,t} = 0$*,* $U^{\text{ac,min}}_{i} < U^{\text{ac}}_{i_1,t} + \left(r^{\text{ac}}_{i i_1} H^{\text{ac}}_{i i_1,t} + x^{\text{ac}}_{i i_1} G^{\text{ac}}_{i i_1,t}\right)/U^{\text{ac}}_0 < U^{\text{ac,max}}_{i}$*;*

4) $\exists t \in \mathcal{T}$*, such that* $\exists j' \in \pi(j_1)$*,* $H^{\text{dc,min}}_{j_1 j'} < H^{\text{dc}}_{j_1 j',t} < H^{\text{dc,max}}_{j_1 j'}$*,* $\beta^{\text{dc2}}_{j',t} = 0$*,* $U^{\text{dc,min}}_{j_1} < U^{\text{dc}}_{j',t} + r^{\text{dc}}_{j_1 j'} H^{\text{dc}}_{j_1 j',t}/U^{\text{dc}}_0 < U^{\text{dc,max}}_{j_1}$*, or* $\exists j \in \delta(j_1)$*,* $H^{\text{dc,min}}_{j j_1} < H^{\text{dc}}_{j j_1,t} < H^{\text{dc,max}}_{j j_1}$*,* $\beta^{\text{dc2}}_{j,t} = 0$*,* $U^{\text{dc,min}}_{j} < U^{\text{dc}}_{j_1,t} + r^{\text{dc}}_{j j_1} H^{\text{dc}}_{j j_1,t}/U^{\text{dc}}_0 < U^{\text{dc,max}}_{j}$*.*

Proof: Since each of the four conditions in *Corollary 1* is derived from the intersection of the conditions presented in *Proposition 1* and *Proposition 3*, it logically follows that any of these conditions can guarantee that $\forall l_1, l_2 \in \mathcal{L}, P^{\text{acdc}}_{i_1 j_1, l_1, t} P^{\text{dcac}}_{i_1 j_1, l_1, t} + P^{\text{acdc}}_{i_1 j_1, l_2, t} P^{\text{dcac}}_{i_1 j_1, l_2, t} = 0$ holds in *Model III*. ■

Remark 3: Theoretically, any combination of a case presented in *Proposition 1* and a case in *Proposition 3* can guarantee that $\forall l_1, l_2 \in \mathcal{L}, P^{\text{acdc}}_{i_1 j_1, l_1, t} P^{\text{dcac}}_{i_1 j_1, l_1, t} + P^{\text{acdc}}_{i_1 j_1, l_2, t} P^{\text{dcac}}_{i_1 j_1, l_2, t} = 0$. Hence, *Corollary 1* could potentially list a total of $C^8_1 \times C^8_1 = 64$ combinations or cases to ensure the non-simultaneous rectification and inversion of BDCs. However, due to space limitations, it is important to note that *Corollary 1* does not present all possible conditions but rather selects a subset of conditions commonly employed in practical operations of hybrid AC/DC NMGs.

Corollary 2. *Including BDCs with* $o_{1,l} = 1$ *in Model III does not alter the sufficient conditions for the remaining BDCs to ensure that* $\forall l \in \mathcal{L}, P^{\text{acdc}}_{i_1 j_1,l,t} P^{\text{dcac}}_{i_1 j_1,l,t} = 0$.

Proof: This can be proven based on *Proposition 2*, which states that BDCs with $o_{1,l} = 1$ do not require any specific operational conditions to ensure their non-simultaneous rectification and inversion. Thus, to guarantee $\forall l \in \mathcal{L}, P^{\text{acdc}}_{i_1 j_1,l,t} P^{\text{dcac}}_{i_1 j_1,l,t} = 0$, it is sufficient to consider only the BDCs with $0 < o_{1,l} < 1$ or $o_{1,l} > 1$ and the corresponding sufficient conditions.■

2.5.3 Condition Extensions with Objective Modifications

In *Model I*, the curtailment costs of WTs and PVs are not included. However, in some studies, these costs are considered to facilitate the consumption of renewable energy. Here, we extend our explored conditions when the objective function is modified by including the penalty costs of curtailing wind power and PV power. Thus, the objective function in (2.1) is modified as follows.

$$\min \sum\nolimits_{t \in \mathcal{T}} \begin{pmatrix} \sum_{i \in \mathcal{I}} c_i^{\text{de}} P_{i,t}^{\text{de}} + \sum_{j \in \mathcal{J}} c_j^{\text{fc}} P_{j,t}^{\text{fc}} \\ + \sum_{l \in \mathcal{L}} c_{i_1 j_1,l} \left(P^{\text{acdc}}_{i_1 j_1,l,t} + P^{\text{dcac}}_{i_1 j_1,l,t} \right) \\ \sum_{i \in \mathcal{I}} c_i^{\text{w}} \rho_{j,t}^{\text{w}} + \sum_{j \in \mathcal{J}} c_j^{\text{pv}} \rho_{j,t}^{\text{pv}} \end{pmatrix} \tag{2.71}$$

Based on the modified objective function (2.71), we can derive the following corollary.

Corollary 3. *Considering the penalty cost of curtailing renewable energy* c_i^{w} *and* c_j^{pv}*, only Condition 1 in the Proposition 1 is required to be extended as follows:*
"1) $\exists t \in \mathcal{T}$*, such that* $0 \le \rho^{\text{w}}_{i_1,t} < P^{\text{w}}_{i_1,t}$, $\underline{\lambda}^{\text{w}}_{i,t} \ge c_i^{\text{w}}$ *or* $0 \le \rho^{\text{pv}}_{j_1,t} < P^{\text{pv}}_{j_1,t}$, $\underline{\lambda}^{\text{pv}}_{j_1,t} \ge c_j^{\text{pv}}$*".*

Proof: Using the stationarity condition of KKT conditions for *Model III*, we have

$$\partial \ell / \partial \rho^{\text{w}}_{i,t} = c_i^{\text{w}} - \underline{\lambda}^{\text{w}}_{i,t} + \bar{\lambda}^{\text{w}}_{i,t} - \beta^{\text{ac2}}_{i,t} = 0 \tag{2.72}$$

Applying the complementary slackness of KKT conditions to $0 \le \rho^{\text{w}}_{i_1,t} < P^{\text{w}}_{i_1,t}$, we obtain

$\bar{\lambda}^{\mathrm{w}}_{i_1,t} = 0$ and $\underline{\lambda}^{\mathrm{w}}_{i_1,t} \geq 0$. With $\underline{\lambda}^{\mathrm{w}}_{i,t} \geq c^{\mathrm{w}}_{i}$, Eq. (2.72) can be reformulated as $\beta^{\mathrm{ac2}}_{i,t} = c^{\mathrm{w}}_{i} - \underline{\lambda}^{\mathrm{w}}_{i,t} + \bar{\lambda}^{\mathrm{w}}_{i,t}$ $= c^{\mathrm{w}}_{i} - \underline{\lambda}^{\mathrm{w}}_{i,t} \leq 0$. Thus, we can obtain $f^{\mathrm{ac}}\left(o_{1,l}, \beta^{\mathrm{ac2}}_{i_1,t}\right) \geq 0$, which contradicts the necessary conditions of Theorem 1. This concludes that a solution with $0 \leq \rho^{\mathrm{w}}_{i_1,t} < P^{\mathrm{w}}_{i_1,t}$ and $\underline{\lambda}^{\mathrm{w}}_{i,t} \geq c^{\mathrm{w}}_{i}$ is suboptimal. Similarly, the condition of $0 \leq \rho^{\mathrm{pv}}_{j_1,t} < P^{\mathrm{pv}}_{j_1,t}$ and $\underline{\lambda}^{\mathrm{pv}}_{j_1,t} \geq c^{\mathrm{pv}}_{j}$ can be proved with the same manner. ■

Corollary 3 indicates that only Condition 1 of *Proposition 1* that is associated with renewable curtailment $\rho^{\mathrm{w}}_{j,t}$ and $\rho^{\mathrm{pv}}_{j,t}$ should be modified, while all other conditions in *Proposition 1* and *Proposition 3* remain unchanged. It is noteworthy that, *Corollary 3* can be seen as the generalization of *Proposition 1*. When ignoring the penalty cost of renewable curtailment, i.e., $c^{\mathrm{w}}_{i} = c^{\mathrm{pv}}_{j} = 0$, we can obtain *Corollary 3* is degenerate to *Proposition 1* since $\underline{\lambda}^{\mathrm{w}}_{i,t}$ and $\underline{\lambda}^{\mathrm{pv}}_{j_1,t}$ are the Lagrange multipliers that are larger than zero.

2.5.4 Condition Analysis with Newly Added Components

Various generation units, e.g., DEs, FCs, WTs, and PVs, have been included in *Model I*. To accommodate the ongoing introduction of new types of components into hybrid AC/DC NMGs, we further analyze the impacts of including new elements, e.g., battery energy storage systems, on our previously proposed conditions. Firstly, we derive the following corollary for any newly added components with convex models.

Corollary 4. *Assume that the model of a newly added component can be described independently from the existing ones and can be represented in convex mathematical forms. Then, all the conditions stated in Proposition 1 and Proposition 3 remain valid, even with the inclusion of the newly added component.*

Proof: Without loss of generality, we express the convex model of the newly added components as follows:

$$\varphi(\boldsymbol{P}^{\mathrm{ne}}, \boldsymbol{Q}^{\mathrm{ne}}, \boldsymbol{Y}^{\mathrm{ne}}) \leq 0 \tag{2.73}$$

where $\boldsymbol{P}^{\mathrm{ne}} = [P^{\mathrm{ne}}_{i,t}, P^{\mathrm{ne}}_{j,t}]$, $\boldsymbol{Q}^{\mathrm{ne}} = [Q^{\mathrm{ne}}_{i,t}]$.

When integrating a new component into the system, the modifications are typically limited to the system constraints, specifically the active and reactive power balance constraints (2.6), (2.7), and

(2.14). These constraints may need to be adjusted by incorporating corresponding variable terms. The modified system constraints can be expressed as follows:

$$\begin{aligned} P_{i,t}^{\mathrm{de}} - L_{i,t}^{\mathrm{ac}} + \sum\nolimits_{j_1 \in \Omega(i_1)} \sum\nolimits_{l \in \mathcal{L}} \left(P_{i_1 j_1, l, t}^{\mathrm{ac}} - P_{i_1 j_1, l, t}^{\mathrm{acdc}} \right) + P_{i,t}^{\mathrm{ne}} \\ = \sum\nolimits_{i' \in \pi(i)} H_{ii',t}^{\mathrm{ac}} - \sum\nolimits_{i'' \in \delta(i)} H_{i''i,t}^{\mathrm{ac}} - \left(P_{i,t}^{\mathrm{w}} - \rho_{i,t}^{\mathrm{w}} \right), \forall i, t : \beta_{i,t}^{\mathrm{ac2}} \end{aligned} \tag{2.74}$$

$$Q_{i,t}^{\mathrm{de}} - Q_{i,t}^{\mathrm{ac}} + Q_{i,t}^{\mathrm{ne}} = \sum\nolimits_{i' \in \pi(i)} G_{ii',t}^{\mathrm{ac}} - \sum\nolimits_{i'' \in \delta(i)} G_{i''i,t}^{\mathrm{ac}}, \forall i, t : \beta_{i,t}^{\mathrm{ac3}} \tag{2.75}$$

$$\begin{aligned} P_{j,t}^{\mathrm{fc}} - L_{j,t}^{\mathrm{dc}} + \sum\nolimits_{i_1 \in \Omega(j_1)} \sum\nolimits_{l \in \mathcal{L}} \left(P_{i_1 j_1, l, t}^{\mathrm{dc}} - P_{i_1 j_1, l, t}^{\mathrm{dcac}} \right) + P_{j,t}^{\mathrm{ne}} \\ = \sum\nolimits_{j' \in \pi(j)} H_{jj',t}^{\mathrm{dc}} - \sum\nolimits_{j'' \in \delta(j)} H_{j''j,t}^{\mathrm{dc}} - \left(P_{j,t}^{\mathrm{pv}} - \rho_{j,t}^{\mathrm{pv}} \right), \forall j, t : \beta_{j,t}^{\mathrm{dc2}} \end{aligned} \tag{2.76}$$

As the model of newly added components is independent of the models of other components, the inclusion of new component variables does not impact the derivation process of the conditions outlined in *Proposition 1* and *Proposition 3*. These conditions remain valid and can be derived without alterations. ■

2.5.5 Iterative Process for Obtaining the Final Solution

While the operation of a hybrid AC/DC NMG generally satisfies one of the conditions outlined in *Propositions 1-3*, there exists a slight possibility that the complementarity constraint (2.19) cannot be fully satisfied. To address this issue, we propose an iterative verification process to dynamically ensure the complete satisfaction of the non-convex constraint (2.19) in the final solution. The steps of this iterative procedure are outlined below.

Step 1: Initialize an empty set $\mathcal{S}$ to represent the pair indices (l', t') for BDCs. This set captures any specific instances where the sufficient conditions derived in Section 2.4 are not met, ensuring the model handles the worst-case scenarios where convex relaxation would otherwise fail.

Step 2: Solve *Model III* with the additional constraints $P_{i_1 j_1, l', t'}^{\mathrm{acdc}} \cdot P_{i_1 j_1, l', t'}^{\mathrm{dcac}} = 0, \forall (l', t') \in \mathcal{S}$.

Step 3: Compute the complementarity gap value $\varepsilon_{l,t}$ for each BDC *l* at every time period *t* as follows:

$$\varepsilon_{l,t} = P_{i_1 j_1, l, t}^{\mathrm{acdc}} \cdot P_{i_1 j_1, l, t}^{\mathrm{dcac}}, \forall l \in \mathcal{L}, t \in \mathcal{T} \tag{2.77}$$

Step 4: Verify that the complementarity gap values $\varepsilon_{l,t}$ are zero for all BDCs at all periods. If all $\varepsilon_{l,t}$ are zero, it indicates that $\mathcal{S}$ requires no additional pair indices (l', t').

Step 5: If the complementarity gap value $\varepsilon_{l,t}$ is greater than zero for certain pair indices (l', t'), update $\mathcal{S}$ by adding the pair indices (l', t') corresponding to $\varepsilon_{l',t'} > 0$, and return to *Step 2*.

Steps 2-5 are repeated until the obtained economic dispatch solution can fully satisfy the non-convex complementarity constraint (2.19). This iterative process dynamically identifies active non-convex complementarity constraints (2.19) during the solution process, minimizing the number of non-convex constraints required in the economic dispatch model. This approach accelerates the solution process while ensuring that BDCs do not rectify and invert simultaneously.

2.6 Case Studies

To verify the rectification and inversion behaviors of the BDC-based models (*Model I*, *Model II* and *Model III*), simulations were conducted using a typical topology of a hybrid AC/DC NMG as shown in Fig. 2.1. We note that a standardized test system for the hybrid AC/DC NMG does not currently exist. The modified IEEE 33-node system, enhanced with PVs and WTs, is commonly used to validate the performance of economic dispatch strategies in microgrids [130]. Therefore, this paper considers a 66-node hybrid AC/DC NMG that employs a modified IEEE 33-node framework for both ACsM and DCsM. The detailed parameters of the hybrid AC/DC NMG are presented in Table 2.1. The simulations were carried out using GAMS 24.4 software on a computer with an Intel Core i7-8700 CPU and 16 GB RAM. For the non-convex *Model I* and *Model II*, multiple solvers including MOSEK, BARON, CONOPT, IPOPT, and SNOPT, were employed to find the optimal solutions. The convex *Model III* was solved using the GUROBI solver. In all cases, the algorithm gap was set to zero to obtain the optimal solution for each model. The daily wind and solar power outputs in the simulations were derived from practical data obtained from the Northwestern China power system [128]. These realistic power outputs were incorporated into the models to accurately evaluate their performance.

Table 2.1: Parameters of the hybrid AC/DC NMG

Parameter	Value	Parameter	Value	Parameter	Value
c_i^{de} (¥/MWh)	1200	$\bar{P}_i^{\text{de}}$ (MW)	0.8	U_0^{ac} (p.u.)	1.05
c_j^{fc} (¥/MWh)	1200	$\bar{P}_j^{\text{fc}}$ (MW)	0.8	$\bar{U}_i^{\text{ac}}$ (p.u.)	1.05
$c_{i_1 j_1,l}$ (¥/MWh)	60	$P_{j,t}^{\text{pv}}$ (MW)	0.5	$\underline{U}_i^{\text{ac}}$ (p.u.)	0.95
$\bar{P}_{i_1 j_1,l}^{\text{acdc}}$ (MW)	0.4	$P_{i,t}^{\text{w}}$ (MW)	1.0	$\bar{U}_j^{\text{dc}}$ (p.u.)	1.05
$\bar{P}_{i_1 j_1,l}^{\text{dcac}}$ (MW)	0.4	$\bar{Q}_i^{\text{de}}$	0.18	$\underline{U}_j^{\text{dc}}$ (p.u.)	0.95

2.6.1 Error Comparisons of Different BDC Efficiency Models

In this section, we compare the proposed BDC approximation model with the typical fixed constant model [23], [24] and the Chebyshev uniform linear approximation-based model[34]. To quantify the performance of these models, we define the absolute error δ^{Abs} and the relative error δ^{Rel} as follows:

$$\delta^{\text{Abs}} = \frac{1}{|\mathcal{L}|} \sum\nolimits_{l \in \mathcal{L}} \int_{[0,1]} \sqrt{\left[\varphi_l(x) - \eta_{\text{BDC},l}^{\text{dcac}}(x)\right]^2} dx \tag{2.78}$$

$$\delta^{\text{Rel}} = \frac{\sum_{l \in \mathcal{L}} \int_{[0,1]} \sqrt{\left[\varphi_l(x) - \eta_{\text{BDC},l}^{\text{dcac}}(x)\right]^2} dx}{\sum_{l \in \mathcal{L}} \int_{[0,1]} \eta_{\text{BDC},l}^{\text{dcac}}(x) dx} \tag{2.79}$$

Here, we have chosen two commonly used types of BDCs in [27] and [28] to evaluate approximation errors. The results comparing different models with these BDCs are presented in Table 2.2. It can be observed that, the fixed constant model of BDC efficiency, which ignores the dynamic characteristics of BDC efficiency, exhibits the highest errors, with relative errors of 5.651% and 33.427% for the BDC parameters in [27] and [28], respectively. This is because the fixed model assumes a constant efficiency, which does not capture the variations in BDC efficiency under different operating conditions. The Chebyshev-based model, while outperforming the fixed model by reducing the maximum value of relative and absolute errors, fails to achieve the lowest overall average error. In contrast, the proposed approximation method is designed to minimize the overall approximation error over the entire operating range of the BDC, achieving a more than threefold reduction in relative error compared to the fixed constant model. Thus, the proposed approximation method can achieve the optimal value of the average error.

Table 2.2: Error comparisons of using different efficiency models

BDC Efficiency Models	BDC Parameters in [27]		BDC Parameters in [28]	
	δ^{Abs}	δ^{Rel}	δ^{Abs}	δ^{Rel}
Fixed Constant[23], [24]	2.756×10^{-2}	5.651%	1.509×10^{-1}	33.427%
Chebyshev[34]	0.760×10^{-2}	1.558%	6.719×10^{-2}	14.887%
Proposed LSA	0.452×10^{-2}	0.926%	4.108×10^{-3}	9.102%

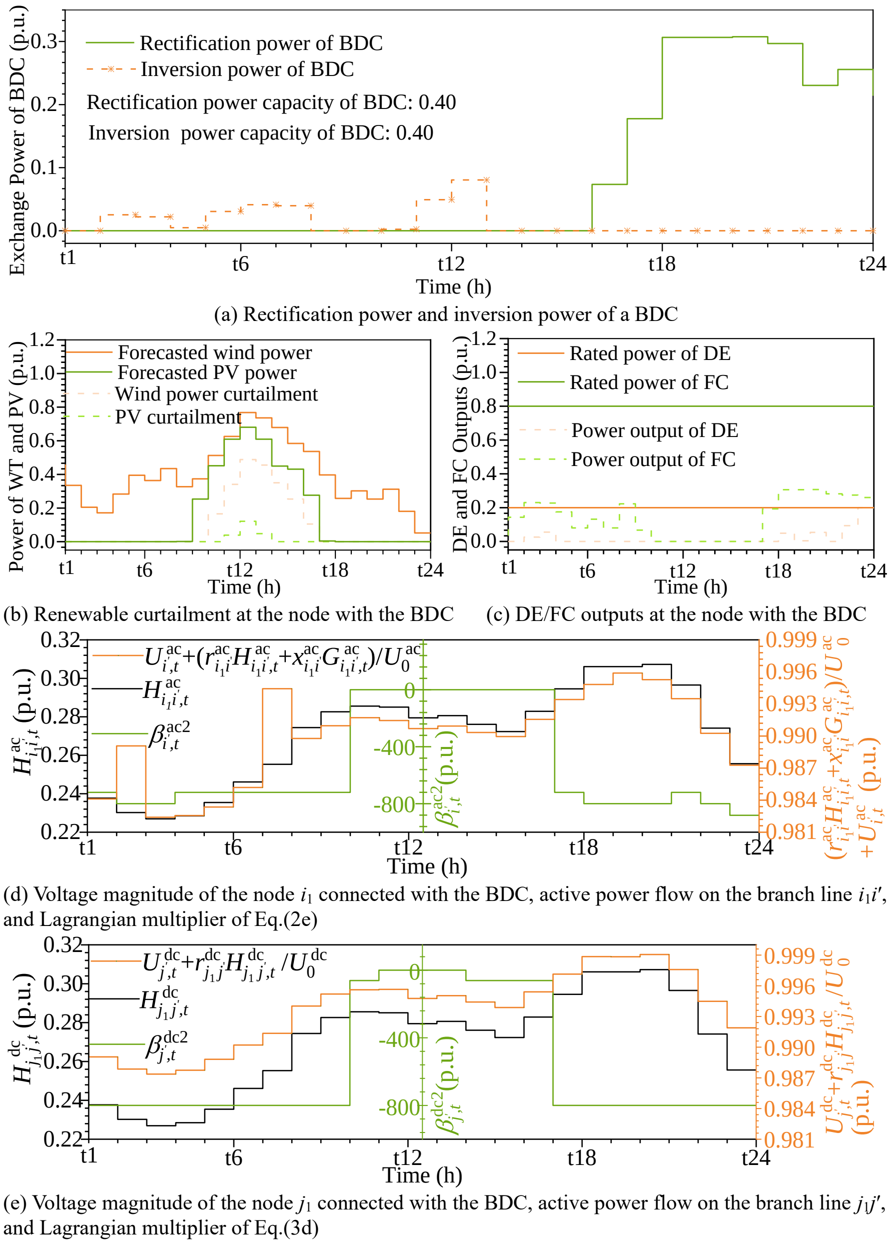
2.6.2 Analysis of BDC Rectifying-Inverting Behaviors

We first examine the non-simultaneously rectifying-inverting behaviors of BDCs that satisfy at least one of the conditions in *Proposition 1*. We then discuss the simultaneously rectifying-inverting behaviors of BDCs in specific time periods that do not satisfy any of the conditions in *Proposition 3*.

2.6.2.1 Non-simultaneously Rectifying-Inverting Behaviors

To investigate the rectifying-inverting behaviors of the BDC, we conduct convex *Model III* with a single BDC $0 < o_{1,l} < 1$. The analysis of these behaviors is shown in Fig. 2.3.

We can see that the BDC ensures non-simultaneously rectifying-inverting operations in all time periods. In Fig. 2.3(a), we observe that the wind power curtailment satisfies *case 1.i)* in Condition 1, while the PV curtailment during periods t11−t18 satisfies *case 1.ii)* in Condition 1. Moreover, the renewable power outputs and curtailment during time periods t1−t10 and t19−t24 correspond to Circumstance 1 in Section V, where the predicted value of PV generation is zero. Additionally, Fig. 2.3(c) fulfills Condition 2, as the power outputs of DEs and FCs during periods t1−t10 and t18−t24 satisfy it. On the other hand, Fig. 2.3(d) represents the case when Condition 3 is met, while Fig. 2.3(e) demonstrates the case when Condition 4 is satisfied. Overall, the power profiles of the BDC in Fig. 2.3 confirm that the non-simultaneous rectification power and inversion power can be achieved in all time periods. This is because, for any given time period, at least one of the conditions in *Proposition 1* is satisfied, as verified by Fig. 2.3(b)−Fig. 2.3(e).



(a) Rectification power and inversion power of a BDC

(b) Renewable curtailment at the node with the BDC

(c) DE/FC outputs at the node with the BDC

(d) Voltage magnitude of the node i_1 connected with the BDC, active power flow on the branch line i_1i', and Lagrangian multiplier of Eq.(2e)

(e) Voltage magnitude of the node j_1 connected with the BDC, active power flow on the branch line j_1j', and Lagrangian multiplier of Eq.(3d)

Figure 2.3: Verification of the four conditions in Proposition 1 to ensure non-simultaneously rectifying-inverting operation of the BDC when $0 < o_{1,l} < 1$.

2.6.2.2 Simultaneously Rectifying-Inverting Behaviors

The behavior of the BDC in simultaneously rectifying and inverting power is investigated using the convex *Model III* with a single BDC $o_{1,l} > 1$. The verification is done by analyzing BDC rectifying-inverting behaviors during specific time periods, as shown in Fig. 2.4. From the power profiles in Fig. 2.4(a) it can be observed that non-simultaneous rectifying-inverting operations of the BDC are guaranteed during periods t11–t17. This is because at least one of the conditions in *Proposition 3* is satisfied, e.g., $0 < \rho^{\mathrm{w}}_{i_1,t} \leq P^{\mathrm{w}}_{i_1,t}$ in Condition 1 and $0 \leq P^{\mathrm{de}}_{i_1,t} < \bar{P}^{\mathrm{de}}_{i_1}, \lambda^{\mathrm{de1}}_{i_1,t} \geq c^{\mathrm{de}}_{i_1}$ in Condition 2. However, it should be noted that during periods t1–t10 and t18–t24, simultaneous rectification and inversion of the BDC occur when none of the conditions in *Proposition 3* are satisfied.

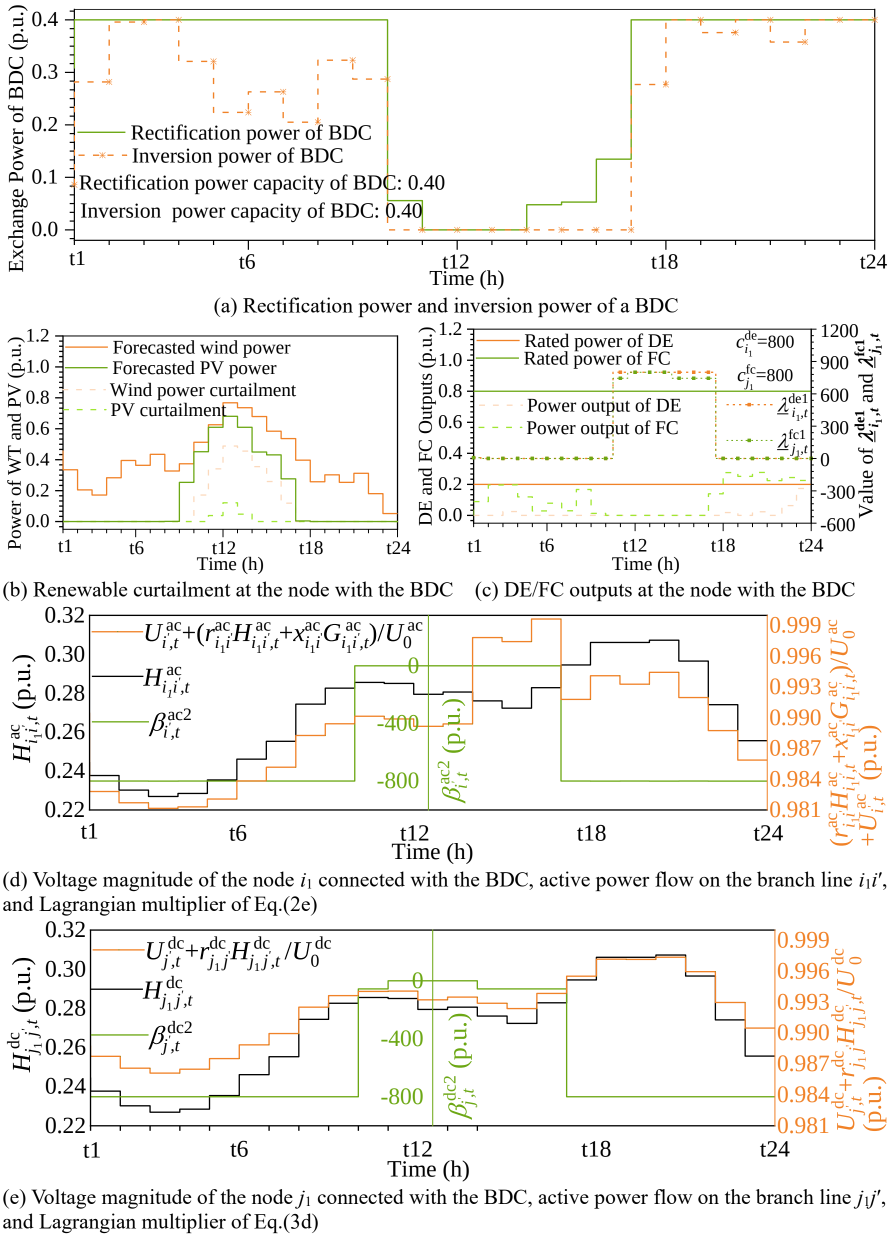


Figure 2.4: Simultaneously rectifying-inverting operation behaviours of the BDC when $o_{1,l} > 1$ in Proposition 3.

2.6.3 Comparative Analysis of Different Models

To quantitatively evaluate the performance of the proposed relaxed model, several performance indexes are defined for the purpose of model comparisons, presented as follows:

2.6.3.1 Simultaneously Rectifying and Inverting Index (SRII)

is defined as a metric to calculate the complementarity gap for all BDCs over the entire time horizon.

$$\mathrm{SRII} = \sum\nolimits_{l\in\mathcal{L},t\in\mathcal{T}} \left(P^{\mathrm{acdc}}_{i_1j_1,l,t} \cdot P^{\mathrm{dcac}}_{i_1j_1,l,t}\right) \tag{2.80}$$

2.6.3.2 Rectifying and Inverting Power Amount Index (RIPAI)

is defined as the sum of the rectifying power and inverting power of all BDCs over the entire time horizon.

$$\mathrm{RIPAI} = \sum\nolimits_{l\in\mathcal{L},t\in\mathcal{T}} \left(P^{\mathrm{acdc}}_{i_1j_1,l,t} + P^{\mathrm{dcac}}_{i_1j_1,l,t}\right) \tag{2.81}$$

2.6.3.3 Rectifying and Inverting Power Difference Index (RIPDI)

is defined to measure the difference between the rectifying and inverting power outputs of *Model II* and *Model III*. If the solutions of *Model II* and *Model III* are identical, the RIPDI should be zero.

$$\begin{aligned}\mathrm{RIPDI} = \sum\nolimits_{t\in\mathcal{T}} \Big(& \Big|\sum\nolimits_{l\in\mathcal{L}} P^{\text{acdc-III}}_{i_1j_1,l,t} - \sum\nolimits_{l\in\mathcal{L}} P^{\text{acdc-II}}_{i_1j_1,l,t}\Big| \\ & + \Big|\sum\nolimits_{l\in\mathcal{L}} P^{\text{dcac-III}}_{i_1j_1,l,t} - \sum\nolimits_{l\in\mathcal{L}} P^{\text{dcac-II}}_{i_1j_1,l,t}\Big|\Big)\end{aligned} \tag{2.82}$$

To assess the performance of different models when dealing with uncertainties in WTs and PVs, we extend the models by incorporating scenario-based stochastic programming techniques [131]. Specifically, we consider 360 daily scenarios, each encompassing 24 hours with one-hour dispatch intervals, to comprehensively capture the power output uncertainty of WTs and PVs. The comparison results between the proposed *Model III*, *Model I*, a simplified version of *Model I* referred to as *S-Model I*, and *Model II* are summarized in Table 2.3. Here, *S-Model I* is obtained by simplifying *Model I* through the replacement of the complex expressions for BDC dynamic conversion efficiency

$\eta_{\mathrm{BDC},l}^{\mathrm{acdc/dcac}}$ in (2.22) with the cubic polynomial expressions from [28].

The results indicate that, *Model I* is not solvable by any off-the-shelf solvers due to its complex combination expressions involving trigonometric and exponential functions, while *S-Model I* allows specialized solvers to solve it by simplifying trigonometric functions with non-convex cubic polynomial expressions. Moreover, the proposed *Model II* accelerates this process by using approximations with convex expressions, broadening the range of solvers capable of handling the problem. However, the solution time can be long, up to 139 hours, and there is a high risk of obtaining locally optimal solutions. By comparison, the proposed *Model III* significantly speeds up the solving process while achieving a globally optimal solution. It achieves speed-up factors of 117,320, and 90,973 for CONOPT, IPOPT, and SNOPT solvers, respectively, compared to *Model II*. The solution obtained by *Model III* is identical to that of *Model II*. Furthermore, the value of SRII in *Model III* is exactly zero, which verifies the exact relaxation of the non-simultaneously inverting-rectifying operations of each BDC.

Table 2.3: Model solution comparisons under single BDC

Models	Solver	SRII	PIPAI (kWh)	RIPDI (kWh)	Cost (¥)	Time (s)
Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	NA	NA	NA	NA	NA
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
S-Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	737961	–	1828198.4	941.80
	IPOPT	0	756035	–	1892778.2	36941.20
	SNOPT	NA	NA	NA	NA	NA
Model II	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	737436	0	1827886.1	644.63
	IPOPT	0	737436	0	1827886.1	1766.88
	SNOPT	0	737436	0	1914453.0	501262.53
Model III	GUROBI	0	737436	0	1827886.1	5.51

2.6.4 Validation of the Proposed Iterative Process Method

To assess the effectiveness of the proposed iterative process method, we conduct a comparative analysis of *Model II*, *Model III*, and *Model III* integrated with the iterative process across varying parameters for two distinct BDCs: one with a parameter $0 < o_{1,l_1} < 1$ and another with $o_{1,l_2} > 1$. The comparative results are detailed in Table 2.4. The comparison reveals that, when compared to *Model III* without the iterative process, *Model III* with the proposed iterative process method can ensure that BDCs do not simultaneously rectify and invert, as indicated by the value of SRII being exactly zero. Furthermore, although the iterative process may require additional computational time, the proposed method still achieves a reduction in solving time of more than 20 times compared to *Model II*. This improvement in efficiency is attributed to the dynamic identification of non-convex constraints in the economic dispatch model of hybrid AC/DC NMGs.

Table 2.4: Verification of the proposed iterative process method

Models	Solver	SRII	PIPAI (kWh)	RIPDI (kWh)	Cost (¥)	Time (s)
Model II	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	1683616	0	1795357.0	996.03
	IPOPT	0	1683616	0	1795357.0	2651.33
	SNOPT	NA	NA	NA	NA	NA
Model III	GUROBI	40945	29661967	0	1779718.0	5.90
Model III-Iterative	GUROBI+ CONOPT	0	1683616	0	1795357.0	44.16

2.6.5 Dynamic Efficiency Validation of BDCs

To validate the efficacy of the dynamic conversion efficiency of BDCs, we conduct a comparative analysis between the proposed dynamic efficiency model and the static conversion efficiency models. These static models include a lossless efficiency model with a presumed 100% conversion efficiency [23], as well as a scale-based model with a fixed 95% conversion efficiency [24]. Additionally, we tested the dynamic efficiency model across different voltage scenarios, specifically at 0.95 p.u., 1.00 p.u., and 1.05 p.u., to assess its performance under varying conditions. Comparative results in terms of the total operation costs and BDC power losses are depicted in Fig. 2.5.

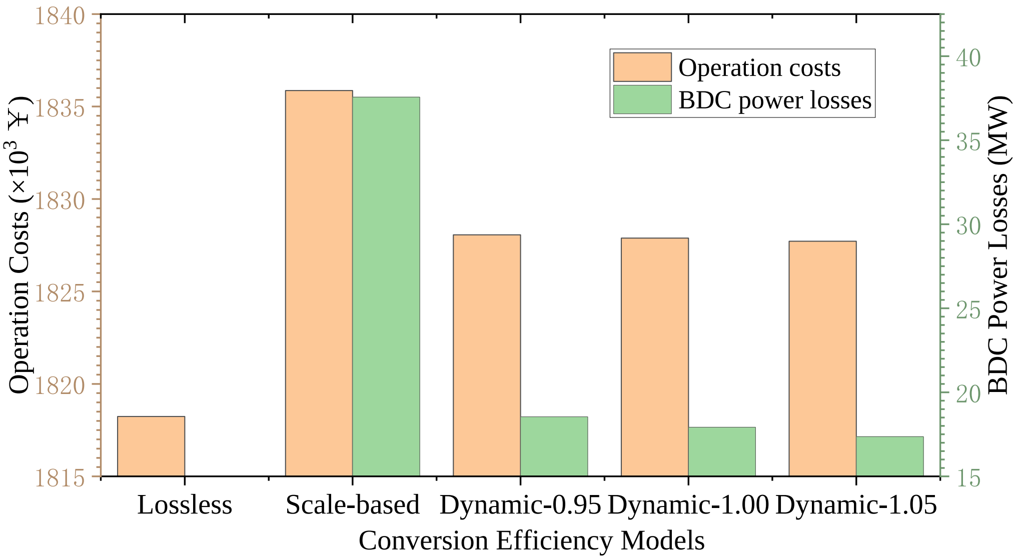


Figure 2.5: Evaluation of BDC efficiency models.

The lossless efficiency model yields the least operational cost because it is overly idealistic and does not account for any losses during BDC conversion. Conversely, the scale-based model results in the highest operational costs, attributed to its static nature and inability to adapt to changing conditions. In contrast, the dynamic model is capable of identifying the most favorable conversion efficiency by considering the current operational state. It is important to highlight that, when assessing economic dispatch costs, the influence of voltage values on the efficiency models is relatively minor compared to the differences between the conversion efficiency models themselves.

2.6.6 Impact of Renewable Penetration on Model Solutions

In this paper, we consider WTs and PVs to have capacities of 1.0 MW and 0.5 MW, respectively. To assess how different levels of renewable energy capacity integration affect our model solutions' performance, we experiment with various renewable energy penetration scenarios within the hybrid AC/DC NMG. Both *S-Model I* and *Model II* are processed using the CONOPT solver, whereas *Model III* is tackled with the GUROBI solver. As shown in Table 2.5, *Model III* maintains the BDC's ability to prevent simultaneous rectification and inversion, similar to *Model II*, regardless of the renewable energy capacity penetration level. Besides, *Model III* presents a significant advantage by slashing the computation time by more than two magnitudes compared to previous models.

Table 2.5: Comparative analysis of model solutions across different levels of renewable energy capacity

Capacity (WTs, PVs)	Models	SRII	PIPAI (kWh)	RIPDI (kWh)	Cost (¥)	Time (s)
(0.5 MW, 0.5 MW)	S-Model I	0	466216	–	2285760.7	678.24
	Model II	0	465771	0	2285456.4	492.26
	Model III	0	465771	0	2285456.4	4.35
(1.0 MW, 0.5 MW)	S-Model I	0	737961	–	1828198.4	941.80
	Model II	0	737436	0	1827886.1	644.63
	Model III	0	737436	0	1827886.1	5.51
(1.0 MW, 1.0 MW)	S-Model I	0	807884	–	1616722.1	654.69
	Model II	0	807331	0	1616456.7	500.06
	Model III	0	807331	0	1616456.7	4.26
(1.5 MW, 1.0 MW)	S-Model I	0	952745	–	1413167.4	679.69
	Model II	0	952143	0	1412907.3	477.36
	Model III	0	952143	0	1412907.3	4.32

2.6.7 Impact of Integrating Energy Storage on Model Solutions

In this section, we further analyze the effects of integrating energy storage into the operational models. The specific energy storage models and parameters are obtained from [24]. The results are detailed in Table 2.6, allowing for a comparison with the previously discussed outcomes in Table 2.3. The inclusion of energy storage systems enables a temporal linkage of microgrid operation, while increasing the complexity and computational demands of four models. Nonetheless, the integration of energy storage can lead to lower costs for microgrid operation. This cost reduction arises from the energy storage system's capability to store excess renewable energy during periods of peak production and release it during periods of high demand. Consequently, energy storage acts as a buffer, optimizing the use of renewable sources and mitigating the fluctuations in renewable energy supply. Ultimately, this leads to more efficient and cost-effective microgrid operations.

2.6.8 Impact of BDC Number on Model Solutions

To identify the influence of the number of BDCs on the solutions of *Model I*, *S-Model I*, *Model II*, and *Model III*, we conduct case testing with two and ten BDCs. The results are summarized in Table 2.7 and Table 2.8. Overall, we observe that the solving time for the three models generally increases as the number of BDCs increases. Specifically, for *Model I*, it remains intractable for all off-the-shelf

Table 2.6: Model solution comparisons with energy storage integration

Models	Solver	SRII	PIPAI (kWh)	RIPDI (kWh)	Cost (¥)	Time (s)
Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	NA	NA	NA	NA	NA
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
S-Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	760239	–	1784896.1	1091.03
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
Model II	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	759726	0	1784579.0	846.11
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
Model III	GUROBI	0	759726	0	1784579.0	7.05

solvers regardless of the number of BDCs. In the case of *Model II*, the computation time required for solving the model increases as the number of BDCs increases. For *Model III*, the solving time remains basically unchanged regardless of the number of BDCs. Thus, in contrast to *Model II*, *Model III* does not impose specific solver requirements and is not significantly affected by the number of BDCs.

Table 2.7: Model solution comparisons under two BDCs

Models	Solver	SRII	PIPAI (kWh)	RIPDI (kWh)	Cost (¥)	Time (s)
Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	NA	NA	NA	NA	NA
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
S-Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	738172	–	1829003.1	965.94
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
Model II	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	737613	0	1828520.6	927.72
	IPOPT	0	737613	0	1828520.6	20369.67
	SNOPT	0	737613	0	1828520.6	117491.88
Model III	GUROBI	0	737613	0	1828520.6	5.72

Table 2.8: Model solution comparisons under 10 BDCs

Models	Solver	SRII	PIPAI (kWh)	RIPDI (kWh)	Cost (¥)	Time (s)
Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	NA	NA	NA	NA	NA
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
S-Model I	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	739593	–	1834085.7	1158.11
	IPOPT	NA	NA	NA	NA	NA
	SNOPT	NA	NA	NA	NA	NA
Model II	MOSEK	NA	NA	NA	NA	NA
	BARON	NA	NA	NA	NA	NA
	CONOPT	0	739033	0	1833601.1	991.83
	IPOPT	0	739033	0	1833601.1	21724.06
	SNOPT	NA	NA	NA	NA	NA
Model III	GUROBI	0	739033	0	1833601.1	5.99

2.6.9 Result Verifications on Various Test Systems

To further validate the effectiveness of the proposed methods, we apply them to various MG test systems, including the practical China-Shangyu MG test system from [115] (TS1) and the 30-node MG test system which consists of two 15-node MG systems from [132] (TS2). We utilize 365 daily scenarios to simulate uncertainties from WTs and PVs. *Model I*, *S-Model I*, and *Model II* are solved using the CONOPT solver, while *Model III* is solved using the GUROBI solver. The results are summarized in Table 2.9.

We observe that *Model I* fails to solve both test systems due to the overly complex trigonometric expressions of BDC efficiency. In contrast, our developed convex *Model III* achieves an identical cost to *Model II* but significantly faster, completing within 5 seconds. Additionally, the SRII value in *Model III* is zero for all 365 scenarios, indicating that our explored sufficient conditions fully ensure the non-simultaneous rectifying and inverting behaviors of BDCs. These results further verify that our developed methods are applicable to various test systems.

Table 2.9: Model solution comparisons under various test systems

Test Systems	Models	SRII	Cost (¥)	Time (s)
TS1	Model I	NA	NA	NA
	S-Model I	0	157917.5	789.57
	Model II	0	157311.0	205.45
	Model III	0	157311.0	3.29
TS2	Model I	NA	NA	NA
	S-Model I	0	16370765.1	712.08
	Model II	0	16377641.3	281.58
	Model III	0	16377641.3	4.77

2.7 Summary

To enhance the computational tractability and efficiency of economic dispatch in BDC-based hybrid AC/DC NMGs, we propose two key techniques to achieve a fully convex representation of the BDC models. Firstly, the LSA method was employed to convexify the highly non-linear and generally non-convex BDC efficiency expressions while minimizing the global approximation error. This enables us to represent these expressions in a convex form, thereby facilitating computation.

Secondly, we relax the non-linear complementary constraints associated with identifying BDC conversion directions by thoroughly exploring sufficient conditions that ensure non-simultaneous rectification and inversion behaviors of BDCs. This exploration further results in a convex model for the high-efficiency operation of hybrid AC/DC NMGs equipped with numerous BDCs. We further tailor these conditions for various real-world scenarios to enhance their practicality. Numerical studies indicate that our approximation model outperforms state-of-the-art BDC efficiency models in terms of accuracy. The convex model, augmented with the explored sufficient conditions, accurately performs non-simultaneous rectification and inversion actions. However, it significantly reduces the solution time by over two orders of magnitude. Our findings are particularly impactful for stochastic programming or robust optimization methods used in hybrid NMGs. These methods rely on convex optimization models to ensure convergence of solution algorithms, e.g., Benders decomposition algorithms or C&CG algorithms.

Chapter 3

Heterogeneous-of-Expert Multi-Task Learning for Flexible Load Identification

Many NILM approaches to decomposing household-level energy consumption into appliance-specific usage patterns using aggregated smart-meter data in conjunction with advanced deep learning techniques are able to leverage the complementary information existing between the power consumption estimation and appliance state detection tasks using multi-task frameworks. However, existing multitask frameworks suffer from three primary limitations, including a reliance on a single shared neural network that is unable to extract features for both tasks fully, particularly under long sampling intervals on the order of 15 min, the application of homogeneous neural network architectures that are unable to extract the diverse characteristics of NILM data, and the reliance on static weights to balance the competing task objectives. The present work addresses these issues by developing a multiple and heterogeneous deep learning expert-based approach to NILM conducted using low sampling frequency smart-meter data. Specifically, the proposed approach leverages a multi-gate mixture-of-experts architecture comprising five different neural network architectures, to capture both local and global time-series features effectively. Additionally, a novel multi-objective optimization strategy is designed in conjunction with the gradient descent algorithm to manage tradeoffs between the appliance state detection and power consumption estimation tasks dynamically by optimizing the weights for each task. The performance of the proposed approach is compared with that of existing state-of-the-art methods. Generalizability is validated through cross-household testing on unseen homes in the public UK-DALE and REDD datasets, while robustness to realistic data sparsity is demonstrated on a real-world dataset sampled at challenging 15-minute intervals. The results obtained for the proposed approach demonstrate significant improvements in state classification accuracy and power estimation error, particularly for appliances with complex power consumption patterns.

3.1 NILM Problem and Objectives

The current NILM problem denotes the power load of the i-th household electrical appliance at uniform time intervals t over a time period T as $\boldsymbol{x}_i = [x_i(1), x_i(2), ..., x_i(T)] \in \mathcal{X}$ and the time-series aggregated input data as $\boldsymbol{Y} = [\boldsymbol{y}(1), \boldsymbol{y}(2), ..., \boldsymbol{y}(T)] \in \mathcal{Y}$. For $\forall t \in T$, NILM problems can be formulated by minimizing the following defined deviation $\varepsilon(t)$ between the aggregated load measured by smart meters $\boldsymbol{y}(t)$ and the combination of $x_i(t)$ and the state $s_i(t)$ of the i-th appliance:

$$\min \varepsilon(t) = |\boldsymbol{y}(t) - \sum\nolimits_{i=1}^{I} s_i(t) x_i(t)|. \tag{3.1}$$

$$\text{s.t.} \quad s_i(t) \in \{0, 1\}, x_i(t) \geq 0, \quad \forall i \in \mathcal{I}, t \in T. \tag{3.2}$$

When the on/off state $s_i(t)$ of the i-th appliance is inferred from estimated power values instead of directly from switch state information, it can be determined using the following rule:

$$s_i(t) = \begin{cases} 0, \ i\text{th appliance is off, } x_i(t) < \varrho_i \\ 1, \ i\text{th appliance is on, } x_i(t) \geq \varrho_i \end{cases} \tag{3.3}$$

Here, ϱ_i denotes the power threshold associated with the i-th appliance. It is important to highlight that, in single-task DLMs, the choice of ϱ_i significantly impacts the inferred appliance state. This threshold is often tailored to specific appliances and datasets, which introduces subjectivity and limits the model's generalization capability. Furthermore, many appliances exhibit low-power idle states or noisy power usage patterns, making it difficult to reliably distinguish between on and off states using power values alone.

In contrast, multitask DLM frameworks can directly predict both the binary on/off states and the corresponding power consumption without requiring manual threshold tuning. Specifically, the NILM process considered herein implicitly involves two simultaneous tasks: 1) determining whether the state of the appliance is on and 2) estimating how much power the appliance consumes. Hence, the objective is to obtain an effective neural network $\mathcal{F}$ that can simultaneously identify the on/off states of appliances by conducting classification tasks, and specify the detailed power consumption of appliances by conducting regression tasks. The mapping process for the two tasks can be expressed as follows:

$$\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, ..., \hat{\boldsymbol{x}}_i, ... = \hat{\mathcal{X}} = \mathcal{F}(\mathcal{Y}; \Theta), \tag{3.4}$$

$$\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, ..., \hat{\boldsymbol{s}}_i, ... = \hat{\mathcal{S}} = \text{softmax}(\mathcal{F}(\mathcal{Y}; \Theta)), \tag{3.5}$$

where Θ represents the parameter set of the neural network, and softmax$(\cdot)$ denotes the softmax function that normalizes the output of $\mathcal{F}$ into a probability distribution. This multitask formulation promotes task synergy, allowing the classification output to support the regression task, particularly during periods of low appliance activity and ambiguous signals.

3.2 HMMOE Framework

3.2.1 Heterogeneous Network Architecture

The proposed HMMOE framework architecture is illustrated in Fig. 3.1. As can be seen, the model includes three main components: information sharing by diverse expert blocks 1–5, multiple gates A and B, and task towers A and B.

Information Sharing: In contrast to a previously proposed HMMOE framework that employs the same fully connected neural network in the information sharing layer [133], the proposed approach extracts expert feature information using various neural networks, including a DNN (Expert 1) and a CNN [134] (Expert 2) to extract local features, and a transformer network (Trm) [135] (Expert 3), a TCN [136] (Expert 4), and an LSTM [137] (Expert 5) to capture global correlation. Specifically,

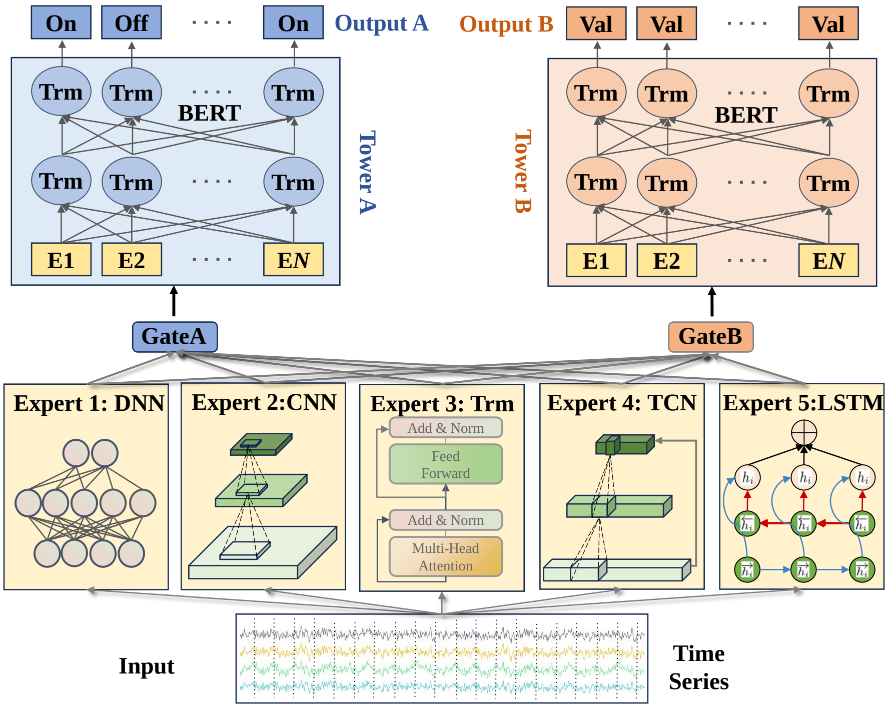


Figure 3.1: HMMOE architecture.

the mapping processes of these five heterogeneous expert networks can be expressed as follows.

$$f_{\mathrm{DNN}}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_{\mathrm{DNN}}) \tag{3.6}$$

$$f_{\mathrm{CNN}}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_{\mathrm{CNN}}) \tag{3.7}$$

$$f_{\mathrm{Trm}}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_{\mathrm{Trm}}) \tag{3.8}$$

$$f_{\mathrm{TCN}}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_{\mathrm{TCN}}) \tag{3.9}$$

$$f_{\mathrm{LSTM}}(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_{\mathrm{LSTM}}) \tag{3.10}$$

For convenience of representation, these five expressions are presented in the following unified form:

$$f_e(\boldsymbol{x}) = \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_e), \forall e \in \mathcal{E}, \tag{3.11}$$

where $\mathcal{E} = \{\mathrm{DNN}, \mathrm{CNN}, \mathrm{Trm}, \mathrm{TCN}, \mathrm{LSTM}\}$.

Additionally, we note that the modular design of the HMMOE framework makes it inherently adaptable and easy to upgrade by allowing new expert models or improved versions of existing models to be incorporated seamlessly, and thereby ensuring that the framework evolves with advancements in artificial intelligence. It is worth pointing out that, multiple experts utilize several parallel processing pathways, whereas in heterogeneous experts, each of these pathways is implemented using a different architectural model. Our HMMOE approach adopts a heterogeneous multiple-experts design, in which each expert employs a unique architecture, such as DNN, CNN, Trm, TCN, and LSTM, each bringing distinct inductive biases. This stands in contrast to homogeneous expert systems, where all experts share the same architecture. By leveraging architectural diversity, our model is able to capture a broader range of appliance behavior: CNNs excel at identifying local power patterns, RNN-based models like LSTM and GRU are effective at modeling temporal dynamics, and DNNs are suited for learning complex nonlinear relationships.

3.2.2 Multi-Gate Mechanism

In contrast to traditional shared-bottom architectures and single-gate MoE models[69], our approach employs independent gating mechanisms for the on/off state classification and energy consumption estimation tasks. This design enables the model to learn task-specific routing strategies, allowing each task to prioritize different combinations of experts based on its unique requirements. Each gating network processes the input features and outputs softmax-normalized weights over the set of expert networks, enabling adaptive computation tailored to the task. For example, the classification task might favor experts that are effective at detecting binary state transitions, while the regression task may rely more on those that capture temporal consumption patterns. By decoupling the expert selection process, the model reduces task interference and mitigates gradient conflicts during training, all while maintaining parameter efficiency through shared expert usage. Additionally, the gates are input-dependent, meaning they are dynamically generated by neural components and vary across data instances. This multi-gate design strikes a balance between shared representation learning and task-specific specialization, making it particularly well-suited for the challenges of NILM. By enabling flexible expert allocation per task, our method delivers improved multi-task performance over fixed-sharing alternatives. For each task k ($\forall k \in \mathcal{K} = \{\mathrm{c}, \mathrm{r}\}$), the output of the gate is a weighted vector, which can be expressed as follows:

$$\sum\nolimits_{e\in\mathcal{E}} g_e^k(\boldsymbol{x}) = 1, 0 \leq g_e^k(\boldsymbol{x}) \leq 1, \forall k \in \mathcal{K}. \tag{3.12}$$

For a given input, the output through a gate is a combination of the selected experts, which is presented as follows:

$$f^k(\boldsymbol{x}) = \sum\nolimits_{e\in\mathcal{E}} g_e^k(\boldsymbol{x}) f_e(\boldsymbol{x}), \forall k \in \mathcal{K}. \tag{3.13}$$

The multi-gate mechanism allows the model to dynamically allocate the expertise of different experts to specific tasks, effectively leveraging their strengths and adaptively combining their outputs, resulting in enhanced performance in multi-task learning for NILM. Note that, to generate the final representation for a specific task, the outputs of all experts are combined via linear integration. This is a weighted sum, where the weights are dynamically supplied by that task's dedicated gating network. The resulting task-specific representation is then fed into the final processing tower.

3.2.3 State Identification (Tower A)

The extracted features in the data, including the generally complex temporal patterns and short-term and long-term dependencies, are integrated through the multi-gate mechanism to facilitate the classification of the on/off state of appliances in Tower A. Thus, multi-head self-attention structures are employed to identify the on/off state of appliances. The self-attention structure of the h-th head is expressed as follows.

$$\mathbf{Q}_h^{\mathrm{c}} = f^{\mathrm{c}}(\boldsymbol{x})\mathbf{W}_h^{\mathbf{Q},\mathrm{c}} \tag{3.14}$$

$$\mathbf{Z}_h^{\mathrm{c}} = f^{\mathrm{c}}(\boldsymbol{x})\mathbf{W}_h^{\mathbf{Z},\mathrm{c}} \tag{3.15}$$

$$\mathbf{V}_h^{\mathrm{c}} = f^{\mathrm{c}}(\boldsymbol{x})\mathbf{W}_h^{\mathbf{V},\mathrm{c}} \tag{3.16}$$

$$\mathbf{F}_h^{\mathrm{c}} = \mathrm{softmax}(\frac{\mathbf{Q}_h^{\mathrm{c}}\mathbf{Z}_h^{\mathrm{c}\,\mathrm{T}}}{\sqrt{d}})\mathbf{V}_h^{\mathrm{c}} \tag{3.17}$$

For the h-th head, $\mathbf{Q}_h^{\mathrm{c}}$, $\mathbf{Z}_h^{\mathrm{c}}$, and $\mathbf{V}_h^{\mathrm{c}}$ are the query, key, and value matrices, respectively, which have dimensions of $T \times d$, where d is the hidden size, $\mathbf{W}_h^{\mathbf{Q},\mathrm{c}} \in \mathbb{R}^{d\times d}$, $\mathbf{W}_h^{\mathbf{Z},\mathrm{c}} \in \mathbb{R}^{d\times d}$, and $\mathbf{W}_h^{\mathbf{V},\mathrm{c}} \in \mathbb{R}^{d\times d}$ are the

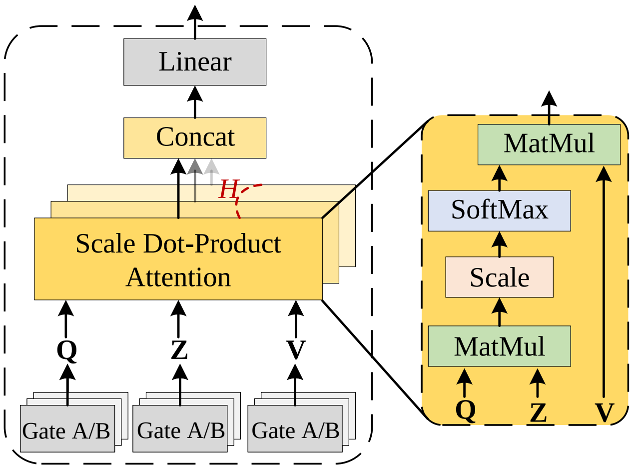


Figure 3.2: Multi-head self-attention architecture.

corresponding learnable weight matrices, and $\mathbf{F}_h^{\mathrm{c}} \in \mathbb{R}^{T \times d}$ is the output, where the softmax function is applied row-wise to normalize the weights, and thereby ensuring that the attention distribution sums to 1. As can be seen, the superscript c has been applied to variables employed in the classification task.

In addition, the multi-head attention architecture illustrated in Fig. 3.2 is employed to enhance the fitting ability of the proposed learning approach, where several different linear transformations of the queries (**Q**), keys (**K**), and values (**V**) are applied to compute multiple independent attention heads. Hence, the final multi-head attention output is a linear integration of the output $\mathbf{F}_h^{\mathrm{c}}$ of all heads with a skip-connection part:

$$\mathbf{F}^{\mathrm{c}} = [\mathbf{F}_1^{\mathrm{c}}, \mathbf{F}_2^{\mathrm{c}}, ..., \mathbf{F}_{H^{\mathrm{c}}}^{\mathrm{c}}]^{\mathrm{T}} \mathbf{W}^{\mathrm{c}}, \tag{3.18}$$

where H^{c} is the number of heads in the multi-head attention architecture. Additionally, $\mathbf{W}^{\mathrm{c}} \in \mathbb{R}^{(H \cdot d) \times d}$ is the learnable weight matrix for the state identification task, and $\mathbf{F}^{\mathrm{c}} \in \mathbb{R}^{T \times d}$ is the integrated output for the state identification task.

3.2.4 Power Estimation (Tower B)

Multi-head self-attention structures are further employed in Tower B to predict the continuous power consumption of appliances. Accordingly, the self-attention structure of the h-th head is expressed as follows.

$$\mathbf{Q}_h^{\mathrm{r}} = f^{\mathrm{r}}(\boldsymbol{x})\mathbf{W}_h^{\mathbf{Q},\mathrm{r}} \tag{3.19}$$

$$\mathbf{Z}_h^{\mathrm{r}} = f^{\mathrm{r}}(\boldsymbol{x})\mathbf{W}_h^{\mathbf{Z},\mathrm{r}} \tag{3.20}$$

$$\mathbf{V}_h^{\mathrm{r}} = f^{\mathrm{r}}(\boldsymbol{x})\mathbf{W}_h^{\mathbf{V},\mathrm{r}} \tag{3.21}$$

$$\mathbf{F}_h^{\mathrm{r}} = \mathrm{softmax}(\frac{\mathbf{Q}_h^{\mathrm{r}}\mathbf{Z}_h^{\mathrm{r}\,\mathrm{T}}}{\sqrt{d}})\mathbf{V}_h^{\mathrm{r}} \tag{3.22}$$

$$\mathbf{F}^{\mathrm{r}} = [\mathbf{F}_1^{\mathrm{r}}, \mathbf{F}_2^{\mathrm{r}}, ..., \mathbf{F}_{H^{\mathrm{r}}}^{\mathrm{r}}]^{\mathrm{T}}\mathbf{W}^{\mathrm{r}} \tag{3.23}$$

Here, $\mathbf{Q}_h^{\mathrm{r}}$, $\mathbf{Z}_h^{\mathrm{r}}$, and $\mathbf{V}_h^{\mathrm{r}}$ are the corresponding query, key, and value matrices, $\mathbf{W}_h^{\mathbf{Q},\mathrm{r}} \in \mathbb{R}^{d\times d}$, $\mathbf{W}_h^{\mathbf{Z},\mathrm{r}} \in \mathbb{R}^{d\times d}$, $\mathbf{W}_h^{\mathbf{V},\mathrm{r}} \in \mathbb{R}^{d\times d}$ are the learnable weight matrices, and $\mathbf{F}_h^{\mathrm{r}} \in \mathbb{R}^{T\times d}$ is the output. Again, the final multi-head attention output in (3.23) is a linear integration of the output $\mathbf{F}_h^{\mathrm{r}}$ of all heads with a skip-connection part, where H^{r} is the number of heads in the multi-head attention architecture. Additionally, $\mathbf{W}^{\mathrm{r}} \in \mathbb{R}^{(H\cdot d)\times d}$ is the learnable weight matrix for the power estimation task, and $\mathbf{F}^{\mathrm{r}} \in \mathbb{R}^{T\times d}$ is the integrated output for the power estimation task. As can be seen, the superscript r has been applied to variables employed in the regression task.

Note that, both Tower A (state classification) and Tower B (power estimation) are implemented using lightweight transformer encoder blocks. This design choice is motivated by the need to model long-term and non-local dependencies in time-series NILM data. While recurrent models (e.g., LSTM) capture sequential patterns, transformers provide parallelized attention mechanisms that can dynamically focus on relevant time steps, an advantage for handling asynchronous and low-activation appliances. Additionally, we use a compact transformer configuration to mitigate the model size in this paper. Each tower is designed with only 3 encoder layers and utilizes 4 attention heads. The feedforward network within each encoder block is limited to a size of 128, and the dropout rate of 0.1 is applied to prevent overfitting. Each Transformer tower receives a task-specific aggregated expert representation $f^k(\boldsymbol{x})$, and produces a sequence-level output used for classification and regression. Because these towers operate on already-refined features, the computational load is significantly lower than that of full-sequence transformers.

3.3 Multi-Objective Optimization Strategy

3.3.1 Loss Function Definition

The dual empirical loss function applied for the joint optimization of the classification and regression tasks is defined as

$$\min_{\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{\mathrm{K}}}\mathcal{L}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{\mathrm{K}}) = \alpha^{\mathrm{c}}\mathcal{L}^{\mathrm{c}}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{\mathrm{c}}) + \alpha^{\mathrm{r}}\mathcal{L}^{\mathrm{r}}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{\mathrm{r}}), \tag{3.24}$$

where $\mathcal{E} = \{\mathrm{DNN}, \mathrm{CNN}, \mathrm{Trm}, \mathrm{TCN{,}LSTM}\}$, $\mathcal{K} = \{\mathrm{r}, \mathrm{c}\}$, θ^{E} and θ^{K} are the network parameters for experts and tasks, respectively, and α^{c} and α^{r} are the weight coefficients applied to facilitate a tradeoff in the objectives of the different tasks, which are subject to the constraints $\alpha^{\mathrm{c}} + \alpha^{\mathrm{r}} = 1$ and $\alpha^{\mathrm{K}} \geq 0$. These weight coefficients are regarded as variables optimized in the training process. Moreover, the component loss functions are defined as follows.

$$\mathcal{L}^{\mathrm{c}}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{\mathrm{c}}) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \log\left(1 + \exp\left(-\hat{s}_i s_i\right)\right) \tag{3.25}$$

$$\begin{aligned}\mathcal{L}^{\mathrm{r}}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{\mathrm{r}}) = \frac{1}{T}\sum\nolimits_{t=1}^{T} \left(\hat{x}_i - x_i\right)^2 + \frac{\lambda}{T}\sum\nolimits_{i\in O} |\hat{x}_i - x_i| \\ + D_{KL}(\mathrm{softmax}(\frac{\hat{x}}{\tau})\| \,\mathrm{softmax}(\frac{x}{\tau}))\end{aligned} \tag{3.26}$$

Here, $\mathcal{L}^{\mathrm{c}}$ defines the soft-margin loss function to evaluate the difference between the predicted states $\hat{s}_i$ and ground-truth states s_i of the i-th electrical appliance during time interval t in the NILM binary classification problem, while $\mathcal{L}^{\mathrm{r}}$ considers deviations between the predicted and ground-truth power loads of electrical appliances, and includes from left to right the mean squared error loss, the L1 loss, where λ is the weighting coefficient that balances the loss terms, and the Kullback–Leibler divergence loss, where τ is the temperature parameter used in the softmax function.

3.3.2 Dynamic Weight Adjustment

The present work addresses the challenge of objectively defining suitable weights for objectives by adopting a novel multi-objective programming method, which can be expressed as follows.

$$\min_{\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{k}}\{\mathcal{L}^{k}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{k}),\forall k\in\mathcal{K}\} \tag{3.27}$$

Definition 1. *Let $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$ be two solutions. Solution $\boldsymbol{\theta}'$ dominates solution $\boldsymbol{\theta}''$ if and only if the following conditions hold:*

1) $\mathcal{L}^{k}(\boldsymbol{\theta}'^{\mathrm{E}},\boldsymbol{\theta}'^{k})\leq\mathcal{L}^{k}(\boldsymbol{\theta}''^{\mathrm{E}},\boldsymbol{\theta}''^{k})$, $\forall k\in\mathcal{K}$;

2) $\mathcal{L}^{k}(\boldsymbol{\theta}'^{\mathrm{E}},\boldsymbol{\theta}'^{k})<\mathcal{L}^{k}(\boldsymbol{\theta}''^{\mathrm{E}},\boldsymbol{\theta}''^{k})$, $\exists k\in\mathcal{K}$.

Definition 2. *A solution $\boldsymbol{\theta}^{*}$ is considered Pareto non-inferior (or efficient) if there is no other feasible solution $\boldsymbol{\theta}$ that dominates $\boldsymbol{\theta}^{*}$.*

The set of all Pareto non-inferior solutions forms the Pareto non-inferior set in a multi-objective optimization problem, and their corresponding objective values form a Pareto frontier. Then, a multi-gradient descent algorithm is applied to determine the optimal solution among the Pareto non-inferior set based on the gradients of various objectives.

Past work has demonstrated that, if a solution is non-inferior, the solution must satisfy the Pareto stationarity condition [138]. Introducing the Pareto stationarity condition to (3.27) generates the following requirements for the task parameters.

1) $\exists\alpha^{k}\geq 0(\forall k\in\mathcal{K})$, such that $\sum_{k\in\mathcal{K}}\alpha^{k}=1$ and $\sum_{\forall k\in\mathcal{K}}\alpha^{k}\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^{k}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{k})=0$;

2) $\forall k\in\mathcal{K}$, $\nabla_{\boldsymbol{\theta}^{k}}\mathcal{L}^{k}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{k})=0$.

For convenience of illustration, we first define the following quadratic programming with linear constraints:

$$\begin{aligned}&\min_{\alpha^{k},\forall k\in\mathcal{K}}\left\|\sum\nolimits_{\forall k\in\mathcal{K}}\alpha^{k}\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^{k}(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^{k})\right\|_{2}^{2}\\ \text{s.t.}\quad&\sum\nolimits_{\forall k\in\mathcal{K}}\alpha^{k}=1,\alpha^{k}\geq 0,\forall k\in\mathcal{K}.\end{aligned} \tag{3.28}$$

Theorem 2. *Solving the quadratic programming problem yields one of two outcomes: 1) the solution is zero and the point satisfies a Pareto stationarity condition, 2) the solution provides a descent direction for all tasks.*

Proof: Let ℓ be the Lagrangian function for the optimization problem (3.28). We introduce the Lagrangian multipliers μ for the equality constraint and β^k for the inequality constraints. The resulting Lagrangian function is expressed as:

$$\begin{aligned}\ell = &\left\|\sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k)\right\|_2^2 \\ &+ \mu\left(\sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k - 1\right) - \sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k \beta^k\end{aligned} \tag{A1}$$

Based on the necessary optimality conditions for optimization in (3.28), we can obtain that

$$\frac{\partial \ell}{\partial \alpha^k} = 0, \alpha^k \beta^k = 0, \forall k \in \mathcal{K}. \tag{A2}$$

That is

$$2\nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k) \sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k) + \mu - \beta^k = 0 \tag{A3}$$

This expression can be further rewritten as follows:

$$\beta^k = 2\nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k) \sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k) + \mu \tag{A4}$$

Since $\alpha^k \beta^k = 0, \forall k \in \mathcal{K}$, we can obtain that

$$\alpha^k \beta^k = 2\alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k) \sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k) + \alpha^k \mu = 0 \tag{A5}$$

Summing over all k:

$$\mu = -2(\sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k))^2 \tag{A6}$$

It reveals a direct relationship between the Lagrange multiplier of the equality constraint μ, and the value of the objective function at the optimal solution. We now analyze the two possible cases for the optimal solution to the optimization problem in (3.28).

First, if the minimum value of the objective function in (3.28) is zero, then

$$\left\|\sum\nolimits_{\forall k\in\mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k)\right\|_2^2 = 0 \tag{A7}$$

This implies that:

$$\sum\nolimits_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)=0 \tag{A8}$$

Furthermore, the primal feasibility conditions require that $\sum_{\forall k\in\mathcal{K}}\alpha^k=1,\alpha^k\geq 0,\forall k\in\mathcal{K}$. These conditions are precisely the definition of a Pareto stationary point.

Alternatively, if the minimum value of the objective function is positive, then

$$\mu=-2\left\|\sum\nolimits_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)\right\|_2^2>0 \tag{A9}$$

This implies that the current point is not Pareto stationary. Substituting μ back into (A4), yields:

$$\beta^k=2\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)\sum\nolimits_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)+\mu\geq 0 \tag{A10}$$

Since $\mu<0$, we can obtain that

$$\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)\sum\nolimits_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)>0 \tag{A11}$$

This inequality demonstrates that $\sum_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)$ has a positive inner product with every task's gradient $\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)$. In multi-objective optimization, the goal is to minimize all functions $\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)$ simultaneously. A vector $\boldsymbol{d}$ is defined as a common descent direction for all tasks if it satisfies:

$$(\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k))^T\boldsymbol{d}<0,\forall k \tag{A12}$$

Let us now consider $\boldsymbol{d}=-\sum_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)$. The inner product with each gradient is:

$$\begin{aligned}&(\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k))^T\boldsymbol{d}\\&=-(\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k))^T\sum\nolimits_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)\\&\leq 0.5\mu<0\end{aligned} \tag{A13}$$

This proves that when the optimal value is non-zero, the vector $-\sum_{\forall k\in\mathcal{K}}\alpha^k\nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}},\boldsymbol{\theta}^k)$ is indeed a common descent direction for all tasks. We can therefore simultaneously decrease all objective functions by applying the update rule $\boldsymbol{\theta}^{\mathrm{E}} \leftarrow \boldsymbol{\theta}^{\mathrm{E}} - \rho \sum_{\forall k \in \mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k)$, where ρ is a positive step size. ■

We note that Theorem 1 can transform the Pareto stationarity condition requirements into a quadratic programming problem. Hence, the optimal value of α^k can be computed by rewriting (3.28) into the following form.

$$\min_{\alpha \in [0,1]} \left\| \alpha \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^{\mathrm{c}}(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^{\mathrm{c}}) + (1-\alpha) \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^{\mathrm{r}}(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^{\mathrm{r}}) \right\|_2^2 \tag{3.29}$$

This optimization model can be rewritten as follows:

$$\min_{\alpha \in [0,1]} \Phi(\alpha) = \left\| \alpha \boldsymbol{\varphi} + (1-\alpha) \boldsymbol{\psi} \right\|_2^2, \tag{3.30}$$

where the gradient notation has been simplified by replacing $\nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^{\mathrm{c}}(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^{\mathrm{c}})$ and $\nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^{\mathrm{r}}(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^{\mathrm{r}})$ in (3.29) with $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$, respectively. The derivative of the function $\Phi(\alpha)$ in (3.30) can be formulated as follows.

$$d\Phi(\alpha)/d\alpha = 2(\alpha \boldsymbol{\varphi} + (1-\alpha)\boldsymbol{\psi}, \boldsymbol{\varphi} - \boldsymbol{\psi}) \tag{3.31}$$

Based on the necessary optimal condition of model (3.30) that $d\Phi(\alpha)/d\alpha = 0$, we can obtain the following.

$$(\alpha(\boldsymbol{\varphi} - \boldsymbol{\psi}) + \boldsymbol{\psi}) \cdot (\boldsymbol{\varphi} - \boldsymbol{\psi}) = 0 \tag{3.32}$$

Since $\boldsymbol{\varphi} \neq \boldsymbol{\psi}$ generally holds, we can solve for the unconstrained critical point α_{crit} as follows.

$$\alpha_{\text{crit}} = \frac{-\boldsymbol{\psi}}{\boldsymbol{\varphi} - \boldsymbol{\psi}} = \frac{\boldsymbol{\psi} \cdot (\boldsymbol{\psi} - \boldsymbol{\varphi})}{\|\boldsymbol{\varphi} - \boldsymbol{\psi}\|^2} = \frac{\|\boldsymbol{\psi}\|^2 - \boldsymbol{\psi} \cdot \boldsymbol{\varphi}}{\|\boldsymbol{\varphi}\|^2 - 2\boldsymbol{\psi} \cdot \boldsymbol{\varphi} + \|\boldsymbol{\psi}\|^2} \tag{3.33}$$

After obtaining α, we must perform a boundary analysis based on the constraint $\alpha \in [0, 1]$. The objective function $\Phi(\alpha)$ is a strictly convex function (since its second derivative $d^2\Phi/d\alpha^2 = 2\|\varphi - \psi\|^2 > 0$). Therefore, the global minimum over the interval $[0, 1]$ is either at the critical point α_{crit} if it falls within the open interval $(0, 1)$, or at one of the boundaries $\alpha = 0$ or $\alpha = 1$.

The optimal solution lies at the boundary $\alpha = 0$ when the critical point α_{crit} is less than or equal to zero. This condition mathematically translates to $\boldsymbol{\psi} \cdot \boldsymbol{\varphi} \geq \|\boldsymbol{\psi}\|^2$. The validity of this boundary minimum is confirmed by the derivative at this point, $\Phi'(0) = 2(\boldsymbol{\psi} \cdot \boldsymbol{\varphi} - \|\boldsymbol{\psi}\|^2)$, which is non-negative under this condition, indicating the function's value increases as α moves into the interval.

Conversely, the optimal solution is found at the other boundary $\alpha = 0$ when the critical point α_{crit} is greater than or equal to one. This corresponds to the condition $\boldsymbol{\psi} \cdot \boldsymbol{\varphi} \geq \|\boldsymbol{\varphi}\|^2$. The non-positive derivative at this boundary $\Phi'(1) = 2(\|\boldsymbol{\varphi}\|^2 - \boldsymbol{\psi} \cdot \boldsymbol{\varphi}) \leq 0$, validates this choice, showing the function's value decreases as it approaches $\alpha = 1$ from within the interval.

In the event that neither boundary condition is met, the critical point α_{crit} must lie strictly between 0 and 1. This occurs when $\boldsymbol{\psi} \cdot \boldsymbol{\varphi}$ is less than both $\|\boldsymbol{\psi}\|^2$ and $\|\boldsymbol{\varphi}\|^2$, which can be compactly written as

$$0 < \|\boldsymbol{\psi}\|^2 - \boldsymbol{\psi} \cdot \boldsymbol{\varphi} < \|\boldsymbol{\varphi}\|^2 - 2\boldsymbol{\psi} \cdot \boldsymbol{\varphi} + \|\boldsymbol{\psi}\|^2, \tag{3.34}$$

Such that

$$\boldsymbol{\psi} \cdot \boldsymbol{\varphi} \leq \min(\|\boldsymbol{\psi}\|^2, \|\boldsymbol{\varphi}\|^2). \tag{3.35}$$

In this scenario, the optimal solution for α is simply the critical point in (3.33).

Finally, the analytic solution of α can be expressed as follows.

$$\alpha = \begin{cases} 0, \text{ if } \boldsymbol{\psi} \cdot \boldsymbol{\varphi} \geq \|\boldsymbol{\psi}\|^2 \\ \dfrac{\boldsymbol{\psi}(\boldsymbol{\psi} - \boldsymbol{\varphi})}{\|\boldsymbol{\varphi} - \boldsymbol{\psi}\|^2}, \text{ if } \boldsymbol{\psi} \cdot \boldsymbol{\varphi} < \|\boldsymbol{\psi}\|^2 \text{ and } \boldsymbol{\psi} \cdot \boldsymbol{\varphi} < \|\boldsymbol{\varphi}\|^2 \\ 1, \text{ if } \boldsymbol{\psi} \cdot \boldsymbol{\varphi} \geq \|\boldsymbol{\varphi}\|^2 \end{cases} \tag{3.36}$$

This analytic solution ensures that the optimization process satisfies the Pareto stationarity condition, where no task can be improved without degrading the performance of the other. Thus, this method ensures that both tasks contribute with maximum effectiveness toward the overall optimization goal.

The optimization process is further illustrated in Fig. 3.3 based on projections of gradients $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$. Mathematically, α is set to 0 when $\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\psi} \geq \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\psi}$, which prioritizes the power estimation task because the gradient of the state identification task is already sufficiently aligned with the shared direction of optimization. Conversely, α is assigned a fractional value if $\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\psi} < \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\psi}$ and $\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\psi} < \boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\varphi}$ based on the projection of $\boldsymbol{\psi}$ onto the line $\dfrac{\boldsymbol{\psi}(\boldsymbol{\psi} - \boldsymbol{\varphi})}{\|\boldsymbol{\varphi} - \boldsymbol{\psi}\|^2}$, which ensures a weighted contribution

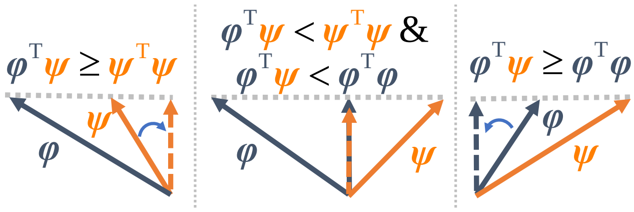


Figure 3.3: Illustration of the task weight optimization process via gradient projections.

of both tasks to the shared objective, balancing their importance dynamically. Finally, $\alpha = 1$ when $\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\psi} \geq \boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\varphi}$, which prioritizes the state identification task.

It is noteworthy that α can be calculated directly from (3.36) under the current conditions involving just two tasks. However, this process becomes infeasible when considering more than two tasks. Therefore, a generalizable solution process is maintained in the proposed HMMOE approach by applying the Frank-Wolfe algorithm based on an iterative first-order optimization to efficiently solve problem (3.36)[139], as shown in **Algorithm 1**. To simplify notation, we define $g^k = \nabla_{\boldsymbol{\theta}^{\mathrm{E}}}\mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k)$. With this, the objective in equation (3.28) can be reformulated as: $\min_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \boldsymbol{M} \boldsymbol{\alpha}$, where $\boldsymbol{M}$ is a positive semidefinite Gram matrix of size $|\mathcal{K}| \times |\mathcal{K}|$, with entries $\boldsymbol{M}_{ij} = (g^i)^T g^j$. The feasible set for $\boldsymbol{\alpha}$ is the standard probability simplex in $\mathbb{R}^{|\mathcal{K}|}$, denoted by Δ. The Frank-Wolfe algorithm follows a three-step procedure at each iteration.

3.3.2.1 Direction-Finding via Linear Minimization Oracle (LMO)

At iteration r, we compute the gradient of the objective function at the current point $\boldsymbol{\alpha}_r$, given by $\nabla f(\boldsymbol{\alpha}_r) = 2\boldsymbol{M}\boldsymbol{\alpha}_r$. Next, the algorithm identifies a descent direction by solving the LMO problem: finding a point $s_r \in \Delta$ that minimizes the linear approximation of the objective:

$$\boldsymbol{s}_r = \arg\min_{\boldsymbol{s} \in \Delta} \nabla f(\boldsymbol{\alpha}_r)^T \boldsymbol{s} = \arg\min_{\boldsymbol{s} \in \Delta} (2\boldsymbol{M}\boldsymbol{\alpha}_r)^T \boldsymbol{s} \tag{3.37}$$

A key property of linear optimization over a convex polytope, such as the probability simplex Δ, is that the minimum is always attained at one of its vertices. In this case, the vertices of Δ are the standard basis vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{|\mathcal{K}|}\}$. Consequently, the LMO step reduces to identifying the index $\hat{k}$ corresponding to the smallest component of the gradient vector:

$$\hat{k} = \arg \min_{k \in \{1,\ldots,|\mathcal{K}|\}} (2\boldsymbol{M}\boldsymbol{\alpha}_r)_k \tag{3.38}$$

The optimal solution to the LMO is then given by $\boldsymbol{s}_r = \boldsymbol{e}_{\hat{k}}$. This computation is efficient, involving a matrix-vector multiplication followed by a simple search for the minimum entry, with an overall time complexity of $O\left(|\mathcal{K}|^2\right)$.

3.3.2.2 Step Size Calculation

To update the iterate, the algorithm moves from the current point $\boldsymbol{\alpha}_r$ toward the selected vertex $\boldsymbol{s}_r$, using a convex combination $\boldsymbol{\alpha}_{r+1} = (1-\gamma)\boldsymbol{\alpha}_r + \gamma \boldsymbol{s}_r$, where $\gamma \in [0,1]$ is the step size. The optimal value of γ_r is obtained by solving the following one-dimensional minimization problem:

$$\gamma_r = \arg \min_{\gamma \in [0,1]} f\left(\boldsymbol{\alpha}_r + \gamma\left(\boldsymbol{s}_r - \boldsymbol{\alpha}_r\right)\right) \tag{3.39}$$

Since $f(\boldsymbol{\alpha})$ is a quadratic function, the line search reduces to minimizing a one-dimensional quadratic in γ, which admits a closed-form solution. Defining the descent direction as $\boldsymbol{d}_r = \boldsymbol{s}_r - \boldsymbol{\alpha}_r$, we differentiate $f\left(\boldsymbol{\alpha}_r + \gamma \boldsymbol{d}_r\right)$ with respect to γ and set the derivative to zero. This yields the unconstrained optimal step size: $\gamma^* = -\frac{\boldsymbol{d}_r^T \boldsymbol{M}\boldsymbol{\alpha}_r}{\boldsymbol{d}_r^T \boldsymbol{M}\boldsymbol{d}_r}$. To ensure feasibility, the step size is projected onto the interval $[0,1]$ by clipping: $\gamma_r = \min\left(\max\left(0, \gamma^*\right), 1\right)$.

3.3.2.3 Update

The weight vector is then updated using the computed step size $\boldsymbol{\alpha}_{r+1} = \left(1-\gamma_r\right)\boldsymbol{\alpha}_r + \gamma_r \boldsymbol{s}_r$. As this update is a convex combination of two points within the simplex Δ, the resulting iterate $\boldsymbol{\alpha}_{r+1}$ remains feasible by construction.

The iterative process is presented in **Algorithm 1**. Specifically, the algorithm commences by computing and applying standard gradient descent updates to the parameters of the two non-shared, task-specific networks. The central component of the algorithm addresses the shared expert networks. It computes the gradient vectors of each task's loss function with respect to the shared param-

Algorithm 1 Process for dynamically updating the weights of task objectives during network training

Input: learning rate ρ, maximum number of iterations R, acceptable solution gap $0 \leq \varepsilon \leq 1$
Initialize: $\boldsymbol{\alpha} = (\alpha^1, \cdots, \alpha^{|\mathcal{K}|}) = (\frac{1}{|\mathcal{K}|}, \cdots, \frac{1}{|\mathcal{K}|})$, current gap value gap $= 10$, and iteration number $r = 0$
Determine the gradient for each individual task network
for $k \in \mathcal{K}$ **do**
    $\boldsymbol{\theta}^k = \boldsymbol{\theta}^k - \rho \nabla_{\boldsymbol{\theta}^k} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k)$
end for
Determine the gradient for the shared networks by Frank-Wolfe algorithm:
$g^k \leftarrow \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k), k \in \mathcal{K}$
Compute Gram matrix $\boldsymbol{M}$: $M_{ij} = (g^i)^{\mathrm{T}} (g^j), \forall i, j \in \mathcal{K}$
while gap $> \varepsilon$, and $r \leq R$ **do**
    $\hat{k} = \arg\min_{k \in \{1, \ldots, |\mathcal{K}|\}} (2\boldsymbol{M}\boldsymbol{\alpha}_r)_k$
    $\boldsymbol{s}_r \leftarrow \boldsymbol{e}_{\hat{k}}$
    Descent direction $\boldsymbol{d}_r = \boldsymbol{s}_r - \boldsymbol{\alpha}_r$
    Closed-form step size $\gamma_r \leftarrow \min\left(\max\left(0, -\frac{\boldsymbol{d}_r^T \boldsymbol{M} \boldsymbol{\alpha}_r}{\boldsymbol{d}_r^T \boldsymbol{M} \boldsymbol{d}_r}\right), 1\right)$
    Update weights $\boldsymbol{\alpha}_{r+1} = (1 - \gamma_r)\, \boldsymbol{\alpha}_r + \gamma_r \boldsymbol{s}_r$
    gap $\leftarrow 2\mathrm{M}\boldsymbol{\alpha}_{\mathrm{r}}(\boldsymbol{\alpha}_{\mathrm{r}} - \mathrm{s}_{\mathrm{r}})$
    if $\mathrm{gap}_{\mathrm{r}} <$ gap **then**
        gap $\leftarrow \mathrm{gap}_{\mathrm{r}}$
    end if
    $r \leftarrow r + 1$
end while
Obtain: $\boldsymbol{\alpha} = (\alpha^1, \cdots, \alpha^{|\mathcal{K}|})$
Output: $\boldsymbol{\theta}^E = \boldsymbol{\theta}^E - \rho \sum_{\forall k \in \mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k)$

eters. A subproblem is then formulated to determine the optimal weights for these gradients, aiming to find an update direction that represents a Pareto-optimal trade-off. This subproblem is solved iteratively using the Frank-Wolfe algorithm. Upon convergence, the resulting optimal weights are used to compute a weighted sum of the individual task gradients. This final, combined gradient is then used to update the parameters of the shared expert networks, thereby achieving a balanced optimization of both the classification and regression objectives at each training iteration.

3.3.3 Convergence Analysis

To prove the convergence of **Algorithm 1**, we establish a set of lemmas and theorems.

Lemma 1. *The weighted-sum objective function* $f(\alpha) = \left\|\sum_{\forall k \in \mathcal{K}} \alpha^k \nabla_{\boldsymbol{\theta}^{\mathrm{E}}} \mathcal{L}^k(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^k)\right\|_2^2 = \left\|\sum_{\forall k \in \mathcal{K}} \alpha^k g^k\right\|_2^2$ *is convex.*

Proof: The objective function can be rewritten as:

$$f(\alpha) = (\sum_{k=1}^{|\mathcal{K}|} \alpha^k g^k)^T (\sum_{j=1}^{|\mathcal{K}|} \alpha^j g^j) = \sum_{k=1}^{|\mathcal{K}|} \sum_{j=1}^{|\mathcal{K}|} \alpha^k \alpha^j \left((g^k)^T g^j\right) \tag{3.40}$$

To determine if $f(\alpha)$ is convex, we need to analyze its Hessian matrix. The Hessian is the matrix of second partial derivatives of f with respect to the components of α. First, we compute the second partial derivative, which forms the (j, k)-th entry of the Hessian matrix, $\boldsymbol{H}$:

$$H_{jk} = \frac{\partial^2 f}{\partial \alpha^j \partial \alpha^k} = 2(g^j)^T g^k \tag{3.41}$$

Thus, the Hessian matrix $\boldsymbol{H}$ can be expressed by $\boldsymbol{H} = 2\boldsymbol{G}^T\boldsymbol{G}$, where $\boldsymbol{G} = [g^1, g^2, ..., g^{|\mathcal{K}|}]$.

A function is convex if and only if its Hessian matrix is positive semidefinite. A matrix $\boldsymbol{A}$ is positive semidefinite if for any vector $\boldsymbol{a}$, $\boldsymbol{a}^T\boldsymbol{A}\boldsymbol{a} \geq 0$. Next, let's test this for the Hessian $\boldsymbol{H}$. For any vector $\boldsymbol{a} \in \mathbb{R}^{|\mathcal{K}|}$

$$\boldsymbol{a}^T\boldsymbol{H}\boldsymbol{a} = \boldsymbol{a}^T \left(2\boldsymbol{G}^T\boldsymbol{G}\right) \boldsymbol{a} = 2(\boldsymbol{G}\boldsymbol{a})^T(\boldsymbol{G}\boldsymbol{a}) = 2\|\boldsymbol{G}\boldsymbol{a}\|_2^2 \tag{3.42}$$

Since the squared Euclidean norm of any vector is always non-negative, we have:

$$\boldsymbol{a}^T\boldsymbol{H}\boldsymbol{a} = 2\|\boldsymbol{G}\boldsymbol{a}\|_2^2 \geq 0 \tag{3.43}$$

This shows that the Hessian matrix $\boldsymbol{H}$ is positive semidefinite. Therefore, the objective function $f(\alpha)$ is a convex function. ■

By proving that $f(\alpha)$ is convex, Lemma 1 ensures that any identified local minimum is also the global minimum. This characteristic transforms the optimization task from a potentially intractable global search into a well-defined problem where iterative descent will provably converge to the global solution. This guarantee of a unimodal search space is foundational to the subsequent convergence analysis.

Lemma 2. *The feasible domain $\Delta = \{\alpha \in \mathbb{R}^{|\mathcal{K}|} \mid \sum_{\forall k \in \mathcal{K}} \alpha^k = 1,\, \alpha^k \geq 0, \forall k \in \mathcal{K}\}$ is a compact convex set.*

Proof: Let $\alpha_1, \alpha_2 \in \Delta$ and let $\omega \in [0,1]$. We prove that their convex combination $\alpha' = \omega\alpha_1 + (1-\omega)\alpha_2$ is also in Δ. First, we can derive that

$$\begin{aligned}&\sum\nolimits_{\forall k\in\mathcal{K}} \alpha'^k = \sum\nolimits_{\forall k\in\mathcal{K}} \left(\omega\alpha_1^k + (1-\omega)\alpha_2^k\right) = \omega \sum\nolimits_{\forall k\in\mathcal{K}} \alpha_1^k \\ &+ (1-\omega)\sum\nolimits_{\forall k\in\mathcal{K}} \alpha_2^k = \omega(1) + (1-\omega)(1) = 1\end{aligned} \tag{3.44}$$

Since $\omega \geq 0, (1-\omega) \geq 0, \alpha_1^k \geq 0$, and $\alpha_2^k \geq 0$, it follows that $\alpha'^k = \omega\alpha_1^k + (1-\omega)\alpha_2^k \geq 0$. Thus, $\alpha' \in \Delta$, which proves that Δ is a convex set.

Then, note that, a set in $\mathbb{R}^{|\mathcal{K}|}$ is compact if it is closed and bounded. The set Δ is defined by linear equalities ($\sum_{\forall k\in\mathcal{K}} \alpha^k = 1$) and non-strict inequalities ($\alpha^k \geq 0$). Such sets are always closed.

Additionally, for any $\alpha \in \Delta$, we have $\alpha^k \geq 0$ and $\sum \alpha^k = 1$. This implies that $0 \leq \alpha^k \leq 1$ for all k. Therefore, $\|\alpha\|_\infty \leq 1$, meaning the set is bounded. Since Δ is both closed and bounded, it is a compact set. ■

The Weierstrass Extreme Value Theorem, states that a continuous function on a compact set must attain its minimum and maximum values [140]. Given that $f(\alpha)$ is continuous and its domain Δ is proven to be compact in Lemma 2, we are guaranteed that a minimum value exists to be found. Without the property of compactness, the function could potentially decrease without bound or approach a minimum value only at a boundary that is not included in the domain, thus never attaining it.

Theorem 3. *If the gradient of $f, \nabla f$, is L-Lipschitz continuous with respect to some norm $\|\cdot\|$ over the domain Δ (i.e., $\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\| \leq L\|\mathbf{y}-\mathbf{x}\|$), then the curvature constant C_f is bounded by $C_f \leq L \cdot \mathrm{diam}(\Delta)^2$, where $L > 0$ and $\mathrm{diam}(\Delta) = \sup_{\mathbf{x},\mathbf{s}\in\Delta} \|\mathbf{s}-\mathbf{x}\|$ is the diameter of the domain.*

Theorem 3 shows that the curvature constant C_f is directly related to the Lipschitz constant L and the size of the domain. Note that, the curvature constant C_f quantifies the maximum possible curvature of the objective function over the entire domain. This property permits the construction of a quadratic model that reliably upper-bounds the function's value across the domain. The guarantee that the true function value will not exceed this quadratic approximation is instrumental in formally proving sufficient descent and convergence in the subsequent theorem. Without a finite bound on curvature, no such guarantee would be possible, as the function could curve upwards with arbitrary steepness, potentially negating the progress of any descent step.

Theorem 4. *At any iteration r, the iterate $\alpha_r + 1$ generated by the Frank-Wolfe algorithm satisfies the following inequality:*

$$f(\alpha_{r+1}) \leq f(\alpha_r) - \gamma_r g(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.45}$$

where $g(\alpha_r)$ is the simply duality gap at an iterate α_r, defined as the maximum possible improvement predicted by the first-order approximation of f at α_r over the entire domain, i.e., $g(\alpha_r) = \langle \nabla f(\alpha_r), \alpha_r - \mathbf{s}_r \rangle$.

Proof: The proof begins with the definition of the curvature constant C_f. This definition provides a quadratic upper bound on the function's value. We apply this definition to the specific update step of the Frank-Wolfe algorithm.

The update is given by $\alpha_{r+1} = (1 - \gamma_r)\alpha_r + \gamma_r \mathbf{s}_r$, which can be rewritten as $\alpha_{r+1} = \alpha_r + \gamma_r(\mathbf{s}_r - \alpha_r)$. This matches the form $\mathbf{y} = \mathbf{x} + \gamma(\mathbf{s} - \mathbf{x})$ used in the definition of C_f, with $\mathbf{x} = \alpha_r$, $\mathbf{s} = \mathbf{s}_r$, and $\gamma = \gamma_r$. Applying the inequality from the definition of C_f:

$$\begin{aligned} f(\alpha_{r+1}) &= f(\alpha_r + \gamma_r(\mathbf{s}_r - \alpha_r)) \\ &\leq f(\alpha_r) + \gamma_r \langle \nabla f(\alpha_r), \mathbf{s}_r - \alpha_r \rangle + \frac{C_f}{2}\gamma_r^2 \end{aligned} \tag{3.46}$$

Next, we recall the definition of the Frank-Wolfe gap $g(\alpha_r) = \langle \nabla f(\alpha_r), \alpha_r - \mathbf{s}_r \rangle$. By negating both sides, we can express the inner product term from the inequality above in terms of the gap:

$$\langle \nabla f(\alpha_r), \mathbf{s}_r - \alpha_r \rangle = -g(\alpha_r) \tag{3.47}$$

Substituting this expression for the inner product back into the inequality from (3.46) yields the desired result:

$$f(\alpha_{r+1}) \leq f(\alpha_r) - \gamma_r g(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.48}$$

This completes the proof of the theorem. ■

This theorem links the function value at the next step to the current function value and the gap, which itself is a measure of suboptimality. It shows that for a sufficiently small step size, the function value is guaranteed to decrease as long as the gap is positive. This theorem provides a formal, quantifiable guarantee that every single step is productive and moves the solution closer to the minimum.

Proposition 4. *Let α^* be a minimizer of f, and define the suboptimality at iteration r as $h\left(\alpha_r\right) := f\left(\alpha_r\right) - f\left(\alpha^*\right)$. The Frank-Wolfe algorithm, with step size γ_r, generates a sequence of iterates $\{\alpha_r\}$ whose suboptimality satisfies the following recursive inequality: $h\left(\alpha_{r+1}\right) \leq \left(1-\gamma_r\right) h\left(\alpha_r\right) + \frac{C_f}{2}\gamma_r^2$.*

Proof: The proof begins with the **Theorem** 4, which provides an upper bound on the function value at the next iterate, $\boldsymbol{\alpha}_{r+1}$:

$$f(\alpha_{r+1}) \leq f(\alpha_r) - \gamma_r g(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.49}$$

where $g(\alpha_r)$ is the Frank-Wolfe gap at iteration r.

To shift the focus from the function values to the suboptimality, we subtract the optimal function value, $f(\alpha^*)$, from both sides of the inequality:

$$f(\alpha_{r+1}) - f(\alpha^*) \leq f(\alpha_r) - f(\alpha^*) - \gamma_r g(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.50}$$

By substituting the definition of the suboptimality, $h(\alpha) = f(\alpha) - f(\alpha^*)$, into the expression, we get:

$$h(\alpha_{r+1}) \leq h(\alpha_r) - \gamma_r g(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.51}$$

A critical property of convex functions is that the suboptimality is bounded by the Frank-Wolfe gap, meaning $h(\alpha_r) \leq g(\alpha_r)$. This implies that $-g(\alpha_r) \leq -h(\alpha_r)$. Substituting this relationship into the inequality allows us to bound the recursion in terms of the suboptimality itself:

$$h(\alpha_{r+1}) \leq h(\alpha_r) - \gamma_r h(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.52}$$

Factoring out $h(\alpha_r)$ yields the final suboptimality recursion:

$$h(\alpha_{r+1}) \leq (1-\gamma_r)h(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.53}$$

This concludes the proof. ■.

Proposition 4 describes the entire trajectory of the error. Specifically, the term $(1-\gamma_r)\,h\,(\alpha_r)$ reveals that at each iteration, the algorithm is guaranteed to eliminate at least a γ_r fraction of the current error, albeit with the addition of the small curvature penalty term. This is analogous to an iterative process where one repeatedly covers a fraction of the remaining distance to a destination. Such a process is intuitively guaranteed to converge. Proposition 4 provides the precise mathematical formalization of this intuition, creating the direct input needed for the final convergence theorem.

Theorem 5. *For the Frank-Wolfe algorithm applied to a convex function f with curvature constant C_f over a compact convex domain Δ, using the step size $\gamma_r = \frac{2}{r+2}$ for $r = 0, 1, 2, \ldots$, the suboptimality at iteration r is bounded by:*

$$h\,(\alpha_r) = f\,(\alpha_r) - f\,(\alpha^*) \leq \frac{2C_f}{r+2} \tag{3.54}$$

Proof: The proof proceeds by establishing a base case and then proving the inductive step. Assume that the theorem holds for some iteration $r \geq 0$. That is, we assume:

$$h\,(\alpha_r) \leq \frac{2C_f}{r+2} \tag{3.55}$$

We must show that, given the hypothesis, the theorem also holds for the next iteration, $r+1$. Our goal is to prove:

$$h\,(\alpha_{r+1}) \leq \frac{2C_f}{(r+1)+2} = \frac{2C_f}{r+3} \tag{3.56}$$

We start from the suboptimality recursion derived in Proposition 4:

$$h\,(\alpha_{r+1}) \leq (1-\gamma_r)\,h\,(\alpha_r) + \frac{C_f}{2}\gamma_r^2 \tag{3.57}$$

Substitute the specific step size $\gamma_r = \frac{2}{r+2}$ and apply the inductive hypothesis $h\,(\alpha_r) \leq \frac{2C_f}{r+2}$:

$$h(\alpha_{r+1}) \leq \left(1 - \frac{2}{r+2}\right)\left(\frac{2C_f}{r+2}\right) + \frac{C_f}{2}\left(\frac{2}{r+2}\right)^2 \tag{3.58}$$

Now, we simplify the algebraic expression. First, simplify the term $(1-\frac{2}{r+2}) = \frac{r}{r+2}$. The second term simplifies to $\frac{C_f}{2}\left(\frac{4}{(r+2)^2}\right) = \frac{2C_f}{(r+2)^2}$. Substitute these simplified parts back into the main inequality:

$$h(\alpha_{r+1}) \leq \frac{2rC_f}{(r+2)^2} + \frac{2C_f}{(r+2)^2} \tag{3.59}$$

Combine the terms over the common denominator:

$$h\left(\alpha_{r+1}\right) \leq \frac{2C_f(r+1)}{(r+2)^2} \tag{3.60}$$

The final part of the proof is to show that this derived bound is less than or equal to our target bound, $\frac{2C_f}{r+3}$. We need to prove the following inequality for all $r \geq 0$:

$$\frac{2C_f(r+1)}{(r+2)^2} \leq \frac{2C_f}{r+3} \tag{3.61}$$

Since $2C_f$ is a positive constant, we can divide both sides by it, reducing the problem to proving:

$$\frac{r+1}{(r+2)^2} \leq \frac{1}{r+3} \tag{3.62}$$

Since $r \geq 0$, both denominators $(r+2)^2$ and $(r+3)$ are positive. We can therefore cross-multiply without changing the direction of the inequality $(r+1)(r+3) \leq (r+2)^2$. Expand both sides of the inequality and subtract $r^2 + 4r$ from both sides, leaving $3 \leq 4$. This inequality is clearly true. Therefore, the inequality in (3.61) holds for all $r \geq 0$.

We have shown that if the hypothesis holds for iteration r, it must also hold for iteration $r+1$. By the principle of mathematical induction, the theorem is true for all iterations $r \geq 0$. This result directly establishes that the suboptimality of the Frank-Wolfe algorithm converges to zero, and the rate of this convergence is of the order $O(1/r)$. ■

This theorem demonstrates that as the number of iterations r tends to infinity, the associated error bound, given by $\frac{2C_f}{r+2}$, converges to zero. Consequently, the algorithm is guaranteed to approximate the optimal weights α of the multi-objective problem to an arbitrarily high degree of accuracy.

The network training procedure employed for the proposed HMMOE approach is presented in

Algorithm 2, which simultaneously optimizes appliance state classification and power estimation tasks using dynamic weights. The procedure iterates over a set number of epochs and processes the training data in batches within each epoch. For each batch of data, the algorithm initiates a forward pass. The aggregated power data is processed through the five heterogeneous expert networks. The outputs of these experts are then selectively combined by two gating networks to create task-specific feature representations. These features are subsequently fed into their respective task towers to produce the final predictions for appliance state and power consumption. It is noteworthy that, the training process iteratively updates expert network parameters using gradient-based optimization while employing the adaptive weighting scheme in **Algorithm 1** to balance task contributions dynamically. Finally, these dynamically calculated weights are used to perform the backward pass and update all network parameters via gradient descent before the process is repeated with the next batch.

It can be seen that, the algorithm features a two-level optimization structure. The outer loop focuses on optimizing the primary network parameters to minimize overall task losses. This process is dependent on an inner optimization loop, carried out by **Algorithm 1**. This inner loop's specific role is to find the ideal meta-parameters α that weight and combine the individual task gradients for the most effective update in the outer loop. This two-tiered approach ensures that each update to the primary network is guided by an optimally balanced combination of task-specific gradients.

3.4 Experiments

In this section, we first briefly describe the three datasets and data preprocessing employed for the experiments. Then, we define the several metrics employed for evaluating NILM performance. Subsequently, the performance of the proposed HMMOE approach is compared with those obtained by existing state-of-the-art NILM approaches employing homogeneous expert structures, including CNN [57], TCN [61], LSTM [59], BiLSTM [60], GRU network [58], sequence-to-point model (Seq2point) [141], decision tree (DT) [142], random forest (RF)[143], and HMM[51], for the three datasets considered. The configurations of the proposed approach are presented in Table 3.1. The input to the model and its transformer towers has a shape of Batch size$\times$Sequence length$\times$Data

Algorithm 2 Network training procedure

Input: M training data with aggregated power labels and individual appliance power labels, learning rate ρ, epoch number M_{ep}, and batch size M_{ba}
Initialize: $\boldsymbol{\alpha} = (\alpha^1, \cdots, \alpha^{|\mathcal{K}|})$
for epoch $m \in [1, M_{ep}]$ **do**
  for batch $n \in [1, M_{ba}]$ **do**
    Heterogeneous Prediction
    Information sharing with diverse experts: Eqs. (3.6)-(3.11)
    Multi-gate Mechanism:
    Gate A: State identification
    $f_e(\boldsymbol{x}) \leftarrow \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_e), \forall e \in \mathcal{E}$
    $f^{\mathrm{c}}(\boldsymbol{x}) \leftarrow \sum_{e \in \mathcal{E}} g_e^{\mathrm{c}}(\boldsymbol{x}) f_e(\boldsymbol{x})$
    Gate B: Power estimation
    $f_e(\boldsymbol{x}) \leftarrow \mathcal{F}(\boldsymbol{x}, \boldsymbol{\Theta}_e), \forall e \in \mathcal{E}$
    $f^{\mathrm{r}}(\boldsymbol{x}) \leftarrow \sum_{e \in \mathcal{E}} g_e^{\mathrm{r}}(\boldsymbol{x}) f_e(\boldsymbol{x})$
    Task Towers
    Tower A: State identification
    $\mathbf{F}_h^{\mathrm{c}} \leftarrow \mathrm{softmax}(\frac{\mathbf{Q}_h^{\mathrm{c}} {\mathbf{Z}_h^{\mathrm{c}}}^{\mathrm{T}}}{\sqrt{d}}) \mathbf{V}_h^{\mathrm{c}}$
    $\mathbf{F}^{\mathrm{c}} \leftarrow [\mathbf{F}_1^{\mathrm{c}}, \mathbf{F}_2^{\mathrm{c}}, ..., \mathbf{F}_{H^{\mathrm{c}}}^{\mathrm{c}}] \mathbf{W}^{\mathrm{c}}$
    Tower B: Power estimation
    $\mathbf{F}_h^{\mathrm{r}} \leftarrow \mathrm{softmax}(\frac{\mathbf{Q}_h^{\mathrm{r}} {\mathbf{Z}_h^{\mathrm{r}}}^{\mathrm{T}}}{\sqrt{d}}) \mathbf{V}_h^{\mathrm{r}}$
    $\mathbf{F}^{\mathrm{r}} \leftarrow [\mathbf{F}_1^{\mathrm{r}}, \mathbf{F}_2^{\mathrm{r}}, ..., \mathbf{F}_{H^{\mathrm{r}}}^{\mathrm{r}}] \mathbf{W}^{\mathrm{r}}$
    Calculate the loss function $\mathcal{L}(\boldsymbol{\theta}^{\mathrm{E}}, \boldsymbol{\theta}^{\mathrm{K}})$ from (3.24)-(3.26)
    Optimize weights $\boldsymbol{\alpha}^*$ using **Algorithm 1**
  end for
end for
Output: Optimal dynamic weights $\boldsymbol{\alpha}^*$ and model parameters $(\boldsymbol{\Theta}_e^*, g_e^{*k}(\boldsymbol{x}), \mathbf{W}_h^{*\mathbf{Q},e}, \mathbf{W}_h^{*\mathbf{Z},e}, \mathbf{W}_h^{*\mathbf{V},e}, \mathbf{W}^{*e})$

dimension, specifically 256×144×1. The corresponding output shape is 256×144×2. The data dimension of the output is 2, rather than 1, because the method simultaneously outputs the on/off state and energy consumption estimation. For CNN, TCN, LSTM, BiLSTM, and a GRU network, their configurations are shown in Fig. 3.2. Moreover, the superiority of the developed heterogeneous expert design is evaluated by comparing the performance of the HMMOE approach with that obtained using the same architecture, but with all of the heterogeneous expert blocks replaced by the same type of network (e.g., only DNNs or only CNNs). Additionally, we verify the superiority of the proposed multi-objective optimization (MOO) method by comparing the obtained results with those obtained using fixed-weighted methods.

Table 3.1: Configuration of the proposed method.

<table>
<tr><th>Parts</th><th>Specific</th><th>General</th></tr>
<tr><td>Tower Transformer</td><td>Number of Layers: 3
Number of Heads: 4
Size of Hidden Layers: 128, 128, 128
Number of Final FC layers: 2
Size of FC Layers: 128, 128, 1
Activation of FC layers: Tanh</td><td rowspan="6">Sequence Length: 144
Data Dimension: 1
Batch Size: 256
Dropout: 0.1
Training Method: Adam
Initialization: Xavier
Number of Towers: 2
Number of Experts: 5</td></tr>
<tr><td>CNN Expert</td><td>Number of Layers: 5
In Channels: 1, 64, 128, 128, 128
Out Channels: 64, 128, 128, 128, 128
Kernel Size: 5, 5, 5, 5, 5
Stride: 2, 2, 2, 2, 2
Activation: ReLU
Number of Final FC Layers: 2
Size of FC Layers: 128, 1024, 1</td></tr>
<tr><td>DNN Expert</td><td>Number of Layers: 3
Size of Layers: 32, 64, 32
Activation: ReLU</td></tr>
<tr><td>Transformer Expert</td><td>Number of Layers: 1
Number of Heads: 1
Size of Hidden Layers: 32
Number of Final FC layers: 2
Size of FC Layers: 128, 128, 1
Activation of FC Layers: Tanh</td></tr>
<tr><td>LSTM Expert</td><td>Number of Layers: 1
Size of Hidden Layers: 128
Activation: Tanh
Number of Final FC layers: 1
Size of FC Layers: 256, 1</td></tr>
<tr><td>TCN Expert</td><td>Number of Channels: 64, 64, 64
Number of Layers: 3
Kernel Size: 2
Number of Final FC layers: 1
Size of FC Layers: 64, 1</td></tr>
</table>

Table 3.2: Configuration of other methods for comparison.

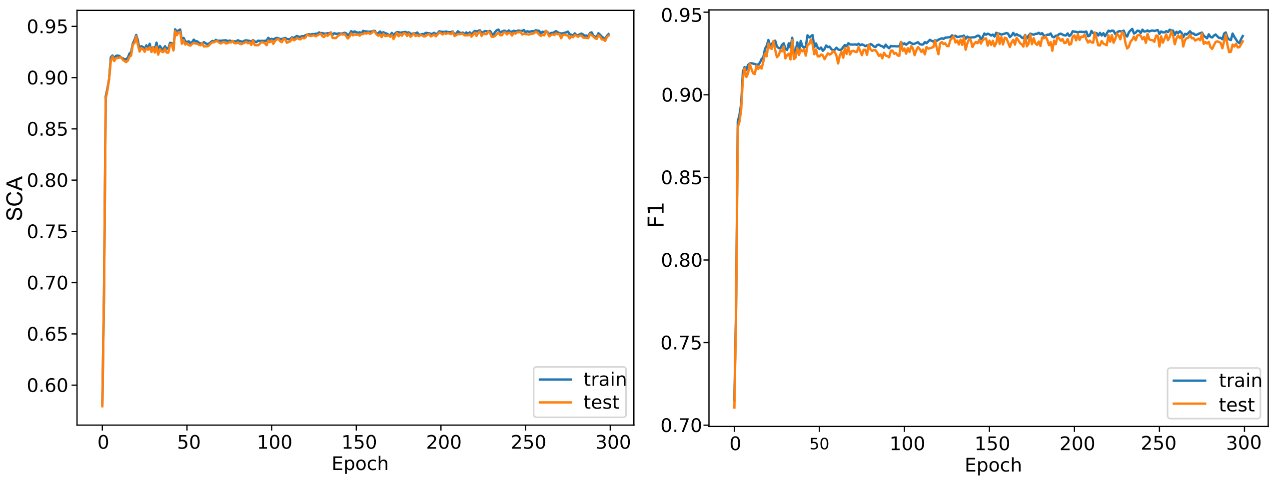
Methods	Hyper-Parameters
CNN	Number of Layers: 5 In Channels: 1, 64, 128, 128, 128 Out Channels: 64, 128, 128, 128, 128 Kernel Size: 5, 5, 5, 5, 5 Stride: 2, 2, 2, 2, 2 Activation: ReLU Number of Final FC Layers: 2 Size of FC Layers: 128, 1024, 1
LSTM	Number of Layers: 3 Size of Layers: 128, 128, 128 Activation: ReLU Number of Final FC layers: 2 Size of FC Layers: 256, 256, 1
TCN	Number of Channels: 64, 64, 64 Number of Layers: 3 Kernel Size: 2 Number of Final FC layers: 1 Size of FC Layers: 64, 1
BiLSTM	Number of Layers: 3 Size of Hidden Layers: 128, 128, 128 Activation: Tanh Number of Final FC layers: 2 Size of FC Layers: 256, 256, 1
GRU	Number of Layers: 3 Size of Layers: 128, 128, 128 Activation: ReLU Number of Final FC layers: 2 Size of FC Layers: 256, 256, 1

3.4.1 Dataset and Preprocessing

The UK-DALE dataset features both aggregated and appliance-level energy consumption records extracted at 6-s intervals from five households, and includes kettle, refrigerator, laundry, and microwave appliances with durations ranging from 36 to 655 days [144]. Similarly, the REDD dataset collects the same data extracted at 6-s intervals from six households, and includes refrigerator, laundry, and microwave appliances with durations greater than 119 days [145]. The real-world Hong Kong dataset was collected to validate the model's performance under realistic, low-frequency conditions. The data collection experiment was designed as follows: power consumption was monitored in a single-family residence in Hong Kong for 152 consecutive days. A primary smart meter at the main electrical panel recorded the aggregate household power at 15-minute intervals, simulating typical utility data. To establish ground truth, a network of appliance-level smart plugs and circuit-level meters was deployed to monitor individual loads. These sub-meters recorded power at 15-minute intervals to align with the aggregate signal. All metering devices were time-synchronized using a central data logger to ensure precise temporal correspondence between the aggregate and ground-truth streams. The dataset includes labeled data for 11 loads: one washer, five air conditioner (AC) units, one kitchen circuit, one bathroom heater (BH), and three water heater (WH) units. The impact of differing magnitudes among the data was eliminated by normalizing the original data to within a uniform range between 0 and 1 as follows: $x_i(t) = \frac{x_i(t) - \min(\boldsymbol{x}_i)}{\max(\boldsymbol{x}_i) - \min(\boldsymbol{x}_i)}, \forall i, t$.

For experiments conducted with the UK-DALE and REDD datasets, the data derived from a single household were designated as the testing dataset, while the data extracted from the remaining households were used as the training dataset. Specifically, for the UK-DALE dataset, data from House 2 were designated as the testing dataset, while data from the remaining four houses (Houses 1, 3, 4, and 5) were used for training. Similarly, for the REDD dataset, data from House 1 were used for testing, with data from the remaining five houses (Houses 2, 3, 4, 5, and 6) constituting the training set. This approach enables the generalization capabilities of DLMs to be assessed spatially across different homes. In contrast, the real-world dataset was divided temporally, where the initial 80% of the time-series data were employed as the training dataset, and the last 20% of the time-series data were reserved as the testing dataset.

To address the potential for overfitting, several mitigation techniques were incorporated, namely dropout, batch normalization, and early stopping criteria based on the best-achieved test performance. We further verified the absence of overfitting by examining the learning curves. The plots of SCA and F1-score for both training and testing phases, illustrated using the refrigerator appliance from the REDD dataset in Fig. 3.4, exhibit similar profiles. This congruence provides strong evidence that our model did not overfit the training data.



(a) SCA curve (b) F1-score curve

Figure 3.4: Curves of SCA and $F1$ score index.

3.4.2 Evaluation Metrics

The state classification performance of the NILM methods was evaluated according to the following state classification accuracy (SCA) of the i-th appliance:

$$SCA_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}, \tag{3.63}$$

where TP_i and TN_i respectively denote the number of correctly classified positive and negative instances, while FP_i and FN_i respectively represent the misclassified positive and negative instances. Accordingly, these terms are defined as follows.

$$TP_i = \sum\nolimits_t \mathbb{I}\left(s_i(t) = 1 \wedge \hat{s}_i(t) = 1\right) \tag{3.64}$$

$$TN_i = \sum\nolimits_t \mathbb{I}\left(s_i(t) = 0 \wedge \hat{s}_i(t) = 0\right) \tag{3.65}$$

$$FP_i = \sum\nolimits_t \mathbb{I}\left(s_i(t) = 0 \wedge \hat{s}_i(t) = 1\right) \tag{3.66}$$

$$FN_i = \sum\nolimits_t \mathbb{I}\left(s_i(t) = 1 \wedge \hat{s}_i(t) = 0\right) \tag{3.67}$$

Here, $\mathbb{I}(\cdot)$ is an indicator function that returns 1 if its argument is true and 0 otherwise. Hence, the classification accuracy increases as the SCA value approaches 100%. While the SCA is a useful metric to evaluate classification performance for individual appliances, it may not fully capture the effectiveness of a NILM model, particularly in cases of class imbalance or when the consequences of false positives and false negatives vary significantly. This limitation was addressed to provide a more balanced evaluation of the classification performance of a DLM by also calculating the F1 score, which is defined as follows as the harmonic mean of the precision (Pre_i) and the recall (Rec_i) for the i-th appliance.

$$F1_i = \frac{2 \cdot (\text{Pre}_i \cdot \text{Rec}_i)}{\text{Pre}_i + \text{Rec}_i} \tag{3.68}$$

These terms can be defined quantitatively as $\text{Pre}_i = \frac{TP_i}{TP_i + FP_i}$, which quantifies the ability of a DLM to avoid false positives, and $\text{Rec}_i = \frac{TP_i}{TP_i + FN_i}$, which quantifies the ability of a DLM to identify all positive samples correctly. Hence, the classification accuracy increases as the F1 score approaches a value of 1.

The power consumption estimation performance of the NILM methods for the i-th appliance was evaluated according to the mean absolute error (MAE) as follows.

$$MAE_i = \frac{1}{T} \sum\nolimits_{t=1}^{T} |\hat{x}_i(t) - x_i(t)| \tag{3.69}$$

In addition, we applied the mean relative error (MRE) to measure the discrepancy between the estimated and ground-truth power consumption values of individual appliances, which is defined as follows.

$$MRE_i = \frac{1}{T} \sum\nolimits_{t=1}^{T} \frac{|\hat{x}_i(t) - x_i(t)|}{\max(\hat{x}_i(t), x_i(t))} \tag{3.70}$$

Hence, power estimation accuracy increases as the MAE and MRE values approach a value of 0.

Due to space limitations, we present the MAE results on the publicly available datasets UK-DALE and REDD, while the MRE results are reported on the real-world Hong Kong dataset to illustrate the percentage error.

Table 3.3: Comparisons of state classification and power load estimation performances of various NILM approaches to two publicly available datasets.

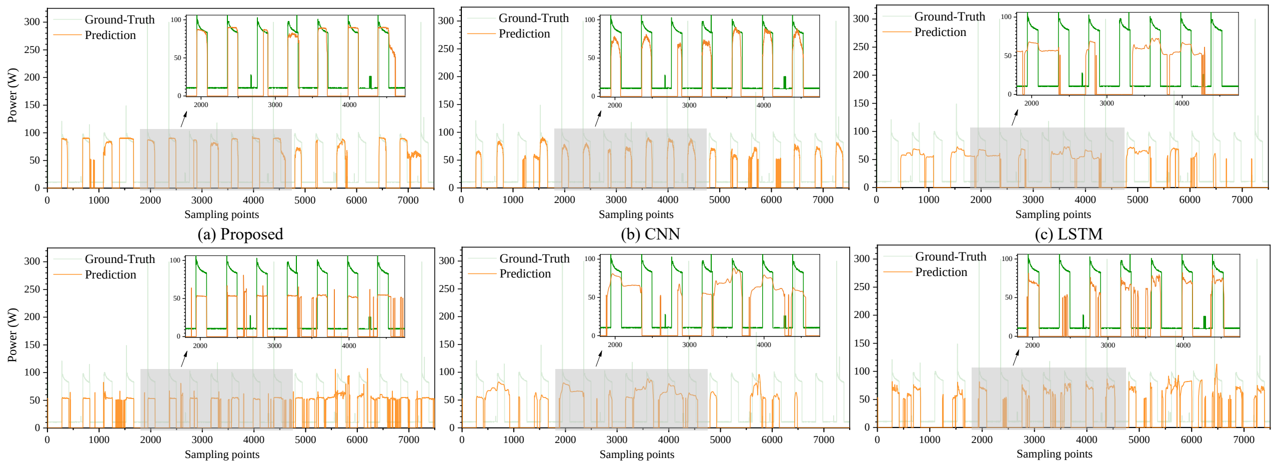
Metrics	Methods	UK-DALE Dataset					REDD Dataset			
		Kettle	Refrigerator	Laundry	Microwave	Average	Refrigerator	Laundry	Microwave	Average
SCA	Proposed	**99.87%**	**87.65%**	**99.69%**	**99.58%**	**96.70%**	**94.88%**	**98.85%**	**99.78%**	**97.55%**
	CNN[57]	98.78%	76.15%	98.70%	99.46%	93.27%	89.34%	97.83%	98.67%	95.28%
	LSTM[59]	99.39%	62.45%	95.22%	99.55%	89.15%	48.26%	97.84%	98.72%	81.61%
	TCN[61]	99.07%	64.27%	96.22%	98.20%	89.44%	50.77%	96.87%	96.97%	81.54%
	BiLSTM[60]	99.83%	72.78%	94.91%	99.50%	91.76%	48.80%	98.83%	98.60%	82.08%
	GRU[58]	99.25%	72.18%	96.41%	99.54%	91.85%	49.45%	98.83%	98.68%	82.32%
	Seq2point[141]	94.23%	61.68%	96.83%	99.62%	89.34%	86.94%	97.99%	87.70%	90.88%
	DT[142]	99.21%	62.17%	95.72%	99.45%	89.13%	83.51%	97.21%	89.01%	89.91%
	RF[143]	99.19%	62.37%	95.53%	99.33%	89.11%	83.92%	97.17%	95.34%	92.14%
	HMM[51]	99.33%	76.15%	96.42%	99.14%	92.76%	86.06%	98.13%	93.25%	92.48%
F1	Proposed	**93.11%**	**85.94%**	**88.96%**	51.94%	**79.99%**	**94.34%**	**60.50%**	**68.88%**	**74.57%**
	CNN[57]	21.52%	66.15%	64.05%	0.88%	38.15%	81.48%	50.94%	0.00%	44.14%
	LSTM[59]	53.96%	44.57%	19.21%	41.63%	39.84%	48.70%	47.19%	50.18%	48.69%
	TCN[61]	8.94%	51.37%	30.00%	5.80%	24.03%	50.10%	38.37%	42.43%	43.63%
	BiLSTM[60]	90.69%	65.07%	19.71%	19.32%	48.70%	48.50%	0.00%	56.21%	34.91%
	GRU[58]	36.57%	61.19%	24.61%	**72.13%**	48.62%	48.65%	0.00%	0.29%	16.31%
	Seq2point[141]	92.34%	62.58%	41.62%	0.00%	49.14%	85.76%	0.00%	0.01%	28.59%
	DT[142]	21.29%	62.56%	38.52%	6.57%	32.24%	80.05%	0.00%	35.90%	38.65%
	RF[143]	22.77%	63.09%	39.02%	7.31%	33.05%	80.91%	0.00%	44.56%	41.82%
	HMM[51]	57.80%	74.59%	66.57%	13.05%	53.00%	90.49%	53.34%	48.26%	64.03%
MAE	Proposed	**8.54**	**13.26**	**8.07**	6.05	**8.98**	**22.64**	**21.53**	**13.07**	**11.92**
	CNN[57]	37.27	31.04	8.99	6.67	20.99	36.98	32.83	18.41	29.41
	LSTM[59]	20.85	39.41	21.75	6.35	22.09	68.78	35.77	17.97	40.84
	TCN[61]	27.76	38.40	18.34	10.96	23.86	59.22	27.88	34.68	38.26
	BiLSTM[60]	10.99	33.48	28.27	6.42	19.79	72.30	35.78	17.45	41.85
	GRU[58]	23.99	33.83	11.61	**5.99**	18.85	61.99	35.78	18.40	38.73
	Seq2point[141]	19.65	37.00	29.99	7.47	23.53	30.26	49.06	15.51	31.61
	DT[142]	29.54	38.37	35.27	7.62	27.70	32.21	30.11	18.52	26.95
	RF[143]	19.58	38.06	35.00	7.61	25.06	31.56	32.21	16.79	26.85
	HMM[51]	14.15	26.22	16.43	7.02	15.96	25.65	26.44	17.82	23.30

3.4.3 Comparisons With Low-Frequency Public Datasets

In this section, we evaluate the proposed model on the UK-DALE and REDD datasets, both sampled at 6-s intervals, which are considered low-frequency in NILM research. Despite the lower temporal resolution, these datasets remain widely used and offer a practical benchmark for real-world NILM applications. The SCA, F1 score, and MAE values obtained by the proposed HMMOE approach and state-of-the-art NILM methods employing homogeneous CNN, TCN, LSTM, BiLSTM, GRU, Seq2point, DT, RF, and HMM for the two publicly available datasets are listed in Table 3.3 for each of the individual types of appliances. As can be seen, the proposed multitask approach significantly outperforms all other methods in terms of all three metrics for almost all appliance types, as well as in terms of the averages of all appliances for both datasets. In comparison, the methods based on homogeneous networks exhibit more varied performance across the different appliance types. In particular, these methods exhibit low state classification and power estimation accuracy for refrigerators in the UK-DALE and REDD datasets, although the CNN-based and HMM approaches yield values close to those of the proposed method for refrigerator loads. Hence, the superior performance of the proposed HMMOE approach can be attributed to the more nuanced prediction obtained by leveraging multiple heterogeneous expert networks, where each network potentially learns specific features relevant to different types of appliances, as well as to the gating mechanism employed for dynamically controlling how much each expert block contributes to the final prediction, which effectively allows the model to adapt its prediction strategy based on real-time data. It is important to note that the relatively high F1 scores and SCA achieved for refrigerators in the REDD and UK-DALE datasets can be attributed, in part, to the higher sampling frequency (every 6 seconds) and the preprocessing techniques applied to the data. These factors enhance the model's ability to distinguish the refrigerator's cyclical usage patterns from background noise. Additionally, the use of heterogeneous expert networks allows the framework to capture both short-term cycling and long-term periodicity, which are key characteristics of refrigerator operation.

It is important to highlight that the low F1 score observed for the microwave appliance does not indicate a standalone model failure, but rather reflects meaningful insights arising from two interconnected factors. First, the microwave operates using rapid magnetron cycling, which generates a complex and non-stationary power signature. This makes it inherently more difficult to model compared to appliances like kettles and refrigerators, which exhibit more stable and predictable usage patterns. Second, the microwave is used infrequently and for short durations, resulting in severe class imbalance within the time-series data. Given that the F1 score is the harmonic mean of precision and recall, it is particularly sensitive to the high number of false negatives that can occur when the model attempts to detect such rare and noisy events. While our proposed HMMOE approach may not achieve the highest F1 score for the microwave appliance, it still outperforms existing methods across other appliances.

The performance of the various NILM methods can be evaluated more intuitively by comparing the predicted disaggregated power consumption results obtained by the various NILM methods for the refrigerator appliance type in the UK-DALE and REDD datasets against the ground-truth values, as presented in Figs. 3.5 and 3.6, respectively. It can be observed that the proposed HMMOE approach tends to capture the timing and the magnitude of the energy consumption spikes in the sampled data without significant overshooting or delays better than the existing CNN, LSTM, TCN, BiLSTM, and GRU methods, which exhibit more distorted predictions, such as predictions that either lead or lag behind the actual usage times. These findings are more clearly indicated by the



(a) Proposed (b) CNN (c) LSTM

(d) TCN (e) BiLSTM (f) GRU

Figure 3.5: Comparison of the disaggregated power consumption results obtained by the various NILM methods for the refrigerator appliance type in UK-DALE dataset.

insets of Fig. 3.5 obtained over a subset of the sampling period for the UK-DALE dataset. Hence, the proposed approach offers greater accuracy and sensitivity to changes in appliance behavior than the existing methods considered. Nonetheless, we note that the CNN-based approach provided a prediction performance close to that of the proposed method.

3.4.4 Comparisons With Real-World Low-Frequency Dataset

We further evaluate our model on a real-world dataset collected in Hong Kong, which uses a 15-min sampling interval, a typical setting for smart meters deployed in residential buildings. This represents a very low-frequency scenario and tests the model's ability to detect appliance usage with sparse input. The prediction performance of the proposed HMMOE approach to the real-world dataset captured at a sampling interval of 15 min was evaluated by comparisons with the performance of the CNN-based approach alone because this homogeneous model typically exhibited a performance second only to that of the proposed model. It is important to note that the 'Kitchen' load represents the aggregated consumption from a single circuit serving multiple kitchen appliances, such as a toaster, a coffee maker and an electric kettle, providing a realistic test of the model's ability to handle complex, superimposed load signatures. The SCA, F1 score, and MRE values obtained by the proposed HMMOE approach the CNN baseline, and two advanced multitask learning models, such as the

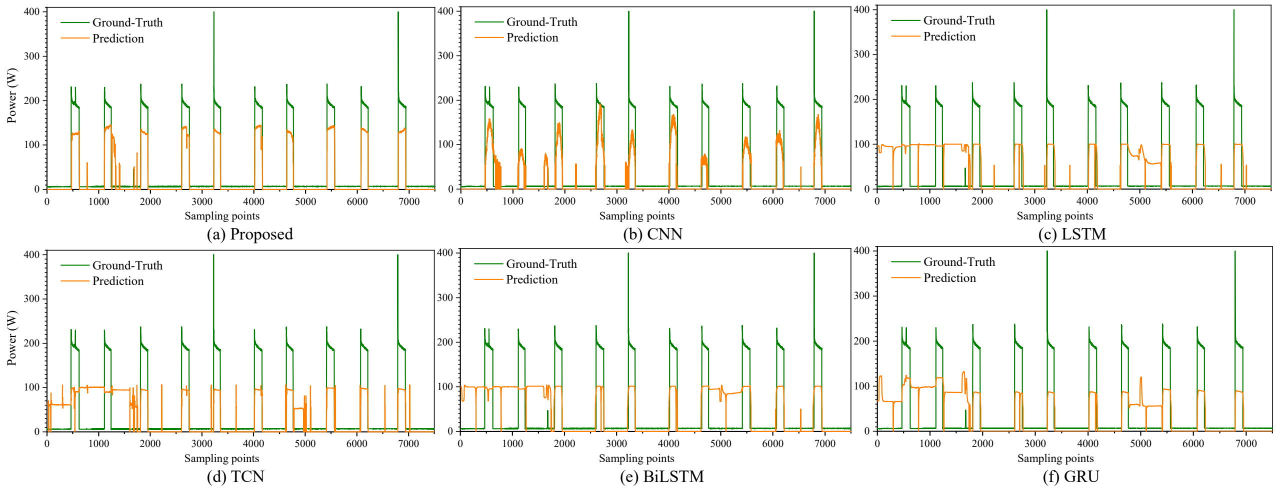


Figure 3.6: Comparison of the disaggregated power consumption results obtained by the various NILM methods for the refrigerator appliance type in REDD dataset.

scale- and attention-experts based multitask neural network (SAMNet) [63] and the CNN-BiLSTM-based multitask neural network (CBMNet) [64], are listed in Table 3.4 for each of the 11 appliances for which data were labeled in the dataset. The results indicate that SAMNet and CBMNet, as advanced multitask architectures, generally outperform conventional single-task deep learning models like CNN, thus validating their effectiveness as strong state-of-the-art baselines. Nevertheless, the proposed HMMOE framework delivers consistently superior performance across different appliances and evaluation metrics, achieving the highest average scores overall. This finding supports our central hypothesis: the architectural diversity of the parallel expert ensemble in our HMMOE model provides a more robust and comprehensive feature representation than a single-pipeline multitask architecture, allowing it to adapt more effectively to the varied characteristics of different appliances.

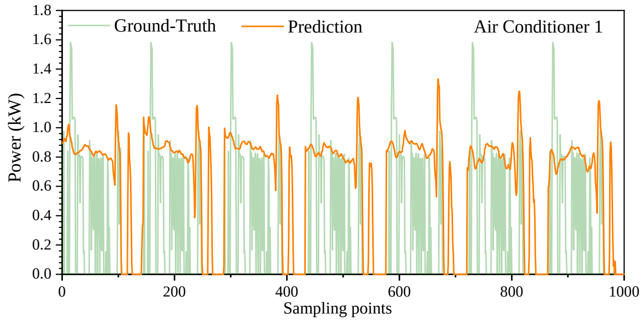
In addition, the predicted disaggregated power consumption results obtained by the two NILM methods for AC1 are presented in Fig. 3.7 in conjunction with the ground-truth values. The consistent superiority of the proposed approach in terms of the SCA, F1 score, and MRE values obtained across a complex mix of appliance types demonstrates the good robustness and adaptability of the proposed approach under conditions of low sampling frequency. We further note from Fig. 3.7 that the proposed HMMOE approach again captures the timing and the magnitude of the energy consumption spikes in the sampled data much better than the existing CNN under the low sampling frequency data.

Note that, the proposed model performs well across most appliances, but lower F1 scores are observed for the washer, kitchen group, and BH, each of which presents unique challenges. The washer operates in multiple phases, such as motor tumbling, draining, and high-speed spinning, with intermittent power usage. These phases often resemble background noise and may overlap with other appliances, making classification more difficult. The kitchen group includes several small appliances (e.g., toaster, coffee maker, and electric kettle) connected to a single circuit. These devices typically produce short, high-power bursts, which can be missed with a 15-minute sampling interval, leading to reduced recall. The shared circuit also obscures individual appliance signatures. As for BH, it is used infrequently and irregularly, resulting in a severe class imbalance with very few positive instances. Its power consumption is also similar to that of WHs, further complicating accurate identification. Importantly, these challenges are inherent to the data and not specific to our approach. In fact, our method improves F1 scores compared to the CNN-based baseline, which was previously the second-best performer on the UK-DALE and REDD datasets.

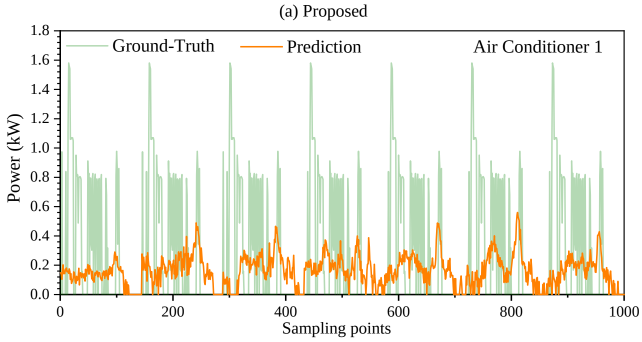
Table 3.4: Comparisons of the state classification and power load estimation performances of the proposed HMMOE approach and the advanced NILM approaches to the real-world dataset.

Metrics	Methods	Real-world Hong Kong Dataset											
		Washer	AC1	AC2	AC3	AC4	AC5	Kitchen	BH	WH1	WH2	WH3	Average
SCA	Proposed	**92.89%**	**90.70%**	**88.07%**	**88.07%**	**87.47%**	**88.57%**	**94.86%**	**98.98%**	**97.69%**	**97.86%**	**97.54%**	**92.97%**
	CNN[57]	89.31%	71.35%	67.06%	80.33%	83.57%	73.41%	91.54%	48.26%	73.42%	77.78%	92.84%	77.17%
	SAMNet [63]	47.13%	87.98%	86.88%	83.11%	87.20%	86.51%	94.49%	46.72%	96.49%	96.58%	96.53%	82.69%
	CBMNet [64]	85.17%	83.56%	80.08%	81.14%	85.35%	84.53%	93.89%	65.67%	97.49%	97.61%	97.24%	86.52%
F1	Proposed	**14.36%**	**74.36%**	**58.93%**	**71.37%**	**62.34%**	**68.57%**	**24.34%**	**2.61%**	**63.52%**	**66.21%**	**42.06%**	**49.61%**
	CNN[57]	5.67%	47.48%	36.87%	32.73%	33.48%	53.08%	4.67%	0.67%	15.22%	17.12%	3.79%	22.80%
	SAMNet [63]	12.06%	65.74%	55.71%	71.28%	57.98%	67.37%	5.45%	1.28%	6.36%	2.94%	7.04%	32.11%
	CBMNet [64]	7.62%	60.06%	49.27%	64.37%	57.86%	66.49%	16.12%	2.09%	60.65%	58.78%	24.09%	42.49%
MRE	Proposed	**10.50%**	**14.49%**	**18.05%**	**16.75%**	**16.31%**	**18.03%**	**6.63%**	**1.32%**	**4.93%**	**4.47%**	**4.21%**	**10.52%**
	CNN[57]	11.01%	36.78%	39.65%	22.68%	19.13%	37.02%	8.63%	51.90%	28.58%	24.00%	7.29%	26.06%
	SAMNet [63]	10.82%	15.59%	27.99%	21.37%	18.42%	28.35%	7.53%	5.53%	13.79%	8.31%	10.75%	15.31%
	CBMNet [64]	11.82%	23.99%	21.68%	31.65%	22.51%	21.47%	8.22%	3.32%	7.60%	10.98%	5.11%	15.30%

Our proposed HMMOE model contains 2,158,415 trainable parameters. This complexity is a direct result of its sophisticated architecture, which includes five heterogeneous expert blocks, two gating networks, and two task-specific towers designed to capture a wide array of features from NILM data. To contextualize this, we benchmark our model against other state-of-the-art multitask NILM frameworks in Table 3.5. As shown, the parameter count of our model is on par with leading approaches such as SAMNet and CBMNet, positioning it well within the competitive range for



(a) Proposed



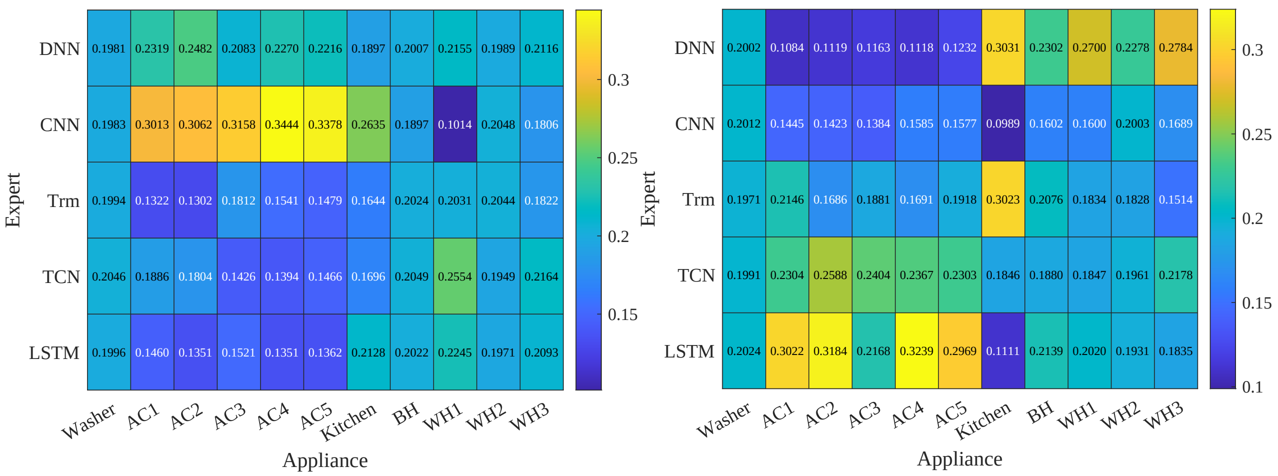
(b) CNN

Figure 3.7: Comparison of the disaggregated power consumption results obtained by the various NILM methods for air conditioner 1 (AC1) in the real-world dataset.

Table 3.5: Model parameter number and testing time

Model	Parameter Number	Testing Time (s)
Proposed	2158415	1.6953
SAMNet[63]	1586417	1.6521
CBMNet[64]	2850588	1.3000

high-performance NILM systems. Importantly, this level of complexity is not excessive but rather necessary to tackle the challenging dual-objective NILM problem, especially when working with low-frequency data. The architectural depth is a deliberate design choice aimed at enhancing the accuracy of disaggregation. To further evaluate the practicality of the model, we measured its computational efficiency. On an NVIDIA GeForce RTX 4090, the model processes a batch of 256 parallel samples in just 1.6953 seconds during testing. This demonstrates that the model is not only accurate but also computationally efficient, making it well-suited for near-real-time deployment in practical applications such as the standard 15-minute interval power monitoring.



(a) State identification task

(b) Power estimation task

Figure 3.8: Gate weights per appliance and task in the real-world dataset.

3.4.5 Analysis of Expert Contribution and Specialization

To justify the heterogeneous architecture and enhance the interpretability of expert contributions, we conducted a quantitative analysis of the internal behavior of the proposed MMOE framework. Specifically, we examined which experts are prioritized across different appliances and tasks. The findings are illustrated in Fig. 3.8. For the AC units, we observe a clear divergence in expert utilization between the classification and regression tasks. As shown in Fig. 3.8(a), the classification gate assigns the highest weight to the CNN expert, highlighting its effectiveness in detecting sharp on/off transitions in aggregated signals. In contrast, the regression task prioritizes temporal experts, primarily the LSTM and TCN, to better capture the fluctuating power consumption during prolonged operation. This suggests that the model adopts a dynamic and task-specific strategy, leveraging different expert combinations not only across appliances but also within the same appliance for distinct objectives.

Similar trends are evident in other appliances. For the kitchen circuit, which includes devices with short-duration, high-power usage, the CNN expert is strongly favored for classification. In the case of WH units, characterized by sustained and distinct power patterns, the regression task relies on the DNN expert, while the TCN is preferred for classification. These insights offer compelling quantitative support that the model performs intelligent, task-aware expert routing rather than merely aggregating expert outputs in a uniform manner.

Table 3.6: Comparison of the state classification and power load estimation performances of the proposed heterogeneous and corresponding homogeneous expert architectures for the real-world dataset.

Metrics	Experts	Real-world Hong Kong Dataset					
		Washer	AC	Kitchen	BH	WH	Average
SCA	Heterogeneous	**92.89%**	**88.58%**	**94.86%**	**98.98%**	**97.70%**	**92.97%**
	Homogeneous	90.45%	87.50%	94.37%	77.32%	96.83%	90.01%
F1	Heterogeneous	**14.36%**	**66.41%**	**24.34%**	**2.61%**	**57.26%**	**49.61%**
	Homogeneous	10.65%	63.97%	2.48%	1.43%	15.46%	34.62%
MRE	Heterogeneous	**10.50%**	**16.73%**	**6.63%**	**1.32%**	**4.54%**	**10.52%**
	Homogeneous	17.90%	19.30%	6.86%	3.18%	3.98%	12.40%

3.4.6 Comparisons Between Heterogeneous and Homogeneous Experts

To conduct a rigorous ablation study isolating the specific impact of architectural diversity, we compared the performance of our proposed HMMOE framework with several homogeneous variants. Specifically, we designed and trained five homogeneous baseline models, each maintaining the overall HMMOE structure but replacing the five distinct expert modules (DNN, CNN, Transformer, TCN, and LSTM) with five identical experts of a single type (e.g., one model with five CNN experts, another with five TCN experts, and so on). Table 3.6 presents the SCA, F1 score, and MRE results for the proposed HMMOE and its homogeneous counterparts, evaluated per appliance on the real-world dataset. The performance metrics reported for homogeneous configurations reflect the best scores achieved among the five homogeneous models. This approach ensures a fair and robust assessment, highlighting the intrinsic advantage of architectural heterogeneity. Analysis of the results indicates that the proposed heterogeneous network architecture provides an average improvement of 2.96% in SCA values, 14.99% increase in F1 scores, and 1.9% reduction in MRE values. These results demonstrate the benefits of applying heterogeneous expert blocks in a multi-gate model, where the strengths of different types of neural networks can be leveraged to extract specific features of time-series load data accurately, as well as extract a wider variety of features. Moreover, the ensemble of different expert models can reduce model bias and variance compared to the homogeneous structure using multiple instances of the same type of model. Furthermore, the designed heterogeneous architecture can be easily extended to include innovative neural networks as they are developed in the future.

Table 3.7: Performance comparisons of the proposed HMMOE approach with dynamic moo versus a baseline with fixed, equal task weights (α^c= 0.5,α^r = 0.5) on the real-world Hong Kong dataset.

Metrics	Weights	Real-world Hong Kong Dataset					
		Washer	AC	Kitchen	BH	WH	Average
SCA	With MOO	**92.89%**	**88.58%**	**94.86%**	**98.98%**	**97.70%**	**92.97%**
	Without MOO	92.58%	86.00%	94.30%	98.10%	97.24%	91.52%
F1	With MOO	**14.36%**	**66.41%**	**24.34%**	**2.61%**	**57.26%**	**49.61%**
	Without MOO	9.68%	33.74%	6.14%	0.67%	26.13%	23.96%
MRE	With MOO	**10.50%**	**16.73%**	**6.63%**	**1.32%**	4.54%	**10.52%**
	Without MOO	16.82%	18.70%	8.65%	1.61%	**3.56%**	11.93%

3.4.7 Comparisons With and Without the Proposed Multi-Objective Optimization Procedure

To validate the superiority of the proposed dynamic optimization strategy, we compared the performance of our full HMMOE framework against a baseline version trained without the MOO process. For this baseline case, we employed a standard static weighting scheme, setting the task weights for both state classification and power estimation to a fixed value of 0.5. This equal weighting represents a common approach in multi-task learning and serves as a neutral benchmark to isolate the benefits of our adaptive strategy. The SCA, F1 score, and MRE values obtained by the proposed HMMOE approach with and without the MOO process are listed in Table 3.7 for each type of appliance in the real-world dataset. As can be seen, applying the MOO procedure increases the average SCA value across all appliances from 91.52% to 92.97%, and the average F1-score improves significantly from 23.96% to 49.61%. Hence, the MOO process provides a better balance between precision and recall in state classification than that obtained without the MOO process. Similarly, the average MRE decreases from 11.93% to 10.52%, representing an enhanced accuracy in power consumption estimation. Moreover, these improvements are consistent across all appliance types, including AC and WH appliance types, which often exhibit complex and variable consumption patterns. Hence, these results demonstrate that the proposed MOO procedure effectively balances the objectives associated with multiple NILM tasks, leading to more robust and generalized model performance. As a result, the MOO procedure increases the effectiveness of the proposed HMMOE approach to NILM applications involving low sampling frequency data.

3.5 Summary

The present work addressed a number of limitations in existing multitask NILM frameworks by proposing an HMMOE framework that combines multiple heterogeneous expert network blocks with a multi-gate integration architecture, and obtains a good tradeoff between the appliance state classification and power consumption estimation tasks by formulating the multitask learning problem as a multi-objective optimization problem. The performance of the proposed NILM approach was evaluated based on numerical results obtained for numerous tests conducted with two publicly available datasets sampled at very short 6-s intervals and a real-world dataset sampled at realistically long 15-min intervals. The results demonstrated that the proposed HMMOE approach significantly outperforms existing state-of-the-art homogenous DLM-based NILM methods, achieving SCA values increased by up to 15% and MAE values decreased by 14%. In addition, comparisons between the proposed heterogeneous expert architecture and corresponding homogenous expert architectures demonstrated that the heterogeneous framework leverages the diverse strengths of the DNN, CNN, transformer, TCN, and LSTM network architectures to capture both local and global patterns in aggregated power consumption data, achieving average SCA values increased by 2.96%, F1 scores increased by 14.99%, and MRE values reduced by 1.9%. Finally, the developed multi-objective optimization procedure was demonstrated to achieve better balance between the separate tasks than fixed-weight approaches, improving the average SCA by 1.45% and F1 score by 25.65%, and reducing the average MRE by 1.41%. Furthermore, the proposed model is inherently scalable, and can be extended easily to include additional or new expert blocks as needed. Specifically, additional experts can be integrated without major redesigns of the entire architecture to accommodate increasing complexity, such as when more appliance types or varied user behaviors must be considered. Also, the model can be customized for specific scenarios or datasets by adjusting the number and nature of network experts and gates.

While the proposed HMMOE framework shows significant performance gains, it is important to acknowledge its limitations, which also suggest avenues for future research. First, the model's enhanced feature extraction capabilities, derived from its ensemble of five heterogeneous experts, come with a trade-off in computational complexity. Future work could explore model compression and knowledge distillation techniques to create a more lightweight version of the framework without significant performance degradation. Additionally, the model's performance on appliances with very short operational cycles is inherently constrained by the information loss at very low sampling frequencies, such as the 15-minute intervals used in our real-world dataset. While our model pushes the boundaries of inference from sparse data, no algorithm can perfectly reconstruct a signal that was not captured. Future research could investigate the fusion of low-frequency power data with other contextual data streams (e.g., time-of-day, occupancy sensors) to help mitigate this fundamental limitation.

Chapter 4

Flexible and Robust Operation of Hybrid Microgrids with Multimodal Multi-Task Load Disaggregation

Despite recent progress in NILM and microgrid energy management, several critical research gaps remain. First, most existing NILM studies focus either on appliance state detection or power regression using single-task formulations. Such approaches often rely solely on electrical signatures and require manual threshold calibration, which leads to limited generalization, particularly for appliances with variable operating modes such as HVAC systems. Moreover, the lack of multimodal data integration means that contextual drivers of energy consumption, such as weather conditions, are often overlooked, resulting in reduced accuracy and reliability in real-world settings.

Second, while multi-task NILM methods have been explored, the problem of balancing competing objectives is typically addressed through heuristic or empirically tuned task weights. This ad hoc treatment lacks theoretical guarantees and can lead to unstable training, where one task dominates at the expense of the other. Finally, there remains a methodological disconnect between smart meter analytics and microgrid-level energy management. Most studies treat NILM and energy scheduling as independent problems, without leveraging appliance-level disaggregation to inform demand response strategies. As a result, the full potential of NILM in enhancing renewable integration, demand flexibility, and cost efficiency in hybrid microgrids remains largely untapped.

To address the aforementioned research gaps, this thesis develops a two-task multimodal NILM framework with an analytical multi-objective balancing strategy and integrates its outputs into robust hybrid microgrid energy management, thereby advancing non-intrusive demand response evaluations, improving task coordination, and bridging the gap between smart meter analytics and operational decision-making. The contributions of this chapter are threefold:

1) Dual-Task Multimodal Learning for Non-Intrusive Demand Response Evaluations: We propose a dual-task multimodal learning approach that jointly performs appliance state classification and power consumption estimation. Unlike conventional NILM models relying only on electrical data in smart meters, our method integrates smart meter measurements with environmental variables such as temperature, humidity, and solar radiation. This multimodal design captures the causal drivers of appliance behavior, particularly for weather-sensitive loads, while the dual-task formulation eliminates the need for manual thresholding and enhances both classification accuracy and regression precision. The proposed framework thus enables reliable non-intrusive evaluations of household demand response capacity in hybrid microgrids.

2) Analytical Multi-Objective Solution for Balancing Two Tasks: We develop a novel analytical solution to dynamically balance the dual objectives of state classification and power estimation during model training. By formulating the optimization as a Pareto-efficient multi-objective problem, we derive a closed-form gradient-based condition that adaptively determines task weights without manual tuning. This ensures that neither task dominates the learning process, mitigates gradient conflicts, and guarantees that both objectives contribute effectively to model convergence. The proposed analytical balancing mechanism provides interpretability and robustness compared to existing empirically chosen weighting schemes.

3) Bridging the Gaps Between Smart Meter Analytics and Microgrid Energy Management: We establish a unified framework that connects appliance-level NILM analytics with robust hybrid microgrid energy management. Specifically, the outputs of the dual-task multimodal NILM model, i.e., estimated appliance states and power consumption, are embedded into a robust optimization framework for scheduling distributed resources and demand-side flexibility. This integration bridges the methodological gap between data-driven analytics and operational decision-making, enabling more accurate evaluation of demand response potential, and cost-effective building microgrid operation.

4.1 Problem Statement of Nonintrusive Load Dissaggregation Based on Smart Meter Data

In microgrid applications, NILM enables detailed energy usage disaggregation and appliance-level monitoring to optimize local energy management and demand response. Specifically, the aggregate electrical consumption data at a single point of measurement can be used to identify and monitor individual appliances and devices within a smart home without requiring separate sensors on each device. Here, we define the electrical consumption profile of the i-th household appliance across discrete time steps t spanning duration T as the vector $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,T}] \in \mathcal{X}$. The aggregated smart meter measurements are represented as $\boldsymbol{Y}^{\mathrm{SM}} = [\boldsymbol{y}_1^{\mathrm{SM}}, \boldsymbol{y}_2^{\mathrm{SM}}, ..., \boldsymbol{y}_{\mathrm{T}}^{\mathrm{SM}}] \in \mathcal{Y}^{\mathrm{SM}}$. Typically, NILM involves minimizing the reconstruction error ε_t between observed aggregate consumption $\boldsymbol{y}_t$ and the estimated sum of each individual appliance load $x_{i,t}$ at each time step $\forall t \in \mathcal{T}$:

$$\min \varepsilon_t = |\boldsymbol{y}_t^{\mathrm{M}} - \sum\nolimits_{i=1}^{I} s_{i,t} x_{i,t}|. \tag{4.1}$$

Subject to the operational constraints:

$$s_{i,t} \in \{0, 1\}, x_{i,t} \geq 0, \quad \forall i \in \mathcal{I}, t \in \mathcal{T}. \tag{4.2}$$

Current approaches determine appliance operational status $s_i(t)$ through power threshold comparison:

$$s_{i,t} = \begin{cases} 0, \ i\text{th appliance inactive when} \quad x_{i,t} < \tau_i \\ 1, \ i\text{th appliance active when } x_{i,t} \geq \tau_i \end{cases} \tag{4.3}$$

The threshold parameter τ_i critically influences state detection accuracy in conventional single-objective models. This approach introduces dataset-specific bias and reduces model transferability. Moreover, appliances with variable standby consumption or noisy operational patterns present significant challenges for threshold-based classification.

It is noteworthy that, the quality of smart meter data, including measurement noise, missing values, and potential communication errors, directly affects the construction of data-driven uncertainty sets and the resulting operational decisions. Inaccurate or biased measurements may distort the estimated bounds and correlations of renewable generation and load profiles, leading either to overly conservative dispatch decisions or to underestimated operational risks. To mitigate such effects, basic data validation and cleaning procedures, including outlier detection and interpolation of missing samples, are applied prior to model construction. Nevertheless, it should be noted that extreme data imperfections may still propagate into the uncertainty sets and influence the robustness–cost trade-off observed in the numerical results.

4.2 Heterogeneous Multitask Learning with Multi-Source Data

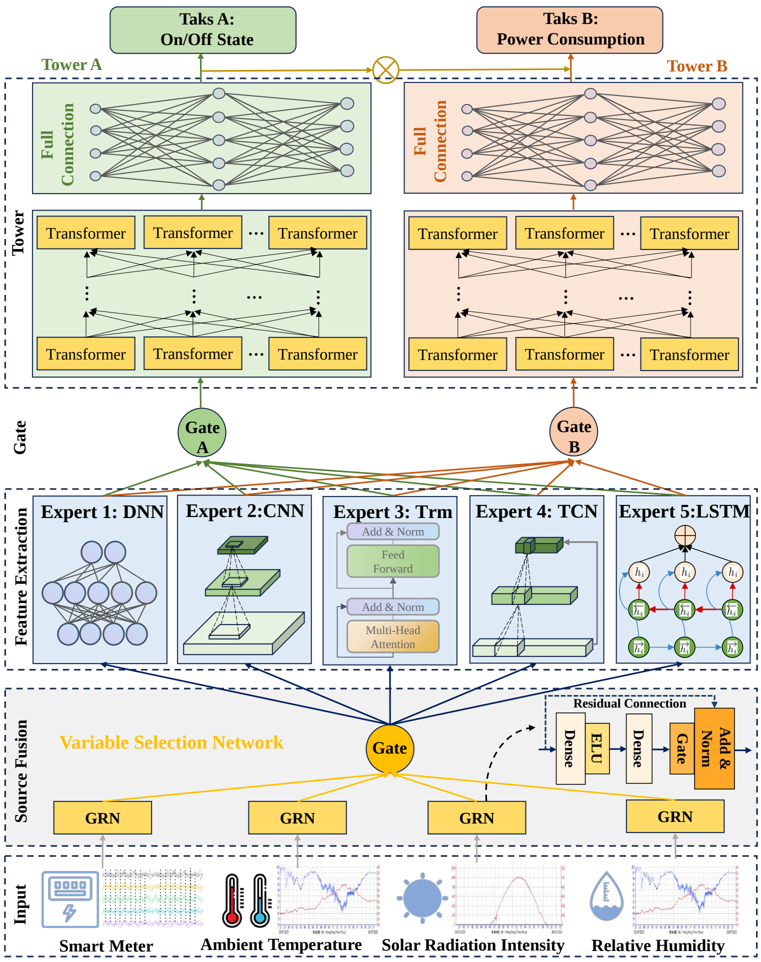


Figure 4.1: Structure of Our Proposed Multitask Neural Network.

Our proposed methodology addresses these limitations through simultaneous two-task learning, eliminating manual threshold calibration. The framework concurrently addresses: (i) binary state classification for appliance operation detection, and (ii) continuous power regression for consumption quantification. We seek an optimal neural mapping function $\mathcal{G}$ that jointly performs both tasks. The dual-objective formulation is expressed as:

$$\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, ..., \hat{\boldsymbol{x}}_i, ... = \hat{\mathcal{X}} = \mathcal{G}(\mathcal{Y}^{\mathrm{SM}}; \Theta), \tag{4.4}$$

$$\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, ..., \hat{\boldsymbol{s}}_i, ... = \hat{\mathcal{S}} = \mathrm{softmax}(\mathcal{G}(\mathcal{Y}^{\mathrm{SM}}; \Theta)), \tag{4.5}$$

where Θ denotes the complete neural network parameter set, and the softmax activation transforms network outputs into valid probability distributions for state classification. This joint learning paradigm enables mutual task reinforcement, where classification outputs inform regression predictions, particularly beneficial during low-activity periods and ambiguous consumption patterns. The incorporation of environmental covariates provides contextual information that correlates with appliance usage behaviors (e.g., temperature-dependent HVAC operation, daylight-influenced lighting patterns).

The structure of our proposed dual-task multi-source data fusion neural network is depicted in Fig. 4.1. Our proposed HMMOE framework comprises four primary components: (i) adaptive multi-source data fusion with variable selection, (ii) diversified expert processing modules, (iii) task-specific gating mechanisms, and (iv) specialized task-oriented towers for final predictions.

4.2.1 Adaptive Multi-Source Data Fusion with Variable Selection

Our model is designed to process four parallel, heterogeneous time-series data streams, as depicted in Fig. 4.1. Unlike current NILM methods that only use the power information of smart meters, we enhance disaggregation performance by incorporating three auxiliary environmental data streams that can be obtained from publicly available meteorological databases. These include an electrical signal from a smart meter and three lower-frequency environmental signals from a public data repository. Specifically, three exogenous data streams, including ambient temperature measurements $\boldsymbol{Y}^{\mathrm{TEM}}$, relative humidity readings $\boldsymbol{Y}^{\mathrm{HUM}}$, and solar radiation intensity data $\boldsymbol{Y}^{\mathrm{SOL}}$, are sourced from a publicly accessible repository, such as the Hong Kong Observatory [146].

$$\boldsymbol{Y}^{\text{TEM}} = [\boldsymbol{y}_1^{\text{TEM}}, \boldsymbol{y}_2^{\text{TEM}}, \ldots, \boldsymbol{y}_T^{\text{TEM}}] \in \mathcal{Y}^{\text{TEM}} \tag{4.6}$$

$$\boldsymbol{Y}^{\text{HUM}} = [\boldsymbol{y}_1^{\text{HUM}}, \boldsymbol{y}_2^{\text{HUM}}, \ldots, \boldsymbol{y}_T^{\text{HUM}}] \in \mathcal{Y}^{\text{HUM}} \tag{4.7}$$

$$\boldsymbol{Y}^{\text{SOL}} = [\boldsymbol{y}_1^{\text{SOL}}, \boldsymbol{y}_2^{\text{SOL}}, \ldots, \boldsymbol{y}_T^{\text{SOL}}] \in \mathcal{Y}^{\text{SOL}} \tag{4.8}$$

Rather than simple concatenation of heterogeneous inputs, we adopt an attention-based variable selection mechanism that dynamically weights each data source based on its instantaneous relevance. This approach is particularly crucial for NILM applications where the causal influence of environmental factors varies significantly across time (e.g., HVAC consumption strongly correlates with temperature during extreme weather but shows weak correlation during mild conditions).

To capture potential interactions between different data modalities, we construct a comprehensive multi-source representation by concatenating all embeddings:

$$\Xi_t = [\boldsymbol{y}_t^{\text{SM}}, \boldsymbol{y}_t^{\text{TEMP}}, \boldsymbol{y}_t^{\text{HUM}}, \boldsymbol{y}_t^{\text{SOL}}] \tag{4.9}$$

This concatenated representation Ξ_t serves as the input to our variable selection network, enabling the model to learn cross-modal dependencies when determining source importance.

Before designing the variable selection mechanism, we introduce the gated residual network (GRN) [147], which serves as the fundamental building block throughout our architecture. The GRN enables flexible non-linear processing with adaptive depth through component gating. Given inputs $\Xi_t = \boldsymbol{y}_t^r, \forall r \in \mathcal{R} = \{\text{SM}, \text{TEMP}, \text{HUM}, \text{SOL}\}$, GRN is formulated as:

$$\text{GRN}_\omega(\Xi_t) = \text{LayerNorm}(\Xi_t + \text{GLU}_\omega(\boldsymbol{\eta}_1)) \tag{4.10}$$

$$\boldsymbol{\eta}_1 = \boldsymbol{W}_{1,\omega}\boldsymbol{\eta}_2 + \boldsymbol{b}_{1,\omega} \tag{4.11}$$

$$\boldsymbol{\eta}_2 = \begin{cases} \boldsymbol{W}_{2,\omega}\Xi_t + \boldsymbol{b}_{2,\omega}, & \text{if } \boldsymbol{W}_{2,\omega}\Xi_t + \boldsymbol{b}_{2,\omega} \geq \mathbf{0}, \\ \alpha\left(\exp\left(\boldsymbol{W}_{2,\omega}\Xi_t + \boldsymbol{b}_{2,\omega}\right) - \mathbf{1}\right), & \text{if } \boldsymbol{W}_{2,\omega}\Xi_t + \boldsymbol{b}_{2,\omega} < \mathbf{0}, \end{cases} \tag{4.12}$$

where $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are intermediate layers, LayerNorm denotes standard layer normalization, $\alpha > 0$ is a hyperparameter controlling the saturation value for negative inputs, and ω is an index for weight sharing. The gated linear unit (GLU) provides adaptive gating:

$$\text{GLU}_\omega(\boldsymbol{\eta}_1) = \sigma(\boldsymbol{W}_{3,\omega}\boldsymbol{\eta}_1 + \boldsymbol{b}_{3,\omega}) \odot (\boldsymbol{W}_{4,\omega}\boldsymbol{\eta}_1 + \boldsymbol{b}_{4,\omega}) \tag{4.13}$$

where $\sigma(\cdot)$ is the sigmoid activation function, $\odot$ denotes element-wise Hadamard product, and all weight matrices $\boldsymbol{W}_{(\cdot)}$ with corresponding biases $\boldsymbol{b}_{(\cdot)}$. The GLU mechanism allows the network to suppress unnecessary components by producing outputs close to zero when non-linear processing is not beneficial.

To dynamically assess the relevance of each data source, we employ a variable selection network that computes instance-wise importance weights. The source selection weights are generated by feeding the concatenated representation through a GRN followed by a softmax normalization layer:

$$\boldsymbol{v}_t^{\text{multi}} = \text{softmax}(\text{GRN}_{v^{\text{multi}}}(\Xi_t)) \tag{4.14}$$

where $\boldsymbol{v}_t^{\text{multi}} = [v_t^{\text{SM}}, v_t^{\text{TEMP}}, v_t^{\text{HUM}}, v_t^{\text{SOL}}]^T$ represents the importance weights for each source at time t, satisfying $\sum_r v_t^r = 1$ and $0 \leq v_t^r \leq 1$. The normalization ensures proper probability distribution while the GRN enables non-linear transformation of the concatenated features to capture complex inter-modal relationships.

To extract source-specific features while maintaining architectural flexibility, each embedded source undergoes independent non-linear processing through dedicated GRNs with unshared parameters:

$$\tilde{\boldsymbol{y}}_t^r = \text{GRN}_{\tilde{\boldsymbol{y}}^r}(\boldsymbol{y}_t^r), \quad \forall r \in \mathcal{R} \tag{4.15}$$

where $\tilde{\boldsymbol{y}}_t^r$ represents the processed feature vector for source r. Each $\text{GRN}_{\tilde{\boldsymbol{y}}^r}$ has distinct learnable parameters, enabling the network to learn source-specific transformations that are most relevant for appliance disaggregation. This design is particularly important given the heterogeneous nature of electrical and environmental measurements.

The final multi-source representation is obtained through the importance-weighted combination of processed source features:

$$\boldsymbol{Y}_t^{\text{multi}} = \sum\nolimits_{r \in \mathcal{R}} v_t^r \cdot \tilde{\boldsymbol{y}}_t^r \tag{4.16}$$

This adaptive fusion mechanism replaces the naive concatenation approach and enables the model to: (1) automatically suppress noisy and irrelevant modalities at specific time steps, (2) amplify informative signals when environmental factors exhibit strong causal relationships with appliance operation, and (3) provide interpretable insights into which data sources contribute most to disaggregation at different temporal contexts. For the complete temporal sequence, we obtain:

$$\tilde{\boldsymbol{Y}} = [\boldsymbol{Y}_1^{\text{multi}}, \boldsymbol{Y}_2^{\text{multi}}, \ldots, \boldsymbol{Y}_T^{\text{multi}}] \tag{4.17}$$

4.2.2 Expert Network Diversity

Unlike typical mixture-of-experts implementations that employ identical architectures across all experts, our approach leverages architectural heterogeneity to capture diverse data patterns. The expert ensemble consists of: (i) Dense neural network (DNN, Expert 1) to capture complex non-linear feature interactions; (ii) Convolutional network (CNN, Expert 2) to extract localized temporal patterns; (iii) Transformer architecture (Trm, Expert 3) to model long-range dependencies through self-attention; (iv) Temporal convolutional network (TCN, Expert 4) to process sequential information with dilated convolutions; (v) Recurrent network (Expert 5) to maintain sequential memory through LSTM cells. The generalization expression of each expert's computational mapping is formally defined as:

$$h_j(\tilde{\boldsymbol{Y}}) = \mathcal{G}_j(\tilde{\boldsymbol{Y}}, \boldsymbol{\Phi}_j), \quad \forall j \in \mathcal{J} \tag{4.18}$$

where $\mathcal{J} = \{\text{DNN}, \text{CNN}, \text{Trm}, \text{TCN}, \text{LSTM}\}$ represents the expert set. It is noteworthy that this modular design enables seamless integration of emerging architectures, ensuring framework adaptability to future developments. Each expert contributes unique inductive biases: convolutional layers excel at local pattern recognition, recurrent structures capture temporal dependencies, dense networks model complex interactions, and attention mechanisms identify relevant temporal relationships.

4.2.3 Task-Specific Gating Strategy

Departing from traditional shared-bottom and single-gate approaches, our framework implements independent gating networks for classification and regression tasks. This design allows each task to develop specialized expert selection strategies based on task-specific requirements. The gating mechanism ensures proper probability distribution over experts for each task $k \in \mathcal{K} = \{\text{cls}, \text{reg}\}$:

$$\sum\nolimits_{j \in \mathcal{J}} w_j^k(\tilde{\boldsymbol{Y}}) = 1, \quad 0 \leq w_j^k(\tilde{\boldsymbol{Y}}) \leq 1, \quad \forall k \in \mathcal{K} \tag{4.19}$$

The task-specific expert combination is computed as:

$$h^k(\tilde{\boldsymbol{Y}}) = \sum\nolimits_{j \in \mathcal{J}} w_j^k(\tilde{\boldsymbol{Y}}) \cdot h_j(\tilde{\boldsymbol{Y}}), \quad \forall k \in \mathcal{K} \tag{4.20}$$

This dual-gate architecture reduces task interference and gradient conflicts while maintaining computational efficiency through shared expert utilization. The gates adapt dynamically to input characteristics, enabling flexible expert allocation that optimizes performance for each specific task.

4.2.4 Classification Tower (Binary State Detection)

The classification tower is responsible for on/off state prediction, which is a binary classification task that determines whether a particular appliance is currently in use (ON) or idle (OFF). The classification module processes aggregated expert representations through multi-head attention mechanisms followed by fully connected layers to generate the final binary predictions. For the m-th attention head:

$$\mathbf{Q}_m^{\text{cls}} = h^{\text{cls}}(\tilde{\boldsymbol{Y}}) \cdot \mathbf{W}_m^{Q,\text{cls}} \tag{4.21}$$

$$\mathbf{K}_m^{\text{cls}} = h^{\text{cls}}(\tilde{\boldsymbol{Y}}) \cdot \mathbf{W}_m^{K,\text{cls}} \tag{4.22}$$

$$\mathbf{V}_m^{\text{cls}} = h^{\text{cls}}(\tilde{\boldsymbol{Y}}) \cdot \mathbf{W}_m^{V,\text{cls}} \tag{4.23}$$

The attention computation follows:

$$\mathbf{A}_m^{\text{cls}} = \text{softmax}(\frac{\mathbf{Q}_m^{\text{cls}} \cdot \mathbf{K}_m^{\text{cls}\top}}{\sqrt{d}}) \cdot \mathbf{V}_m^{\text{cls}} \tag{4.24}$$

Multi-head integration combines all attention outputs:

$$\mathbf{h}_{\text{attn}}^{\text{cls}} = \text{Concat}\left[\mathbf{A}_1^{\text{cls}}, \mathbf{A}_2^{\text{cls}}, \ldots, \mathbf{A}_{M^{\text{cls}}}^{\text{cls}}\right] \cdot \mathbf{W}^{\text{cls}} \tag{4.25}$$

where $\mathbf{W}^{\text{cls}}$ denotes the number of attention heads, and $\mathbf{W}^{\text{cls}}$ represents the projection matrix. Unlike traditional recurrent architectures such as LSTMs or GRUs, transformers are highly parallelizable and can capture complex temporal relationships efficiently.

Following the multi-head attention mechanism, the processed representations are fed through fully connected layers to generate the final classification outputs:

$$\mathbf{h}_{\text{fc}}^{\text{cls}} = \text{ReLU}(\mathbf{W}_{\text{fc}}^{\text{cls}} \cdot \mathbf{h}_{\text{attn}}^{\text{cls}} + \mathbf{b}_{\text{fc}}^{\text{cls}}) \tag{4.26}$$

$$\hat{\boldsymbol{s}} = \text{softmax}(\mathbf{W}_{\text{out}}^{\text{cls}} \cdot \mathbf{h}_{\text{fc}}^{\text{cls}} + \mathbf{b}_{\text{out}}^{\text{cls}}) \tag{4.27}$$

where $\mathbf{W}_{\text{fc}}^{\text{cls}}$, $\mathbf{b}_{\text{fc}}^{\text{cls}}$ are the weight matrix and bias vector of the fully connected layer, and $\mathbf{W}_{\text{out}}^{\text{cls}}$, $\mathbf{b}_{\text{out}}^{\text{cls}}$ are the parameters of the final output layer that transforms the FC layer outputs into probability distributions over appliance states.

4.2.5 Regression Tower (Power Consumption Estimation)

The regression tower focuses on power consumption estimation, a regression task that predicts the actual power usage of an appliance at a given time. Similar to the classification tower, it employs multi-head attention followed by fully connected layers for continuous power prediction.

$$\mathbf{Q}_m^{\text{reg}} = h^{\text{reg}}(\tilde{\boldsymbol{Y}}) \cdot \mathbf{W}_m^{Q,\text{reg}} \tag{4.28}$$

$$\mathbf{K}_m^{\text{reg}} = h^{\text{reg}}(\tilde{\boldsymbol{Y}}) \cdot \mathbf{W}_m^{K,\text{reg}} \tag{4.29}$$

$$\mathbf{V}_m^{\text{reg}} = h^{\text{reg}}(\tilde{\boldsymbol{Y}}) \cdot \mathbf{W}_m^{V,\text{reg}} \tag{4.30}$$

The attention computation follows:

$$\mathbf{A}_m^{\text{reg}} = \text{softmax}(\frac{\mathbf{Q}_m^{\text{reg}} \cdot \mathbf{K}_m^{\text{reg}\top}}{\sqrt{d}}) \cdot \mathbf{V}_m^{\text{reg}} \tag{4.31}$$

Multi-head integration combines all attention outputs:

$$\boldsymbol{x}^{\text{reg}} = \text{ Concat } [\mathbf{A}_1^{\text{reg}}, \mathbf{A}_2^{\text{reg}}, \ldots, \mathbf{A}_{M^{\text{reg}}}^{\text{reg}}] \cdot \mathbf{W}^{\text{reg}} \tag{4.32}$$

where $\mathbf{W}^{\text{reg}}$ denotes the number of attention heads and $\mathbf{W}^{\text{reg}} \in \mathbb{R}^{(M \cdot d) \times d}$ represents the projection matrix.

The attention outputs are then processed through fully connected layers to generate the final power consumption estimates:

$$\mathbf{h}_{\text{fc}}^{\text{reg}} = \text{ReLU}(\mathbf{W}_{\text{fc}}^{\text{reg}} \cdot \mathbf{h}_{\text{attn}}^{\text{reg}} + \mathbf{b}_{\text{fc}}^{\text{reg}}) \tag{4.33}$$

$$\hat{\boldsymbol{x}} = \mathbf{W}_{\text{out}}^{\text{reg}} \cdot \mathbf{h}_{\text{fc}}^{\text{reg}} + \mathbf{b}_{\text{out}}^{\text{reg}} \tag{4.34}$$

where $\mathbf{W}_{\text{fc}}^{\text{reg}}$, $\mathbf{b}_{\text{fc}}^{\text{reg}}$ are the weight matrix and bias vector of the fully connected layer, and $\mathbf{W}_{\text{out}}^{\text{reg}}$, $\mathbf{b}_{\text{out}}^{\text{reg}}$ are the parameters of the final output layer that generates the continuous power consumption predictions. The towers receive task-specific expert representations and generate sequence-level outputs for their respective objectives. Operating on pre-processed expert features and utilizing the FC layers significantly reduces computational overhead compared to full-scale transformer architectures while preserving the benefits of attention-based temporal modeling and enabling precise task-specific predictions.

4.3 Loss Function Design and Optimization

4.3.1 Loss Function Design

We formulate a comprehensive loss function that jointly optimizes binary state detection and continuous power estimation through a weighted combination of task-specific objectives:

$$\begin{aligned}&\min_{\boldsymbol{\Theta}_{\text{expert}},\boldsymbol{\Theta}_{\text{task}}} \mathcal{J}(\boldsymbol{\Theta}_{\text{expert}},\boldsymbol{\Theta}_{\text{task}}) = \\ &\alpha^{\text{cls}}\mathcal{J}_{\text{state}}(\boldsymbol{\Theta}_{\text{expert}},\boldsymbol{\Theta}_{\text{state}}) \quad +\alpha^{\text{reg}}\mathcal{J}_{\text{power}}(\boldsymbol{\Theta}_{\text{expert}},\boldsymbol{\Theta}_{\text{power}})\end{aligned} \tag{4.35}$$

where $\boldsymbol{\Theta}_{\text{expert}}$ represents parameters of the heterogeneous expert ensemble DNN, CNN, Transformer, TCN, LSTM, $\boldsymbol{\Theta}_{\text{task}}$ encompasses task-specific parameters, and the weighting coefficients satisfy $\alpha^{\text{cls}} + \alpha^{\text{reg}} = 1$ with $\alpha^{\text{cls}} \geq 0, \alpha^{\text{reg}} \geq 0$. Specifically, for appliance operational state prediction, we employ a robust logistic-based formulation:

$$\mathcal{J}_{\text{state}}(\boldsymbol{\Theta}_{\text{expert}},\boldsymbol{\Theta}_{\text{state}}) = \frac{1}{T\cdot I}\sum_{t=1}^{T}\sum_{i=1}^{I}\log(1+\exp(-s_i\cdot\hat{s}_i)) \tag{4.36}$$

where s_i and $\hat{s}_i$ represent true and predicted binary states respectively. The logistic-based formulation is essentially a soft-margin hinge loss, providing smooth gradients. The power estimation combines three-component regression loss terms, i.e., mean squared error term, L1 robust term, and Kullback-Leibler (KL) divergence term:

$$\begin{aligned}\mathcal{J}_{\text{power}}(\boldsymbol{\Theta}_{\text{expert}},\boldsymbol{\Theta}_{\text{power}}) &= \frac{1}{T}\textstyle\sum_{t=1}^{\mathcal{T}}(\hat{x}_i - x_i)^2 \\ + \frac{\gamma}{T}\textstyle\sum_{i=1}^{\mathcal{I}}|\hat{x}_i - x_i| &+ \mathcal{D}_{\text{KL}}(\sigma(\frac{\hat{\boldsymbol{x}}}{\delta}) \parallel \sigma(\frac{\boldsymbol{x}}{\delta}))\end{aligned} \tag{4.37}$$

where δ is the temperature scaling parameter for the KL divergence, converting power values to pseudo-probabilities. The mean squared error term penalizes large prediction errors heavily due to the quadratic penalty, while the L1 robust term creates constant gradient magnitude regardless of error size, making the model less sensitive to extreme values. The KL divergence term forces the predicted power distribution to match the true power distribution across all appliances.

4.3.2 Analytical Solution for Optimally Balancing Weight

Note that, the weight $(\alpha^{\text{cls}}, \alpha^{\text{reg}})$ directly determines the importance of the state detection task and the power consumption, and heavily affects the final solutions. To address the challenge of determining appropriate weights between classification and regression tasks, we formulate the dual-task optimization as a Pareto-optimal solution finding problem to dynamically balance competing objectives through gradient-based analysis. Consider the bi-objective optimization problem:

$$\min_{\boldsymbol{\Theta}_{\text{expert}}, \boldsymbol{\Theta}_{\text{task}}} \left\{ \mathcal{J}_{\text{state}} \left(\boldsymbol{\Theta}_{\text{expert}}, \boldsymbol{\Theta}_{\text{state}} \right), \mathcal{J}_{\text{power}} \left(\boldsymbol{\Theta}_{\text{expert}}, \boldsymbol{\Theta}_{\text{power}} \right) \right\} \tag{4.38}$$

Definition 3. *A solution $\boldsymbol{\Theta}^*$ achieves Pareto efficiency when no other feasible solution can improve one objective without degrading the other.*

This leads to the Pareto stationarity condition requiring:

- Existence of non-negative weights $\alpha^{\text{cls}}, \alpha^{\text{reg}} \geq 0$ with $\alpha^{\text{cls}} + \alpha^{\text{reg}} = 1$ such that:

$$\alpha^{\text{cls}} \nabla_{\boldsymbol{\Theta}_{\text{expert}}} \mathcal{J}_{\text{state}} + \alpha^{\text{reg}} \nabla_{\boldsymbol{\Theta}_{\text{expert}}} \mathcal{J}_{\text{power}} = \mathbf{0} \tag{4.39}$$

- Individual task optimality conditions:

$$\nabla_{\boldsymbol{\Theta}_{\text{state}}} \mathcal{J}_{\text{state}} = \mathbf{0}, \quad \nabla_{\boldsymbol{\Theta}_{\text{power}}} \mathcal{J}_{\text{power}} = \mathbf{0} \tag{4.40}$$

Then, the optimal weight determination reduces to solving:

$$\begin{aligned} &\min_{\alpha^{\text{cls}}, \alpha^{\text{reg}}} \left\| \alpha^{\text{cls}} \nabla_{\boldsymbol{\Theta}_{\text{expert}}} \mathcal{J}_{\text{state}} + \alpha^{\text{reg}} \nabla_{\boldsymbol{\Theta}_{\text{expert}}} \mathcal{J}_{\text{power}} \right\|_2^2 \\ &s.t. \quad \alpha^{\text{cls}} + \alpha^{\text{reg}} = 1, \alpha^{\text{cls}}, \alpha^{\text{reg}} \geq 0 \end{aligned} \tag{4.41}$$

Denote $\mathbf{g}_{\text{state}} = \nabla_{\boldsymbol{\Theta}_{\text{expert}}} \mathcal{J}_{\text{state}}$ and $\mathbf{g}_{\text{power}} = \nabla_{\boldsymbol{\Theta}_{\text{expert}}} \mathcal{J}_{\text{power}}$, the formulation (4.41) can be simplified to:

$$\min_{\beta \in [0,1]} \left\| \beta \mathbf{g}_{\text{state}} + (1-\beta) \mathbf{g}_{\text{power}} \right\|_2^2 \tag{4.42}$$

Proposition 5. *Let* $\mathbf{g}_{state}$ *and* $\mathbf{g}_{power}$ *be the gradients of the state classification and power regression losses with respect to the shared expert parameters. The optimal weight* β^* *for the state classification task is given by:*

$$\beta^* = \begin{cases} 0, & \textit{if } \mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}} \geq |\mathbf{g}_{\text{power}}|^2 \\ \dfrac{|\mathbf{g}_{\text{power}}|^2 - \mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}}}{|\mathbf{g}_{\text{state}} - \mathbf{g}_{\text{power}}|^2}, & \textit{if } \mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}} < \mathbf{g}^{\text{min}} \\ 1, & \textit{if } \mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}} \geq |\mathbf{g}_{\text{state}}|^2 \end{cases}$$

where $\mathbf{g}^{\text{min}} = \min(|\mathbf{g}_{\text{power}}|^2,\ |\mathbf{g}_{\text{state}}|^2)$.

Proof: We first define the objective function:

$$\Psi(\beta) = \|\beta \mathbf{g}_{\text{state}} + (1-\beta)\mathbf{g}_{\text{power}}\|_2^2 \tag{4.43}$$

Taking the derivative with respect to β:

$$\frac{d\Psi(\beta)}{d\beta} = 2(\beta \mathbf{g}_{\text{state}} + (1-\beta)\mathbf{g}_{\text{power}}) \cdot (\mathbf{g}_{\text{state}} - \mathbf{g}_{\text{power}}) \tag{4.44}$$

Setting the derivative to zero for the critical point:

$$\beta_{\text{critical}} = \frac{\|\mathbf{g}_{\text{power}}\|^2 - \mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}}}{\|\mathbf{g}_{\text{state}}\|^2 - 2\mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}} + \|\mathbf{g}_{\text{power}}\|^2} \tag{4.45}$$

Since the second derivative is $\frac{d^2\Psi(\beta)}{d\beta^2} = 2\left\|\mathbf{g}_{\text{state}} - \mathbf{g}_{\text{power}}\right\|^2 > 0$ (assuming $\mathbf{g}_{\text{state}} \neq \mathbf{g}_{\text{power}}$), $\Psi(\beta)$ is strictly convex. Therefore, the global minimum over $[0,1]$ occurs either at β_{critical} (if $\beta_{\text{critical}} \in [0,1]$ or at a boundary point.

Case 1: If $\mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}} \geq |\mathbf{g}_{\text{power}}|^2$, then $\beta_{\text{critical}} \leq 0$. Since $\Psi(\beta)$ is convex and decreasing on $[0,1]$, the minimum occurs at $\beta^* = 0$.

Case 2: If $\mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}} \geq |\mathbf{g}_{\text{state}}|^2$, then:

$$\beta_{\text{critical}} = \frac{\left\|\mathbf{g}_{\text{power}}\right\|^2 - \mathbf{g}_{\text{state}} \cdot \mathbf{g}_{\text{power}}}{\left\|\mathbf{g}_{\text{state}} - \mathbf{g}_{\text{power}}\right\|^2} \geq \frac{\left\|\mathbf{g}_{\text{state}}\right\|^2 - \mathbf{g}_{\text{state}} \cdot \mathbf{g}_{\text{power}}}{\left\|\mathbf{g}_{\text{state}} - \mathbf{g}_{\text{power}}\right\|^2} = 1 \tag{4.46}$$

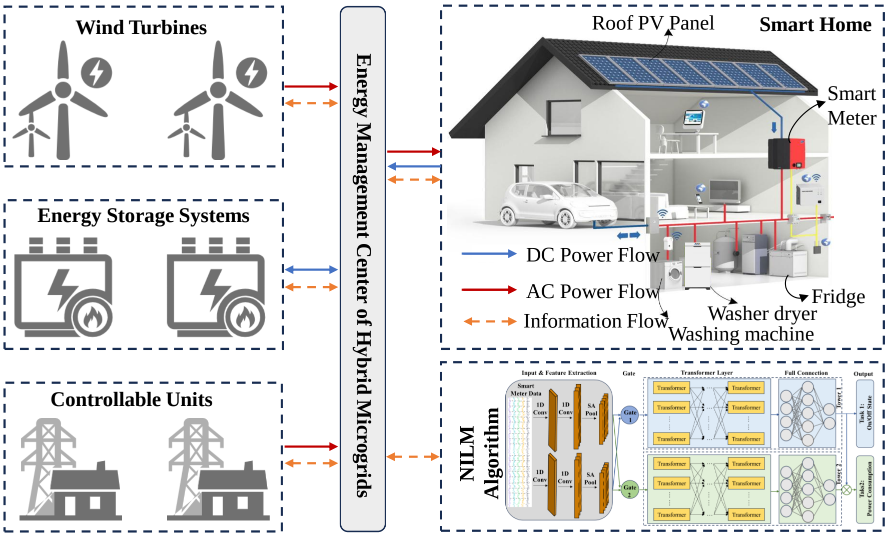


Figure 4.2: Hybrid building microgrid system with smart meter analytics

Since $\Psi(\beta)$ is convex and increasing on $[0, 1]$, the minimum occurs at $\beta^* = 1$.

Case 3: If $\mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}} < \mathbf{g}^{\text{min}}$, then $\beta_{\text{critical}} \leq (0, 1)$, and the global minimum occurs at $\beta^* = \beta_{\text{critical}}$. ■

Corollary 5. *The optimal weight allocation has the following interpretations: 1) When $\beta^* = 0$: State gradients are well-aligned with power gradients; prioritize power estimation. 2) When $\beta^* = 1$: Power gradients are well-aligned with state gradients; prioritize state classification. 3) When $\beta^* = \beta_{critical}$: Gradients conflict significantly; balanced weighting is required.*

The comprehensive dual-task training procedure with multimodal information is presented in **Algorithm 3**, which dynamically balances state classification and power estimation through the application of **Proposition 5** for optimal weight determination. This adaptive balancing ensures that both classification and regression objectives contribute effectively to the overall learning process without manual tuning, ultimately improving dual-task performance in NILM applications.

Algorithm 3 Dual-Task Training with Adaptive Balance

Input: Training dataset $\mathcal{D} = \{(\mathbf{Y}^{(n)}_{\text{multi}}, \mathbf{x}^{(n)}_i, \sigma^{(n)}_i)\}_{n=1}^{N}$, learning rate η, epoch number E_{max}, and batch size B_{size}
Initialize: Task balance weights $\boldsymbol{\alpha} = (\alpha^{\text{cls}}, \alpha^{\text{reg}})$
for $e = 1$ to E_{max} **do**
  for $b = 1$ to B_{size} **do**
    ▷ **Expert Feature Generation**
    $h_{\text{DNN}}(\mathbf{z}) \leftarrow \mathcal{G}_{\text{DNN}}(\mathbf{z}, \mathbf{\Phi}_{\text{DNN}})$
    $h_{\text{CNN}}(\mathbf{z}) \leftarrow \mathcal{G}_{\text{CNN}}(\mathbf{z}, \mathbf{\Phi}_{\text{CNN}})$
    $h_{\text{Trm}}(\mathbf{z}) \leftarrow \mathcal{G}_{\text{Trm}}(\mathbf{z}, \mathbf{\Phi}_{\text{Trm}})$
    $h_{\text{TCN}}(\mathbf{z}) \leftarrow \mathcal{G}_{\text{TCN}}(\mathbf{z}, \mathbf{\Phi}_{\text{TCN}})$
    $h_{\text{LSTM}}(\mathbf{z}) \leftarrow \mathcal{G}_{\text{LSTM}}(\mathbf{z}, \mathbf{\Phi}_{\text{LSTM}})$
    ▷ **Task-Specific Routing & Multi-Head Attention**
    Compute classification gates $w_j^{\text{cls}}(\mathbf{z})$
    $h^{\text{cls}}(\mathbf{z}) \leftarrow \sum_j w_j^{\text{cls}}(\mathbf{z}) \cdot h_j(\mathbf{z})$
    Compute regression gates $w_j^{\text{reg}}(\mathbf{z})$
    $h^{\text{reg}}(\mathbf{z}) \leftarrow \sum_j w_j^{\text{reg}}(\mathbf{z}) \cdot h_j(\mathbf{z})$
    $\mathbf{A}_m^{\text{cls}} \leftarrow \text{softmax}\left(\frac{\mathbf{Q}_m^{\text{cls}} \mathbf{K}_m^{\text{cls}\top}}{\sqrt{d}}\right) \mathbf{V}_m^{\text{cls}}$
    $\mathbf{h}_{\text{attn}}^{\text{cls}} \leftarrow \text{Concat}[\mathbf{A}_1^{\text{cls}}, \ldots, \mathbf{A}_{M_{\text{cls}}}^{\text{cls}}]\mathbf{W}^{\text{cls}}$
    $\mathbf{A}_m^{\text{reg}} \leftarrow \text{softmax}\left(\frac{\mathbf{Q}_m^{\text{reg}} \mathbf{K}_m^{\text{reg}\top}}{\sqrt{d}}\right) \mathbf{V}_m^{\text{reg}}$
    $\mathbf{h}_{\text{attn}}^{\text{reg}} \leftarrow \text{Concat}[\mathbf{A}_1^{\text{reg}}, \ldots, \mathbf{A}_{M_{\text{reg}}}^{\text{reg}}]\mathbf{W}^{\text{reg}}$
    ▷ **Weight Optimization & Parameter Updates**
    Calculate $\mathbf{g}_{\text{power}} \cdot \mathbf{g}_{\text{state}}$, $\|\mathbf{g}_{\text{power}}\|^2$, $\|\mathbf{g}_{\text{state}}\|^2$
    Apply **Proposition 5** to determine β^*
    Optimal weights $\boldsymbol{\alpha}^*$
    $\mathcal{J}_{\text{power}}(\mathbf{\Theta}_{\text{expert}}, \mathbf{\Theta}_{\text{power}}) \leftarrow \alpha^{\text{cls}} \mathcal{J}_{\text{state}} + \alpha^{\text{reg}} \mathcal{J}_{\text{power}}$
    $\nabla_{\mathbf{\Phi}} \mathcal{J} \leftarrow \frac{\partial \mathcal{J}}{\partial \mathbf{\Phi}}$
    $\mathbf{\Phi} \leftarrow \mathbf{\Phi} - \eta \cdot \nabla_{\mathbf{\Phi}} \mathcal{J}$
  end for
end for

4.4 NILM-Aided Robust Energy Management for Hybrid Building Microgrids

The framework in Fig. 4.2 presents the energy management system for a hybrid AC/DC building microgrid architecture, integrating NILM to identify the potential of demand-side response. It consists of RES units (i.e., wind turbines and PV panels), energy storage systems (ESS), and controllable units, such as diesel generators (DGs) and fuel cell units (FCs), all managed by a central energy management center that optimizes AC and DC power flow. A smart home equipped with a smart meter collects aggregated real-time energy consumption data, while our NILM algorithm disaggregates appliance-level consumption from aggregated smart meter data, identifying and monitoring the detailed state of appliances on the demand side without requiring intrusive sub-sensors. This demand-side information can be leveraged to properly implement demand response strategies, thereby mitigating the intermittency of RES units and minimizing operational costs.

4.4.1 NILM-aided Modelling of Household Electric Appliances

Unlike conventional NILM techniques that focus solely on power regression, the proposed dual-task approach yields two distinct outputs for each monitored appliance i within a household at every time step t:

1) Estimated Operational State ($\hat{s}_{i,t}$): A binary variable derived from the model's classification tower, where $\hat{s}_{i,t} = 1$ signifies that appliance i is active and $\hat{s}_{i,t} = 0$ signifies it is inactive.

2) Estimated Power Consumption ($\hat{x}_{i,t}$): A continuous output $\hat{x}_{i,t} \in \mathbb{R}^+$ from the model's regression tower, quantifying the electrical power draw of appliance i when operational.

The synergistic output of both state and power provides a complete and less ambiguous characterization of the appliance behavior required for scheduling and dispatch decisions. To construct a modular and accurate mathematical model, household electric appliances are classified based on their operational characteristics and inherent flexibility for demand response. The appliances can be categorized into two primary types:

4.4.1.1 Fixed Loads

The indeferrable base loads, such as freezers and essential lighting, present inflexible consumption patterns whose operation is critical and cannot be shifted or curtailed without significant disruption. The power consumption of these loads, $\hat{x}_{i,t}^{\text{base}}$ is treated as an uncontrollable, deterministic parameter.

4.4.1.2 Deferrable Loads

These appliances represent the primary source of demand-side flexibility and are subdivided into deferrable non-interruptible loads and deferrable interruptible loads. For deferrable non-interruptible loads $\mathcal{I}_{\text{NI}}$, appliances such as washing machines, clothes dryers, and dishwashers, operate in discrete, uninterruptible cycles D_i^{NI} with flexibility in start time scheduling within user-specified windows $\left[T_i^{\text{start}}, T_i^{\text{end}}\right]$. Their operation constraints are modeled in (4.47)-(4.50).

$$u_{i,t}^{\text{NI}} = 0, \quad \forall t \notin \left[T_i^{\text{start}}, T_i^{\text{end}} - D_i^{\text{NI}} + 1\right], \quad \forall i \in \mathcal{I}^{\text{NI}} \tag{4.47}$$

$$\sum\nolimits_{t \in \mathcal{T}} u_{i,t}^{\text{NI}} = 1, \quad \forall i \in \mathcal{I}^{\text{NI}} \tag{4.48}$$

$$\sigma_{i,t} = \sum\nolimits_{\tau = t - D_i^{\text{NI}} + 1}^{t} u_{i,\tau}^{\text{NI}}, \quad \forall i \in \mathcal{I}^{\text{NI}}, \forall t \in \mathcal{T} \tag{4.49}$$

$$x_{i,t}^{\text{NI}} = P_i^{\text{NI,rated}} \cdot \sigma_{i,t}, \quad \forall i \in \mathcal{I}^{\text{NI}}, \forall t \in \mathcal{T} \tag{4.50}$$

Here, $u_{i,t}^{\text{NI}}$ is the binary start variable to indicate whether the appliance starts at t, while $\sigma_{i,t}$ is the binary operational state to indicate whether the appliance operates at t. $x_{i,t}^{\text{NI}}$ and $P_i^{\text{NI,rated}}$ are the power consumption and the rated value of appliance i, respectively. Constrain (4.47) prevents appliances from starting outside their user-defined operational window, ensuring the appliance can complete its full cycle within the allowable time frame. Constraint (4.48) ensures each appliance is scheduled to run exactly once during the optimization horizon, guaranteeing task completion. Constraint (4.49) links the operational state to the start decision, ensuring the appliance remains ON for its complete cycle duration once started. Constraint (4.50) sets the power consumption to the rated value when the appliance is operational and zero when inactive.

As the primary deferrable interruptible loads, EVs offer greater flexibility through deferrability and interruptibility. Their operation constraints are modeled in (4.51)-(4.55).

$$S_{i,t+1}^{\text{EV}} = x_{i,t}^{\text{EV+}} \eta_i^{\text{EV+}} - x_{i,t}^{\text{EV-}} / \eta_i^{\text{EV-}} + S_{i,t}^{\text{EV}}, \forall i \in \mathcal{I}^{\text{I}}, \forall t \in \mathcal{T} \tag{4.51}$$

$$\underline{S}_i^{\text{EV}} \leq S_{i,t}^{\text{EV}} \leq \bar{S}_i^{\text{EV}}, \quad \forall i \in \mathcal{I}^{\text{I}}, \forall t \in \mathcal{T} \tag{4.52}$$

$$S_{i,t}^{\text{EV}} = S_i^{\text{EV,in}}, \forall i \in \mathcal{I}^{\text{I}}, \forall t = t_i^{\text{in}} \tag{4.53}$$

$$S_{i,t}^{\text{EV}} \geq S_i^{\text{EV,req}}, \forall i \in \mathcal{I}^{\text{I}}, \forall t = t_i^{\text{out}} \tag{4.54}$$

$$0 \le x_{i,t}^{\mathrm{EV}+}, \eta_i^{\mathrm{EV}-} \le \hat{s}_{i,t}^{EV} \bar{P}_i^{\mathrm{EV}}, \forall t_i^{\mathrm{in}} \le t \le t_i^{\mathrm{out}} \tag{4.55}$$

Here, $S_{i,t}^{\mathrm{EV}}$ is the energy level of EV i at time t, while $\underline{S}_i^{\mathrm{EV}}$ and $\bar{S}_i^{\mathrm{EV}}$ are its lower and upper limits. $x_{i,t}^{\mathrm{EV}+}$ is the charging power of EV i, and $\eta_i^{\mathrm{EV}+}/\eta_i^{\mathrm{EV}-}$ is its charging efficiency. $\bar{P}_i^{\mathrm{EV}+}$ is the maximum charging power of EV i. Equation (4.51) models the battery energy balance over time, while Constraint (4.52) maintains the battery energy level within safe operating limits to prevent damage and ensure longevity. Constraint (4.53) sets the initial battery energy level $S_i^{\mathrm{EV,in}}$ at the time t_i^{in} when the EV arrives at home, while Constraint (4.54) ensures the battery has the required energy $S_i^{\mathrm{EV,req}}$ to meet the user's driving needs upon departure time t_i^{out}. Constraint (4.55) bounds the charging power within physical limits.

4.4.2 Energy Management of Hybrid Building Microgrids with NILM

With the identified deferrable loads from the dual-task NILM approach, we formulate a NILM-aided robust model for the energy management of hybrid microgrids by evaluating the demand response potential of appliances using both estimated operational states $\hat{s}_{i,t}$ and estimated power consumption $\hat{x}_{i,t}$. Here, we first present the NILM-aided deterministic model as follows:

$$\min C^{\mathrm{de}} + C^{\mathrm{fc}} + C^{\mathrm{r}} + C^{\mathrm{bdc}} \tag{4.56}$$

$$C^{\mathrm{de}} = \sum_{t=1}^{\mathcal{T}} \sum_{a=1}^{\mathcal{A}} (n_a^{\mathrm{de}} U_{a,t}^{\mathrm{de}} + f_a^{\mathrm{de}} D_{a,t}^{\mathrm{de}} + c_a^{\mathrm{de}} P_{a,t}^{\mathrm{de}}) \tag{4.57}$$

$$C^{\mathrm{fc}} = \sum_{t=1}^{\mathcal{T}} \sum_{a=1}^{\mathcal{A}} (n_a^{\mathrm{fc}} U_{a,t}^{\mathrm{fc}} + f_a^{\mathrm{fc}} D_{a,t}^{\mathrm{fc}} + c_a^{\mathrm{fc}} P_{a,t}^{\mathrm{fc}}) \tag{4.58}$$

$$C^{\mathrm{r}} = \sum_{t=1}^{\mathcal{T}} \sum_{a=1}^{\mathcal{A}} (c_a^{\mathrm{w}} \hat{\rho}_{a,t}^{\mathrm{w}} + c_a^{\mathrm{pv}} \hat{\rho}_{a,t}^{\mathrm{pv}}) \tag{4.59}$$

$$C^{\mathrm{bdc}} = \sum_{t=1}^{\mathcal{T}} \sum_{l=1}^{\mathcal{L}} c_l^{\mathrm{loss}} (P_{a_1 b_1,l,t}^{\mathrm{acdc}} + P_{a_1 b_1,l,t}^{\mathrm{dcac}}) \tag{4.60}$$

s.t.

$$I_{a,t}^{\mathrm{de}} = I_{a,t-1}^{\mathrm{de}} + U_{a,t}^{\mathrm{de}} - D_{a,t}^{\mathrm{de}}, \forall a, t \tag{4.61}$$

$$I_{a,t}^{\mathrm{de}} \ge \sum_{\tau=1}^{T_a^{\mathrm{de,on}}} U_{a,t-\tau}^{\mathrm{de}}, I_{a,t}^{\mathrm{de}} \le 1 - \sum_{\tau=1}^{T_a^{\mathrm{de,off}}} D_{a,t-\tau}^{\mathrm{de}}, \forall a, t \tag{4.62}$$

$$I_{a,t}^{\mathrm{de}}\underline{P}_a^{\mathrm{de}} \le P_{a,t}^{\mathrm{de}} \le I_{a,t}^{\mathrm{de}}\bar{P}_a^{\mathrm{de}}, \forall a,t \tag{4.63}$$

$$I_{a,t}^{\mathrm{de}}\underline{Q}_a^{\mathrm{de}} \le Q_{a,t}^{\mathrm{de}} \le I_{a,t}^{\mathrm{de}}\bar{Q}_a^{\mathrm{de}}, \forall a,t \tag{4.64}$$

$$-R_a^{\mathrm{P,de}} \le P_{a,t}^{\mathrm{de}} - P_{a,t-1}^{\mathrm{de}} \le R_a^{\mathrm{P,de}}, \forall a,t \tag{4.65}$$

$$-R_a^{\mathrm{Q,de}} \le Q_{a,t}^{\mathrm{de}} - Q_{a,t-1}^{\mathrm{de}} \le R_a^{\mathrm{Q,de}}, \forall a,t \tag{4.66}$$

$$0 \le \hat{\rho}_{a,t}^{\mathrm{w}} \le \hat{P}_{a,t}^{\mathrm{w}}, \forall a,t \tag{4.67}$$

$$V_{b,t}^{\mathrm{ac}} = V_{a,t}^{\mathrm{ac}} - \left(r_{ab}^{\mathrm{ac}} H_{ab,t}^{\mathrm{ac}} + x_{ab}^{\mathrm{ac}} G_{ab,t}^{\mathrm{ac}}\right) / V_{0,t}^{\mathrm{ac}}, \forall ab,t \tag{4.68}$$

$$\underline{H}_{ab,t}^{\mathrm{ac}} \le H_{ab,t}^{\mathrm{ac}} \le \bar{H}_{ab,t}^{\mathrm{ac}}, \forall ab,t \tag{4.69}$$

$$\underline{G}_{ab,t}^{\mathrm{ac}} \le G_{ab,t}^{\mathrm{ac}} \le \bar{G}_{ab,t}^{\mathrm{ac}}, \forall ab,t \tag{4.70}$$

$$\underline{V}_{a,t}^{\mathrm{ac}} \le V_{a,t}^{\mathrm{ac}} \le \bar{V}_{a,t}^{\mathrm{ac}}, \forall a,t \tag{4.71}$$

$$\begin{aligned} & P_{a,t}^{\mathrm{de}} + \left(\hat{P}_{a,t}^{\mathrm{w}} - \hat{\rho}_{a,t}^{\mathrm{w}}\right) + \sum\nolimits_{b_1 \in \pi(a_1), l \in \mathcal{L}} (1 - \eta_{\mathrm{BDC},l}^{\mathrm{acdc}} P_{a_1 b_1,l,t}^{\mathrm{acdc}}) \\ & = \hat{x}_{a,t}^{\mathrm{ac,base}} + x_{a,t}^{\mathrm{NI}} + \sum\nolimits_{b \in \pi(a)} H_{ab,t}^{\mathrm{ac}} - \sum\nolimits_{k \in \delta(a)} H_{ka,t}^{\mathrm{ac}}, \forall a,t \end{aligned} \tag{4.72}$$

$$Q_{a,t}^{\mathrm{de}} - Q_{a,t}^{\mathrm{ac}} = \sum\nolimits_{b \in \pi(a)} G_{ab,t}^{\mathrm{ac}} - \sum\nolimits_{k \in \delta(a)} G_{ka,t}^{\mathrm{ac}}, \forall a,t \tag{4.73}$$

$$I_{a,t}^{\mathrm{fc}} = I_{a,t-1}^{\mathrm{fc}} + U_{a,t}^{\mathrm{fc}} - D_{a,t}^{\mathrm{fc}}, \forall a,t \tag{4.74}$$

$$I_{a,t}^{\mathrm{fc}} \ge \sum\nolimits_{\tau=1}^{T_a^{\mathrm{fc,on}}} U_{a,t-\tau}^{\mathrm{fc}}, I_{a,t}^{\mathrm{fc}} \le 1 - \sum\nolimits_{\tau=1}^{T_a^{\mathrm{fc,off}}} D_{a,t-\tau}^{\mathrm{fc}}, \forall a,t \tag{4.75}$$

$$I_{a,t}^{\mathrm{fc}}\underline{P}_a^{\mathrm{fc}} \le P_{a,t}^{\mathrm{fc}} \le I_{a,t}^{\mathrm{fc}}\bar{P}_a^{\mathrm{fc}}, \forall a,t \tag{4.76}$$

$$-R_a^{\mathrm{fc}} \le P_{a,t}^{\mathrm{fc}} - P_{a,t-1}^{\mathrm{fc}} \le R_a^{\mathrm{fc}}, \forall a,t \tag{4.77}$$

$$0 \le \hat{\rho}_{a,t}^{\mathrm{pv}} \le \hat{P}_{a,t}^{\mathrm{pv}}, \forall a,t \tag{4.78}$$

$$u_{a,t}^{\mathrm{ES+}} P_a^{\mathrm{ES,min}} \le P_{a,t}^{\mathrm{ES+}} \le u_{a,t}^{\mathrm{ES+}} P_a^{\mathrm{ES,max}}, \forall a,t \tag{4.79}$$

$$u_{a,t}^{\mathrm{ES-}} P_a^{\mathrm{ES,min}} \le P_{a,t}^{\mathrm{ES-}} \le u_{a,t}^{\mathrm{ES-}} P_a^{\mathrm{ES,max}}, \forall a,t \tag{4.80}$$

$$S_{a,1}^{\mathrm{ES}} = S_{a,T}^{\mathrm{ES}} = 0.5, \forall a \tag{4.81}$$

$$S_a^{\text{ES,min}} \leq S_{a,t}^{\text{ES}} \leq S_a^{\text{ES,max}}, \forall a, t \tag{4.82}$$

$$S_{a,t+1}^{\text{ES}} = S_{a,t}^{\text{ES}} + P_{a,t}^{\text{ES+}} \eta_a^{\text{ES+}} - P_{a,t}^{\text{ES-}} / \eta_a^{\text{ES-}}, \forall a, t \tag{4.83}$$

$$V_{b,t}^{\text{dc}} = V_{a,t}^{\text{dc}} - \left(r_{ab}^{\text{dc}} H_{ab,t}^{\text{dc}} \right) / V_{0,t}^{\text{dc}}, \forall ab, t \tag{4.84}$$

$$\underline{H}_{ab,t}^{\text{dc}} \leq H_{ab,t}^{\text{dc}} \leq \bar{H}_{ab,t}^{\text{dc}}, \forall ab, t \tag{4.85}$$

$$\underline{V}_{a,t}^{\text{dc}} \leq V_{a,t}^{\text{dc}} \leq \bar{V}_{a,t}^{\text{dc}}, \forall a, t \tag{4.86}$$

$$\begin{aligned} & P_{a,t}^{\text{fc}} + \left(\hat{P}_{a,t}^{\text{pv}} - \hat{\rho}_{a,t}^{\text{pv}} \right) + \sum\nolimits_{b_1 \in \pi(a_1), l \in \mathcal{L}} (1 - \eta_{\text{BDC},l}^{\text{dcac}} P_{a_1 b_1, l, t}^{\text{dcac}}) \\ & = \hat{x}_{a,t}^{\text{dc,base}} + x_{a,t}^{\text{EV+}} + \sum\nolimits_{b \in \pi(a)} H_{ab,t}^{\text{dc}} - \sum\nolimits_{k \in \delta(a)} H_{ka,t}^{\text{dc}}, \forall a, t \end{aligned} \tag{4.87}$$

$$0 \leq P_{a_1 b_1, l, t}^{\text{acdc}} \leq I_{l,t}^{\text{acdc}} \bar{P}_{a_1 b_1, l}^{\text{acdc}}, \forall l, t \tag{4.88}$$

$$0 \leq P_{a_1 b_1, l, t}^{\text{dcac}} \leq I_{l,t}^{\text{dcac}} \bar{P}_{a_1 b_1, l}^{\text{dcac}}, \forall l, t \tag{4.89}$$

$$I_{l,t}^{\text{acdc}} + I_{l,t}^{\text{dcac}} \leq 1, \forall l, t \tag{4.90}$$

The objective (4.56)-(4.60) minimizes the total operating cost, including those of DE units, FC units, renewable energy curtailment, and BDC losses. DE unit constraints (4.61)-(4.66) model the logical, power output, and ramping constraints of DE units with minimum on/off time requirements. Wind power constraint (4.67) ensures the curtailed wind generation does not exceed the forecasted available wind power. AC power flow Constraints (4.68)–(4.71) represent AC power flow equations, line flow limits, and voltage magnitude limits at AC buses. AC power balance (4.72)–(4.73) enforce active and reactive power balance at each AC bus considering generation, loads, and interconnections. FC unit constraints (4.74)–(4.77) specify logical, output, and ramping constraints of FC units with minimum on/off times. PV power constraint (4.78) ensures PV curtailment is bounded by the available solar generation. Energy storage constraints (4.79)–(4.83) govern charging/discharging limits, efficiency, state-of-energy bounds, and terminal conditions of storage units. DC power flow constraints (4.84)–(4.86) model active power flow and voltage magnitude limits in the DC subsystem. DC power balance (4.87) enforces active power balance at each DC bus considering FCs, PVs, EVs, loads, and inter-bus power exchange. BDC operation constraints (4.88)–(4.90) specify power flow limits of BDCs and ensure mutual exclusiveness of AC–DC and DC–AC operating modes.

The deterministic model under the predicted value $\widehat{\boldsymbol{u}}$ of renewable outputs and load demand can be rewritten in the following matrix form:

$$\begin{aligned} \min_{\boldsymbol{x}_1\in\{\mathbf{0},\mathbf{1}\},\boldsymbol{x}_2\in\{\mathbf{0},\mathbf{1}\},\boldsymbol{y}} \quad & \boldsymbol{a}^\top\boldsymbol{x}_1+\boldsymbol{b}^\top\boldsymbol{y} \\ \text{s.t.}\ & \boldsymbol{C}\boldsymbol{x}_1\le\boldsymbol{c},\quad \boldsymbol{x}_1\in\{\mathbf{0},\mathbf{1}\}, \\ & \boldsymbol{D}_1\boldsymbol{x}_1+\boldsymbol{D}_2\boldsymbol{x}_2+\boldsymbol{E}\boldsymbol{y}\le\widehat{\boldsymbol{u}},\quad \boldsymbol{x}_2\in\{\mathbf{0},\mathbf{1}\}, \\ & \boldsymbol{x}_2'+\boldsymbol{x}_2''\le 1,\quad \boldsymbol{x}_2=\left(\boldsymbol{x}_2',\boldsymbol{x}_2''\right),\quad \boldsymbol{x}_2',\boldsymbol{x}_2''\in\{\mathbf{0},\mathbf{1}\} \end{aligned} \tag{4.91}$$

where $\boldsymbol{a}^\top,\boldsymbol{b}^\top,\boldsymbol{C},\boldsymbol{D}_1,\boldsymbol{D}_2,\boldsymbol{E}$ are the matrices assembled from the constraints (4.56)-(4.90). $\boldsymbol{x}_1$ denotes the here-and-now binary variables, while $\boldsymbol{x}_2$ are the mutual-exclusion binary variables. Furthermore, $\boldsymbol{y}$ are the second-stage continuous variables. To address the uncertainties in the above model, we employ the robust min–max–min energy management model as follows:

$$\min_{\boldsymbol{x}_1\in\{\mathbf{0},\mathbf{1}\}}\ \max_{\boldsymbol{u}\in\mathcal{U}}\ \min_{\boldsymbol{x}_2,\boldsymbol{y}\in\mathcal{X}(\boldsymbol{x}_1,\boldsymbol{u})} \boldsymbol{a}^\top\boldsymbol{x}_1+\boldsymbol{b}^\top\boldsymbol{y} \tag{4.92}$$

where $\mathcal{X}(\boldsymbol{x}_1,\boldsymbol{u}):=\{\boldsymbol{x}:\boldsymbol{D}_1\boldsymbol{x}_1+\boldsymbol{D}_2\boldsymbol{x}_2+\boldsymbol{E}\boldsymbol{y}\le\boldsymbol{u},\boldsymbol{x}_2'+\boldsymbol{x}_2''\le 1\}$. The robust energy management model can produce a solution that is robust against any fluctuation within the uncertainty set $\mathcal{U}$.

4.4.3 Customized C&CG Algorithm with a Semi-relaxation Embedded Sub-model

Standard C&CG algorithms assume that the recourse model in robust energy management is a convex linear program. However, the introduction of EV-related binary variables $\boldsymbol{x}_2$ obtained from NILM transforms the recourse into a mixed-integer program (MIP). This destroys convexity and turns the tri-level min–max–min problem into a more challenging bilevel MIP. Direct application of classical nested C&CG becomes computationally prohibitive, especially as the number of binary variables increases. To overcome this, we propose a customized C&CG algorithm with a semi-relaxation embedded sub-model. The whole solution process can be divided into an MIP master model and a semi-relaxation embedded sub-model.

4.4.3.1 Master Problem

The master model approximates the robust optimization by iteratively incorporating worst-case realizations from the subproblem. At iteration V, the master problem is formulated as a MILP:

$$\begin{aligned}
&\min_{\boldsymbol{x}_1\in\{\mathbf{0},\mathbf{1}\},\boldsymbol{x}_2\in\{\mathbf{0},\mathbf{1}\},\boldsymbol{y}} \boldsymbol{a}^\top\boldsymbol{x}_1+\Psi\\
\text{s.t. }&\boldsymbol{C}\boldsymbol{x}_1\le\boldsymbol{c},\quad \boldsymbol{x}_1\in\{\mathbf{0},\mathbf{1}\},\\
&\boldsymbol{D}_1\boldsymbol{x}_1+\boldsymbol{D}_2\boldsymbol{x}_2+\boldsymbol{E}\boldsymbol{y}\le\widehat{\boldsymbol{u}},\boldsymbol{x}_2\in\{\mathbf{0},\mathbf{1}\},\\
&\boldsymbol{x}_2'+\boldsymbol{x}_2''\le 1,\boldsymbol{x}_2=\left(\boldsymbol{x}_2',\boldsymbol{x}_2''\right),\boldsymbol{x}_2',\boldsymbol{x}_2''\in\{\mathbf{0},\mathbf{1}\},\\
&\Psi\ge\boldsymbol{b}^\top\boldsymbol{y}^V,\forall V,\\
&\boldsymbol{D}_1\boldsymbol{x}_1+\boldsymbol{D}_2\boldsymbol{x}_2^V+\boldsymbol{E}\boldsymbol{y}^V\le\boldsymbol{u}^V,\boldsymbol{x}_2^V\in\{\mathbf{0},\mathbf{1}\},\forall V,\\
&\boldsymbol{x}_2^{\prime,V}+\boldsymbol{x}_2^{\prime\prime,V}\le 1,\boldsymbol{x}_2^V=\left(\boldsymbol{x}_2^{\prime,V},\boldsymbol{x}_2^{\prime\prime,V}\right),\boldsymbol{x}_2^{\prime,V},\boldsymbol{x}_2^{\prime\prime,V}\in\{\mathbf{0},\mathbf{1}\},\forall V
\end{aligned}\tag{4.93}$$

where Ψ approximates the worst-case recourse cost across scenarios generated by the subproblem.

4.4.3.2 Subproblem

Given $\boldsymbol{x}_1^*$, the subproblem identifies the worst-case uncertainty realization:

$$\max_{\boldsymbol{u}\in\mathcal{U}}\min_{\boldsymbol{x}_2^V,\boldsymbol{y}^V\in\mathcal{X}(\boldsymbol{x}_1,\boldsymbol{u})}\boldsymbol{b}^\top\boldsymbol{y}^V\tag{4.94}$$

This bilevel problem is non-convex due to binary recourse variables. To handle this issue, we customize two techniques: 1) semi-relaxation based iterative solution procedure; 2) block coordinate descent theory.

4.4.3.3 Semi-Relaxation Embedded Iteration

To ensure tractability, binary variables $\boldsymbol{x}_2^V \in \{0,1\}$ are relaxed to $\boldsymbol{x}_2^V \in [0,1]$. The relaxed subproblem at iteration v is:

$$\min_{\boldsymbol{x}_2^{V,v}, \boldsymbol{y}^{V,v} \in \mathcal{X}(\boldsymbol{x}_1, \boldsymbol{u})} \quad \boldsymbol{b}^\top \boldsymbol{y}^{V,v} \tag{4.95}$$

$$\boldsymbol{D}_1 \boldsymbol{x}_1^{*V} + \boldsymbol{D}_2 \boldsymbol{x}_2^{V,v} + \boldsymbol{E} \boldsymbol{y}^{V,v} \leq \boldsymbol{u}^{V,v}, \boldsymbol{x}_2^{V,v} \in [\boldsymbol{0}, \boldsymbol{1}], \tag{4.96}$$

$$\boldsymbol{x}_2^{\prime,V,v} + \boldsymbol{x}_2^{\prime\prime,V,v} \leq 1, \boldsymbol{x}_2^{\prime,V,v}, \boldsymbol{x}_2^{\prime\prime,V,v} \in [\boldsymbol{0}, \boldsymbol{1}], \tag{4.97}$$

$$\boldsymbol{u}^{V,v} = \boldsymbol{u}^{V,v-1} : \boldsymbol{\pi}^{V,v} \tag{4.98}$$

A violated mutual-exclusion constraint is detected when

$$VS_{e,t} = \boldsymbol{x}_{2,e,t}^{*\prime,V,v} \cdot \boldsymbol{x}_{2,e,t}^{*\prime\prime,V,v} > 0, \forall t, e \in \{\mathrm{EV}, \mathrm{ES}, \mathrm{BDC}\} \tag{4.99}$$

which indicates simultaneous activation of exclusive states. Only the violated constraints are added back as binary restrictions, thereby identifying the minimal number of required integer constraints. The tightened model becomes:

$$\begin{aligned}
& (4.95) - (4.97) \\
\text{s.t. } & \boldsymbol{D}_1 \boldsymbol{x}_1^V + \boldsymbol{D}_2 \boldsymbol{x}_{2,e,t}^{V,v} + \boldsymbol{E} \boldsymbol{y}^{V,v} \leq \boldsymbol{u}^{V,v}, \\
& \boldsymbol{x}_{2,e,t}^{\prime,V,v} + \boldsymbol{x}_{2,e,t}^{\prime\prime,V,v} \leq 1, \\
& \boldsymbol{x}_{2,e,t}^{V,v}, \boldsymbol{x}_{2,e,t}^{\prime,V,v}, \boldsymbol{x}_{2,e,t}^{\prime\prime,V,v} \in \{\boldsymbol{0}, \boldsymbol{1}\}, \forall (e,t) \in \mathcal{S},
\end{aligned} \tag{4.100}$$

where $\mathcal{S} = (e,t)$ is the set of violated indices.

The solution of the model (4.100) produces the optimal value of the binary variables $\boldsymbol{x}_2^{*V,v}$, $\boldsymbol{x}_2^{*\prime,V,v}$ and $\boldsymbol{x}_2^{*\prime\prime,V,v}$, which enables to obtain the optimal value of the marginal variable $\boldsymbol{\pi}^{V,v}$ by solving the following model:

$$\begin{aligned}
& (4.95) - (4.97) \\
\text{s.t. } & \boldsymbol{D}_1 \boldsymbol{x}_1^{*V} + \boldsymbol{D}_2 \boldsymbol{x}_{2,e,t}^{V,v} + \boldsymbol{E} \boldsymbol{y}^{V,v} \leq \boldsymbol{u}^{V,v}, \\
& \boldsymbol{u}^{V,v} = \boldsymbol{u}^{V,v-1} : \boldsymbol{\pi}^{V,v}
\end{aligned} \tag{4.101}$$

4.4.3.4 Block Coordinate Descent Update

The marginal variable $\boldsymbol{\pi}^{*V,v}$ is obtained by solving the tightened model. Then the outer maximization is approximated using a first-order Taylor expansion:

$$\max_{\boldsymbol{u}^{V,v}\in\mathcal{U}} \boldsymbol{b}^\top \boldsymbol{y}^{*V,v} + (\boldsymbol{\pi}^{*V,v})^\top(\boldsymbol{u}^{V,v} - \boldsymbol{u}^{*V,v-1}) \tag{4.102}$$

This block coordinate descent procedure iteratively updates $\boldsymbol{u}^{V,v}$ until convergence. The solution process of the customized C&CG algorithm with a semi-relaxation embedded sub-model is summarized in **Algorithm** 4.

Algorithm 4 Customized C&CG with Semi-relaxation Embedded Sub-model

Initialize $V \leftarrow 1$, $LB \leftarrow -\infty$, $UB \leftarrow +\infty$, cuts$\leftarrow \varnothing$, $\boldsymbol{u}^{(V,0)} \leftarrow \widehat{\boldsymbol{u}}$, $\varepsilon \leftarrow 10^{-6}$
while $(UB - LB)/UB > \varepsilon$ **do**
    Solve MILP Master Problem (4.93) $\rightarrow \boldsymbol{x}_1^*, \Psi^*$
    $LB \leftarrow \boldsymbol{a}^\top \boldsymbol{x}_1^* + \Psi^*$
    ▷ Subproblem: find worst-case cost for $\boldsymbol{x}_1^*$
    $v \leftarrow 1$, $\boldsymbol{u}^{(V,0)} \leftarrow \widehat{\boldsymbol{u}}$
    repeat
        Solve relaxation model (4.95)-(4.98) $\rightarrow (\boldsymbol{x}_2^{V,v}, \boldsymbol{y}^{V,v})$
        Compute violation set $VS_{e,t}$ by (4.99)
        if $VS_{e,t}$ nonempty **then**
            Identify indices $\mathcal{S}$, solve (4.100) $\rightarrow (\boldsymbol{x}_2^{*V,v}, \boldsymbol{y}^{*V,v})$
        else
            $(\boldsymbol{x}_2^{*V,v}, \boldsymbol{y}^{*V,v}) \leftarrow (\boldsymbol{x}_2^{V,v}, \boldsymbol{y}^{V,v})$
        end if
        Fix $\boldsymbol{x}_2^{*V,v}$, solve LP to get duals $\boldsymbol{\pi}^{*V,v}$ (4.101)
        Solve Taylor-approximated max (4.102) $\rightarrow \boldsymbol{u}^{V,v}$
        $v \leftarrow v + 1$
    until inner convergence $\|\boldsymbol{u}^{V,v} - \boldsymbol{u}^{V,v-1}\| < \varepsilon$
    Receive worst-case $W^{*V} = \boldsymbol{b}^\top \boldsymbol{y}^{*V}$ and $(\boldsymbol{u}^{*V}, \boldsymbol{x}_2^{*V})$
    $UB \leftarrow \min(UB,\ \boldsymbol{a}^\top \boldsymbol{x}_1^* + W^{*V})$
    Generate cuts from $(\boldsymbol{u}^{*V}, \boldsymbol{x}_2^{*V}, \boldsymbol{y}^{*V})$ and add to Master:

$$\Psi \geq \boldsymbol{b}^\top \boldsymbol{y}^V, \quad \boldsymbol{D}_1\boldsymbol{x}_1 + \boldsymbol{D}_2\boldsymbol{x}_2^V + \boldsymbol{E}\boldsymbol{y}^V \leq \boldsymbol{u}^V$$

    $V \leftarrow V + 1$
end while
return final $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{y}^*, \boldsymbol{u}^*)$

4.5 Case Studies

4.5.1 System Description

In this thesis, a building microgrid system with practical CLP Hong Kong data is used to verify the effectiveness and superiority of the proposed method. The CLP dataset is a residential energy monitoring dataset spanning six months from May 1 to October 30, 2023, with active power measurements sampled at 15-minute intervals. The building includes six categories of appliances, such as EVs, washers, air conditioning (AC) units, bathroom heaters (BH), water heating (WH) elements, and kitchen outlets. The data is chronologically split with training data from May-September and testing data from October, providing 80% training and 20% testing samples. The maximum charging power of EVs is set to 11 kW according to Tesla Model 3 [148], while the weather data is publicly obtained from Hong Kong Observatory [146]. The other data and parameters for the building microgrid are provided in [149].

4.5.2 Comparisons of Nonintrusive Load Monitoring Approaches

To assess the accuracy and effectiveness of the proposed dual-task and multimodal approaches, three evaluation metrics are employed: state classification accuracy (SCA), F1 score, and mean relative error (MRE), as defined in (4.103)–(4.105). The performance of our models is compared with representative state-of-the-art methods, including CNN [57], SAMNet [63], and CBMNet [64], with results summarized in Tables 4.2 and 4.3.

$$SCA_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{4.103}$$

$$F1_i = \frac{2 \cdot (\text{Pre}_i \cdot \text{Rec}_i)}{\text{Pre}_i + \text{Rec}_i} \tag{4.104}$$

$$MRE_i = \frac{1}{T} \sum\nolimits_{t=1}^{T} \frac{|\hat{x}_i(t) - x_i(t)|}{\max(\hat{x}_i(t), x_i(t))} \tag{4.105}$$

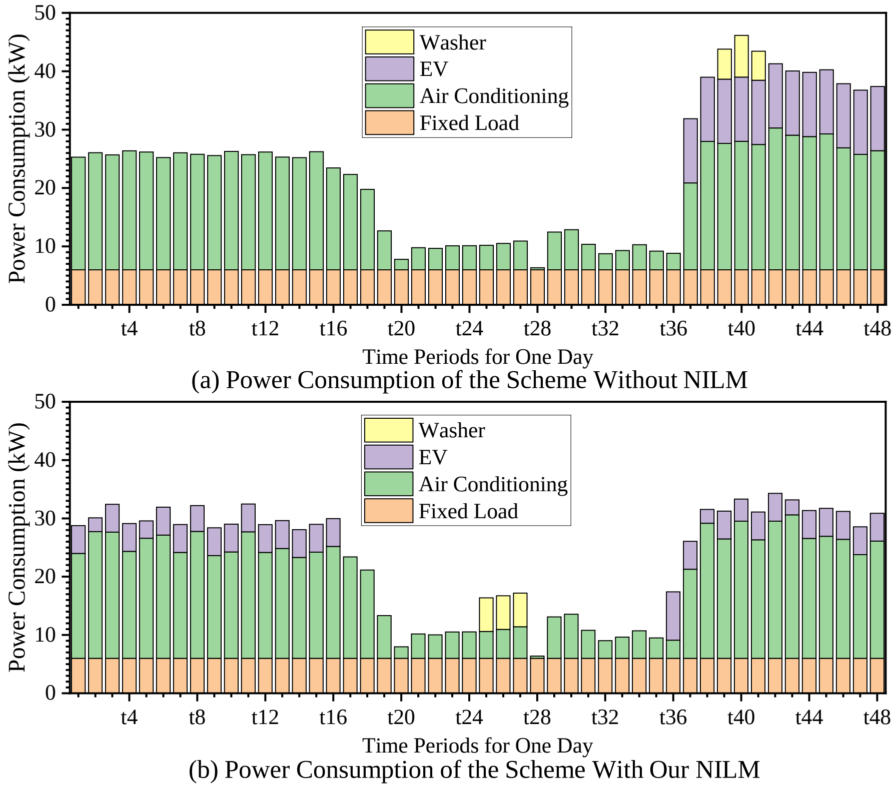
Here, TP_i and TN_i denote the counts of true positives and true negatives, corresponding to the time steps when appliance i is correctly detected as ON or OFF. Conversely, FP_i and FN_i represent false positives and false negatives, where the appliance state is misclassified. A higher SCA reflects more reliable state identification. Precision, defined as $\text{Pre}_i = \frac{TP_i}{TP_i+FP_i}$, measures the proportion of predicted ON states that are correct, while recall, $\text{Rec}_i = \frac{TP_i}{TP_i+FN_i}$, captures the fraction of actual ON states that are successfully identified.

As shown in Table 4.2, the proposed method achieves an average SCA of 95.02%, considerably higher than CNN (77.17%), SAMNet (82.69%), and CBMNet (86.52%). This gain is particularly pronounced for complex appliances such as air conditioners, whose variable modes and dependence on external factors often challenge traditional models. Similarly, the F1 score of our method reaches 68.73% on average, significantly surpassing CNN (22.80%), SAMNet (32.11%), and CBMNet (42.49%). Furthermore, the MRE of 9.07% highlights the robustness of our framework in load estimation, demonstrating strong generalization across diverse appliances.

Table 4.3 further illustrates the benefits of multimodal inputs by comparing the performance of the proposed method with and without environmental covariates. Incorporating temperature, humidity, and solar radiation yields consistent improvements across all metrics. Specifically, average SCA rises from 92.97% to 95.02%, while the F1 score nearly doubles from 49.61% to 68.73%. Likewise, MRE decreases from 10.52% to 9.07%, confirming the importance of auxiliary weather data for appliances sensitive to environmental conditions, such as air conditioners and water heaters. Taken together, these findings demonstrate that integrating multitask learning with multimodal features significantly enhances NILM performance compared to conventional single-task, single-modal approaches, positioning the proposed HMMOE as a state-of-the-art solution for appliance-level monitoring in building microgrids.

4.5.3 Comparisons of Various Energy Management Approaches

To validate the effectiveness of the proposed NILM-aided robust energy management strategy, we benchmark it against deterministic and non-NILM-based approaches. The comparative results in Table 4.4 clearly demonstrate the benefits of integrating NILM into robust energy management. Compared with the cases without NILM, both BbUS and CHUS strategies achieve significant cost reductions when appliance-level disaggregation is considered. In particular, the proposed NILM-aided



(a) Power Consumption of the Scheme Without NILM

(b) Power Consumption of the Scheme With Our NILM

Figure 4.3: Power consumption of appliances with various approaches

CHUS reduces the total cost (TC) to ¥267.6, which is approximately 23% lower than the corresponding case without NILM (¥347.7). This improvement mainly arises from the sharp decrease in renewable curtailment (RC) and second-stage imbalance cost (SSC), reflecting the model's ability to exploit appliance flexibility for aligning demand with renewable availability. Although the deterministic approach with NILM shows the lowest total cost, its lower NVS value (64.51%) indicates weak robustness under uncertainty, which underscores the advantage of our robust NILM-aided formulation.

Fig. 4.3 provides additional insights into the load scheduling outcomes. Without NILM, the system fails to differentiate between fixed and flexible loads, resulting in inefficient peak demand and greater reliance on balancing resources. By contrast, the NILM-aided strategy successfully captures the deferrability of washers and EVs, shifting their operation to off-peak hours and smoothing overall consumption. This demand reshaping not only lowers imbalance and curtailment costs but also improves renewable utilization. As a result, the NILM-aided robust CHUS strategy achieves a favorable trade-off between cost efficiency and robustness, highlighting the crucial role of appliance-level analytics in enhancing microgrid operation under uncertainty.

4.6 Summary

This thesis develops a unified NILM-aided robust energy management framework for hybrid microgrids, addressing key limitations of conventional single-task and single-modal approaches. By proposing a dual-task multimodal learning model, the framework simultaneously improves appliance state detection and power estimation, capturing causal relationships between environmental conditions and appliance operation. The analytical multi-objective balancing mechanism ensures stable and interpretable training, while the integration of NILM outputs into the robust scheduling model bridges the gap between appliance-level analytics and system-level decision making. Case studies using real-world Hong Kong data verify the superior performance of the proposed method over benchmark NILM approaches, demonstrating significant improvements in classification accuracy, F1 score, and load estimation. Furthermore, comparative studies on energy management approaches reveal that the proposed NILM-aided strategy reduces total operating costs, enhances renewable energy utilization, and ensures robust performance under uncertainty. The results confirm that accurate appliance-level disaggregation provides actionable demand-side flexibility, enabling better coordination of distributed resources and mitigating renewable variability.

Table 4.1: Notation for the building microgrid model

Indices and sets	
$\mathcal{A}/a, b, k$	Set and index of buses
$\mathcal{L}/l$	Set and index of BDCs
$\pi(a)/\delta(a)$	Set of father/son buses connected to bus a
Parameters	
$f_a^{\text{de}}/f_a^{\text{fc}}$	Start-up cost of DE/FC units
$n_a^{\text{de}}/n_a^{\text{fc}}$	Constant cost coefficient of DE/FC units
$c_a^{\text{de}}/c_a^{\text{fc}}$	Variable cost coefficient of DE/FC units
$c_a^{\text{w}}/c_a^{\text{pv}}$	Cost coefficient for wind/PV power curtailment
c_l^{loss}	Cost coefficient for BDC power loss
$T_a^{\text{de,on}}/T_a^{\text{fc,on}}$	Minimum online time for DE/FC units
$T_a^{\text{de,off}}/T_a^{\text{fc,off}}$	Minimum offline time for DE/FC units
$\underline{P}_a^{\text{de}}/\bar{P}_a^{\text{de}}$	Minimum/maximum active power output of DE units
$\underline{Q}_a^{\text{de}}/\bar{Q}_a^{\text{de}}$	Minimum/maximum reactive power output of DE units
$R_a^{\text{P,de}}/R_a^{\text{Q,de}}$	Maximum active/reactive power rate of DE units
$\hat{P}_{a,t}^{\text{w}}/\hat{P}_{a,t}^{\text{pv}}$	Forecasted wind/PV power
$V_{0,t}^{\text{ac}}/V_{0,t}^{\text{dc}}$	Reference voltage magnitude for AC/DC microgrids
$\underline{H}_{ab,t}^{\text{ac}}/\bar{H}_{ab,t}^{\text{ac}}$	Minimum/maximum active power flow on AC line ab
$\underline{G}_{ab,t}^{\text{ac}}/\bar{G}_{ab,t}^{\text{ac}}$	Minimum/maximum reactive power flow on AC line ab
$\underline{V}_{a,t}^{\text{ac}}/\bar{V}_{a,t}^{\text{ac}}$	Minimum/maximum voltage magnitude at AC bus
$\hat{x}_{a,t}^{\text{ac,base}}/\hat{x}_{a,t}^{\text{dc,base}}$	Fixed base AC/DC active power load
$Q_{a,t}^{\text{ac}}$	Fixed base reactive power load
$\underline{P}_a^{\text{fc}}/\bar{P}_a^{\text{fc}}$	Minimum/maximum active power output of FC unit
R_a^{fc}	Maximum ramp-up/down rate for active power of FC unit
$P_a^{\text{ES,min}}/P_a^{\text{ES,max}}$	Minimum/maximum power of storage unit
$S_a^{\text{ES,min}}/S_a^{\text{ES,max}}$	Minimum/maximum energy level of storage unit
$\eta_a^{\text{ES}+}/\eta_a^{\text{ES}-}$	Charging/discharging efficiency of storage unit
$r_{ab}^{\text{ac}}/r_{ab}^{\text{dc}}$	Resistance of AC/DC line
x_{ab}^{ac}	Reactance of AC/DC line
$\underline{H}_{ab,t}^{\text{dc}}/\bar{H}_{ab,t}^{\text{dc}}$	Minimum/maximum active power flow on DC line ab
$\underline{V}_{a,t}^{\text{dc}}/\bar{V}_{a,t}^{\text{dc}}$	Minimum/maximum voltage magnitude at DC bus
$\bar{P}_{a_1b_1,l}^{\text{acdc}}/\bar{P}_{a_1b_1,l}^{\text{dcac}}$	Power flow capacity of BDC l (AC-DC/DC-AC)
$\eta_{\text{BDC},l}^{\text{acdc}}/\eta_{\text{BDC},l}^{\text{dcac}}$	Efficiency of BDC l for AC-DC/DC-AC conversion
Variables	
$U_{a,t}^{\text{de}}/U_{a,t}^{\text{fc}}$	Binary start-up status of DE/FC unit
$D_{a,t}^{\text{de}}/D_{a,t}^{\text{fc}}$	Binary shut-down status of DE/FC unit
$P_{a,t}^{\text{de}}/P_{a,t}^{\text{fc}}$	Active power output of DE/FC unit
$\hat{\rho}_{a,t}^{\text{w}}/\hat{\rho}_{a,t}^{\text{pv}}$	Curtailed wind/PV power
$P_{a_1b_1,l,t}^{\text{acdc}}/P_{a_1b_1,l,t}^{\text{dcac}}$	Power flow through BDC l from AC-DC / DC-AC
$I_{a,t}^{\text{de}}/I_{a,t}^{\text{fc}}$	Binary online status of DE/FC unit
$Q_{a,t}^{\text{de}}$	Reactive power output of DE unit
$I_{a,t}^{\text{de}}/I_{a,t}^{\text{fc}}$	Binary online status of DE/FC units
$V_{a,t}^{\text{ac}}/V_{a,t}^{\text{dc}}$	Voltage magnitude at AC/DC bus
$H_{ab,t}^{\text{ac}}/H_{ab,t}^{\text{dc}}$	Active power flow from bus a to bus b on AC/DC line
$G_{ab,t}^{\text{ac}}$	Reactive power flow from bus a to bus b on AC line
$u_{a,t}^{\text{ES}+}/u_{a,t}^{\text{ES}-}$	Binary charging/discharging status of storage units
$S_{a,t}^{\text{ES}}$	Energy level of storage units
$P_{a,t}^{\text{ES}+}/P_{a,t}^{\text{ES}-}$	Charging/discharging power of storage units
$I_{l,t}^{\text{acdc}}/I_{l,t}^{\text{dcac}}$	Binary status of BDC l for AC-DC/DC-AC conversion
$P_{a_1b_1,l,t}^{\text{acdc}}/P_{a_1b_1,l,t}^{\text{dcac}}$	Power flow through BDC l from AC-DC/DC-AC

Table 4.2: Performance comparisons of the proposed HMMOE approach and the advanced NILM approaches.

Metrics	Methods	Real-world Hong Kong Dataset											
		Washer	AC1	AC2	AC3	AC4	AC5	Kitchen	BH	WH1	WH2	WH3	Average
SCA	Proposed	**93.16%**	**91.54%**	**95.59%**	**96.31%**	**87.80%**	**88.70%**	**96.68%**	**97.22%**	**99.54%**	**99.53%**	**99.15%**	**95.02%**
	CNN[57]	89.31%	71.35%	67.06%	80.33%	83.57%	73.41%	91.54%	48.26%	73.42%	77.78%	92.84%	77.17%
	SAMNet [63]	47.13%	87.98%	86.88%	83.11%	87.20%	86.51%	94.49%	46.72%	96.49%	96.58%	96.53%	82.69%
	CBMNet [64]	85.17%	83.56%	80.08%	81.14%	85.35%	84.53%	93.89%	65.67%	97.49%	97.61%	97.24%	86.52%
F1	Proposed	**22.19%**	**84.85%**	**93.71%**	**94.84%**	**77.00%**	**79.94%**	**24.88%**	**13.81%**	**92.26%**	**92.29%**	**80.24%**	**68.73%**
	CNN[57]	5.67%	47.48%	36.87%	32.73%	33.48%	53.08%	4.67%	0.67%	15.22%	17.12%	3.79%	22.80%
	SAMNet [63]	12.06%	65.74%	55.71%	71.28%	57.98%	67.37%	5.45%	1.28%	6.36%	2.94%	7.04%	32.11%
	CBMNet [64]	7.62%	60.06%	49.27%	64.37%	57.86%	66.49%	16.12%	2.09%	60.65%	58.78%	24.09%	42.49%
MRE	Proposed	**10.18%**	**14.23%**	**10.98%**	**9.16%**	**15.98%**	**17.90%**	**4.75%**	**8.31%**	**2.93%**	**2.88%**	**2.45%**	**9.07%**
	CNN[57]	11.01%	36.78%	39.65%	22.68%	19.13%	37.02%	8.63%	51.90%	28.58%	24.00%	7.29%	26.06%
	SAMNet [63]	10.82%	15.59%	27.99%	21.37%	18.42%	28.35%	7.53%	5.53%	13.79%	8.31%	10.75%	15.31%
	CBMNet [64]	11.82%	23.99%	21.68%	31.65%	22.51%	21.47%	8.22%	3.32%	7.60%	10.98%	5.11%	15.30%

Table 4.3: Performance comparisons of the proposed HMMOE approach with multi-modal data versus a baseline with single smart meter data.

Metrics	Information	Real-world Hong Kong Dataset					
		Washer	AC	Kitchen	BH	WH	Average
SCA	Smart Meter	92.89%	88.58%	94.86%	**98.98%**	97.70%	92.97%
	Multi-Modal	**93.16%**	**91.99%**	**96.68%**	97.22%	**99.41%**	**95.02%**
F1	Smart Meter	14.36%	66.41%	24.34%	2.61%	57.26%	49.61%
	Multi-Modal	**22.19%**	**86.07%**	**24.88%**	**13.81%**	**88.26%**	**68.73%**
MRE	Smart Meter	10.50%	16.73%	6.63%	**1.32%**	4.54%	10.52%
	Multi-Modal	**10.18%**	**13.65%**	**4.75%**	8.31%	**2.75%**	**9.07%**

Table 4.4: Comparisons with different energy management methods (¥)

Methods		**TC**	**GC**	**TrC**	**RC**	**LC**	**SSC**	**NVS**
Deterministic [150], [151]		206.1	179.2	6.3	9.8	2.3	8.5	64.51%
Without NILM	BbUS	467.7	223.8	6.2	58.4	2.6	174.1	100%
	CHUS	347.7	227.5	1.3	45.6	1.6	70.1	100%
With Our NILM	BbUS	396.6	219.5	1.6	48.0	2.6	122.4	100%
	CHUS	267.6	217.4	15.4	16.8	1.6	14.7	100%

TC: total cost; FSC: first-stage cost including generation cost (GC) + trading cost (TrC) + renewable curtailment cost (RC) + loss cost on BDCs (LC); SSC: second-stage unbalanced cost.

Chapter 5

Probability-Driven Intersected Uncertainty Set Modelling of Renewable Energy Sources for Microgrid Operation

5.1 Nomenclature

5.1.1 Abbreviations

ACsM/DCsM	AC/DC sub-microgrid.
BbUS/CHUS	Box-based/convex hull uncertainty set (US).
BDC	Bi-directional converter.
C&CG	Column & constraint generation.
EUS/PUS	Ellipsoidal/polyhedral US.
HMG	Hybrid AC/DC microgrid (MG).
ICHUS	Intersected pair CHUS.
KKT	Karush-Kuhn-Tucker.
RES	Renewable energy source.
SRII	Simultaneous rectifier and inverter index.

5.1.2 Sets and Indices

$i, \mathcal{I}$	Index and set of buses in an HMG.
i_1, j_1	Index of buses in an ACsM or DCsM linking BDCs.
$l, \mathcal{L}$	Index and set of BDCs.
$m, \mathcal{M}$	Index and set of vertices in a convex hull.
r, R	Index and number of RES units.
$t, \mathcal{T}$	Index and set of time periods.
$\pi(i), \delta(i)$	Set of parents/children of bus i.

5.1.3 Parameters

$c_i^{\mathrm{ac}\pm}/c_i^{\mathrm{dc}\pm}$	Unbalanced penalty costs in ACsM/DCsM.
$c_i^{\mathrm{de}}/c_i^{\mathrm{fc}}$	Fuel cost coefficient of diesel engine/fuel cell units.
$c_i^{\mathrm{w}}/c_i^{\mathrm{pv}}$	Penalty cost coefficient of wind turbine/photovoltaic units.
c_l^{loss}	Loss cost coefficient of BDCs.
$f_i^{\mathrm{de}}/f_i^{\mathrm{fc}}$	Shut-down cost coefficient of diesel engine/fuel cell units.
$\underline{G}_{ij,t}^{\mathrm{ac}}/\bar{G}_{ij,t}^{\mathrm{ac}}$	Lower/upper reactive power of ACsM branch.
$\underline{H}_{ij,t}^{\mathrm{ac}}/\underline{H}_{ij,t}^{\mathrm{dc}}$	Lower active power of ACsM/DCsM branch.
$\bar{H}_{ij,t}^{\mathrm{ac}}/\bar{H}_{ij,t}^{\mathrm{dc}}$	Upper active power of ACsM/DCsM branch.
$L_{i,t}^{\mathrm{ac}}/L_{i,t}^{\mathrm{dc}}$	Load demand in an ACsM/DCsM.
$n_i^{\mathrm{de}}/n_i^{\mathrm{fc}}$	Start-up cost coefficient of diesel engine/fuel cell units.
$o_{0,l}/o_{1,l}$	Coefficients of approximation.
$\underline{P}_i^{\mathrm{de}}/\underline{P}_i^{\mathrm{fc}}$	Lower power output of diesel engine/fuel cell units.
$\bar{P}_i^{\mathrm{de}}/\bar{P}_i^{\mathrm{fc}}$	Upper power output of diesel engine/fuel cell units.
$\hat{P}_{i,t}^{\mathrm{w}}/\hat{P}_{i,t}^{\mathrm{pv}}$	Predicted power output of wind turbine/photovoltaic units.
$\bar{P}_{i_1j_1,l}^{\mathrm{acdc}}/\bar{P}_{i_1j_1,l}^{\mathrm{dcac}}$	Upper rectifier/inverter power of BDCs.
$\underline{Q}_i^{\mathrm{de}}/\bar{Q}_i^{\mathrm{de}}$	Lower/upper reactive power of diesel engine units.
$Q_{i,t}^{\mathrm{ac}}$	Reactive load demand in an ACsM.
$r_{ij}^{\mathrm{ac}}/r_{ij}^{\mathrm{dc}}$	Resistance of ACsM/DCsM branch.
$R_i^{\mathrm{P,de}}/R_i^{\mathrm{fc}}$	Ramping rate of diesel engine/fuel cell units.
$R_i^{\mathrm{Q,de}}/R_i^{\mathrm{Q,de}}$	Reactive ramping rate of diesel engine units.
R_i^{W}	RES power curtailment at node i.
$T_i^{\mathrm{de,on}}/T_i^{\mathrm{fc,on}}$	Minimum up time of diesel engine/fuel cell units.
$T_i^{\mathrm{de,off}}/T_i^{\mathrm{fc,off}}$	Minimum down time of diesel engine/fuel cell units.
$\underline{V}_{i,t}^{\mathrm{ac}}/\underline{V}_{i,t}^{\mathrm{dc}}$	Lower voltage amplitude in an ACsM/DCsM.
$\bar{V}_{i,t}^{\mathrm{ac}}/\bar{V}_{i,t}^{\mathrm{dc}}$	Upper voltage amplitude in an ACsM/DCsM.
x_{ij}^{ac}	Reactance of ACsM branch.
χ	Allowed adjustment time from predicted scenario to worst-case scenario.
$\hat{\lambda}_r$	Predicted RES output value.
$\lambda_r^{\mathrm{u}}/\lambda_r^{\mathrm{d}}$	Upper/lower bounds of RES outputs.

5.1.4 Variables

$C^{\text{de}}/C^{\text{fc}}$	Generation costs of diesel engine/fuel cell units.
$C^{\text{w}}/C^{\text{pv}}$	Penalty costs of wind turbine/photovoltaic power curtailment.
C^{bdc}	Power loss cost of BDCs.
$C^{\text{ac}}/C^{\text{dc}}$	Unbalanced costs in an ACsM/DCsM.
$D^{\text{de}}_{i,t}/D^{\text{fc}}_{i,t}$	Shut-down variable of diesel engine/fuel cell units.
$G^{\text{ac}}_{ij,t}$	Reactive power flow of ACsM branch.
$H^{\text{ac}}_{ij,t}/H^{\text{dc}}_{ij,t}$	Active power flow of ACsM/DCsM branch.
$I^{\text{de}}_{i,t}/I^{\text{fc}}_{i,t}$	State variable of diesel engine/fuel cell units.
$I^{\text{acdc}}_{l,t}/I^{\text{dcac}}_{l,t}$	Rectifier/inverter state variable of BDCs.
$P^{\text{de}}_{i,t}/P^{\text{fc}}_{i,t}$	Generation power outputs of diesel engine/fuel cell units.
$P^{\text{acdc}}_{i_1j_1,l,t}/$ $P^{\text{dcac}}_{i_1j_1,l,t}$	Inverter and rectifier power flow through BDCs linking i_1 in an ACsM and j_1 in a DCsM.
$Q^{\text{de}}_{i,t}$	Reactive power output of diesel engine units.
$U^{\text{de}}_{i,t}/U^{\text{fc}}_{i,t}$	Start-up variable of diesel engine/fuel cell units.
$V^{\text{ac}}_{i,t}/V^{\text{dc}}_{i,t}$	Voltage amplitude in an ACsM/DCsM.
$\hat{\rho}^{\text{w}}_{i,t}/\hat{\rho}^{\text{pv}}_{i,t}$	Power curtailment of wind turbine/photovoltaic units in a predicted scenario.
$\eta^{\text{acdc}}_{\text{BDC},l}/\eta^{\text{dcac}}_{\text{BDC},l}$	Rectifier/inverter efficiency of BDCs.
$\xi^{\text{ac}\pm}_{i,t}/\xi^{\text{dc}\pm}_{i,t}$	Unbalanced power in an ACsM/DCsM.
$\varepsilon^{\text{u}}_{r}/\varepsilon^{\text{d}}_{r}$	Binary variables indicating whether an RES output fluctuates upward/downward.

5.2 Existing Uncertainty Sets

Existing approaches for modeling renewable generation uncertainty mainly include stochastic programming, classical robust optimization, and convex-hull-based uncertainty sets. Stochastic programming relies on accurate probabilistic descriptions and scenario generation, which often requires large volumes of historical data and assumes stationarity of renewable outputs. These assumptions may not hold in microgrid-level applications, where data are limited and renewable generation exhibits strong non-stationary and location-specific characteristics. Classical robust optimization avoids probabilistic assumptions but typically employs fixed polyhedral and interval-based uncertainty sets, which may lead to overly conservative solutions and unnecessary operational costs. Full-dimensional convex hull uncertainty sets constructed directly from historical samples can capture data-driven uncertainty more realistically; however, their computational complexity increases rapidly with the number of time periods and renewable units, making them difficult to scale for practical hybrid AC/DC microgrid applications.

Details regarding the above-discussed uncorrelated uncertainty sets (i.e., the BbUS and PUS) and correlated uncertainty sets (i.e., the EUS and CHUS) are first introduced to support subsequent formulations of the proposed ICHUS and facilitate comparisons among case study results.

5.2.1 Uncorrelated Uncertainty Sets

The BbUS $\mathcal{S}^{\mathrm{Bb}}$ is expressed as follows.

$$\mathcal{S}^{\mathrm{Bb}}=\left\{\lambda_r \;\middle|\; \begin{array}{l} \lambda_r = \hat{\lambda}_r+\varepsilon_r^{\mathrm{u}}(\lambda_r^{\mathrm{u}}-\hat{\lambda}_r)+\varepsilon_r^{\mathrm{d}}(\lambda_r^{\mathrm{d}}-\hat{\lambda}_r) \\ \varepsilon_r^{\mathrm{u}}+\varepsilon_r^{\mathrm{d}} \le 1, \varepsilon_r^{\mathrm{u}}, \varepsilon_r^{\mathrm{d}} \in [0,1], \forall r \end{array}\right\} \tag{5.1}$$

The excessive conservativeness of the solutions derived from the BbUS is addressed by controlling the size of the uncertainty set via the application of an uncertainty budget Γ to constrain the allowed fluctuations in the BbUS. Therefore, the PUS $\mathcal{S}^{\mathrm{P}}$ is constructed as follows.

$$\mathcal{S}^{\mathrm{P}}=\left\{\lambda_r \;\middle|\; \begin{array}{l} \lambda_r = \hat{\lambda}_r+\varepsilon_r^{\mathrm{u}}(\lambda_r^{\mathrm{u}}-\hat{\lambda}_r)+\varepsilon_r^{\mathrm{d}}(\lambda_r^{\mathrm{d}}-\hat{\lambda}_r), \\ \varepsilon_r^{\mathrm{u}}+\varepsilon_r^{\mathrm{d}} \le 1, \varepsilon_r^{\mathrm{u}}, \varepsilon_r^{\mathrm{d}} \in [0,1], \forall r, \\ \sum_{r=1}^{R}\left(\varepsilon_r^{\mathrm{u}}+\varepsilon_r^{\mathrm{d}}\right) \le \Gamma \end{array}\right\} \tag{5.2}$$

A comparison of the BbUS in (5.1) and the PUS in (5.2) indicates that the PUS degrades into the BbUS when $\Gamma \geq R$.

5.2.2 Correlated Uncertainty Sets

The CHUS $\mathcal{S}^{\mathrm{CH}}$ is expressed as follows.

$$\mathcal{S}^{\mathrm{CH}}=\left\{\lambda_r \left| \begin{array}{l} \lambda_r=\hat{\lambda}_r+\sum_{m\in\mathcal{M}}\varepsilon_{rm}(\lambda_{rm}-\hat{\lambda}_r),\forall r, \\ 0\le\varepsilon_{rm}\le 1,\forall r,m, \\ \sum_{m\in\mathcal{M}}\varepsilon_{rm}=1,\forall r \end{array}\right.\right\} \tag{5.3}$$

In the EUS, the terms λ_r and σ_r denote the mean value and the standard deviation of the historical data of the r-th RES unit, respectively. Accordingly, an ellipse with mean normalization can be expressed as follows.

$$\mathcal{S}^{\mathrm{E}}=\left\{\boldsymbol{\lambda}\left|(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})\le\chi_R^2(\alpha)\right.\right\} \tag{5.4}$$

$$\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_{R\times R}=\begin{bmatrix} \sigma(\lambda_1,\lambda_1) & \cdots & \sigma(\lambda_1,\lambda_R) \\ \vdots & \ddots & \vdots \\ \sigma(\lambda_R,\lambda_1) & \cdots & \sigma(\lambda_R,\lambda_R) \end{bmatrix} \tag{5.5}$$

Here, $\boldsymbol{\lambda}=[\lambda_1\lambda_2...\lambda_R]$, $\hat{\boldsymbol{\lambda}}=[\hat{\lambda}_1\hat{\lambda}_2...\hat{\lambda}_R]$, and $\boldsymbol{\Sigma}$ is a positive-definite matrix with R elements, and is a symmetric matrix because $\sigma(\lambda_r,\lambda_{r'})=\sigma(\lambda_{r'},\lambda_r)$. In addition, $\chi_R^2(\alpha)$ defines the size of the ellipse with an α% confidence level, which enables MG dispatchers to flexibly select the size of the ellipse based on their operation experiences. Thus, a crucial issue in EUS $\mathcal{S}^{\mathrm{E}}$ is how to determine the value of $\chi_R^2(\alpha)$ in $\mathcal{S}^{\mathrm{E}}$ based on the confidence level given by MG dispatchers.

The relationships between the described uncertainty sets for the two RES units are illustrated in Fig.5.1. The BbUS is the simplest approach, representing independent RES generation variations within fixed bounds. However, it is highly conservative, as each RES unit can independently reach its extreme generation limits. The PUS improves upon BbUS by using linear inequalities to form convex polytopes, making it less conservative. In comparison, EUS and CHUS capture the correlation between RES generation, with EUS relying on non-linear modeling and CHUS using linear forms. While BbUS and PUS are computationally simpler, EUS and CHUS offer more realistic representations of uncertainty at the cost of increased complexity. To balance conservativeness and computational efficiency, we propose a confidence-level controllable ICHUS in the following sec-

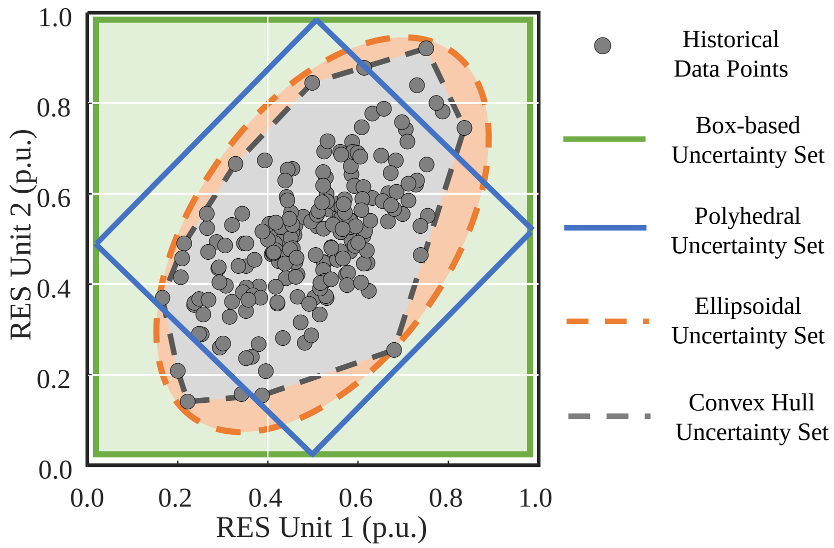


Figure 5.1: Illustrations of various uncertainty sets.

tion.

5.3 Proposed Intersected Convex Hull Uncertainty Set with Controllable Confidence Level

As discussed above, we first formulate an EUS to determine flexible confidence levels, and thereby remove outlier data samples. Then, we construct the ICHUS to tightly approximate the original CHUS with much fewer vertices.

5.3.1 Determination of Confidence Level

In an effort to determine the value of $\chi^2_R(\alpha)$ in (5.4) based on the confidence level given by MG dispatchers, we assume that deviations in RES unit outputs follow normal distributions (i.e., $\boldsymbol{\lambda} \sim \mathrm{N}_R(\hat{\boldsymbol{\lambda}}, \boldsymbol{\Sigma})$), which yields the following.

$$f(\boldsymbol{\lambda}) = (2\pi)^{-R/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\lambda} - \hat{\boldsymbol{\lambda}}) \right\} \tag{5.6}$$

Defining $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})$, where $\boldsymbol{Z} = [z_1 z_2 ... z_R]^{\mathrm{T}}$, yields the expression $\boldsymbol{\lambda} = h(\boldsymbol{Z}) = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}+\hat{\boldsymbol{\lambda}}$. As a result, (5.6) can be rewritten as follows.

$$\begin{aligned} & f(\boldsymbol{Z}) = (2\pi)^{-R/2}|\boldsymbol{\Sigma}|^{-1/2} \\ & \times \exp\left\{-\tfrac{1}{2}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z})\right\} \\ & = (2\pi)^{-R/2}\exp\left\{-\tfrac{1}{2}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\right\} \\ & = \prod_{r=1}^{R}(2\pi)^{-1/2}\exp\left\{-\tfrac{z_r^2}{2}\right\} \end{aligned} \tag{5.7}$$

Equation (5.7) indicates that all elements $z_1, z_2, ..., z_R$ are stochastic variables that are independent of each other and follow standard normal distributions, (i.e., $z_r \sim \mathrm{N}(0,1), \forall r$).

In addition, we can obtain the following relationship:

$$y = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} = (\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\lambda}-\hat{\boldsymbol{\lambda}}), \tag{5.8}$$

which indicates that y follows a Chi-square distribution with R degrees of freedom based on Cochran's theorem [152]. Therefore, $\chi_R^2(\alpha)$ in (5.4) can be determined for a given confidence level using the gamma function $\Gamma(R/2) = \int_0^{+\infty} x^{R/2-1}e^{-x}dx$ and incomplete gamma function $\gamma\left(R/2, \chi_R^2(\alpha)/2\right) = \int_{\chi_R^2(\alpha)/2}^{+\infty} x^{R/2-1}e^{-x}dx$ as follows.

$$F_R(\chi_R^2(\alpha)) = P(y \leq \chi_R^2(\alpha)) = \alpha \tag{5.9}$$

$$F_R(\chi_R^2(\alpha)) = \gamma\left(R/2, \chi_R^2(\alpha)/2\right)/\Gamma\left(R/2\right) \tag{5.10}$$

Remark 1: Empirical evidence from practical cases suggests that the predicted error of RES outputs typically follows a normal (Gaussian) distribution [88], [153]. This observation is supported by the Central Limit Theorem, which states that the sum of a large number of identically distributed random variables tends to approximate a normal distribution, regardless of the original distribution of the variables. This assumption facilitates the derivation of key relationships, such as the Chi-square distribution of y, which is crucial for determining confidence levels. It is generally applicable in practice, as large sample sizes of RES outputs are typically available in practice, satisfying the requirements of the Central Limit Theorem. Additionally, it is noteworthy that historical data of RES outputs may include a few abnormal values. Our method incorporates a confidence-level-enabled ellipse to exclude outliers from historical RES data. This ensures that the obtained energy management solution is not adversely influenced by abnormal or erroneous data points, leading to more realistic and economical strategies.

5.3.2 Intersected Pair Convex Hull Uncertainty Set

The ICHUS is constructed by dividing the RES units into subsets of a pair unit and the remaining units. Specifically, we can derive $C_R^2 = R(R-1)/2$ pairs for R RES units. For each pair of RES units (i.e., (p, q)), we can form a pair CHUS $\mathcal{S}_{p,q}^{\mathrm{PCH}}$ based on the historical data of the RES unit pair (p, q) as follows.

$$\mathcal{S}_{p,q}^{\mathrm{PCH}} = \left\{ \lambda_r \;\middle|\; \begin{array}{l} \lambda_r = \hat{\lambda}_r + \sum_{m \in \mathcal{M}_{p,q}} \varepsilon_{rm} (\lambda_{rm} - \hat{\lambda}_r), \forall r = p, q \\ 0 \leq \varepsilon_{rm} \leq 1, \forall r = p, q; m \in \mathcal{M}_{p,q} \\ \sum_{m \in \mathcal{M}_{p,q}} \varepsilon_{rm} = 1, \forall r = p, q \\ \lambda_r^{\mathrm{d}} \leq \lambda_r \leq \lambda_r^{\mathrm{u}}, \forall r \neq p, q \end{array} \right\} \tag{5.11}$$

Here, $\mathcal{M}_{p,q}$ is the set of vertices for the unit pair $(p, q), \forall 1 \leq p < q \leq I$. Then, we intersect all pairs of $\mathcal{S}_{p,q}^{\mathrm{PCH}}$ to obtain the smallest volume of the convex set as follows.

$$\mathcal{S}^{\mathrm{ICH}} = \cap_{1 \leq p < q \leq I} \mathcal{S}_{p,q}^{\mathrm{PCH}} \tag{5.12}$$

The procedure applied herein for generating the ICHUS is detailed in **Algorithm** 5 and illustrated in Fig. 5.2. This algorithm constructs a compact and accurate uncertainty set for RES units by dividing the R RES units into all possible pairs and creating a pair CHUS for each pair using the Quickhull algorithm [154], which efficiently determines the convex hull of historical RES data points. Each pair CHUS captures the correlations between the paired units while ensuring their outputs remain within realistic bounds.

The algorithm iterates through all pairs, collects their pair CHUSs, and intersects them to form the ICHUS. This process minimizes the volume of the uncertainty set while maintaining robustness against RES uncertainty. By reducing computational complexity and excluding unrealistic outlier scenarios, our ICHUS offers a scalable and less conservative representation of RES uncertainty, making it well-suited for robust energy management in HMGs with extensive RES units.

Algorithm 5 Construction of the ICHUS

Input number of RES units R
Initial $p = 1, q = 2$
Divide R RES units into $R(R-1)/2$ pairs
while $p \leq R-1$ **do**
  while $p < q \leq R$ **do**
    Construct $\mathcal{S}^{\mathrm{PCH}}_{p,q}$ using the Quick-hull algorithm[154]
    $q \leftarrow q+1$
  end while
  $p \leftarrow p+1$
  $q \leftarrow p+2$
end while
Obtain the set of all pairs $\mathcal{S}^{\mathrm{PCH}}_{p,q}, \forall 1 \leq p < q \leq R$
Form the ICHUS based on (5.12)

Furthermore, we derive the following proposition to investigate the relationship among the standard CHUS $\mathcal{S}^{\mathrm{CH}}$, pair CHUS $\mathcal{S}^{\mathrm{PCH}}_{p,q}$, and the proposed ICHUS $\mathcal{S}^{\mathrm{ICH}}$ sets:

Proposition 6. *Given a set of historical RES data, sets $\mathcal{S}^{\mathrm{CH}}$, $\mathcal{S}^{\mathrm{ICH}}$, and $\mathcal{S}^{\mathrm{PCH}}$ are characterized by the following inclusion relationship: $\mathcal{S}^{\mathrm{CH}} \subseteq \mathcal{S}^{\mathrm{ICH}} \subseteq \mathcal{S}^{\mathrm{PCH}}_{p,q}, \forall p, q$.*

Proof. Let $\lambda_r(\forall r)$ represent any given RES data point, and (p, q) denote a unit pair. Initially, $\mathcal{S}^{\mathrm{ICH}} \subseteq \mathcal{S}^{\mathrm{PCH}}_{p,q}, \forall p, q$ follows explicitly from (5.12). Then, the inclusion $\mathcal{S}^{\mathrm{CH}} \subseteq \mathcal{S}^{\mathrm{ICH}}, \forall p, q$ is proven for the case of any $\forall r = p, q$, where the outputs of RES units in $\mathcal{S}^{\mathrm{CH}}$ are given by a linear combination of vertices: $\lambda^{\mathrm{CH}}_r = \hat{\lambda}^{\mathrm{CH}}_r + \sum_{m\in\mathcal{M}} \varepsilon^{\mathrm{CH}}_{rm}(\lambda^{\mathrm{CH}}_{rm} - \hat{\lambda}^{\mathrm{CH}}_r)$, which have expressions in $\mathcal{S}^{\mathrm{PCH}}$. Conversely, the outputs of RES units in $\mathcal{S}^{\mathrm{CH}}$ for $\forall r \neq p, q$ are formulated as $\lambda^{\mathrm{CH}}_r = \hat{\lambda}^{\mathrm{CH}}_r + \sum_{m\in\mathcal{M}} \varepsilon^{\mathrm{CH}}_{rm}(\lambda^{\mathrm{CH}}_{rm} - \hat{\lambda}^{\mathrm{CH}}_r) = \sum_{m\in\mathcal{M}} \varepsilon^{\mathrm{CH}}_{rm}\lambda^{\mathrm{CH}}_{rm} \geq \sum_{m\in\mathcal{M}} \varepsilon^{\mathrm{CH}}_{rm}\lambda^{\mathrm{d}}_r = \lambda^{\mathrm{d}}_r$ because $\sum_{m\in\mathcal{M}} \varepsilon_{rm} = 1$ and $0 \leq \varepsilon_{rm} \leq 1, \forall r, m$. Similarly, we can also deduce that $\lambda^{\mathrm{CH}}_r = \sum_{m\in\mathcal{M}} \varepsilon^{\mathrm{CH}}_{rm}\lambda^{\mathrm{CH}}_{rm} \leq \sum_{m\in\mathcal{M}} \varepsilon^{\mathrm{CH}}_{rm}\lambda^{\mathrm{u}}_r = \lambda^{\mathrm{u}}_r$. Hence, $\lambda^{\mathrm{d}}_r \leq \lambda^{\mathrm{CH}}_r \leq \lambda^{\mathrm{u}}_r, \forall r \neq p, q$. Consequently, we can conclude that $\mathcal{S}^{\mathrm{CH}} \subseteq \mathcal{S}^{\mathrm{PCH}}_{p,q}$ for all $1 \leq p < q \leq R$. This establishes that $\mathcal{S}^{\mathrm{CH}} \subseteq \cap_{1\leq p<q\leq R}\mathcal{S}^{\mathrm{PCH}}_{p,q} = \mathcal{S}^{\mathrm{ICH}}$. □

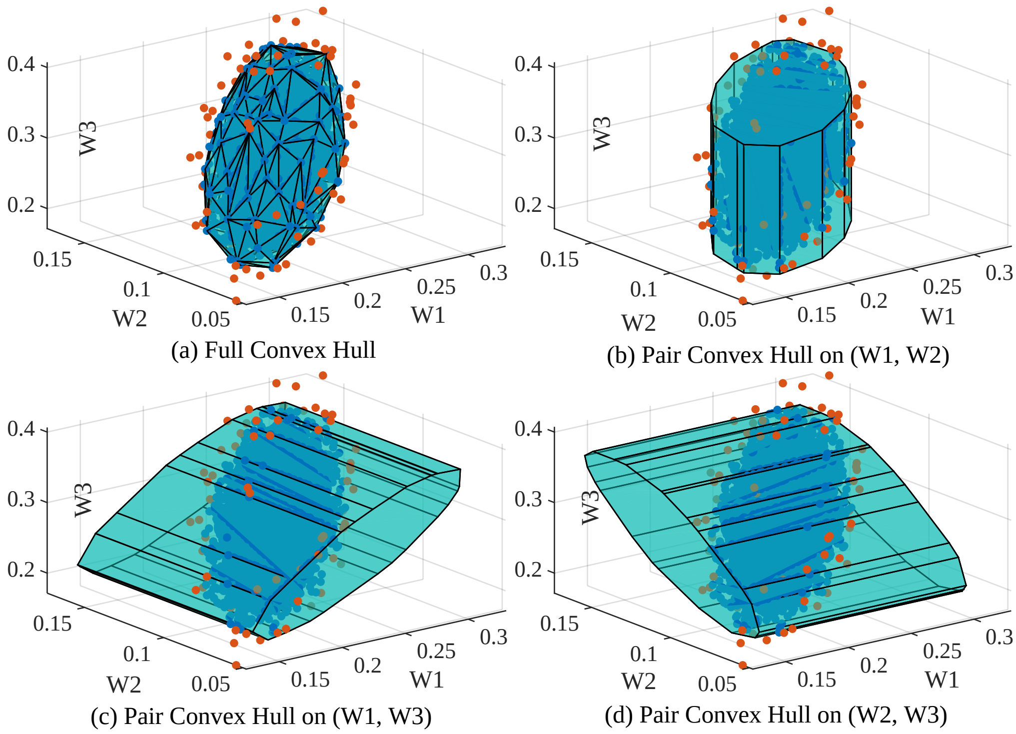


Figure 5.2: Illustrations of the proposed uncertainty sets based on historical RES output data.

Proposition 6 indicates that the proposed ICHUS achieves a favorable balance between the standard CHUS and the pair CHUS. Firstly, the proposed ICHUS includes a much smaller number of vertices than the CHUS, and thereby enhances the computational efficiency of the solution process. Secondly, the ICHUS eliminates a substantial amount of the unnecessary space encompassed by the overly conservative pair CHUS, and thereby decreases the size of the uncertainty set. Consequently, the ICHUS can ensure robustness against RES output uncertainties with improved computational efficiency in a mutually beneficial manner.

Moreover, the proposed ICHUS significantly reduces the inclusion of unnecessary space in RES output uncertainty sets compared with the BbUS and PUS. Therefore, the proposed ICHUS avoids the overly conservative solutions obtained using the BbUS and PUS, while ensuring the robustness of solutions against uncertain RES outputs.

5.4 Robust Energy Management of HMGs

A standard HMG system consists of an ACsM containing diesel engine units, wind turbines, and AC loads, and a DCsM comprising fuel cell units, photovoltaic panels, and DC loads with power flow between these two grids facilitated by BDCs. Accordingly, the HMG energy management model applied herein involves the three levels outlined below.

$$\min_{\boldsymbol{x}}(\vartheta_1 + \max_{\boldsymbol{s}}\min_{\boldsymbol{y}}\vartheta_2) \tag{5.13}$$

$$\vartheta_1 = C^{\mathrm{de}} + C^{\mathrm{fc}} + C^{\mathrm{w}} + C^{\mathrm{pv}} + C^{\mathrm{bdc}} \tag{5.14}$$

$$C^{\mathrm{de}} = \sum_{t=1}^{\mathcal{T}}\sum_{i=1}^{\mathcal{I}}(n_i^{\mathrm{de}}U_{i,t}^{\mathrm{de}} + f_i^{\mathrm{de}}D_{i,t}^{\mathrm{de}} + c_i^{\mathrm{de}}P_{i,t}^{\mathrm{de}}) \tag{5.15}$$

$$C^{\mathrm{fc}} = \sum_{t=1}^{\mathcal{T}}\sum_{i=1}^{\mathcal{I}}(n_i^{\mathrm{fc}}U_{i,t}^{\mathrm{fc}} + f_i^{\mathrm{fc}}D_{i,t}^{\mathrm{fc}} + c_i^{\mathrm{fc}}P_{i,t}^{\mathrm{fc}}) \tag{5.16}$$

$$C^{\mathrm{w}} = \sum_{t=1}^{\mathcal{T}}\sum_{i=1}^{\mathcal{I}}c_i^{\mathrm{w}}\hat{\rho}_{i,t}^{\mathrm{w}} \tag{5.17}$$

$$C^{\mathrm{pv}} = \sum_{t=1}^{\mathcal{T}}\sum_{i=1}^{\mathcal{I}}c_i^{\mathrm{pv}}\hat{\rho}_{i,t}^{\mathrm{pv}} \tag{5.18}$$

$$C^{\mathrm{bdc}} = \sum_{t=1}^{\mathcal{T}}\sum_{l=1}^{\mathcal{L}}c_l^{\mathrm{loss}}(P_{i_l j_l,l,t}^{\mathrm{acdc}} + P_{i_1 j_1,l,t}^{\mathrm{dcac}}) \tag{5.19}$$

$$\vartheta_2 = C^{\mathrm{ac}} + C^{\mathrm{dc}} + \breve{C}^{\mathrm{bdc}} \tag{5.20}$$

$$C^{\mathrm{ac}} = \sum_{t=1}^{\mathcal{T}}\sum_{i=1}^{\mathcal{I}}(c_i^{\mathrm{ac}+}\xi_{i,t}^{\mathrm{ac}+} + c_i^{\mathrm{ac}-}\xi_{i,t}^{\mathrm{ac}-}) \tag{5.21}$$

$$C^{\mathrm{dc}} = \sum_{t=1}^{\mathcal{T}}\sum_{i=1}^{\mathcal{I}}(c_i^{\mathrm{dc}+}\xi_{i,t}^{\mathrm{dc}+} + c_i^{\mathrm{dc}-}\xi_{i,t}^{\mathrm{dc}-}) \tag{5.22}$$

$$\breve{C}^{\mathrm{bdc}} = \sum_{t=1}^{\mathcal{T}}\sum_{l=1}^{\mathcal{L}}c_l^{\mathrm{loss}}(\breve{P}_{i_l j_l,l,t}^{\mathrm{acdc}} + \breve{P}_{i_1 j_1,l,t}^{\mathrm{dcac}}) \tag{5.23}$$

s.t.

$$I_{i,t}^{\mathrm{de}} = I_{i,t-1}^{\mathrm{de}} + U_{i,t}^{\mathrm{de}} - D_{i,t}^{\mathrm{de}}, \forall i,t \tag{5.24}$$

$$I_{i,t}^{\mathrm{de}} \geq \sum_{\tau=1}^{T_i^{\mathrm{de,on}}}U_{i,t-\tau}^{\mathrm{de}}, I_{i,t}^{\mathrm{de}} \leq 1 - \sum_{\tau=1}^{T_i^{\mathrm{de,off}}}D_{i,t-\tau}^{\mathrm{de}}, \forall i,t \tag{5.25}$$

$$I_{i,t}^{\mathrm{de}}\underline{P}_i^{\mathrm{de}} \leq P_{i,t}^{\mathrm{de}} \leq I_{i,t}^{\mathrm{de}}\bar{P}_i^{\mathrm{de}}, \forall i,t \tag{5.26}$$

$$I_{i,t}^{\text{de}}\underline{Q}_{i}^{\text{de}} \leq Q_{i,t}^{\text{de}} \leq I_{i,t}^{\text{de}}\bar{Q}_{i}^{\text{de}}, \forall i,t \tag{5.27}$$

$$-R_{i}^{\text{P,de}} \leq P_{i,t}^{\text{de}} - P_{i,t-1}^{\text{de}} \leq R_{i}^{\text{P,de}}, \forall i,t \tag{5.28}$$

$$-R_{i}^{\text{Q,de}} \leq Q_{i,t}^{\text{de}} - Q_{i,t-1}^{\text{de}} \leq R_{i}^{\text{Q,de}}, \forall i,t \tag{5.29}$$

$$0 \leq \hat{\rho}_{i,t}^{\text{w}} \leq \hat{P}_{i,t}^{\text{w}}, \forall i,t \tag{5.30}$$

$$V_{j,t}^{\text{ac}} = V_{i,t}^{\text{ac}} - \left(r_{ij}^{\text{ac}}H_{ij,t}^{\text{ac}} + x_{ij}^{\text{ac}}G_{ij,t}^{\text{ac}}\right)/V_{0,t}^{\text{ac}}, \forall ij,t \tag{5.31}$$

$$\underline{H}_{ij,t}^{\text{ac}} \leq H_{ij,t}^{\text{ac}} \leq \bar{H}_{ij,t}^{\text{ac}}, \forall ij,t \tag{5.32}$$

$$\underline{G}_{ij,t}^{\text{ac}} \leq G_{ij,t}^{\text{ac}} \leq \bar{G}_{ij,t}^{\text{ac}}, \forall ij,t \tag{5.33}$$

$$\underline{V}_{i,t}^{\text{ac}} \leq V_{i,t}^{\text{ac}} \leq \bar{V}_{i,t}^{\text{ac}}, \forall i,t \tag{5.34}$$

$$\begin{aligned} P_{i,t}^{\text{de}} - L_{i,t}^{\text{ac}} + \sum\nolimits_{j_1 \in \pi(i_1)} \sum\nolimits_{l \in \mathcal{L}} \left(P_{i_1 j_1,l,t}^{\text{ac}} - P_{i_1 j_1,l,t}^{\text{acdc}}\right) \\ = \sum\nolimits_{j \in \pi(i)} H_{ij,t}^{\text{ac}} - \sum\nolimits_{k \in \delta(i)} H_{ki,t}^{\text{ac}} - \left(\hat{P}_{i,t}^{\text{w}} - \hat{\rho}_{i,t}^{\text{w}}\right), \forall i,t \end{aligned} \tag{5.35}$$

$$Q_{i,t}^{\text{de}} - Q_{i,t}^{\text{ac}} = \sum\nolimits_{j \in \pi(i)} G_{ij,t}^{\text{ac}} - \sum\nolimits_{k \in \delta(i)} G_{ki,t}^{\text{ac}}, \forall i,t \tag{5.36}$$

$$I_{i,t}^{\text{fc}} = I_{i,t-1}^{\text{fc}} + U_{i,t}^{\text{fc}} - D_{i,t}^{\text{fc}}, \forall i,t \tag{5.37}$$

$$I_{i,t}^{\text{fc}} \geq \sum\nolimits_{\tau=1}^{T_i^{\text{fc,on}}} U_{i,t-\tau}^{\text{fc}}, I_{i,t}^{\text{fc}} \leq 1 - \sum\nolimits_{\tau=1}^{T_i^{\text{fc,off}}} D_{i,t-\tau}^{\text{fc}}, \forall i,t \tag{5.38}$$

$$I_{i,t}^{\text{fc}}\underline{P}_{i}^{\text{fc}} \leq P_{i,t}^{\text{fc}} \leq I_{i,t}^{\text{fc}}\bar{P}_{i}^{\text{fc}}, \forall i,t \tag{5.39}$$

$$-R_{i}^{\text{fc}} \leq P_{i,t}^{\text{fc}} - P_{i,t-1}^{\text{fc}} \leq R_{i}^{\text{fc}}, \forall i,t \tag{5.40}$$

$$0 \leq \hat{\rho}_{i,t}^{\text{pv}} \leq \hat{P}_{i,t}^{\text{pv}}, \forall i,t \tag{5.41}$$

$$V_{j,t}^{\text{dc}} = V_{i,t}^{\text{dc}} - \left(r_{ij}^{\text{dc}}H_{ij,t}^{\text{dc}}\right)/V_{0,t}^{\text{dc}}, \forall ij,t \tag{5.42}$$

$$\underline{H}_{ij,t}^{\text{dc}} \leq H_{ij,t}^{\text{dc}} \leq \bar{H}_{ij,t}^{\text{dc}}, \forall ij,t \tag{5.43}$$

$$\underline{V}_{i,t}^{\text{dc}} \leq V_{i,t}^{\text{dc}} \leq \bar{V}_{i,t}^{\text{dc}}, \forall i,t \tag{5.44}$$

$$P_{i,t}^{\mathrm{fc}} - L_{i,t}^{\mathrm{dc}} + \sum\nolimits_{j_1\in\pi(i_1)} \sum\nolimits_{l\in\mathcal{L}} \left(P_{i_1j_1,l,t}^{\mathrm{dc}} - P_{i_1j_1,l,t}^{\mathrm{dcac}}\right) = \sum\nolimits_{j\in\pi(i)} H_{ij,t}^{\mathrm{dc}} - \sum\nolimits_{k\in\delta(i)} H_{ki,t}^{\mathrm{dc}} - \left(\hat{P}_{i,t}^{\mathrm{pv}} - \hat{\rho}_{i,t}^{\mathrm{pv}}\right), \forall i,t \tag{5.45}$$

$$0 \le P_{i_1j_1,l,t}^{\mathrm{acdc}} \le I_{l,t}^{\mathrm{acdc}} \bar{P}_{i_1j_1,l}^{\mathrm{acdc}}, \forall l,t \tag{5.46}$$

$$0 \le P_{i_1j_1,l,t}^{\mathrm{dcac}} \le I_{l,t}^{\mathrm{dcac}} \bar{P}_{i_1j_1,l}^{\mathrm{dcac}}, \forall l,t \tag{5.47}$$

$$I_{l,t}^{\mathrm{acdc}} + I_{l,t}^{\mathrm{dcac}} \le 1, \forall l,t \tag{5.48}$$

$$P_{i_1j_1,l,t}^{\mathrm{ac}} = P_{i_1j_1,l,t}^{\mathrm{dcac}} \cdot \eta_{\mathrm{BDC},l}^{\mathrm{dcac}}(P_{i_1j_1,l,t}^{\mathrm{dcac}}), \forall l,t \tag{5.49}$$

$$P_{i_1j_1,l,t}^{\mathrm{dc}} = P_{i_1j_1,l,t}^{\mathrm{acdc}} \cdot \eta_{\mathrm{BDC},l}^{\mathrm{acdc}}(P_{i_1j_1,l,t}^{\mathrm{acdc}}), \forall l,t \tag{5.50}$$

$$P_{i_1j_1,l,t}^{\mathrm{ac}} = o_{0,l} + o_{1,l} P_{i_1j_1,l,t}^{\mathrm{dcac}}, \forall l,t \tag{5.51}$$

$$P_{i_1j_1,l,t}^{\mathrm{dc}} = o_{0,l} + o_{1,l} P_{i_1j_1,l,t}^{\mathrm{acdc}}, \forall l,t \tag{5.52}$$

s.t.

$$\boldsymbol{s} \in \mathcal{S}^{\mathrm{ICH}} \tag{5.53}$$

s.t.

$$I_{i,t}^{\mathrm{de}*} \underline{P}_i^{\mathrm{de}} \le \breve{P}_{i,t}^{\mathrm{de}} \le I_{i,t}^{\mathrm{de}*} \bar{P}_i^{\mathrm{de}} : \underline{\nu}_{i,t}^1, \bar{\nu}_{i,t}^1, \forall i,t \tag{5.54}$$

$$I_{i,t}^{\mathrm{de}*} \underline{Q}_i^{\mathrm{de}} \le \breve{Q}_{i,t}^{\mathrm{de}} \le I_{i,t}^{\mathrm{de}*} \bar{Q}_i^{\mathrm{de}} : \underline{\nu}_{i,t}^2, \bar{\nu}_{i,t}^2, \forall i,t \tag{5.55}$$

$$-R_i^{\mathrm{P,de}} \le \breve{P}_{i,t}^{\mathrm{de}} - \breve{P}_{i,t-1}^{\mathrm{de}} \le R_i^{\mathrm{P,de}} : \underline{\nu}_{i,t}^3, \bar{\nu}_{i,t}^3, \forall i,t \tag{5.56}$$

$$-R_i^{\mathrm{P,de}}\chi \le P_{i,t}^{\mathrm{de}} - \breve{P}_{i,t}^{\mathrm{de}} \le R_i^{\mathrm{P,de}}\chi : \underline{\nu}_{i,t}^4, \bar{\nu}_{i,t}^4, \forall i,t \tag{5.57}$$

$$-R_i^{\mathrm{Q,de}} \le \breve{Q}_{i,t}^{\mathrm{de}} - \breve{Q}_{i,t-1}^{\mathrm{de}} \le R_i^{\mathrm{Q,de}} : \underline{\nu}_{i,t}^5, \bar{\nu}_{i,t}^5, \forall i,t \tag{5.58}$$

$$-R_i^{\mathrm{Q,de}}\chi \le Q_{i,t}^{\mathrm{de}} - \breve{Q}_{i,t}^{\mathrm{de}} \le R_i^{\mathrm{Q,de}}\chi : \underline{\nu}_{i,t}^6, \bar{\nu}_{i,t}^6, \forall i,t \tag{5.59}$$

$$0 \le \xi_{i,t}^{\mathrm{ac}+} \le \lambda_i^{\mathrm{w}} : \underline{\nu}_{i,t}^7, \bar{\nu}_{i,t}^7, \forall i,t \tag{5.60}$$

$$0 \le \xi_{i,t}^{\mathrm{ac}-} \le L_{i,t}^{\mathrm{ac}} : \underline{\nu}_{i,t}^8, \bar{\nu}_{i,t}^8, \forall i,t \tag{5.61}$$

$$\breve{V}_{j,t}^{\mathrm{ac}} = \breve{V}_{i,t}^{\mathrm{ac}} - \left(r_{ij}^{\mathrm{ac}} \breve{H}_{ij,t}^{\mathrm{ac}} + x_{ij}^{\mathrm{ac}} \breve{G}_{ij,t}^{\mathrm{ac}} \right) \Big/ \breve{V}_{0,t}^{\mathrm{ac}} : \upsilon_{ij,t}^{1}, \forall ij, t \tag{5.62}$$

$$\underline{H}_{ij,t}^{\mathrm{ac}} \le \breve{H}_{ij,t}^{\mathrm{ac}} \le \bar{H}_{ij,t}^{\mathrm{ac}} : \underline{\upsilon}_{ij,t}^{2}, \bar{\upsilon}_{ij,t}^{2}, \forall ij, t \tag{5.63}$$

$$\underline{G}_{ij,t}^{\mathrm{ac}} \le \breve{G}_{ij,t}^{\mathrm{ac}} \le \bar{G}_{ij,t}^{\mathrm{ac}} : \underline{\upsilon}_{ij,t}^{3}, \bar{\upsilon}_{ij,t}^{3}, \forall ij, t \tag{5.64}$$

$$\underline{V}_{i,t}^{\mathrm{ac}} \le \breve{V}_{i,t}^{\mathrm{ac}} \le \bar{V}_{i,t}^{\mathrm{ac}} : \underline{\upsilon}_{i,t}^{4}, \bar{\upsilon}_{i,t}^{4}, \forall i, t \tag{5.65}$$

$$\begin{aligned} &\breve{P}_{i,t}^{\mathrm{de}} - L_{i,t}^{\mathrm{ac}} + \sum\nolimits_{j_1 \in \Omega(i_1)} \sum\nolimits_{l \in \mathcal{L}} (\breve{P}_{i_1 j_1, l, t}^{\mathrm{ac}} - \breve{P}_{i_1 j_1, l, t}^{\mathrm{acdc}}) + \xi_{i,t}^{\mathrm{ac}-} \\ &= \sum\nolimits_{j \in \pi(i)} \breve{H}_{ij,t}^{\mathrm{ac}} - \sum\nolimits_{k \in \delta(i)} \breve{H}_{ki,t}^{\mathrm{ac}} - \lambda_i^{\mathrm{w}} + \xi_{i,t}^{\mathrm{ac}+} : \upsilon_{i,t}^{5}, \forall i, t \end{aligned} \tag{5.66}$$

$$\breve{Q}_{i,t}^{\mathrm{de}} - Q_{i,t}^{\mathrm{ac}} = \sum\nolimits_{j \in \pi(i)} \breve{G}_{ij,t}^{\mathrm{ac}} - \sum\nolimits_{k \in \delta(i)} \breve{G}_{ki,t}^{\mathrm{ac}} : \upsilon_{i,t}^{6}, \forall i, t \tag{5.67}$$

$$I_{i,t}^{\mathrm{fc}*} \underline{P}_i^{\mathrm{fc}} \le \breve{P}_{i,t}^{\mathrm{fc}} \le I_{i,t}^{\mathrm{fc}*} \bar{P}_i^{\mathrm{fc}} : \underline{\phi}_{i,t}^{1}, \bar{\phi}_{i,t}^{1}, \forall i, t \tag{5.68}$$

$$-R_i^{\mathrm{fc}} \le \breve{P}_{i,t}^{\mathrm{fc}} - \breve{P}_{i,t-1}^{\mathrm{fc}} \le R_i^{\mathrm{fc}} : \underline{\phi}_{i,t}^{2}, \bar{\phi}_{i,t}^{2}, \forall i, t \tag{5.69}$$

$$-R_i^{\mathrm{fc}} \chi \le P_{i,t}^{\mathrm{fc}} - \breve{P}_{i,t}^{\mathrm{fc}} \le R_i^{\mathrm{fc}} \chi : \underline{\phi}_{i,t}^{3}, \bar{\phi}_{i,t}^{3}, \forall i, t \tag{5.70}$$

$$0 \le \xi_{i,t}^{\mathrm{dc}+} \le \lambda_i^{\mathrm{pv}} : \underline{\phi}_{i,t}^{4}, \bar{\phi}_{i,t}^{4}, \forall i, t \tag{5.71}$$

$$0 \le \xi_{i,t}^{\mathrm{dc}-} \le L_{i,t}^{\mathrm{dc}} : \underline{\phi}_{i,t}^{5}, \bar{\phi}_{i,t}^{5}, \forall i, t \tag{5.72}$$

$$\breve{V}_{j,t}^{\mathrm{dc}} = \breve{V}_{i,t}^{\mathrm{dc}} - \left(r_{ij}^{\mathrm{dc}} \breve{H}_{ij,t}^{\mathrm{dc}} \right) \Big/ \breve{V}_{0,t}^{\mathrm{dc}} : \varphi_{ij,t}^{1}, \forall ij, t \tag{5.73}$$

$$\underline{H}_{ij,t}^{\mathrm{dc}} \le \breve{H}_{ij,t}^{\mathrm{dc}} \le \bar{H}_{ij,t}^{\mathrm{dc}} : \underline{\varphi}_{ij,t}^{2}, \bar{\varphi}_{ij,t}^{2}, \forall ij, t \tag{5.74}$$

$$\underline{V}_{i,t}^{\mathrm{dc}} \le \breve{V}_{i,t}^{\mathrm{dc}} \le \bar{V}_{i,t}^{\mathrm{dc}} : \underline{\varphi}_{i,t}^{3}, \bar{\varphi}_{i,t}^{3}, \forall i, t \tag{5.75}$$

$$\begin{aligned} &\breve{P}_{i,t}^{\mathrm{fc}} - L_{i,t}^{\mathrm{dc}} + \sum\nolimits_{j_1 \in \Omega(i_1)} \sum\nolimits_{l \in \mathcal{L}} (\breve{P}_{i_1 j_1, l, t}^{\mathrm{dc}} - \breve{P}_{i_1 j_1, l, t}^{\mathrm{dcac}}) + \xi_{i,t}^{\mathrm{dc}-} \\ &= \sum\nolimits_{j \in \pi(i)} \breve{H}_{ij,t}^{\mathrm{dc}} - \sum\nolimits_{k \in \delta(i)} \breve{H}_{ki,t}^{\mathrm{dc}} - \lambda_i^{\mathrm{pv}} + \xi_{i,t}^{\mathrm{dc}+} : \varphi_{i,t}^{4}, \forall i, t \end{aligned} \tag{5.76}$$

$$0 \le \breve{P}_{i_1 i_1', l, t}^{\mathrm{acdc}} \le \bar{P}_{i_1 i_1', l}^{\mathrm{acdc}} : \underline{\psi}_{l,t}^{1}, \bar{\psi}_{l,t}^{1}, \forall l, t \tag{5.77}$$

$$0 \le \breve{P}_{i_1 i_1', l, t}^{\mathrm{dcac}} \le \bar{P}_{i_1 i_1', l}^{\mathrm{dcac}} : \underline{\psi}_{l,t}^{2}, \bar{\psi}_{l,t}^{2}, \forall l, t \tag{5.78}$$

$$\breve{P}_{i_1 i_1', l, t}^{\mathrm{acdc}} \cdot \breve{P}_{i_1 i_1', l, t}^{\mathrm{dcac}} = 0, \forall l, t \tag{5.79}$$

$$\breve{P}^{\mathrm{ac}}_{i_1 i'_1,l,t} = o_{0,l} + o_{1,l}\breve{P}^{\mathrm{dcac}}_{i_1 i'_1,l,t}, \forall l, t \tag{5.80}$$

$$\breve{P}^{\mathrm{dc}}_{i_1 i'_1,l,t} = o_{0,l} + o_{1,l}\breve{P}^{\mathrm{acdc}}_{i_1 i'_1,l,t}, \forall l, t \tag{5.81}$$

Here, $\boldsymbol{x} = \{I^{\mathrm{de}}_{i,t}, U^{\mathrm{de}}_{i,t}, D^{\mathrm{de}}_{i,t}, P^{\mathrm{de}}_{i,t}, Q^{\mathrm{de}}_{i,t}, \hat{P}^{\mathrm{w}}_{i,t}, V^{\mathrm{ac}}_{i,t}, H^{\mathrm{ac}}_{ij,t}, G^{\mathrm{ac}}_{ij,t}, I^{\mathrm{fc}}_{i,t}, U^{\mathrm{fc}}_{i,t}, D^{\mathrm{fc}}_{i,t}, P^{\mathrm{fc}}_{i,t}, \hat{\rho}^{\mathrm{pv}}_{i,t}, V^{\mathrm{dc}}_{i,t}, H^{\mathrm{dc}}_{ij,t}, P^{\mathrm{acdc}}_{i_1 j_1,l,t}, P^{\mathrm{dcac}}_{i_1 j_1,l,t}, I^{\mathrm{acdc}}_{l,t}, I^{\mathrm{dcac}}_{l,t}, P^{\mathrm{ac}}_{i_1 j_1,l,t}, P^{\mathrm{dc}}_{i_1 j_1,l,t}\}$, $\boldsymbol{s} = \{\lambda_i\}$, and $\boldsymbol{y} = \{\breve{P}^{\mathrm{de}}_{i,t}, \breve{Q}^{\mathrm{de}}_{i,t}, \xi^{\mathrm{ac}+}_{i,t}, \xi^{\mathrm{ac}-}_{i,t}, \breve{V}^{\mathrm{ac}}_{i,t}, \breve{H}^{\mathrm{ac}}_{ij,t}, \breve{G}^{\mathrm{ac}}_{ij,t}, \breve{P}^{\mathrm{fc}}_{i,t}, \xi^{\mathrm{dc}+}_{i,t}, \xi^{\mathrm{dc}-}_{i,t}, \breve{V}^{\mathrm{dc}}_{i,t}, \breve{H}^{\mathrm{dc}}_{ij,t}, \breve{G}^{\mathrm{dc}}_{ij,t}, \breve{P}^{\mathrm{acdc}}_{i_1 i'_1,l,t}, \breve{P}^{\mathrm{dcac}}_{i_1 i'_1,l,t}, \breve{P}^{\mathrm{ac}}_{i_1 i'_1,l,t}, \breve{P}^{\mathrm{dc}}_{i_1 i'_1,l,t}\}$, respectively represent the decision variables in upper-level model (5.24)-(5.52), middle-level model (5.53), and lower-level model (5.54)-(5.81). The variables or parameters with the hat symbol " $\widehat{*}$ " observed in the constraints of the lower-level model correspond to the variables in the worst-case scenario of the ICHUS, while the variables following the colon in each constraint represent the corresponding Lagrange multipliers for these constraints. The tri-level robust model is a hierarchical framework for energy management. The upper-level model focuses on minimizing the overall operational cost and finding the most cost-effective strategy under the predicted scenario of RES generation, reflecting the decision-making process of the HMG operator. Next, the middle-level model identifies the worst-case scenario for RES fluctuations within the proposed ICHUS framework. Finally, the lower-level model ensures the HMG operates feasibly and optimally, even under the worst-case fluctuation of RES generation. The details of the tri-level model are as follows.

5.4.1 Upper Level

The upper-level model in (5.13) minimizes energy management costs ϑ_1 in (5.14) under the predicted scenario of uncertainties and unbalanced costs in (5.20) under their worst-case scenario, and all individual cost components that make up the sum of energy management costs are depicted in (5.15)-(5.19). For the ACsM, constraint (5.24) presents the relationship between the status and the start-up/shut-down actions of diesel engine units, while constraint (5.25) enforces the minimum time that a diesel engine unit must stay online or offline after being activated or deactivated. Constraints (5.26) and (5.27) define the permissible ranges for the active and reactive power outputs of diesel engine units. The rate at which these power outputs can change is bounded by ramping constraints (5.28) and (5.29). Constraint (5.30) limits the maximum wind turbine power that can be curtailed.

The AC power flow within the ACsM is described in (5.31)-(5.34)based on the LinDistFlow model. The active and reactive power balances are maintained through constraints (5.35) and (5.36), respectively. For the DCsM, (5.37) and (5.38) represent the state transition constraints of fuel cell units, while (5.39) and (5.40) define the active power output and ramping constraints for these units, respectively. Constraint (5.41) sets the allowable power curtailment of photovoltaic units. The DC power flow within the DCsM is given by (5.42)-(5.44) based on the LinDistFlow model. Constraint (5.45) ensures balance between electricity demand and supply. Lastly, for BDCs, constraints (5.46) and (5.47) impose limits on the power conversion capacity of BDCs. Constraint (5.48) enforces the non-simultaneous inverting and rectifying operation constraints of BDCs using non-convex bilinear forms. Equations (5.49) and (5.50) calculate the power after being converted by BDCs. It is worth noting that, although previous studies suggest that the conversion efficiency of BDCs involves complex combinations of trigonometric and polynomial functions, this efficiency can be approximated accurately using inverse forms [17]. Thus, we strike a balance between model accuracy and solution efficiency herein by employing inverse expressions to represent $\eta^{\text{dcac}}_{\text{BDC},l}(\cdot)$ and $\eta^{\text{acdc}}_{\text{BDC},l}(\cdot)$ in (5.49) and (5.50) respectively. This allows us to express $P^{\text{ac}}_{i_1j_1,l,t}$ and $P^{\text{dc}}_{i_1j_1,l,t}$ in (5.49) and (5.50) linearly through (5.51) and (5.52). It is important to note that the conversion efficiency of BDCs in practical applications is less than 100%, meaning the value of $o_{1,l}$ is less than one. As a result, we primarily consider the range $0 < o_{1,l} < 1$ for case analysis. However, cases where $o_{1,l} \geq 1$ can also be extended using similar formulations.

5.4.2 Middle Level

The middle-level model maximizes the worst-case impact of RES fluctuations within the proposed ICHUS. It represents the adversarial nature of RES fluctuations, identifying the most challenging scenario for the HMG to handle.

5.4.3 Lower Level

As presented in (5.20), the lower-level model aims to minimize the adjustment costs required to maintain power balance under the worst-case realization of RES outputs obtained in the middle level, and the adjustment costs are themselves calculated using Eqs. (5.21)–(5.23). Constraints (5.54)-(5.81) are similar to constraints (5.26)-(5.36), (5.39)-(5.45), (5.46)-(5.50), except that they explicitly correspond to the worst-case scenario within the ICHUS. These constraints guarantee that the physical operation limits for various components of the HMG system, such as diesel engine units, fuel cell units, and BDCs, remain within acceptable bounds even under the worst-case scenario.

5.5 Customized Solution Algorithm

The primary challenge for solving the proposed tri-level robust model is that existing algorithms, such as Benders-based algorithms, typical C&CG algorithms, or nested C&CG algorithms, are unable to solve the lower-level model effectively when non-convex constraint (5.79) is incorporated therein. As discussed above, this issue is addressed by first exploring the sufficient conditions enabling the exact relaxation of non-convex constraint (5.79) based on KKT conditions, which transforms the lower-level model into a convex form. Then, a checking loop process is proposed to ensure compliance with a minimum number of those constraints. Finally, an improved C&CG algorithm is designed to avoid becoming trapped in locally optimal solutions by applying multiple vertices of uncertainty sets as starting points.

5.5.1 Sufficient Conditions for a Convex Lower-level Model

As a preliminary, we reformulate the relaxed convex lower-level model as follows:

$$\text{Objective}: \ (5.20) - (5.23) \tag{5.82}$$

$$s.t. \ (5.54) - (5.78), (5.80), (5.81) \tag{5.83}$$

In **Proposition** 7, we demonstrate that relaxing the non-convex constraints (5.79) in the non-convex lower-level model (5.20)–(5.23) and (5.54)–(5.81) does not alter the optimal solution, provided certain conditions are met. Consequently, the solution of the relaxed convex lower-level model (5.82)–(5.83) is equivalent to the solution of the original non-convex lower-level model (5.20)–(5.23) and (5.54)–(5.81).

Proposition 7. *The solution of the lower-level model with* $P^{\mathrm{acdc}}_{i_1 j_1,l,t} \cdot P^{\mathrm{dcac}}_{i_1 j_1,l,t} > 0$ *is suboptimal for* $\forall t \in \mathcal{T}$ *as long as any of the following four conditions is satisfied:*

Cond 1): $I^{\mathrm{de}*}_{i_l,t} \underline{P}^{\mathrm{de}}_{i_1} < \breve{P}^{\mathrm{de}}_{i_1,t} \leq I^{\mathrm{de}*}_{i_l,t} \bar{P}^{\mathrm{de}}_{i_1}$, $-R^{\mathrm{P,de}}_{i_1} < \breve{P}^{\mathrm{de}}_{i_1,t-1} - \breve{P}^{\mathrm{de}}_{i_1,t} < R^{\mathrm{P,de}}_{i_1}$, *and* $-R^{\mathrm{P,de}}_{i_1}\chi \leq P^{\mathrm{de}}_{i_1,t} - \breve{P}^{\mathrm{de}}_{i_1,t} < R^{\mathrm{P,de}}_{i_1}\chi$*;*

Cond 2): $I^{\mathrm{fc}*}_{i_l,t} \underline{P}^{\mathrm{fc}}_{i_1} < \breve{P}^{\mathrm{fc}}_{i_1,t} \leq I^{\mathrm{fc}*}_{i_l,t} \bar{P}^{\mathrm{fc}}_{i_1}$, $-R^{\mathrm{P,fc}}_{i_1} < \breve{P}^{\mathrm{fc}}_{i_1,t-1} - \breve{P}^{\mathrm{fc}}_{i_1,t} < R^{\mathrm{P,fc}}_{i_1}$, *and* $-R^{\mathrm{P,fc}}_{i_1}\chi \leq P^{\mathrm{fc}}_{i_1,t} - \breve{P}^{\mathrm{fc}}_{i_1,t} < R^{\mathrm{P,fc}}_{i_1}\chi$*;*

Cond 3): $0 < \xi^{\mathrm{ac}-}_{i_1,t} \leq L^{\mathrm{ac}}_{i_1,t}$*;*

Cond 4): $0 < \xi^{\mathrm{dc}-}_{i_1,t} \leq L^{\mathrm{dc}}_{i_1,t}$.

Proof. Consider the Lagrangian function ℓ for the relaxed convex lower-level model (5.82) and (5.83). Applying the stationarity requirement from KKT conditions to the relaxed model yields the following.

$$\begin{aligned} \partial\ell/\partial\breve{P}^{\mathrm{acdc}}_{i_1 i'_1,l,t} &= c^{\mathrm{loss}}_l - \underline{\psi}^1_{l,t} + \bar{\psi}^1_{l,t} \\ &- \sum\nolimits_{i_1 \in i_1 i'_1} \upsilon^5_{i_1,t} + \sum\nolimits_{i'_1 \in i_1 i'_1} o_{1,l}\varphi^4_{i'_1,t} = 0 \end{aligned} \tag{5.84}$$

$$\begin{aligned} \partial\ell/\partial\breve{P}^{\mathrm{dcac}}_{i_1 i'_1,l,t} &= c^{\mathrm{loss}}_l - \underline{\psi}^2_{l,t} + \bar{\psi}^2_{l,t} \\ &- \sum\nolimits_{i'_1 \in i_1 i'_1} \varphi^4_{i'_1,t} + \sum\nolimits_{i_1 \in i_1 i'_1} o_{1,l}\upsilon^5_{i_1,t} = 0 \end{aligned} \tag{5.85}$$

Based on complementary slackness requirements in KKT conditions, the constraint where $P^{\mathrm{acdc}}_{i_1 j_1,l,t} \cdot P^{\mathrm{dcac}}_{i_1 j_1,l,t} > 0$ yields the conditions $\underline{\psi}^1_{l,t} = \underline{\psi}^2_{l,t} = 0$ and $\bar{\psi}^1_{l,t}, \bar{\psi}^2_{l,t} \geq 0$. Combining Eqs. (5.84) and (5.85) yields the following.

$$(1+o_{1,l})c^{\mathrm{loss}}_l + \bar{\psi}^1_{l,t} + o_{1,l}\bar{\psi}^2_{l,t} + (o^2_{1,l} - 1)\sum\nolimits_{i_1 \in i_1 i'_1} \upsilon^5_{i_1,t} = 0 \tag{5.86}$$

$$(1+o_{1,l})c^{\mathrm{loss}}_l + o_{1,l}\bar{\psi}^1_{l,t} + \bar{\psi}^2_{l,t} + (o^2_{1,l} - 1)\sum\nolimits_{i'_1 \in i_1 i'_1} \varphi^4_{i'_1,t} = 0 \tag{5.87}$$

These then yield the inequalities $(o^2_{1,l} - 1)\sum_{i_1 \in i_1 i'_1} \upsilon^5_{i_1,t} < 0$ and $(o^2_{1,l} - 1)\sum_{i'_1 \in i_1 i'_1} \varphi^4_{i'_1,t} < 0$. Because $o_{1,l} \in (0,1)$, we know that $\sum_{i_1 \in i_1 i'_1} \upsilon^5_{i_1,t} > 0$ and $\sum_{i'_1 \in i_1 i'_1} \varphi^4_{i'_1,t} > 0$.

Applying the stationarity requirement from KKT conditions to the relaxed model yields the following.

$$\begin{aligned}\partial\ell/\partial\breve{P}^{\mathrm{de}}_{i_1,t} &= -\underline{\nu}^1_{i_1,t} + \bar{\nu}^1_{i_1,t} - \underline{\nu}^3_{i_1,t} + \bar{\nu}^3_{i_1,t} \\ &+ \underline{\nu}^3_{i_1,t+1} - \bar{\nu}^3_{i_1,t+1} + \underline{\nu}^4_{i_1,t} - \bar{\nu}^4_{i_1,t} + \upsilon^5_{i_1,t} = 0\end{aligned} \tag{5.88}$$

$$\begin{aligned}\partial\ell/\partial\breve{P}^{\mathrm{fc}}_{i_1,t} &= -\underline{\phi}^1_{i_1,t} + \bar{\phi}^1_{i_1,t} - \underline{\phi}^2_{i_1,t} + \bar{\phi}^2_{i_1,t} \\ &+ \underline{\phi}^2_{i_1,t+1} - \bar{\phi}^2_{i_1,t+1} + \underline{\phi}^3_{i_1,t} - \bar{\phi}^3_{i_1,t} + \varphi^4_{i_1,t} = 0\end{aligned} \tag{5.89}$$

$$\partial\ell/\partial\xi^{\mathrm{ac}-}_{i_1,t} = c^{\mathrm{ac}-}_{i_1} - \underline{\nu}^8_{i_1,t} + \bar{\nu}^8_{i_1,t} + \upsilon^5_{i_1,t} = 0 \tag{5.90}$$

$$\partial\ell/\partial\xi^{\mathrm{dc}-}_{i_1,t} = c^{\mathrm{dc}-}_{i_1} - \underline{\phi}^5_{i_1,t} + \bar{\phi}^5_{i_1,t} + \varphi^4_{i_1,t} = 0 \tag{5.91}$$

The four conditions in **Proposition** 7 can be proven by contradiction, as outlined below.

1): Based on the complementary slackness of KKT conditions, $\underline{\nu}^1_{i_1,t} = \underline{\nu}^3_{i_1,t} = \bar{\nu}^3_{i_1,t+1} = \bar{\nu}^4_{i_1,t} = 0$, $\bar{\nu}^1_{i_1,t} \geq 0$, $\bar{\nu}^3_{i_1,t} \geq 0$, $\underline{\nu}^3_{i_1,t+1} \geq 0$, and $\underline{\nu}^4_{i_1,t} \geq 0$. Thus, $\upsilon^5_{i_1,t} = -\bar{\nu}^1_{i_1,t} - \bar{\nu}^3_{i_1,t} - \underline{\nu}^3_{i_1,t+1} - \underline{\nu}^4_{i_1,t} \leq 0$, which contradicts $\sum_{i_1 \in i_1 i_1'} \upsilon^5_{i_1,t} > 0$. Therefore, the solution of the lower-level model with $P^{\mathrm{acdc}}_{i_1 j_1,l,t} \cdot P^{\mathrm{dcac}}_{i_1 j_1,l,t} > 0$ must be suboptimal under Condition 1.

2): Similar to the above proof of Condition 1, the complementary slackness of KKT conditions indicates that $\underline{\phi}^1_{i_1,t} = \underline{\phi}^2_{i_1,t} = \bar{\phi}^2_{i_1,t+1} = \bar{\phi}^3_{i_1,t} = 0$, $\bar{\phi}^1_{i_1,t} \geq 0$, $\bar{\phi}^2_{i_1,t} \geq 0$, $\underline{\phi}^2_{i_1,t+1} \geq 0$, and $\underline{\phi}^3_{i_1,t} \geq 0$. Thus, $\varphi^4_{i_1,t} = -\bar{\phi}^1_{i_1,t} - \bar{\phi}^2_{i_1,t} - \underline{\phi}^2_{i_1,t+1} - \underline{\phi}^3_{i_1,t} \leq 0$, which contradicts $\sum_{i_1' \in i_1 i_1'} \varphi^4_{i_1',t} > 0$. Therefore, the solution of the lower-level model with $P^{\mathrm{acdc}}_{i_1 j_1,l,t} \cdot P^{\mathrm{dcac}}_{i_1 j_1,l,t} > 0$ must be suboptimal under Condition 2.

3): Since $0 < \xi^{\mathrm{ac}-}_{i_1,t} \leq L^{\mathrm{ac}}_{i_1,t}$, we can derive that $\underline{\nu}^8_{i_1,t} = 0$ and $\bar{\nu}^8_{i_1,t} \geq 0$. Then, $\upsilon^5_{i_1,t} = -c^{\mathrm{ac}-}_{i_1} - \bar{\nu}^8_{i_1,t} < 0$, which contradicts $\sum_{i_1 \in i_1 i_1'} \upsilon^5_{i_1,t} > 0$. Thus, the solution of the lower-level model with Condition 3 is sub-optimal.

4): Similar to Condition 3, we can derive that $\underline{\phi}^5_{i_1,t} = 0$ and $\bar{\phi}^5_{i_1,t} \geq 0$. Then, $\varphi^4_{i_1,t} = -c^{\mathrm{dc}-}_{i_1} - \bar{\phi}^5_{i_1,t} < 0$. This contradicts to $\sum_{i_1' \in i_1 i_1'} \varphi^4_{i_1',t} > 0$. Thus, the solution of the lower-level model with $P^{\mathrm{acdc}}_{i_1 j_1,l,t} \cdot P^{\mathrm{dcac}}_{i_1 j_1,l,t} > 0$ is suboptimal under Condition 4. □

The above analysis verifies that **Proposition** 7 provides a set of conditions that are sufficient for relaxing non-convex constraint (5.79) in the lower-level model. Specifically, Condition 1 pertains to the diesel engine units connected with BDCs in the ACsM. Analysis indicates that Condition 1 can be divided into three sub-conditions. The first sub-condition asserts that the diesel engine units must operate above their lower power limits. The second sub-condition states that the change in the power outputs of diesel engine units from time periods $t-1$ to t should not reach their lower and upper ramping rates $R_{i_1}^{\mathrm{P,de}}$. The final sub-condition requires that the power reduction of diesel engine units from the day-ahead referenced power $P_{i_1,t}^{\mathrm{de}}$ to the real-time power $\breve{P}_{i_1,t}^{\mathrm{de}}$ should not exceed the adjustable limits. Condition 2 conveys similar constraints to those of Condition 1, but which are applied to fuel cell units in the DCsM. Condition 3 pertains to the occurrence of AC load shedding, while Condition 4 pertains to the occurrence of DC load shedding.

It is noteworthy that load shedding typically occurs in solutions of the lower-level model because the solution is obtained under the worst-case scenario within the ICHUS at this level. This implies that solutions of the lower-level model generally satisfy Conditions 3 and 4. These conditions effectively relax non-convex lower-level model (5.20)-(5.23) and (5.54)-(5.81) into convex lower-level model (5.82)-(5.83), which is amendable to globally optimal solutions using off the shelf commercial solvers, such as Cplex and Gurobi.

5.5.2 Checking Loop Procedure

While sufficient conditions 1)–4) are generally fulfilled in the majority of worst-case operational scenarios, they do not provide a theoretical guarantee that all worst-case scenarios must satisfy at least one of these conditions. Thus, a dynamic checking loop procedure is applied to identify and update the required non-convex constraints dynamically in the solution process, which can minimize the number of non-convex constraints needed to ensure the non-simultaneous rectifying and inverting operations of BDCs. The specific steps of this procedure are executed according to **Algorithm** 6.

Specifically, the checking loop procedure begins by solving a relaxed linear model and calculating the SRII for each BDC and time period to identify violations of the complementary constraint that prevents simultaneous rectifying and inverting operations. When violations are detected ($\mathrm{SRII}_{l,t} \neq 0$), the corresponding indices are added to a dynamic set ($\mathcal{S}^{\mathrm{D}}$) and the complementary constraint is formulated and added to the model. The updated model is then resolved, and SRII is recalculated. This process repeats until no violations remain, ensuring that all necessary constraints are applied while minimizing the computational burden.

Algorithm 6 Dynamic Checking Loop Procedure

Initialize the dynamic set $\mathcal{S}^{\mathrm{D}} = \{l, t\} = \emptyset$
Solve relaxed linear model (5.82), (5.83)
Calculate the value of SRII for each BDC l at each period t:

$$\mathrm{SRII}_{l,t} = P^{\mathrm{acdc}}_{i_1 j_1, l, t} \cdot P^{\mathrm{dcac}}_{i_1 j_1, l, t}, \forall l \in \mathcal{L}, t \in \mathcal{T} \tag{5.92}$$

while $\exists l \in \mathcal{L}, t \in \mathcal{T}, \mathrm{SRII}_{l,t} \neq 0$ **do**
    Add all (l, t) with $\mathrm{SRII}_{l,t} \neq 0$ into dynamic set $\mathcal{S}^{\mathrm{D}}$
    Form complementary constraints:

$$\breve{P}^{\mathrm{acdc}}_{i_1 j_1, l, t} \cdot \breve{P}^{\mathrm{dcac}}_{i_1 j_1, l, t} = 0, \forall (l, t) \in \mathcal{S}^{\mathrm{D}} \tag{5.93}$$

    Solve model (5.82), (5.83) with updated constraints (5.93)
    Calculate the value of $\mathrm{SRII}_{l,t}, \forall l \in \mathcal{L}, t \in \mathcal{T}$
end while

The algorithm's dynamic constraint handling reduces computational complexity by selectively identifying and enforcing only the necessary non-convex constraints during the solution process. Its iterative refinement ensures that all detected violations are resolved while avoiding over-constraining the solution. The approach guarantees physical realizability by preventing simultaneous rectifying and inverting operations of BDCs, ensuring operational feasibility in HMGs. Additionally, its scalability makes it suitable for HMGs with a large number of BDCs and time periods, as it dynamically adjusts constraints without overwhelming the optimization model. This makes the algorithm efficient, accurate, and practical for robust energy management in renewable energy-penetrated HMGs.

5.5.3 Customized C&CG Algorithm Framework

For ease of algorithmic expression, the tri-level robust model presented in Section 5.4 is rewritten in the following matrix form.

$$\min_{\boldsymbol{x}} \left(\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} + \max_{\boldsymbol{s} \in \mathcal{S}^{\mathrm{ICH}}} \min_{\boldsymbol{y} \in \Phi(\boldsymbol{x}, \boldsymbol{s})} \boldsymbol{b}^{\mathrm{T}} \boldsymbol{y} \right) \tag{5.94}$$

$$\boldsymbol{C}\boldsymbol{x} \leq \boldsymbol{c}, \Phi(\boldsymbol{x}, \boldsymbol{s}) = \{\boldsymbol{y} | \boldsymbol{D}\boldsymbol{x} + \boldsymbol{E}\boldsymbol{s} + \boldsymbol{F}\boldsymbol{y} \leq \boldsymbol{d}, \boldsymbol{y} \cdot \boldsymbol{y} = 0\} \tag{5.95}$$

Here, $\boldsymbol{x}$, $\boldsymbol{s}$, and $\boldsymbol{y}$ respectively represent the decision variables in upper-level model (5.13)-(5.52), middle-level model (5.53), and lower-level model (5.54)-(5.81), while $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{C}$, $\boldsymbol{c}$, $\boldsymbol{D}$, $\boldsymbol{d}$, $\boldsymbol{E}$, and $\boldsymbol{F}$

are constant coefficient matrices. It is important to highlight that the lower-level model in (5.94)-(5.95) constitutes a non-convex optimization problem due to the inclusion of the constraint $\boldsymbol{y} \cdot \boldsymbol{y} = 0$, which can easily lead to a local optimal solution. However, as demonstrated in **Proposition** 7, if any one of the four specified conditions is satisfied, the constraint $\boldsymbol{y} \cdot \boldsymbol{y} = 0$ can be removed without compromising accuracy. Consequently, the original non-convex lower-level model (5.55)-(5.81) can be precisely relaxed into a convex lower-level model (5.82)-(5.83) in Subsection 5.5.1, as follows:

$$\begin{gathered} \min_{\boldsymbol{x}} \left(\boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} + \max_{\boldsymbol{s} \in \mathcal{S}^{\mathrm{ICH}}} \min_{\boldsymbol{y} \in \Phi(\boldsymbol{x}, \boldsymbol{s})} \boldsymbol{b}^{\mathrm{T}} \boldsymbol{y} \right) \\ \boldsymbol{C}\boldsymbol{x} \le \boldsymbol{c}, \Phi'(\boldsymbol{x}, \boldsymbol{s}) = \{ \boldsymbol{y} \mid \boldsymbol{D}\boldsymbol{x} + \boldsymbol{E}\boldsymbol{s} + \boldsymbol{F}\boldsymbol{y} \le \boldsymbol{d} \} \end{gathered} \tag{5.96}$$

In the following section, the challenge of solution algorithms becoming trapped in locally optimal solutions owing to the large number of vertices in the ICHUS is addressed by designing a modified C&CG algorithm that leverages multiple vertices as starting points for achieving solutions. Hence, the solution process can be divided into the following master problem and slave problem.

5.5.3.1 Master Problem

The master problem relaxes model (5.96) by iteratively adding active constraints of the lower-level model. Specifically, the master problem is expressed at the K-th iteration as follows:

$$\min_{\boldsymbol{x}, \beta, \boldsymbol{y}^k} \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} + \beta, \tag{5.97}$$

$$\boldsymbol{C}\boldsymbol{x} \le \boldsymbol{c}, \beta \ge 0, \tag{5.98}$$

$$\beta \ge \boldsymbol{b}^{\mathrm{T}} \boldsymbol{y}^k, \boldsymbol{y}^k \in \Phi'(\boldsymbol{x}, \boldsymbol{s}^k), k = 1, 2, \cdots, K-1, \tag{5.99}$$

where $\boldsymbol{y}^k$ is the decision variable of the lower-level model under worst-case scenario $\boldsymbol{s}^k$.

5.5.3.2 Salve Problem

The salve problem is expressed for a given value of $\boldsymbol{x}^{k*}$ in the master problem as follows:

$$\max_{\boldsymbol{s}\in\mathcal{S}^{\text{ICH}}}\min_{\boldsymbol{y}\in\Phi'(\boldsymbol{x}^{k*},\boldsymbol{s})}\boldsymbol{b}^{\text{T}}\boldsymbol{y}, \tag{5.100}$$

$$\Phi'(\boldsymbol{x}^{k*},\boldsymbol{s})=\left\{\boldsymbol{y}|\boldsymbol{D}\boldsymbol{x}^{k*}+\boldsymbol{E}\boldsymbol{s}+\boldsymbol{F}\boldsymbol{y}\leq\boldsymbol{d}\right\}. \tag{5.101}$$

Then, objective (5.100) is reformulated to avoid falling into locally optimal solutions as follows:

$$\max_{\boldsymbol{s}_n\in\mathbb{N}}\max_{\boldsymbol{s}\in\mathcal{S}^{\text{ICH}}}\min_{\boldsymbol{y}\in\Phi'(\boldsymbol{x}^{k*},\boldsymbol{s})}\boldsymbol{b}^{\text{T}}\boldsymbol{y}, \tag{5.102}$$

where $\mathbb{N}=\{\boldsymbol{s}_n,\forall n\in\mathcal{N}\}$ is the set of vertices for the starting points of the algorithm.

A nested block coordinate descent method is developed to solve the tri-level max-max-min problem in (5.102). Here, the inner-loop salve problem in the K'-th iteration is formulated for each given starting point $\boldsymbol{s}_n$ as follows:

$$\Upsilon_{\boldsymbol{s}_n}=\min_{\boldsymbol{s},\boldsymbol{y}}\boldsymbol{b}^{\text{T}}\boldsymbol{y}, \tag{5.103}$$

$$\boldsymbol{D}\boldsymbol{x}^{K'*}+\boldsymbol{E}\boldsymbol{s}+\boldsymbol{F}\boldsymbol{y}\leq\boldsymbol{d}, \tag{5.104}$$

$$\boldsymbol{s}=\boldsymbol{s}_n^{K'}:\boldsymbol{\pi}, \tag{5.105}$$

where $\boldsymbol{\pi}$ in (5.105) denotes the Lagrangian multiplier expressing the sensitivities of the objective in (5.103). After solving the inner-loop salve problem, the inner-loop master problem is stated as follows:

$$\max_{\boldsymbol{s}\in\mathcal{S}^{\text{ICH}}}\Upsilon_{\boldsymbol{s}_n}+(\boldsymbol{\pi}^{K'})^{\text{T}}(\boldsymbol{s}-\boldsymbol{s}^{K'}). \tag{5.106}$$

This problem is modeled based on the first-order Taylor series expansion of the optimal solution obtained in the inner-loop salve problem. After finding the convergent solution of the inner-loop model of the salve problem, the outer-loop model of the salve problem is formed as follows:

$$\max_{\boldsymbol{s}_n\in\mathbb{N}}\Upsilon_{\boldsymbol{s}_n} \tag{5.107}$$

The outer-loop model of the salve problem identifies the worst-case scenario obtained from all starting points, and can therefore escape locally optimal solutions. Notably, the speed of the outer-loop solution process can be accelerated greatly because all inner-loop processes are independent from each other for a given value of $\boldsymbol{x}$, and thus can be solved in parallel.

It should be noted that the solution to model (5.96) is equivalent to that of model (5.94)-(5.95) if any of the four conditions specified in **Proposition** 7 are satisfied. However, in practical applications, particularly under extreme operating scenarios, these conditions may not always hold. In such cases, the non-convex non-simultaneous inverter and rectifier constraint $\boldsymbol{y} \cdot \boldsymbol{y} = 0$ cannot be relaxed without compromising the accuracy of the solution. We notice that, the non-simultaneous inverter and rectifier constraints for BDCs must be satisfied for each BDC and each time period. Consequently, the total number of such constraints is determined by the number of BDCs and the time periods considered. The dynamic checking loop procedure in **Algorithm** 6 is integrated into the solution process of the inner-loop slave problem. This procedure identifies the minimal set of non-convex constraints that need to be included in the inner-loop slave problem to ensure solution accuracy. Using the dynamic set $\mathcal{S}^{\mathrm{D}} = \{l, t\}$ identified in **Algorithm** 6, the inner-loop salve problem (5.103)-(5.105) is modified as follows:

$$
\begin{gathered}
\Upsilon_{\boldsymbol{s}_n} = \min_{\boldsymbol{s},\boldsymbol{y}} \boldsymbol{b}^{\mathrm{T}} \boldsymbol{y} \\
\boldsymbol{D}\boldsymbol{x}^{K'*} + \boldsymbol{E}\boldsymbol{s} + \boldsymbol{F}\boldsymbol{y} \leq \boldsymbol{d}, \quad \boldsymbol{y}_{l,t} \cdot \boldsymbol{y}_{l,t} = 0, \{l,t\} \in \mathcal{S}^{\mathrm{D}} \\
\boldsymbol{s} = \boldsymbol{s}_n^{K'} : \boldsymbol{\pi}
\end{gathered}
\tag{5.108}
$$

Remark 2: For tri-level min-max-min optimization problems where the inner minimization problem involves the non-convex complementary constraint $\boldsymbol{y} \cdot \boldsymbol{y} = 0$, fully relaxed methods [86] and nested C&CG algorithms [155] are commonly used to address this challenge. However, these approaches have notable limitations. On one hand, the fully relaxed method cannot guarantee that the obtained solution will satisfy the complementary constraint under all operating conditions. On the other hand, while the nested C&CG algorithm ensures compliance with the complementary constraint, it imposes a significant computational burden. In contrast, our proposed checking loop procedure minimizes the number of non-convex complementary constraints that need to be enforced

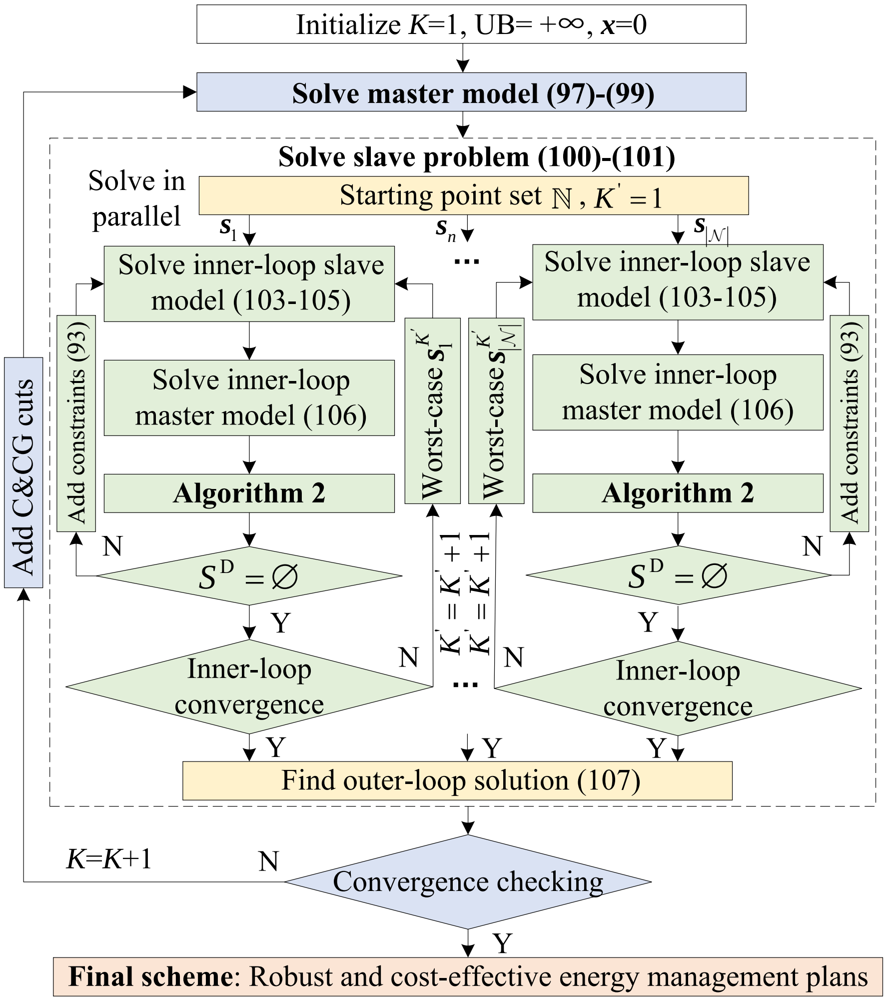


Figure 5.3: Framework of the customized C&CG algorithm.

during the solution process, without sacrificing solution accuracy. As a result, our method reduces the computational burden associated with solving the model while ensuring that the solutions strictly satisfy all constraints.

The algorithm is guaranteed to converge within a finite number of iterations, as the master problem incrementally incorporates active constraints, and the total number of such constraints is finite. Upon convergence, the master problem includes all necessary constraints to accurately represent the worst-case cost. The slave problem ensures that each added constraint corresponds to the true worst-case scenario. By solving the slave problem from multiple starting points and employing a nested block coordinate descent method, the algorithm avoids local optima and enhances its ability to find globally optimal solutions in non-convex cases.

5.6 Case Studies

The efficacy and advantages of the proposed uncertainty set and solution algorithm are demonstrated based on the results of numerical computations conducted with the IEEE 66-bus HMG [84] in conjunction with the corresponding results obtained by existing state-of-the-art methods. The detailed parameters of the HMG system are presented in Table 5.1. The reference values for the system are set with a base voltage of 12.66 kV and a base power of 10 MW. Other parameters of the HMG system used in this work are detailed elsewhere [156]. All experiments were conducted using the CPLEX and IPOPT solvers within the general algebraic modeling system (GAMS) software operating on a personal computer equipped with an Intel® Core™ i7-1260P processor operating at 2.10 GHz, and 16 GB of RAM. The tolerances of all solvers and the C&CG algorithm were set to 0.01%.

Table 5.1: Key parameters of the HMG system

Parameter	Value	Parameter	Value	Parameter	Value
c_i^{de} (¥/MWh)	12000	$\bar{P}_i^{\text{de}}$ (MW)	8	$\bar{P}_{i_1 j_1, l}^{\text{acdc}}$ (MW)	10
c_j^{fc} (¥/MWh)	12000	$\bar{P}_i^{\text{fc}}$ (MW)	8	$\bar{P}_{i_1 j_1, l}^{\text{dcac}}$ (MW)	10
$c_i^{\text{ac}\pm}$ (¥/MWh)	600000	$\underline{P}_i^{\text{de}}$ (MW)	0	$V_{0,t}^{\text{ac/dc}}$ (p.u.)	1.05
$c_i^{\text{dc}\pm}$ (¥/MWh)	600000	$\underline{P}_i^{\text{fc}}$ (MW)	0	$\bar{V}_{i,t}^{\text{ac}}$ (p.u.)	1.05
c_l^{loss} (¥/MWh)	600	$P_{j,t}^{\text{pv}}$ (MW)	10	$\underline{V}_{i,t}^{\text{ac}}$ (p.u.)	0.95
$o_{0,l}$	-0.002	$P_{i,t}^{\text{w}}$ (MW)	10	$\bar{V}_{i,t}^{\text{dc}}$ (p.u.)	1.05
$o_{1,l}$	0.982	$\bar{Q}_i^{\text{de}}$ (MVar)	1.8	$\underline{V}_{i,t}^{\text{dc}}$ (p.u.)	0.95

5.6.1 Analysis of RES Generation Curtailment

To evaluate the effectiveness of our approach, we compared our ICHUS-based energy management approach with the traditional deterministic energy management approach in terms of RES generation curtailment. To this end, we randomly generated 10,000 RES output scenarios, and optimized the RES curtailment for each scenario accordingly. The results, shown in Fig. 5.4, reveal that the deterministic approach (Fig. 5.4(a)) results in a total RES curtailment of 1003.1 p.u., with an average value of 0.1003 p.u. per scenario. This indicates significant underutilization of available renewable energy sources. In contrast, our proposed ICHUS-based energy management approach (Fig. 5.4(b)) demonstrates a substantial improvement in reducing RES curtailment, with a total value of 502.8 p.u. and an average of 0.0503 p.u. per scenario. This represents nearly a 50% reduction in RES

Table 5.2: Comparisons with different uncertainty set (unit:¥)

Uncertainty Set (US)		TC	GC	RC	LC	UBC	NVS
BbUS [79]		1,259,369	573,000	41	774	685,554	0
PUS [84]	$\Gamma = 0.25$	765,680	572,841	33	774	192,031	31
	$\Gamma = 0.50$	947,526	573,101	46	774	373,604	0
	$\Gamma = 0.75$	1,102,826	573,640	74	774	528,338	0
CHUS [95], [98]		901,350	573,000	41	774	327,535	0
ICHUS		901,350	573,000	41	774	327,535	0
CHUS-99% Confidence		789,394	573,000	41	772	215,578	12
ICHUS-99% Confidence		789,561	573,159	49	774	215,578	11
CHUS-95% Confidence		722,250	573,196	51	774	148,229	58
ICHUS-95% Confidence		722,748	573,306	57	774	148,611	57
CHUS-90% Confidence		681,686	573,259	55	774	107,598	127
ICHUS-90% Confidence		681,642	573,225	53	774	107,590	126
CHUS-80% Confidence		638,455	573,558	70	774	64,053	319
ICHUS-80% Confidence		638,712	573,589	72	774	64,277	319

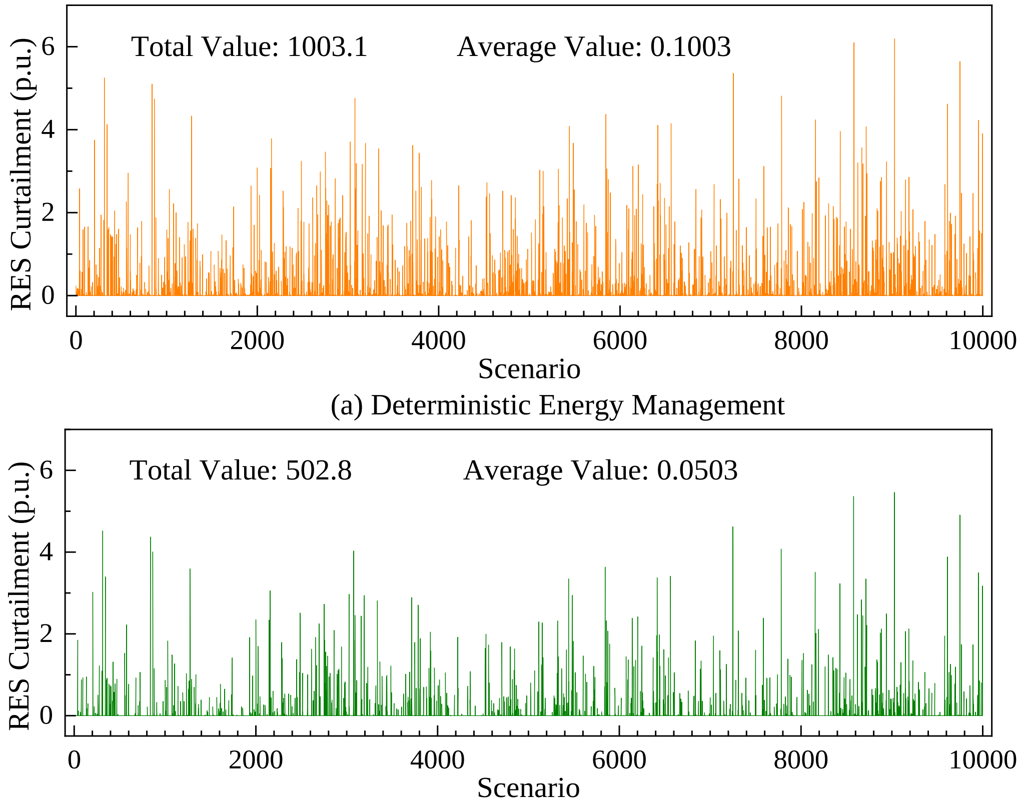
TC: total cost; GC: generation cost; RC: renewable curtailment cost; LC: loss cost on BDCs; UBC: unbalanced cost.

curtailment compared to the deterministic approach, highlighting the effectiveness of our approach in better integrating RES into the energy system.

The variability of RES curtailment is also visibly lower in the ICHUS-based approach, as evidenced by the reduced amplitude and frequency of higher curtailment values in Fig. 5.4(b). This suggests that our approach not only minimizes total curtailment but also stabilizes the energy management process, ensuring more consistent utilization of RES. The significant reduction in RES curtailment achieved by the ICHUS-based approach can be attributed to its ability to handle uncertainties in RES generation and load demands more effectively. By incorporating cost-effective ICHUS-based robust optimization techniques, our approach dynamically adjusts energy dispatch decisions in response to fluctuations in RES availability, thereby maximizing RES utilization and minimizing reliance on conventional energy sources.

5.6.2 Comparisons of Uncertainty Sets

The superiority of the proposed ICHUS was evaluated by comparing the solution results obtained using this uncertainty set with those obtained using the BbUS[79], PUS[84], [86] with different uncertainty budgets Γ of 0.25, 0.50, and 0.75, and the CHUS[95], [98] with various confidence levels



(a) Deterministic Energy Management

(b) Our ICHUS-based Energy Management

Figure 5.4: RES generation curtailment.

of 100% 99%, 95%, 90%, and 80% The obtained results are listed in Table 5.2, including the total cost (TC), generation cost (GC), RES curtailment (RC), loss cost (LC) of BDCs, and unbalanced cost (UC). We further evaluated the robustness of the solutions obtained using the different uncertainty sets against uncertainties under out-of-sample scenarios. To this end, we randomly generated 10,000 RES output scenarios, and optimized the solution for each scenario by fixing the first-stage decision variables accordingly. The results are presented in the last row of Table 5.2 according to the number of violation scenarios (NVS) quantifying the instances in which the constraints of the HMG system were not satisfied for a given dispatch solution due to RES uncertainties.

As expected, the BbUS provides the highest TC value because it employs the most conservative intervals to depict RES output uncertainty. Moreover, we note that the PUS can reduce the TC by adjusting the value of Γ. However, as the TC value decreases with decreasing uncertainty budget Γ, we note that the NVS value increases substantially, indicating that the robustness of the solutions is negatively impacted by the selected value of Γ, as expected. In contrast, the CHUS and proposed ICHUS methods conducted with 100% confidence levels ensure that the constraints of the HMG system are satisfied for 100% of the dispatch solutions (i.e., NVS = 0). Similarly, this is also observed for solutions obtained using the BbUS and the PUS with $\Gamma = 0.75$ and $\Gamma = 0.50$. However, the TC values were decreased significantly by about 28% relative to that obtained with the BbUS, and about 18% and 5% relative to that obtained using the PUS with $\Gamma = 0.75$ and $\Gamma = 0.5$, respectively. We further note that minor reductions in the confidence levels of the CHUS and ICHUS methods from 100% to 99% significantly reduce TC, while increasing NVS only slightly, indicating that the conservativeness of our ICHUS method can be heavily affected by the presence of outliers in the outermost data points. This verifies the necessity of the proposed method for limiting outliers, and demonstrates that the approach provides a flexible tool for dispatchers to adjust the conservativeness of the obtained energy management strategies.

In terms of out-of-sample performance, as shown in Fig. 5.5 and Table 5.2, the proposed ICHUS outperforms BbUS, PUS, and CHUS in feasibility and economic efficiency for robust energy management in hybrid microgrids. While BbUS ensures feasibility (NVS = 0) by using highly conservative intervals, it leads to the highest TC, making it economically impractical. Similarly, PUS reduces TC by lowering the uncertainty budget (e.g., Γ= 0.25), but this significantly increases NVS, reflecting poor robustness in out-of-sample scenarios. In contrast, both CHUS and ICHUS achieve NVS = 0 at 100% confidence while maintaining lower TC compared to BbUS and PUS, making them more cost-effective. Notably, the proposed ICHUS closely matches the out-of-sample performance of CHUS while requiring significantly fewer vertices for constructing the uncertainty set. These findings confirm the feasibility and effectiveness of the proposed ICHUS method.

5.6.3 Verification of Checking Loop Procedure

The effectiveness of the proposed checking loop procedure method was verified by comparing the SRII results obtained from Eq. (5.109) for the HMG energy management model relaxed with and without the proposed checking loop procedure method. The results are presented in Table 5.3. Specif-

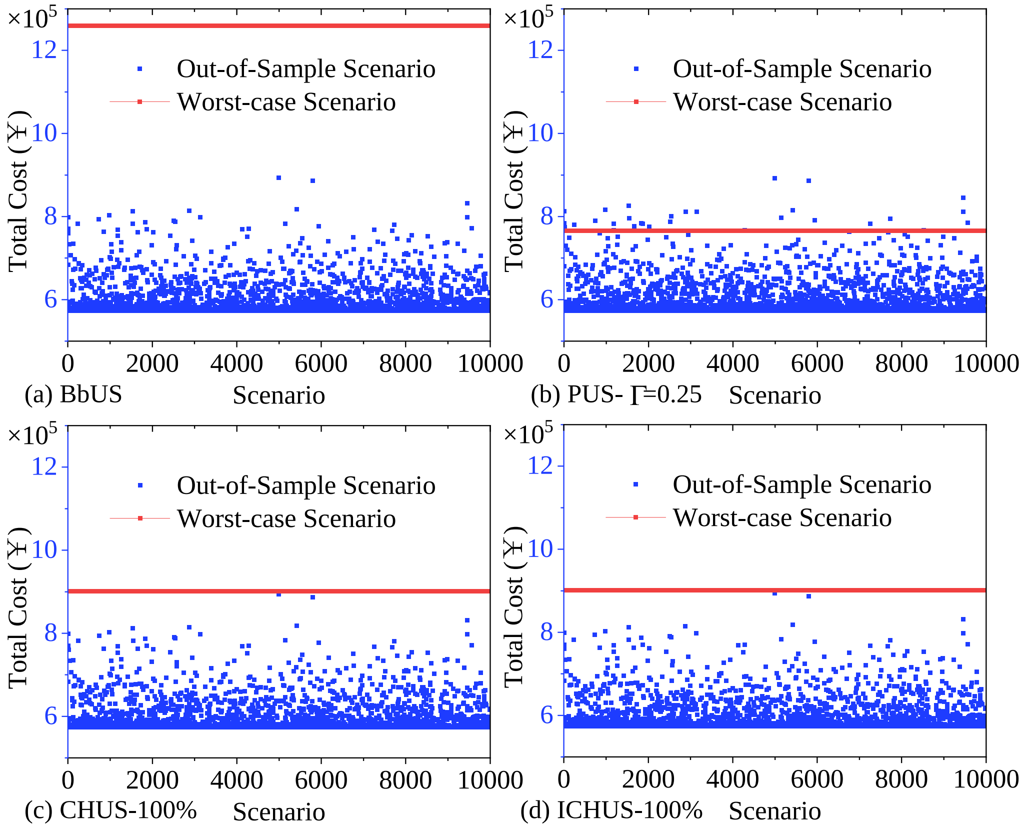


Figure 5.5: Out-of-sample performance of different uncertainty sets.

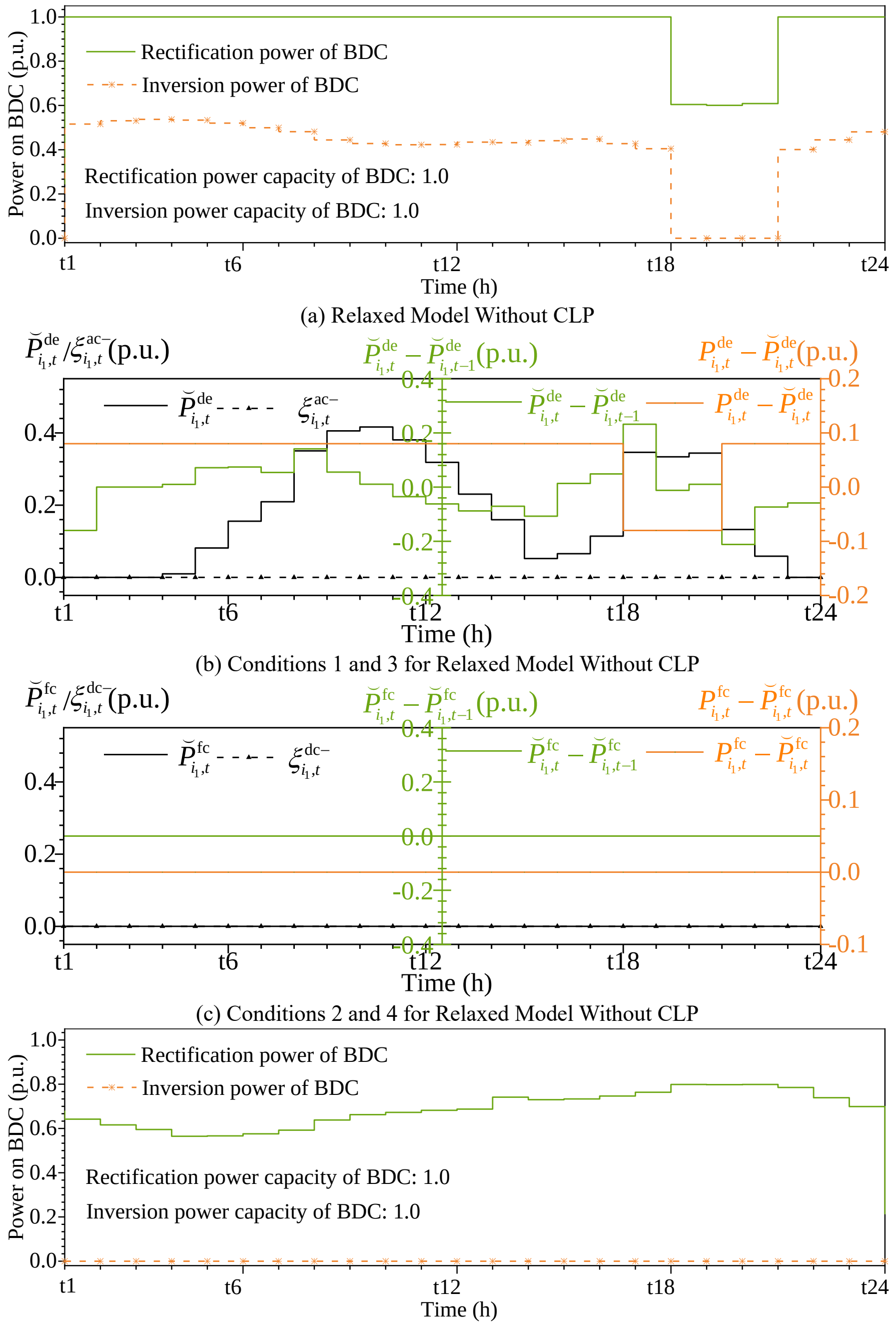
ically, SRII was calculated for all BDCs over the entire time horizon as follows.

$$\mathrm{SRII} = \sum\nolimits_{l \in \mathcal{L}, t \in \mathcal{T}} \left(\breve{P}^{\mathrm{acdc}}_{i_1 i_1', l, t} \cdot \breve{P}^{\mathrm{dcac}}_{i_1 i_1', l, t} \right) \tag{5.109}$$

It can be seen that the model relaxed without the checking loop procedure can achieve a lower objective value because it fails to guarantee non-simultaneous rectifying and inverting operations for BDCs. Hence, the SRII values reside between 9.6 and 10.9. Consequently, solutions obtained without the checking loop procedure violate the physical operation requirements of BDCs, and are therefore impractical for real-world applications. In contrast, the HMG energy management model relaxed with the proposed checking loop procedure can guarantee non-simultaneous rectifier and inverter operations for BDCs, as indicated by SRII values that are exactly equal to zero.

Table 5.3: Verification of the checking loop procedure

Methods	Relaxed Model without Checking Loop Procedure		Relaxed Model with Our Checking Loop Procedure		Exact Model with Full Non-Convex Constraints	
	SRII	Objective	SRII	Objective	SRII	Objective
ICHUS	10.9	881,155	0	901,350	0	901,350
ICHUS-99% Confidence	10.9	776,136	0	789,561	0	789,561
ICHUS-95% Confidence	9.6	710,392	0	722,748	0	722,748
ICHUS-90% Confidence	10.9	669,418	0	681,642	0	681,642
ICHUS-80% Confidence	10.9	626,472	0	638,712	0	638,713



(a) Relaxed Model Without CLP

(b) Conditions 1 and 3 for Relaxed Model Without CLP

(c) Conditions 2 and 4 for Relaxed Model Without CLP

(d) Relaxed Model with CLP or Exact Non-convex Model

Figure 5.6: Rectifying and inverting behaviors of BDCs in different models.

5.6.4 Analysis of Rectifying and Inverting Operations of BDCs

The effectiveness of the sufficient conditions explored in **Proposition** 7 are verified by analyzing the rectifying and inverting operations of BDCs with and without the checking loop procedure. The results obtained over a 24-hour period from t1 to t24 are illustrated in Fig. 5.6. We note from Fig. 5.6(a) that non-simultaneous rectifier and inverter operations of BDCs can be guaranteed during periods t1 and t19–t21 because Condition 1 in **Proposition** 7 is satisfied during these periods. In contrast, the results in Fig. 5.6(a) demonstrate that the relaxed model without the checking loop procedure leads to simultaneous rectifying and inverting operations for BDCs during the periods t2–t18 and t22–t24. This is because none of the conditions in **Proposition** 7 are satisfied, as shown in Fig. 5.6(b) and 5.6(c). In comparison, the results in Fig. 5.6(d) demonstrate that the proposed checking loop procedure can fully ensure non-simultaneous rectifying and inverting operations for BDCs in all time periods, which verifies the effectiveness of the proposed approach.

5.6.5 Verifications of the Developed Nested Block Coordinate Descent Method

The performance of the nested block coordinate descent method in the modified C&CG algorithm is demonstrated by comparing the solutions obtained by the nested block coordinate descent method with those obtained using the big-M method [86] in typical C&CG algorithms at confidence levels of 99%, 95%, 90%, and 80%. Since the big-M method requires that the vertices of uncertainty sets can be explicitly represented, the big-M method cannot be employed to handle our proposed ICHUS with implicit vertices. Thus, we use CHUS with explicit vertices for this case comparison, and the results are summarized in Table 5.6. The nested block coordinate descent method consistently achieves higher objective values than the big-M method, demonstrating its superior ability to identify global worst-case scenarios within uncertainty sets. Additionally, the proposed method significantly reduces computational time. For instance, at a 99% confidence level, the proposed method achieves an objective value of ¥776,121 in only 89.7 seconds, whereas the big-M method requires 20,516.4 seconds to reach its solution. This trend persists across all confidence levels, with the proposed method consistently reducing computation times by an order of magnitude or more. These results clearly highlight the superiority of the nested block coordinate descent method in terms of both solution quality and computational efficiency.

Table 5.4: Verifications of the nested block coordinate descent method

Confidence Level	Big-M Method [86]		Our Method	
	Objective Value (¥)	Computing Time (s)	Objective Value (¥)	Computing Time (s)
99%	773,751	20,516.4	776,121	89.7
95%	701,122	5,011.9	710,581	227.4
90%	657,136	14,153.2	668,901	660.7
80%	622,204	10,055.4	626,807	1,683.1

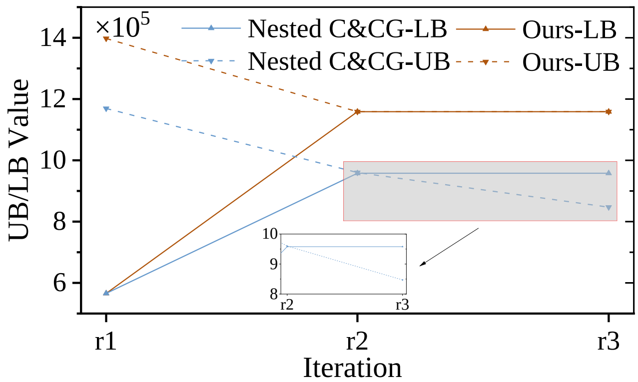


Figure 5.7: Convergence process of nested C&CG algorithm and the customized algorithm.

5.6.6 Comparisons with Nested C&CG Algorithms

To demonstrate the superiority of the proposed modified C&CG algorithm with the checking loop procedure for solving the three-level robust energy management model, the nested C&CG algorithm was also used as a benchmark for comparison [155]. Experiments were conducted to evaluate both methods in terms of objective values, computational time, and NVS for systems with three RES units and seven RES units, assessing their applicability across varying RES unit scales. The results are summarized in Table 5.5.

The results show that the proposed algorithm consistently achieves higher objective values than the nested C&CG algorithm. This indicates the ability of the proposed method to identify more adverse worst-case scenarios and effectively conduct a global search for such scenarios. This advantage becomes particularly significant with larger systems, such as the case of seven RES units, where the nested C&CG algorithm produces objective values up to 17% lower than those of the proposed algorithm. The cause of this difference is illustrated in Fig. 5.7, which shows the iteration process of the upper and lower bounds for the seven RES unit case at an 80% confidence level. In the nested C&CG algorithm, subproblems prematurely generate lower upper bounds compared to those derived from the master problem, as they fail to identify the true worst-case scenarios, leading to premature termination of the iterations. In terms of computational time, the nested C&CG algorithm is significantly slower, taking 7.6 to 717.5 times longer than the proposed algorithm. This highlights a considerable efficiency advantage of the proposed method. Additionally, during robustness testing, the nested C&CG algorithm fails in more scenarios compared to the proposed algorithm. This is due to its inability to consistently find optimal worst-case scenario solutions, a disparity that becomes even more pronounced in the case of seven RES units. Overall, the proposed algorithm demonstrates clear superiority over the nested C&CG algorithm, not only in terms of solution accuracy but also in computational efficiency.

5.6.7 Verifications of the Proposed ICHUS

The computational efficiency of the proposed ICHUS is assessed by comparing its solutions to those of CHUS at confidence levels of 99%, 95%, 90%, and 80%. This comparison is conducted using seven RES units, with the results summarized in Table 5.6. First, when comparing the objective function values of CHUS and ICHUS, the optimal objective values obtained with the proposed ICHUS method are consistently 2.5% to 7.8% higher across different confidence levels. While the ICHUS may produce slightly more conservative solutions, the difference remains acceptable for practical applications. In terms of construction efficiency, the proposed ICHUS offers a significant advantage. The time required to generate the ICHUS ranges from 0.3 to 0.4 seconds, which is approximately one-thousandth of the time needed to construct the CHUS (238.4 to 443.4 seconds). These results demonstrate that the proposed ICHUS closely matches the performance of CHUS in terms of robustness while significantly enhancing computational efficiency. This confirms the effectiveness and superiority of the proposed ICHUS method.

Table 5.5: Comparisons with of the nested C&CG algorithm

RES Number	Confi-dence Level	Nested C&CG Algorithm [155]			Our Method		
		Objective Value (¥)	Computing Time (s)	NVS	Objective Value (¥)	Computing Time (s)	NVS
3	99%	789,561	98,300	17	789,394	137	17
	95%	722,063	53,102	57	722,250	452	58
	90%	676,353	27,710	164	681,686	1,303	147
	80%	634,800	11,078	366	638,455	1,455	330
7	99%	1,374,408	30,888	11	1,404,970	1,176	8
	95%	1,215,454	28,865	67	1,386,483	1,162	40
	90%	1,215,454	13,092	57	1,324,519	1,139	42
	80%	957,955	61,326	557	1,158,632	867	107

Table 5.6: Comparisons between CHUS and the proposed ICHUS

Uncertainty Set	Objective Value (¥)	Computing Time (s)		
		Generating	Solving	Total
CHUS-99% Confidence	1,404,970	238.4	1176.0	1414.4
ICHUS-99% Confidence	1,459,466	0.3	321.5	321.8
CHUS-95% Confidence	1,386,483	355.7	1162.3	1518.0
ICHUS-95% Confidence	1,421,231	0.4	231.7	232.1
CHUS-90% Confidence	1,324,519	412.3	1139.1	1551.4
ICHUS-90% Confidence	1,405,380	0.3	140.9	141.2
CHUS-80% Confidence	1,158,632	443.4	866.7	1310.1
ICHUS-80% Confidence	1,257,305	0.4	160.5	160.9

5.6.8 Impacts of the Number of RES Units on the Model Solution

To analyze the impact of the number of RES units on the model solution, we compared the TC and computational time of solving the model using our customized C&CG algorithms under different numbers of RES units and confidence levels. The results are shown in Fig. 5.8. As the number of RES units increases, the TC generally rises. This indicates that integrating more RES units into the HMG system leads to higher operational costs, primarily due to the increased uncertainty in RES generation. Similarly, the TC increases with higher confidence levels, reflecting a more conservative solution designed to enhance robustness against RES uncertainty. Our confidence-level-adjustable approach offers HMG operators a flexible tool to balance solution conservativeness based on their operational experience.

In addition, the computational time also increases with the number of RES units. This is because the number of pair CHUSs follows the relationship $C_R^2 = R(R-1)/2$ for R RES units, as discussed

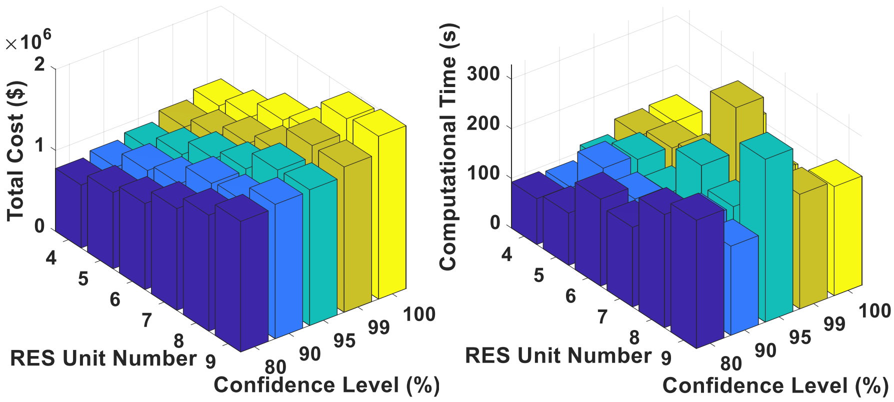


Figure 5.8: Impact of RES unit number on model solution.

in Section 5.3.2. Consequently, as more RES units are included, the number of pair CHUSs grows, leading to an increase in the vertex count of the ICHUS and, therefore, a heavier computational burden. Although the computational time rises from an average of 115 seconds to 243 seconds as the number of RES units increases from four to nine, this level of computational effort remains acceptable for the practical day-ahead energy management of HMGs.

5.7 Summary

The present work addressed the severe challenges associated with the output uncertainty of massive RES units and the non-convex characteristics of BDC modeling in the energy management of HMGs with high RES penetration. To address these challenges, an ICHUS with adjustable confidence levels was developed to eliminate extreme worst-case scenarios while balancing the compactness of the uncertainty set with computational efficiency. Additionally, a series of sufficient conditions were explored to relax the non-convex operational constraints of BDCs. The checking loop procedure was also proposed to ensure the complete satisfaction of all non-convex constraints. Finally, a parallel iterative C&CG-based solution algorithm was developed in conjunction with the checking loop procedure to solve the robust energy management model of HMGs efficiently when including massive numbers of vertices in the ICHUS, and the non-convex constraints of BDCs. The results of case studies demonstrated that the developed confidence level solution method can significantly reduce the conservativeness of the obtained strategies, while maintaining robust performance against the most severe scenarios of RES output uncertainty. Furthermore, the constructed ICHUS closely matched the performance of a standard CHUS but with much greater computational efficiency. Additionally, the explored sufficient conditions and a designed checking loop procedure were demonstrated to ensure non-simultaneous rectifier and inverter operations for BDCs exactly. Finally, the modified C&CG algorithm was demonstrated to achieve globally optimal solutions with a notable decrease in computation time.

By introducing the ICHUS and minimizing the number of non-convex constraints, the proposed method significantly reduced the computational burden while ensuring the accuracy of the obtained solution, making it applicable for large-scale microgrid systems or those with numerous RES units.

Chapter 6

Data-Driven Pair Convex Hull Modeling of Renewable Energy Sources for Microgrid Operation

6.1 Typical Robust Dispatch for Hybrid Microgrids

Typical REMM for HMGs aims to determine an optimal solution that can be hedged against any RES fluctuations within defined uncertainty sets [157]. The matrix form is presented as follows:

$$\min_{\boldsymbol{x}} \boldsymbol{a}^T \boldsymbol{x} + \max_{\boldsymbol{u}} \min_{\boldsymbol{y}} \boldsymbol{b}^T \boldsymbol{y} \tag{6.1}$$

$$\mathrm{s.t.Cx} \leq \mathrm{c}, \mathrm{x} \in \{0, 1\} \tag{6.2}$$

$$\boldsymbol{u} \in \mathcal{U} \tag{6.3}$$

$$\boldsymbol{D}\boldsymbol{x} + \boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{u} \tag{6.4}$$

Here, $\boldsymbol{x}$, $\boldsymbol{u}$ and $\boldsymbol{y}$ represent the decision variables of the first, second, and third-level models in energy management, respectively. $\mathcal{U}$ signifies the uncertainty set encompassing all possible realizations of RES units. For detailed expressions of the model, readers can refer to [158]. The first-level model determines the day-ahead on/off states of controllable units, aiming to make the HMG system as robust as possible against the worst-case scenario within RES uncertainty sets. Given a first-level solution, the second-level model identifies the worst-case fluctuation within the uncertainty sets. Finally, the third-level model minimizes the unbalanced cost under this worst-case scenario.

The size and shape of uncertainty sets directly influence both the conservativeness and computational efficiency of REMM solutions, which can be classified into four types based on their size and shape: IUSs $\mathcal{U}^{\mathrm{I}}$, BIUSs $\mathcal{U}^{\mathrm{BI}}$, EUSs $\mathcal{U}^{\mathrm{E}}$, and CHUSs $\mathcal{U}^{\mathrm{CH}}$. The relationships between these uncertainty sets for a scenario with two RES units are illustrated in Fig. 5.1. Their specific expressions can be presented as:

$$\mathcal{U}^{\mathrm{I}}=\left\{ u_r \left| \begin{array}{l} u_r = \hat{u}_r+\overline{\varepsilon}_r(\overline{u}_r - \hat{u}_r)+\underline{\varepsilon}_r(\underline{u}_r - \hat{u}_r) \\ \overline{\varepsilon}_r + \underline{\varepsilon}_r \leq 1, \overline{\varepsilon}_r, \underline{\varepsilon}_r \in [0,1], \forall r \end{array} \right. \right\} \tag{6.5}$$

$$\mathcal{U}^{\mathrm{BI}}=\left\{ u_r \left| \begin{array}{l} u_r = \hat{u}_r+\overline{\varepsilon}_r(\overline{u}_r - \hat{u}_r)+\underline{\varepsilon}_r(\underline{u}_r - \hat{u}_r) \\ \overline{\varepsilon}_r + \underline{\varepsilon}_r \leq 1, \overline{\varepsilon}_r, \underline{\varepsilon}_r \in [0,1], \forall r \\ \sum_{r=1}^{|\mathcal{R}|} (\overline{\varepsilon}_r + \underline{\varepsilon}_r) \leq \Gamma \end{array} \right. \right\} \tag{6.6}$$

$$\mathcal{U}^{\mathrm{E}}=\left\{ \boldsymbol{u} \left| (\boldsymbol{u}-\hat{\boldsymbol{u}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{u}-\hat{\boldsymbol{u}}) \leq \Gamma \right. \right\} \tag{6.7}$$

$$\mathcal{U}^{\mathrm{CH}}=\left\{ u_r \left| \begin{array}{l} u_r = \hat{u}_r+\sum_{d\in\mathcal{D}} \varepsilon_{rd}(u_{rd}-\hat{u}_r), \forall r \\ 0 \leq \varepsilon_{rd} \leq 1, \forall r, d \\ \sum_{d\in\mathcal{D}} \varepsilon_{rd} = 1, \forall r \end{array} \right. \right\} \tag{6.8}$$

Here, u_r represents the generation output of the r-th RES unit, and $\boldsymbol{u} = [u_1, ..., u_{|\mathcal{R}|}]$ is its vector form (where $|\mathcal{R}|$ denotes the number of RES units). $\hat{u}_r$ signifies the predicted output value of the r-th RES unit, while $\overline{u}_r$ and $\underline{u}_r$ represent the upper and lower fluctuation power of RES outputs, respectively. $\overline{\varepsilon}_r$ and $\underline{\varepsilon}_r$ are binary variables indicating the upper and lower fluctuation direction of RES outputs. ε_{rd} represents the weighting coefficient for the d-th vertex of the convex hull associated with the r-th RES unit, and u_{rd} is the historical fluctuation output value at the d-th vertex. The set $\mathcal{D}$ contains all vertices defining the convex hull, ensuring ε_{rd} forms a valid convex combination ($0 \leq \varepsilon_{rd} \leq 1$, $\sum_{d\in\mathcal{D}} \varepsilon_{rd} = 1$). Γ is the budget parameter that limits the allowable fluctuations of RES outputs, thereby controlling the conservativeness of the solution in BIUS and EUS. $\boldsymbol{\Sigma}$ represents the covariance matrix that characterizes the correlation degree of multivariate outputs of RES units.

Unlike IUS and BIUS, which do not account for the inherent correlations among RES outputs, EUS and CHUS incorporate these correlations into their models. However, it is crucial to note that EUS and CHUS achieve this incorporation either through non-linear representations or by using a high number of vertices, particularly when dealing with numerous RES units. In the subsequent section, the construction of a conservativeness-controllable PCHUS is introduced.

6.2 Uncertainty Set Construction with Confidence Level Approach

This section introduces two key developments. First, a confidence level method that leverages a GNN is proposed to flexibly control the size of RES data-driven uncertainty sets. Subsequently, a PCHUS is designed to significantly reduce the number of vertices required for constructing uncertainty sets. This approach aims to mitigate the computational complexity associated with an increasing number of RES units.

6.2.1 Confidence Level Approach to Immune Outliers

Inspired by recent advancements in GNNs [159], the RES uncertainty dataset is represented as a graph structure. In this structure, each node corresponds to a RES data point, with directed edges connecting each target node to its nearest neighbors. This arrangement enables the network to compute an anomaly score for each target node by integrating information from its neighboring nodes. Fig. 6.1 illustrates the framework of the employed GNN. The learning process in this network is based on a message-passing scheme that consists of four key steps:

6.2.1.1 Graph Representation

To begin, a graph $G = (\Lambda, E)$, where Λ represents the set of nodes, is defined. This graph is composed of nodes (Λ) that represent individual RES data points, interconnected by directed edges (E). Each edge (j, i) denotes a relationship between two nodes, j and i, indicating their connection within

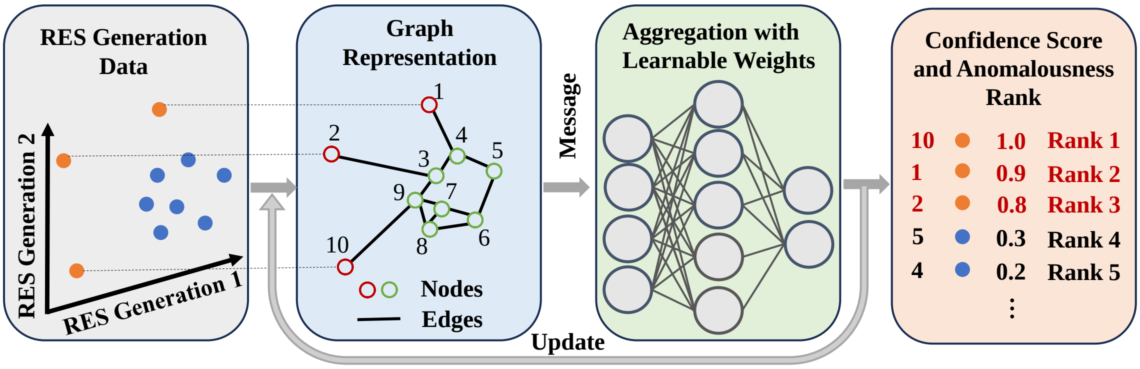


Figure 6.1: Framework of the employed GNN

the graph structure. Both nodes (i) and edges (j, i) are characterized by their respective feature vectors.

More specifically, the feature vector $\boldsymbol{h}_i^{(0)}$ of each node is represented by a RES output point $\boldsymbol{ui}$. Meanwhile, the feature vector of an edge $\boldsymbol{e}_{j,i}$ is derived from the Euclidean distance between the target node i and its source node j, where j is one of the k-nearest neighbors of node i. This process is computed as follows:

$$\boldsymbol{h}_i^{(0)} = \boldsymbol{u}_i \tag{6.9}$$

$$\boldsymbol{e}_{j,i} = \begin{cases} d\left(\boldsymbol{u}_i, \boldsymbol{u}_j\right) = \left\|\boldsymbol{u}_i - \boldsymbol{u}_j\right\|_2, j \in \mathcal{N}_i. \\ 0, j \notin \mathcal{N}_i. \end{cases} \tag{6.10}$$

Here, $\mathcal{N}_i$ represents the set of neighbors for node i. For any two nodes in the graph that are not neighbors, their edge feature is set to 0, signifying the absence of direct message passing between them. The hidden representation of a node in the l-th layer of the GNN is calculated using the following formula:

$$\begin{aligned} \boldsymbol{h}_{\mathcal{N}_i}^{(l)} &= \square_{j\in\mathcal{N}_i}\phi^{(l)}\left(\boldsymbol{h}_i^{(l-1)}, \boldsymbol{h}_j^{(l-1)}, \boldsymbol{e}_{j,i}\right) \\ \boldsymbol{h}_i^{(l)} &= \gamma^{(l)}\left(\boldsymbol{h}_i^{(l-1)}, \boldsymbol{h}_{\mathcal{N}_i}^{(l)}\right) \end{aligned} \tag{6.11}$$

where $\phi^{(l)}$ is the message function that determines the information a node shares with its neighbors. After receiving messages from all neighbors, the aggregation function $\square_{j\in\mathcal{N}_i}$ consolidates these messages into a single, condensed form. The update function $\gamma^{(l)}$ then combines this summarized message with the current information of the node, thereby updating the understanding of the node in the graph structure.

6.2.1.2 Message Passing

Within this graph-based communication framework, the message transmitted from a source node j to a target node i along their connecting edge (j, i) is essentially the distance between their corresponding data points. This distance, encapsulated by the edge feature $\boldsymbol{e}_{j,i}$, constitutes the content of the message and can be expressed as:

$$\phi^{(l)} := \boldsymbol{e}_{j,i}. \tag{6.12}$$

6.2.1.3 Aggregation

To aggregate messages from neighboring nodes, a flexible aggregation function is implemented. Specifically, these messages are consolidated by concatenating them into a single k-dimensional vector $\boldsymbol{e}_i$. Each element within this vector represents the distance between the target node $\boldsymbol{ui}$ and one of its k nearest neighbors.

$$\boldsymbol{e}_i := [\boldsymbol{e}_{1,i}, \ldots, \boldsymbol{e}_{k,i}] \in \mathbb{R}^k \tag{6.13}$$

This k-dimensional vector $\boldsymbol{e}_{i,t}$, which captures the distances to neighboring nodes, is subsequently converted into a scalar anomaly score for node i. This conversion is accomplished through a neural network, denoted as $\mathcal{F}$, which employs learnable weights represented by Θ.

$$\boldsymbol{h}_{\mathcal{N}_i}^{(l)} := \mathcal{F}\left(\boldsymbol{e}_i, \Theta\right), \tag{6.14}$$

where the output of this neural network, denoted as $\boldsymbol{h}_{\mathcal{N}_i}^{(l)}$, quantifies the degree of anomalousness for node i. This adaptive approach is well-suited to RES data, enabling the network to maintain high performance and robustness across varying values of k (the number of neighbors considered in the analysis).

6.2.1.4 Update

In the update phase, each node refines its representation of the graph. This refined representation transcends the initial features of the node, incorporating aggregated information from its neighbors. The fusion of the original features of the node with this newly acquired information yields a more comprehensive and nuanced understanding of the graph structure. This update process can be formulated as follows:

$$\boldsymbol{h}_i^{(l)} = \gamma^{(l)} := \boldsymbol{h}_{\mathcal{N}_i}^{(l)}. \tag{6.15}$$

Based on a given confidence level, the proposed algorithm is designed to generate an outlier score for each data point. This scoring mechanism is constructed such that when $\boldsymbol{u}_i$ represents normal data, the score is low, and conversely, when $\boldsymbol{u}_i$ represents outlier data, the score is high.

Given a dataset of RES generation $\{\boldsymbol{u}_i\}_{i=1}^{|\mathcal{I}|}$, the anomaly score at a RES data point $\boldsymbol{u}$ is defined as $s\left(\boldsymbol{u}; \{\boldsymbol{u}_i\}_{i=1}^{|\mathcal{I}|}\right)$. From this, the following proposition regarding RES data transformation can be formulated:

Proposition 8. *The anomaly score generated by the employed method remains invariant under rotation, translation, or reflection applied uniformly to all data points in the RES generation dataset:*

$$s\left(\boldsymbol{u}; \{\boldsymbol{u}_i\}_{i=1}^{|\mathcal{I}|}\right) = s\left(f(\boldsymbol{u}); \{f\left(\boldsymbol{u}_i\right)\}_{i=1}^{|\mathcal{I}|}\right) \tag{6.16}$$

Proof. For any given two points of RES data $\boldsymbol{u}_i = (u_{i,1}, ..., u_{i,|\mathcal{R}|})$ and $\boldsymbol{u}_j = (u_{j,1}, ..., u_{j,|\mathcal{R}|})$, it is proved that the Euclidean distance between these points remains invariant under translations, rotations, and reflections. Let $\boldsymbol{u}_i' = (u_{i,1}', ..., u_{i,|\mathcal{R}|}')$ and $\boldsymbol{u}_j' = (u_{j,1}', ..., u_{j,|\mathcal{R}|}')$ represent the transformed points after applying these operations.

i) Translations: Let $\boldsymbol{a} = (a_1, ..., a_{|\mathcal{R}|})$ be a translation vector. The transformed points are given by $\boldsymbol{u}_i' = f(\boldsymbol{u}_i) = \boldsymbol{u}_i + \boldsymbol{a}$ and $\boldsymbol{u}_j' = f(\boldsymbol{u}_j) = \boldsymbol{u}_j + \boldsymbol{a}$. The Euclidean distance between the transformed points is: $d\left(\boldsymbol{u}_i', \boldsymbol{u}_j'\right) = \left\|\boldsymbol{u}_i' - \boldsymbol{u}_j'\right\|_2 = \|(\boldsymbol{u}_i + \boldsymbol{a}) - (\boldsymbol{u}_j + \boldsymbol{a})\| = \|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2$. Thus, $d\left(\boldsymbol{u}_i', \boldsymbol{u}_j'\right) = d\left(\boldsymbol{u}_i, \boldsymbol{u}_j\right)$, demonstrating that translation preserves the distance between points.

ii) Rotations: In $|\mathcal{R}|$-dimensional space, a rotation is represented by an orthogonal matrix $\boldsymbol{Q}^{|\mathcal{R}|\times|\mathcal{R}|}$. Orthogonal matrices satisfy the property $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix.

Then, the rotated points are given by: $\boldsymbol{u}_i' = f(\boldsymbol{u}_i) = \boldsymbol{Q}\boldsymbol{u}_i$ and $\boldsymbol{u}_j' = f(\boldsymbol{u}_j) = \boldsymbol{Q}\boldsymbol{u}_j$. The Euclidean distance between the rotated points is calculated as follows:

$$
\begin{aligned}
d\left(\boldsymbol{u}_i', \boldsymbol{u}_j'\right) &= \left\|\boldsymbol{u}_i' - \boldsymbol{u}_j'\right\|_2 = \left\|\boldsymbol{Q}\boldsymbol{u}_j - \boldsymbol{Q}\boldsymbol{u}_i\right\|_2 \\
&= \sqrt{\left(\boldsymbol{Q}\left(\boldsymbol{u}_j - \boldsymbol{u}_i\right)\right)^T \left(\boldsymbol{Q}\left(\boldsymbol{u}_j - \boldsymbol{u}_i\right)\right)} \\
&= \sqrt{\left(\boldsymbol{u}_j - \boldsymbol{u}_i\right)^T \boldsymbol{Q}^T\boldsymbol{Q}\left(\boldsymbol{u}_j - \boldsymbol{u}_i\right)}
\end{aligned}
\tag{6.17}
$$

Utilizing the property $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$, it is concluded that $d\left(\boldsymbol{u}_i', \boldsymbol{u}_j'\right) = \left\|\boldsymbol{u}_i - \boldsymbol{u}_j\right\|_2 = d\left(\boldsymbol{u}_i, \boldsymbol{u}_j\right)$.

iii) Reflections: In $|\mathcal{R}|$-dimensional space, a reflection can be characterized by a Householder plane. Let $\boldsymbol{w} = (\boldsymbol{w}_1, ..., \boldsymbol{w}_{|\mathcal{R}|})$ denote the unit normal vector of the hyperplane. The reflection matrix $\boldsymbol{A}$ is defined as $\boldsymbol{A} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T$, where $\boldsymbol{w}$ is a unit vector.

Then, the reflected points are given by $\boldsymbol{u}_i' = f(\boldsymbol{u}_i) = \boldsymbol{A}\boldsymbol{u}_i$ and $\boldsymbol{u}_j' = f(\boldsymbol{u}_i) = \boldsymbol{A}\boldsymbol{u}_j$.

The Euclidean distance d between the reflected points $\boldsymbol{u}_i'$ and $\boldsymbol{u}_j'$ is calculated as follows:

$$
\begin{aligned}
d\left(\boldsymbol{u}_i', \boldsymbol{u}_j'\right) &= \left\|\boldsymbol{u}_i' - \boldsymbol{u}_j'\right\|_2 = \left\|\boldsymbol{A}\boldsymbol{u}_j - \boldsymbol{A}\boldsymbol{u}_i\right\|_2 \\
&= \sqrt{(\boldsymbol{u}_j - \boldsymbol{u}_i)^T\boldsymbol{A}^T\boldsymbol{A}(\boldsymbol{u}_j - \boldsymbol{u}_i)}
\end{aligned}
\tag{6.18}
$$

Observe that $\boldsymbol{A}^T\boldsymbol{A} = (\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T)^T(\boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}^T) = \boldsymbol{I}$, since $\boldsymbol{w}^T\boldsymbol{w} = 1$. Consequently, $d\left(\boldsymbol{u}_i', \boldsymbol{u}_j'\right) = \|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2 = d\left(\boldsymbol{u}_i, \boldsymbol{u}_j\right)$.

It can be concluded that the Euclidean distance between two points remains invariant under rotations, translations, or reflections in space. The proposed anomaly detection method is fundamentally based on these inter-point distances. Consequently, any combination of these transformations applied to the data points will not alter the distances between them, and thus, the computed anomaly scores will remain constant. □

This proposition establishes the invariance of the proposed anomaly detection method to isometric transformations (rotations, translations, reflections). This property ensures robust anomaly detection across diverse real-world scenarios. Such robustness is crucial when dealing with imperfect or noisy RES data, where sensor readings might be subject to shifts or other distortions that do not inherently signify anomalous behavior.

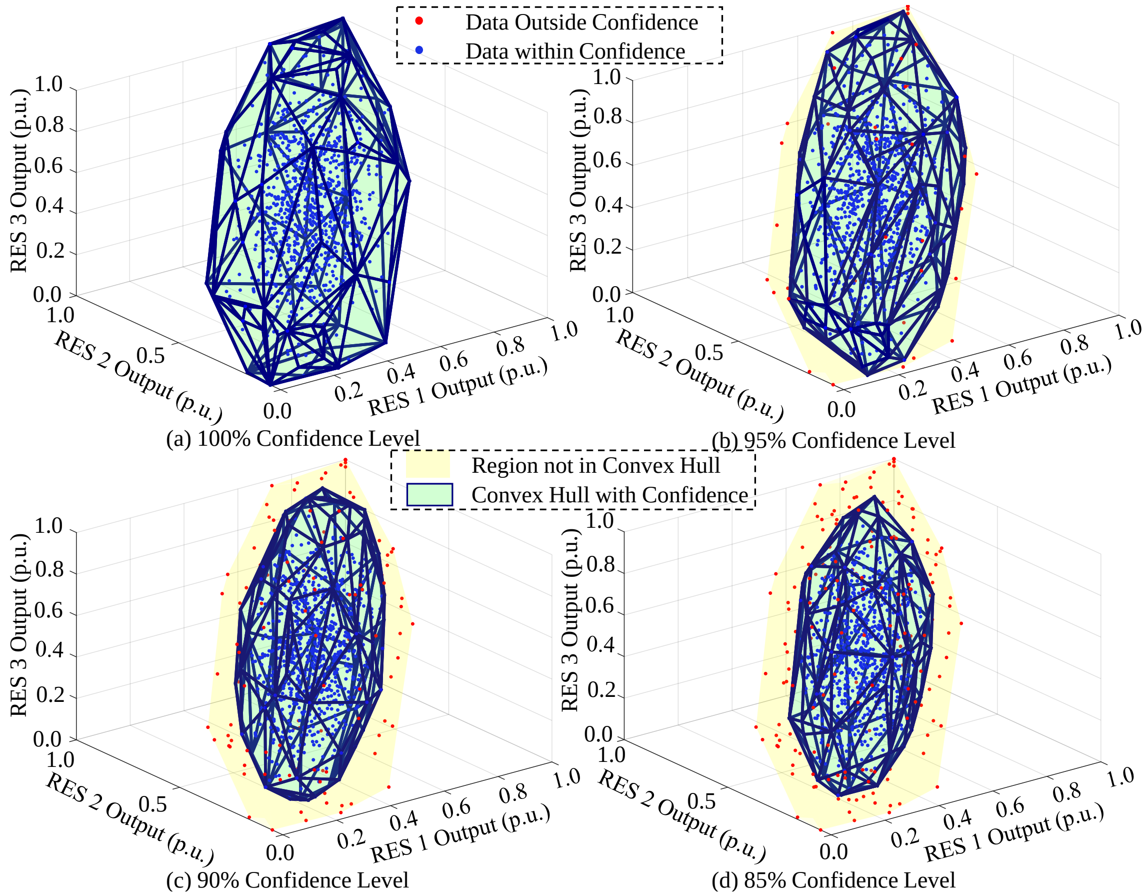


Figure 6.2: Illustration of convex hull with various confidence levels

Figure 6.2 illustrates the relationship between confidence levels and uncertainty set dimensions using a three-RES unit example. As confidence levels decrease, indicating a greater risk tolerance, the corresponding uncertainty sets become more compact. This demonstrates the adaptability of the proposed GNN-based approach, allowing microgrid operators to customize uncertainty sets according to their rich operational experience. The step-by-step implementation of this process is outlined in Algorithm 7.

Algorithm 7 GNN-based Confidence Level Algorithm

Input: Dataset $\mathcal{X} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{|\mathcal{I}|}\}$, confidence level ρ $(0 < \rho < 1)$
Initialize: Confidence level dataset $\mathcal{O}_\rho = \emptyset$
for all $\boldsymbol{u}_i \in \mathcal{X}$ **do**
    Compute anomaly score s_i using GNN model
end for
Create score set $S \leftarrow \{s_1, s_2, \ldots, s_{|\mathcal{I}|}\}$
$N \leftarrow$ argsort(S, descending) ▷ Sort S and get indices
Calculate $k \leftarrow \lfloor |\mathcal{I}| \cdot \rho \rfloor$
$\theta \leftarrow s_{N[k]}$ ▷ Set threshold
for all $\boldsymbol{u}_i \in \mathcal{X}$ **do**
    if $s_i \leq \theta$ **then**
        $\mathcal{O}_\rho \leftarrow \mathcal{O}_\rho \cup \{\boldsymbol{u}_i\}$
    end if
end for
return $\mathcal{O}_\rho$

6.2.2 Pair Convex Hull Uncertainty Set

To address the curse of dimensionality arising from numerous RES units, a PCHUS for a given confidence level is introduced. The PCHUS is defined as follows:

$$\mathcal{U}_{p,q}^{\text{PCH}} = \left\{ u_r \;\middle|\; \begin{array}{l} u_r = \hat{u}_r + \sum_{m \in \mathcal{M}_{p,q}} \varepsilon_{rm}(u_{rm} - \hat{u}_r), \forall r = p, q \\ 0 \leq \varepsilon_{rm} \leq 1, \forall r = p, q; m \in \mathcal{M}_{p,q} \\ \sum_{m \in \mathcal{M}_{p,q}} \varepsilon_{rm} = 1, \forall r = p, q \\ \overline{\nu}_r \leq u_r \leq \underline{\nu}_r, \forall r \neq p, q \\ \overline{\nu}_r = \min(u_{1r}, u_{2r}, ..., u_{|\mathcal{I}|r}), \forall r \neq p, q \\ \underline{\nu}_r = \max(u_{1r}, u_{2r}, ..., u_{|\mathcal{I}|r}), \forall r \neq p, q \end{array} \right\} \tag{6.19}$$

where $\mathcal{M}_{p,q}$ denotes the vertex set of the convex hull enclosing data points from the p-th and q-th RES units. This convex hull represents the uncertainty region between these two units. By systematically pairing each RES unit with every other unit in $\mathcal{R}$, a corresponding CHUS is constructed for each pair. For $|\mathcal{R}|$ RES units, this approach yields $C_{|\mathcal{R}|}^2 = |\mathcal{R}|(|\mathcal{R}| - 1)/2$ CHUS pairs in total.

For a given dataset $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and confidence level ρ, the construction of PCHUS using the GNN-based confidence level is outlined in Algorithm 8. This PCHUS construction algorithm effectively addresses the curse of dimensionality arising from multiple RES units by focusing on pairs of units. The proposed PCHUS approach achieves a balance between computational efficiency and accurate uncertainty representation, although it may yield more conservative estimates compared to a full CHUS (FCHUS)[160]. Consequently, this algorithm provides a practical solution for managing uncertainty in large-scale RES systems, offering a trade-off between accuracy and computational complexity. Moreover, although the proposed PCHUS is initially formulated based on spatial correlations, the underlying construction methodology is general and readily adaptable to incorporate spatio-temporal correlations. However, due to space limitations, we do not delve into this modification. When accounting for both spatial and temporal correlations of RES output, the original spatial PCHUS a dimension of $|\mathcal{R}|$ can be extended to a spatio-temporal PCHUS with a dimension of $2|\mathcal{R}|$. This enhancement allows the uncertainty set to more accurately capture the joint variations of RES outputs over space and time, thereby further reducing the conservativeness of the resulting energy management strategies.

Algorithm 8 PCHUS Construction with Confidence Level

Input: Set of RES units $\mathcal{R}$, dataset with confidence level ρ
Output: Set of all PCHUSs $\mathcal{U}^{\text{PCH}}$
Initialize $\mathcal{U}^{\text{PCH}} \leftarrow \emptyset$
for $p \leftarrow 1$ to $|\mathcal{R}| - 1$ **do**
    for $q \leftarrow p + 1$ to $|\mathcal{R}|$ **do**
        $\mathcal{M}_{p,q} \leftarrow \text{ConvexHull}(\text{data from units } p \text{ and } q)$
        Initialize $\mathcal{U}_{p,q}^{\text{PCH}} \leftarrow \emptyset$
        for $r \in \{p, q\}$ **do**
            Compute $\hat{u}_r$ (e.g., mean of historical data)
            for all vertex $m \in \mathcal{M}_{p,q}$ **do**
                Compute u_{rm} (value at vertex m)
            end for
        end for
        for all $r \in \mathcal{R} \setminus \{p, q\}$ **do**
            $\overline{\nu}_r \leftarrow \min\{u_{ir} : i \in \mathcal{I}\}$
            $\underline{\nu}_r \leftarrow \max\{u_{ir} : i \in \mathcal{I}\}$
        end for
        Construct $\mathcal{U}_{p,q}^{\text{PCH}}$ using Equation (6.19)
        $\mathcal{U}^{\text{PCH}} \leftarrow \mathcal{U}^{\text{PCH}} \cup \{\mathcal{U}_{p,q}^{\text{PCH}}\}$
    end for
end for
return $\mathcal{U}^{\text{PCH}}$

Remark 1: The size and shape of PCHUS can vary significantly depending on the characteristics of the paired RES units, which considerably impacts the conservativeness of robust solutions. PCHUSs of larger size, typically resulting from weakly correlated RES unit pairs, tend to produce more conservative solutions due to their accommodation of a broader range of potential scenarios. In contrast, smaller PCHUSs, often arising from strongly positively correlated RES unit pairs, generally yield less conservative solutions. Notably, any single PCHUS is capable of encompassing all given RES outputs, ensuring that a model utilizing any PCHUS can provide a fully robust solution against RES uncertainty. Consequently, identifying the least conservative uncertainty set among all PCHUSs can improve the economic efficiency of solutions without compromising robustness. This challenge will be addressed in the subsequent section through the development of a quad-level optimization model.

6.3 Quad-Level Energy Management for HMGs and Tailored Algorithm Design

This section introduces a quad-level REMM for HMGs. Subsequently, a tailored TOAT based C&CG algorithm is developed to efficiently solve the proposed quad-level optimization problem.

6.3.1 Quad-Level Energy Management Model

As previously discussed, an HMG with $|\mathcal{R}|$ RES units generates a total of $C^2_{|\mathcal{R}|} = |\mathcal{R}|(|\mathcal{R}| - 1)/2$ PCHUS pairs. Though each PCHUS covers all possible data of RES outputs, the size and shape of each PCHUS are different, which directly affects the conservativeness of the energy management solution. The objective is to identify the least conservative uncertainty set among these $C^2_{|\mathcal{R}|}$ PCHUSs, formulated as follows.

$$\min_{\boldsymbol{x}} \boldsymbol{a}^T \boldsymbol{x} + \min_{(p,q)} \max_{\mathcal{U}^{\text{PCH}}_{p,q}} \min_{\boldsymbol{y}} \boldsymbol{b}^T \boldsymbol{y} \tag{6.20}$$

$$\text{s.t.Constraints}(6.2), (6.19), \text{and}(6.4) \tag{6.21}$$

The structure of this quad-level optimization problem is illustrated in Fig. 6.3. Unlike the existing

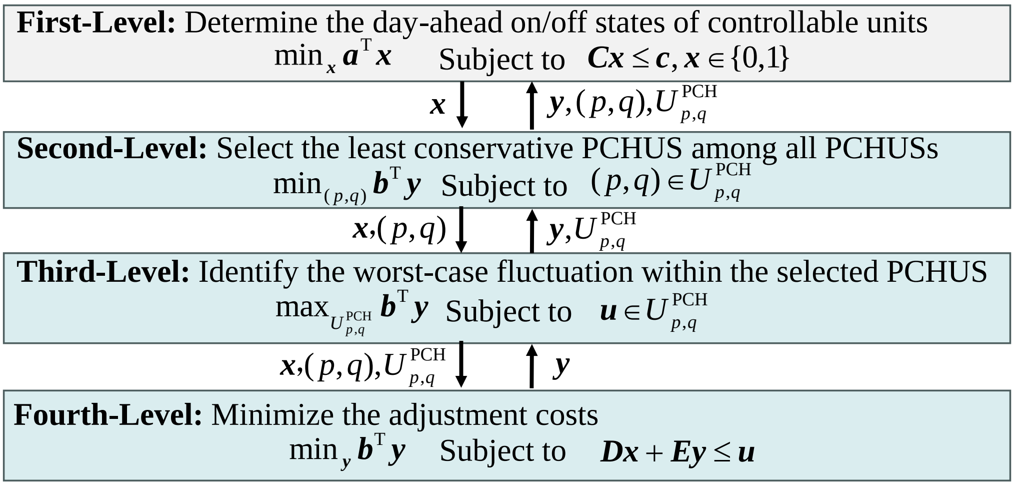


Figure 6.3: Illustration of the quad-level structure

tri-level robust models (6.1)-(6.4), the developed energy management model is structured as a quad-level problem. Specifically, after determining the on/off states in the first-level model, the second-level model identifies the least conservative PCHUS among all possibilities. The third-level model then explores the worst-case fluctuation within this selected PCHUS, while the fourth-level model minimizes the adjustment costs required to maintain power balance under the worst-case realization of RES outputs. This quad-level approach offers dual benefits: it reduces the conservativeness of the REMM solution by selecting the least conservative PCHUS, while simultaneously ensuring the solution remains fully robust against all possible RES fluctuation scenarios. Thus, the proposed model achieves a balance between minimizing conservativeness and maintaining robust performance against all possible data points of RES fluctuations.

Remark 2: It is important to emphasize that the least conservative PCHUS among the $C^2_{|\mathcal{R}|}$ possible combinations is not fixed but instead depends on the decision variables $\boldsymbol{x}$ determined at the first level of optimization. Thus, this optimal PCHUS cannot be predetermined before the optimization begins. Typical tri-level models, which rely on predefined and static uncertainty sets, may simplify the problem formulation but often lead to overly conservative solutions. This lack of adaptability can hinder economic performance by not accurately reflecting the least conservative uncertainty set. In contrast, the proposed quad-level model overcomes this limitation by incorporating a second-level optimization that dynamically identifies the least conservative PCHUS from all possible combinations. This adaptive approach ensures the selected uncertainty set is as tight as possible without sacrificing robustness, thereby reducing conservatism. Additionally, the model eliminates the need for exhaustive offline evaluation of all possible PCHUS pairs, which grows combinatorially with the number of RES units ($C^2_{|\mathcal{R}|}$ for $|\mathcal{R}|$ units), making it computationally efficient. By embedding the uncertainty set selection within the optimization process, the quad-level model enhances both scalability and practicality for microgrid energy management.

6.3.2 Customized TOAT-based C&CG Algorithm

Although the proposed PCHUS significantly reduces the number of vertices in uncertainty sets, the remaining vertices can still lead to a locally optimal solution. To avoid settling for local optima, a customized TOAT-based C&CG algorithm is developed, which strategically selects starting points for the algorithm to enhance its effectiveness in finding global optimal solutions.

6.3.2.1 Starting Point Selection by Extending TOAT Techniques

TOAT, originally developed by Genichi Taguchi, is a method that enables systematic evaluation of numerous test scenarios using a minimal set of representative cases. This technique is adapted to the context to identify the extreme vertices that most significantly influence the solution, and to select initial points for the proposed solution algorithms. As an illustration, in a system with three RES units, each having two levels, four vertices are designated as TOAT vertices, as depicted in Fig. 6.4.

It is noted that the current TOAT technique requires a fixed number of levels for each RES unit, essentially assuming interval-form uncertainty sets. This limitation prevents its direct application to the developed PCHUS. To overcome this constraint, an extension of the TOAT technique specifically tailored for the PCHUS is proposed, referred to as the quasi-TOAT (q-TOAT) technique. This adaptation involves replacing the original TOAT vertices with q-TOAT vertices. The detailed process of this modification is outlined as follows.

The Euclidean distance $d^{z-z'}_{p,q}$ is computed between two sets of vertices: the selected TOAT vertices $\boldsymbol{u}_z$ from IUS $\mathcal{U}^{\mathrm{I}}$, and any vertex $\boldsymbol{u}_{z'}$ from PCHUS $\mathcal{U}^{\mathrm{PCH}}_{p,q}$. This distance can be calculated using

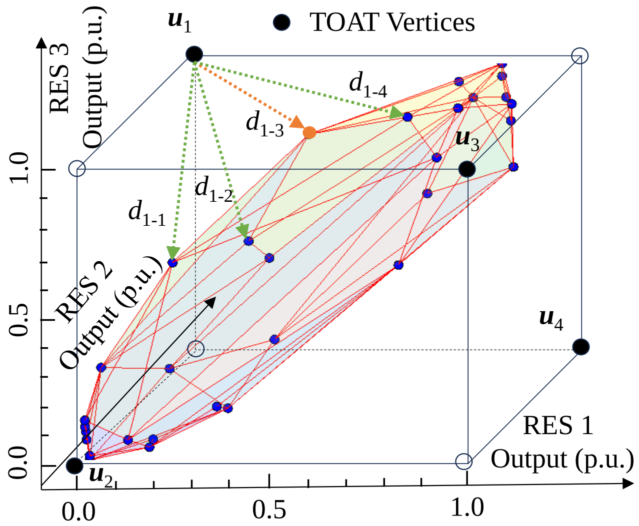


Figure 6.4: Illustration of the extended TOAT vertices

the following formula:

$$d_{p,q}^{z-z'} = \|\boldsymbol{u}_z - \boldsymbol{u}_{z'}\|, \forall u_{z'} \in \mathcal{U}_{p,q}^{\text{PCH}} \tag{6.22}$$

Next, the vertex $\boldsymbol{u}_{z^*}$ in the PCHUS $\mathcal{U}_{p,q}^{\text{PCH}}$ that yields the minimum distance $d_{p,q}^{z-z'}$ when compared to all vertices $\boldsymbol{u}_{z'}$ within the same PCHUS is identified. This minimum distance is determined as follows:

$$\boldsymbol{u}_{z^*} = \underset{\boldsymbol{u}_{z'} \in \mathcal{U}_{p,q}^{\text{PCH}}}{\arg\min} \|d_{p,q}^{z-z'}\| \tag{6.23}$$

The final step involves replacing the original TOAT vertices $\boldsymbol{u}_z$ with their corresponding $\boldsymbol{u}_{z^*}$. This process is visually represented in Fig. 6.4. According to TOAT theory, $N_{\text{TOAT}} = 2^{m_{\text{TOAT}}}$ TOAT vertices $\boldsymbol{u}_z$ are needed to be replaced with the vertices $\boldsymbol{u}_{z'}$, where $m_{\text{TOAT}} \geq \lceil \log 2(|\mathcal{R}| + 1) \rceil$. This condition ensures that m_{TOAT} is large enough to encompass all $|\mathcal{R}|$ uncertainties in the system.

6.3.2.2 Decomposition Algorithm

Utilizing the q-TOAT vertices as initial points, a tailored version of the C&CG algorithm is developed. While C&CG is traditionally applied to tri-level models, the modification adapts it to the proposed quad-level optimization model. In this approach, the complete model is decomposed into two components: one master sub-model (MSM) and N_{TOAT} slave sub-models (SSMs).

MSM: The MSM corresponds to the outer minimization problem in the quad-level formulation

(6.20) and (6.21). It aims to minimize the objective function under the worst-case scenario induced by RES uncertainty. The MSM is solved iteratively, with each iteration incorporating optimal and feasible cuts derived from the solution of the SSMs. This process ensures the feasibility of the solution under the worst-case scenario. The mathematical formulation of the MSM at the V-th iteration can be expressed as follows:

$$\min_{\boldsymbol{x},\theta,\boldsymbol{y}^v} \boldsymbol{a}^T\boldsymbol{x} + \theta \tag{6.24}$$

$$\boldsymbol{C}\boldsymbol{x} \leq \boldsymbol{c}, \boldsymbol{x} \in \{0,1\}, \theta \geq 0 \tag{6.25}$$

$$\theta \geq \boldsymbol{b}^T\boldsymbol{y}^v, v = 1,2,...,V-1 \tag{6.26}$$

$$\boldsymbol{D}\boldsymbol{x} + \boldsymbol{E}\boldsymbol{y}^v \leq \boldsymbol{u}^v, \boldsymbol{u}^v \in \mathcal{U}, v = 1,2,...,V-1 \tag{6.27}$$

where $\boldsymbol{y}^v$ represents the variable in SSMs corresponding to the extreme scenario $\boldsymbol{u}^v$ that was identified during the v-th iteration of the algorithm.

SSMs: The SSMs are formulated based on the decision $\boldsymbol{x}^V$ obtained from the MSM. The mathematical representation of these SSMs is as follows:

$$\min_{(p,q)} \max_{\mathcal{U}^{\text{PCH}}_{p,q}} \min_{\boldsymbol{y}} \boldsymbol{b}^T\boldsymbol{y} \tag{6.28}$$

$$\boldsymbol{u} \in \mathcal{U}^{\text{PCH}}_{p,q}, \forall p,q \in \mathcal{R} \tag{6.29}$$

$$\boldsymbol{D}\boldsymbol{x}^* + \boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{u} \tag{6.30}$$

Subsequently, the previously selected q-TOAT vertices $\boldsymbol{u}_{z'}$ are utilized to aid in determining the global worst-case scenario for each pair (p,q) within the PCHUS $\mathcal{U}^{\text{PCH}}_{p,q}$. This process can be mathematically formulated as follows:

$$\min_{(p,q)} \max_{\boldsymbol{u}_{z'}} \max_{\mathcal{U}^{\text{PCH}}_{p,q}} \min_{\boldsymbol{y}} \boldsymbol{b}^T\boldsymbol{y} \tag{6.31}$$

where $z' = 1,2,...,N_{\text{TOAT}}$. It is important to note that each PCHUS $\mathcal{U}^{\text{PCH}}_{p,q}$, where $p,q \in \mathcal{R}$, is independent of the others. Consequently, for each individual $\mathcal{U}^{\text{PCH}}_{p,q}$ where $p,q \in \mathcal{R}$, the problem

(6.31) can be reformulated as follows:

$$\forall (p,q), \quad \max_{\boldsymbol{u}_{z'}} \max_{\mathcal{U}_{p,q}^{\text{PCH}}} \min_{\boldsymbol{y}} \boldsymbol{b}^T \boldsymbol{y} \tag{6.32}$$

This formulation allows for parallel computation across all pairs (p, q). Moreover, for each value of z' from 1 to N_{TOAT}, the problem (6.32) can be further decomposed and expressed as follows:

$$\forall \boldsymbol{u}_{z'}, \quad \max_{\mathcal{U}_{p,q}^{\text{PCH}}} \min_{\boldsymbol{y}} \boldsymbol{b}^T \boldsymbol{y} \tag{6.33}$$

The individual models represented in (6.33) are amenable to parallel computation, allowing for simultaneous solution.

To address this tri-level optimization structure, a nested block coordinate descent method is developed. This approach decomposes the entire problem (6.33) into two components: an inner-loop MSM and an inner-loop SSM. Specifically, at the V'-th iteration, the inner-loop SSM is formulated as follows:

$$\psi_{\boldsymbol{u}_{z'}} = \min_{\boldsymbol{u},\boldsymbol{y},\boldsymbol{\pi}} \boldsymbol{b}^T \boldsymbol{y} \tag{6.34}$$

$$\boldsymbol{D}\boldsymbol{x}^{V*} + \boldsymbol{E}\boldsymbol{y} \leq \boldsymbol{u}, \boldsymbol{u} \in \mathcal{U}_{p,q}^{\text{PCH}} \tag{6.35}$$

$$\boldsymbol{u} = \boldsymbol{u}^{V'} : \boldsymbol{\pi} \tag{6.36}$$

where $\boldsymbol{\pi}$ denotes the dual variable that quantifies the sensitivities of the constraints in (6.34). This value is derived from the solution of the optimization problem defined by (6.34)-(6.36). Subsequently, the inner-loop MSM is formulated by applying a first-order Taylor series expansion to (6.34)-(6.36), resulting in the following expression:

$$\max_{\mathcal{U}_{p,q}^{\text{PCH}}} (\boldsymbol{\pi}^{V'})^{\text{T}} (\boldsymbol{u} - \boldsymbol{u}^{V'}) + \psi_{\boldsymbol{u}_{z'}} \tag{6.37}$$

It is important to highlight that each q-TOAT vertex is associated with its own MSM and the corresponding SSM. Upon convergence of all inner-loop models, the global worst-case scenario is determined by identifying the maximum value of $\psi_{\boldsymbol{u}_{z'}}$ across all vertices $\boldsymbol{u}_{z'}$ within the PCHUS $\mathcal{U}_{p,q}^{\text{PCH}}$.

This process can be mathematically expressed as follows:

$$\varphi_{(p,q)} = \max\nolimits_{\boldsymbol{u}_{z'} \in \mathcal{U}_{p,q}^{\mathrm{PCH}}} \psi_{\boldsymbol{u}_{z'}} \tag{6.38}$$

For each pair (p, q), the inner-loop MSM and inner-loop SSM are solved iteratively until they meet the convergence threshold $\varepsilon^{\mathrm{In}}$. Given that each PCHUS $\mathcal{U}_{p,q}^{\mathrm{PCH}}$ covers all possible historical fluctuations in RES outputs, the least conservative PCHUS is identified by determining the minimum value of $\varphi_{(p,q)}$. This process can be mathematically expressed as:

$$\phi = \min\nolimits_{(p,q)} \varphi_{(p,q)} \tag{6.39}$$

The comprehensive workflow of the algorithm, which integrates the proposed q-TOAT approach, is detailed in **Algorithm** 9.

It is noteworthy that the computational complexity of uncertainty set construction and optimization varies significantly across the different approaches presented in this paper. For the FCHUS, the complexity grows exponentially with the number of RES units. Specifically, constructing a convex hull for $|\mathcal{R}|$ RES units and n data points has a complexity of $\mathcal{O}(n^{\lceil |\mathcal{R}|/2 \rceil})$ using standard algorithms, such as Quickhull, Beneath-Beyond, or Incremental algorithms. This results in exponential growth in the number of vertices that define the uncertainty set, potentially reaching $\mathcal{O}(n^{\lceil |\mathcal{R}|/2 \rceil})$ in the worst case. When applying the C&CG algorithm to problems involving FCHUS, the overall complexity becomes

$$\mathcal{O}(V \times (\mathrm{MP} + n^{\lceil |\mathcal{R}|/2 \rceil} \times \mathrm{SP})),$$

where MP is the complexity of the master problem, and SP is the complexity of the subproblem.

In contrast, the PCHUS approach significantly reduces computational complexity. For each pair of RES units, convex hull construction has a complexity of only $\mathcal{O}(n \log n)$. With $C_{|\mathcal{R}|}^2 = |\mathcal{R}|(|\mathcal{R}| - 1)/2$ total pairs, the overall construction complexity becomes $\mathcal{O}(|\mathcal{R}|^2 \cdot n \log n)$. The number of vertices per pair grows linearly in the worst case, approximately $\mathcal{O}(n)$. This results in a C&CG

Algorithm 9 Customized TOAT-Based C&CG Algorithm

Input: Convergence threshold $\varepsilon^{\text{Out}} = \varepsilon^{\text{In}} = 10^{-5}$, confidence level ρ.
Initialize: $UB^{\text{Out}} \leftarrow \infty$, $LB^{\text{Out}} \leftarrow -\infty$, $UB^{\text{In}} \leftarrow \infty$, $LB^{\text{In}} \leftarrow -\infty$, $V \leftarrow 1$.
Generate RES dataset $\mathcal{O}_\rho$ using ρ. ▷ **Algorithm 7**
while $\left|UB^{\text{Out}} - LB^{\text{Out}}\right| / UB > \varepsilon^{\text{Out}}$ **do**
    Compute $\boldsymbol{x}^V$ by solving outer-loop MSM (6.24)-(6.27)
    for $p \neq q \in \mathcal{R}$ **do** ▷ Parallel processing
        Initialize $V' \leftarrow 1$
        Construct PCHUS $\mathcal{U}^{\text{PCH}}_{p,q}$ using $\mathcal{O}_\rho$ ▷ **Algorithm 8**
        Set initial $\boldsymbol{u}_{z'}$ to predicted value $\hat{\boldsymbol{u}}$
        for $z' \leftarrow 1$ **to** N_{TOAT} **do**
            Determine q-TOAT vertices $\boldsymbol{u}_{z'}$ using (6.23)
            while $\left|UB^{\text{In}} - LB^{\text{In}}\right| / UB^{\text{In}} > \varepsilon^{\text{In}}$ **do**
                Solve inner-loop SSM (6.34)-(6.36) $\rightarrow \psi_{\boldsymbol{u}_{z'}}$ and $\boldsymbol{\pi}$
                Solve inner-loop MSM (6.37) $\rightarrow \boldsymbol{u}^{V'+1}$
                $V' \leftarrow V' + 1$
            end while
        end for
        Determine $\boldsymbol{u}_{z'*}$ by solving model (6.38)
    end for
    Identify $\mathcal{U}^{\text{PCH}}_{p*,q*}$ by solving model (6.39)
    Increment $V \leftarrow V + 1$
end while
Return optimized $\boldsymbol{x}, \mathcal{U}^{\text{PCH}}_{p*,q*}$

algorithm complexity of

$$\mathcal{O}(V \times (\text{MP} + |\mathcal{R}|^2 \cdot n \cdot \text{SP})),$$

This quadratic, rather than exponential, growth with respect to $|\mathcal{R}|$ explains why PCHUS remains computationally tractable even as the number of RES units increases.

Further efficiency gains are achieved through the proposed parallel processing implementation. With P_{para} processors, the complexity of PCHUS with parallel processing reduces to approximately

$$\mathcal{O}\left(V \times \left(\text{MP} + \frac{|\mathcal{R}|^2 \cdot n \cdot \text{SP}}{P_{\text{para}}}\right)\right).$$

This theoretical improvement aligns with the substantial reductions in computing time observed in Table 6.5, where parallel processing enabled solutions for cases that previously required prohibitively large computational resources.

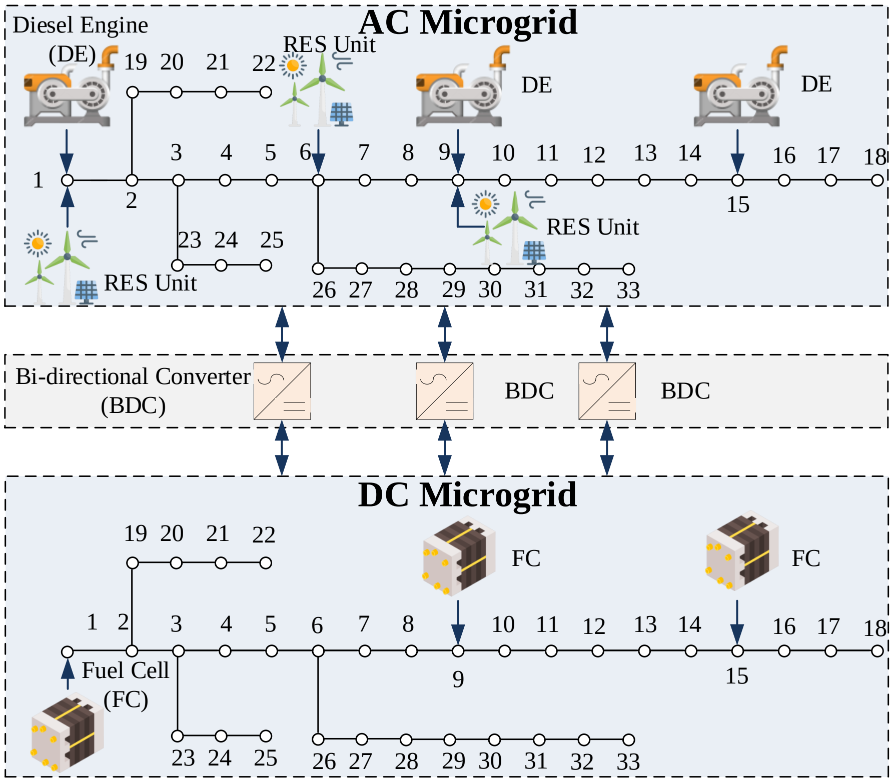


Figure 6.5: Structure of the HMG system

6.4 Case Studies

6.4.1 Case Description

To verify the effectiveness and superiority of the developed data-driven quad-level approach, a comprehensive case study is conducted using a practical hybrid HMG [17], as shown in Fig.6.5. The system comprises two integrated subsystems: an AC microgrid and a DC microgrid, linked by multiple BDCs that facilitate power exchange between the two domains. The AC microgrid features three diesel engines (DEs) at buses 1, 9, and 15, serving as controllable dispatchable units, alongside three RES units installed at buses 1, 6, and 9, representing variable wind and solar generation. The DC microgrid, on the other hand, comprises three fuel cells (FCs) placed at buses 1, 9, and 15. Both AC and DC subsystems are modeled using an IEEE 33-bus radial topology, rather than a simplified single-bus structure [157], to closely resemble the configuration of realistic low-voltage microgrids. For further details on system parameters and configurations, readers are referred to [158].

All comparative analyses are executed using Matlab 2023b and General Algebraic Modeling System (GAMS) 24.4 on a high-performance workstation featuring 64 GB RAM and an I9-7900X CPU.

6.4.2 Comparative Analysis of Energy Management Solutions

To demonstrate the effectiveness of our proposed PCHUS, a comparative assessment against other advanced uncertainty set formulations is conducted. These include IUS [161], BIUS [157], and FCHUS [160]. The analysis focuses on two key aspects: associated costs and the influence of varying confidence levels. Table 6.1 presents a comprehensive summary of these comparative results. Here, 10,000 scenarios are randomly generated for out-of-sample robustness testing, and non-robust scenarios are defined as those where the solution fails to maintain robustness.

Table 6.1 reveals that the IUS-based energy management strategy results in the highest total cost (TC) for microgrid operation, despite its robust performance across all stochastic scenarios. BIUS incorporates a budget parameter Γ to adjust the conservativeness of the solution. As Γ increases, there is a corresponding rise in both total cost and conservativeness. However, it is noteworthy that this method fails to account for the correlation of RES historical data.

In comparison, our developed GNN-based method incorporates the majority of practical RES data within its uncertainty set. It selectively excludes the limited data points or outliers based on an operator-defined confidence level, thereby offering flexible control over the conservativeness of the solution. This adjustable feature facilitates a balanced trade-off between robustness and operational cost. Both the FCHUS and the proposed PCHUS exhibit a gradual increase in total cost as confidence levels rise. Notably, at equivalent confidence levels, the total cost associated with the proposed PCHUS is only marginally higher than that of the FCHUS. This marginal cost difference suggests that PCHUS closely approximates the performance of FCHUS while utilizing significantly fewer vertices, thus offering computational advantages.

Regarding out-of-sample performance, each subplot in Fig.6.6 illustrates a comparison between the realized TC for each scenario (blue dots) and the worst-case cost determined during optimization (red line). The IUS method exhibits strong conservativeness, achieving zero violations but resulting in a substantially inflated worst-case cost. In contrast, BIUS reduces both conservativeness and cost,

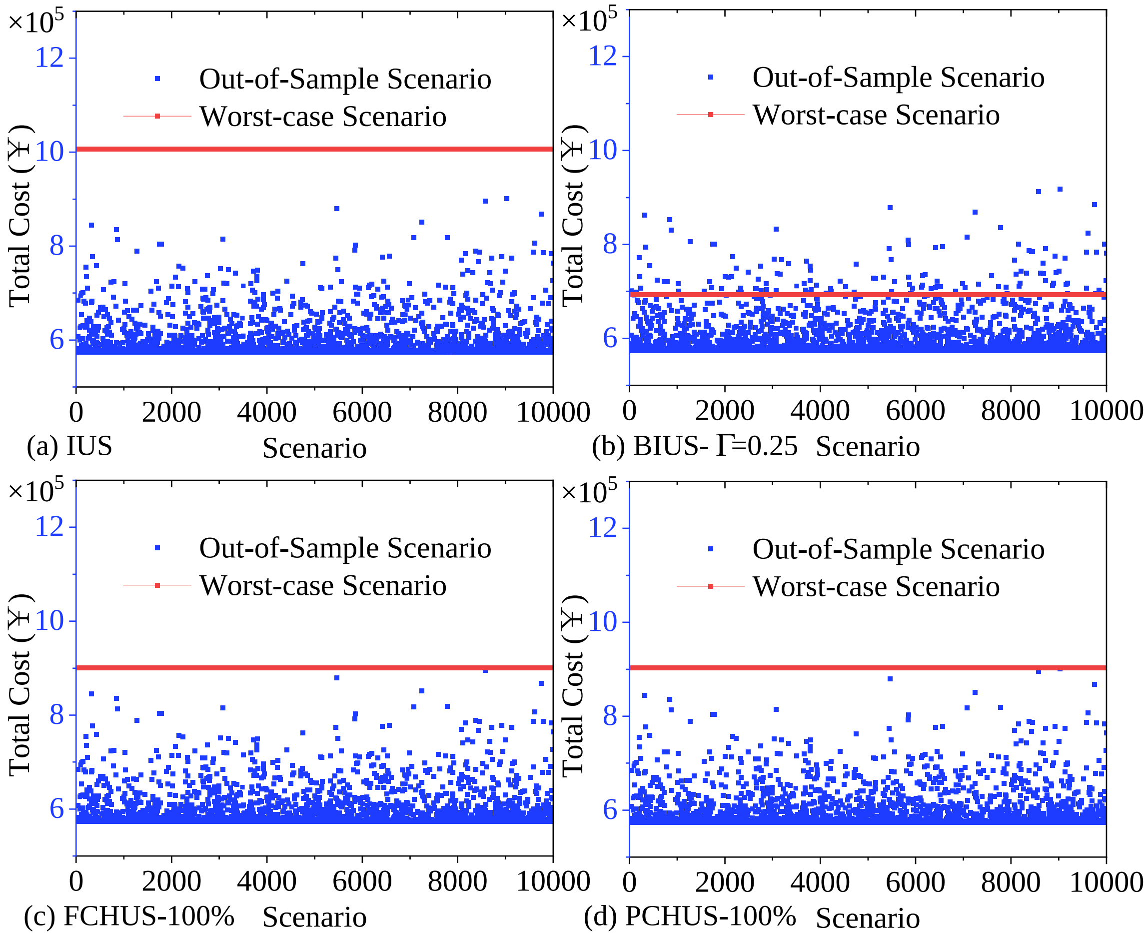


Figure 6.6: Out-of-sample performance of different uncertainty sets

but at the cost of 229 non-robust outcomes, reflecting diminished resilience to uncertainty. Both FCHUS-100% and the proposed PCHUS-100% maintain full robustness while being less conservative than IUS. This highlights the ability of the proposed PCHUS-100% to accurately capture the true uncertainty space without unnecessary conservativeness. Combined with its significantly enhanced computational efficiency and scalability, PCHUS emerges as a practical and high-performing approach for robust energy management in HMGs.

6.4.3 Comparative Analysis of Algorithmic Performance

To verify the superiority of the customized q-TOAT C&CG algorithm in solving the REMM with FCHUS and PCHUS, a comprehensive performance comparison across various algorithms under different confidence levels is conducted. The results of this comparison are presented in Table 6.2

Table 6.1: Solution comparisons of various uncertainty sets

Uncertainty Set		**Total Cost (¥)**	**First-Stage Cost (¥)**	**Second-Stage Cost (¥)**	**Non-Robust Scenarios**
IUS[161]		1006976.0	573000.4	433975.6	0
BIUS [157]	Γ =0.75	850096.7	573015.2	277081.5	6
	Γ =0.50	815338.1	572863.6	242474.5	34
	Γ =0.25	693114.7	573044.9	120069.8	229
FCHUS-100%[160]		901350.1	573000.4	328349.6	0
FCHUS-99%		759182.6	573000.0	186182.5	35
FCHUS-97%		723056.1	573189.4	149866.7	57
FCHUS-95%		704434.0	573064.8	131369.1	95
FCHUS-90%		678419.6	572873.8	105545.8	165
FCHUS-85%		667028.4	573153.1	93875.3	195
FCHUS-80%		649395.7	573272.8	76122.9	280
Proposed PCHUS-100%		902810.6	573000.4	329810.1	0
Proposed PCHUS-99%		763888.2	573000.2	190888.0	33
Proposed PCHUS-97%		725446.7	572648.1	152798.6	53
Proposed PCHUS-95%		708120.5	573049.8	135070.7	86
Proposed PCHUS-90%		681181.5	572931.4	108250.1	158
Proposed PCHUS-85%		669440.8	573268.6	96172.2	188
Proposed PCHUS-80%		652289.9	573240.6	79049.3	266

and illustrated in Fig. 6.7. The algorithms evaluated in this comparison include the C&CG algorithm with a standard single initial point (SIP) [38], the C&CG algorithm with 50 randomly selected initial points (RIPs)[109], and the C&CG algorithm incorporating the proposed quasi-TOAT initial points (TOATPs).

The results in Table 6.2 demonstrate that the developed TOATP-based method consistently outperforms both SIP and RIP methods across all confidence levels and for both uncertainty set types. Specifically, the typical C&CG algorithm utilizing SIP sets a single initial point to initialize the algorithm, aiming to identify the worst-case scenario of RES fluctuations within the defined uncertainty sets. However, this approach often leads to suboptimal solutions, as it has a high likelihood of converging to local optima when determining the worst-case scenario, and frequently encounters non-convergence issues. While the RIP method attempts to mitigate these limitations by randomly selecting multiple initial points, it still tends to identify locally worst-case scenarios rather than global ones, and continues to face convergence challenges. In contrast, the proposed TOATP-based method demonstrates superior capabilities. It efficiently identifies the global worst-case scenario of RES

Table 6.2: Objective comparisons under various algorithms (¥)

Confidence Level	**FCHUS**			**PCHUS**		
	SIP [38]	**RIPs**[109]	**TOATPs**	**SIP** [38]	**RIPs**[109]	**TOATPs**
100%	901350.1	901350.1	901350.1	728989.0	902810.6	902810.6
99%	740451.5	753768.3	759182.6	667385.4	755972.5	763888.2
97%	659282.1	720465.0	723056.1	641451.4	708793.4	725446.7
95%	639278.5	694624.8	704434.0	623033.3	699584.6	708120.5
90%	595353.9	677987.0	678419.6	592188.1	667891.0	681181.5
85%	628099.1	663217.7	667028.4	602587.9	659498.0	669440.8
80%	579845.6	642876.4	649395.7	599388.4	643346.8	652289.9

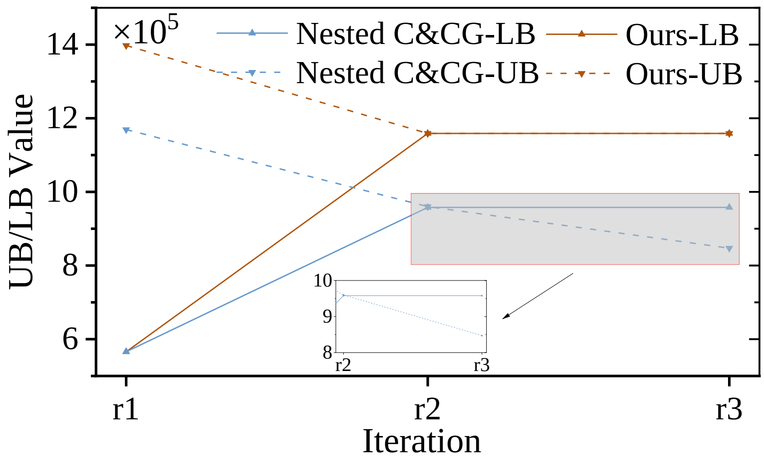


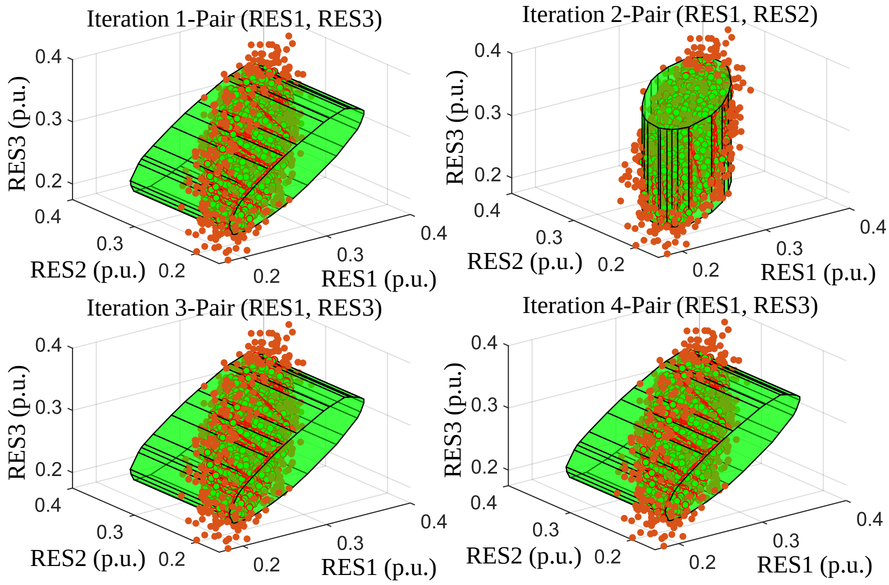
Figure 6.7: Convergence curve with various C&CG algorithms

fluctuations while requiring significantly fewer initial points. Moreover, the proposed TOATP-based approach ensures algorithmic convergence, addressing a key limitation of previous methods.

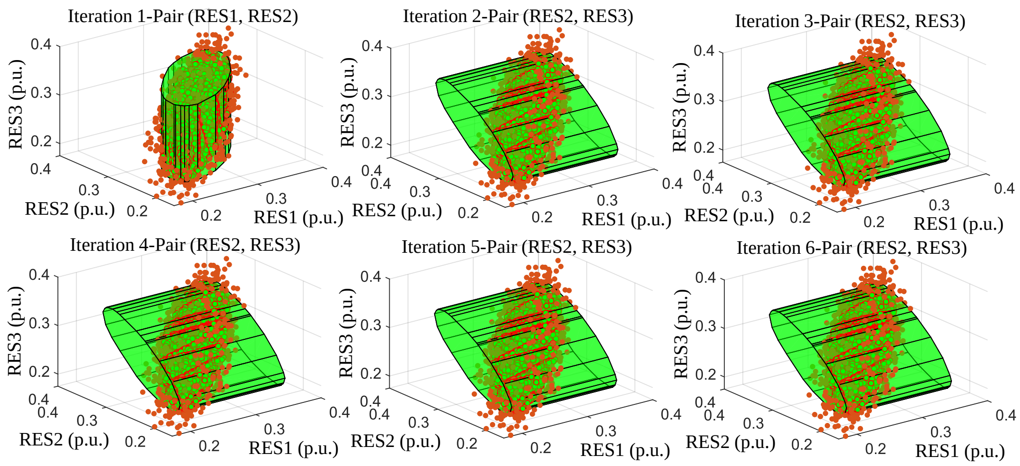
6.4.4 Role Verification of the Second-Level Model

The proposed quad-level model introduces a novel second-level model that autonomously identifies the optimal PCHUS from the available options. To evaluate the impact of this second-level model, Table 6.3 presents a comparison between the quad-level model and conventional tri-level models that use predefined PCHUS pairs. For this analysis, a 100% confidence level is used as a representative example.

When using static tri-level models with fixed PCHUS pairs, the selection of suboptimal pairs, e.g., (RES1, RES2) and (RES2, RES3), leads to significantly higher TCs. Only the manually selected



(a) Proposed TOATPs



(b) TIPs

Figure 6.8: Worst-case PCHUS in each iteration

pair (RES1, RES3) achieves a competitive cost of ¥902,810.6. In contrast, the adaptive quad-level model achieves the lowest TC for microgrid operation, matching the best static PCHUS. However, it is crucial to note that the optimal RES pair (RES1, RES3) cannot be known in advance without the adaptive second-level model. The quad-level model inherently excludes overly conservative pairs (e.g., weakly correlated RES units) and autonomously identifies this optimal pair, without requiring any prior assumptions and exhaustive testing. This demonstrates the model's ability to dynamically optimize uncertainty set selection and avoid the pitfalls of suboptimal fixed configurations. Despite minimizing conservatism, the quad-level model guarantees robustness even as it reduces

Table 6.3: Role verifications of the second-level model

Uncertainty Set		Total Cost (¥)	First-stage Cost (¥)	Second-stage Cost (¥)	Non-Robust Scenarios
Static PCHUS	(RES1, RES2)	911347.2	573000.4	338346.8	0
	(RES1, RES3)	902810.6	573000.4	329810.2	0
	(RES2, RES3)	916184.3	573000.4	343183.9	0
Proposed Adaptive PCHUS		902810.6	573000.4	329810.2	0

conservatism. Every solution generated by this approach covers 100% of RES scenarios, with zero non-robust outcomes reported in Table 6.3. This is achieved because the constructed PCHUS always encloses all historical data points, ensuring comprehensive scenario coverage without relying on heuristic tuning and trial-and-error methods common in tri-level frameworks.

6.4.5 Impact Analysis of Temporal Correlations on Solutions

In real-world applications, RES outputs exhibit both spatial correlations across different units and temporal correlations over time. To capture these uncertainties more accurately, the proposed PCHUS framework is extended to incorporate spatio-temporal correlations, enabling a more comprehensive representation of RES uncertainty. Table 6.4 compares the performance of the spatial-only and spatio-temporal PCHUS models in terms of TC and computing time (CT). The results indicate that accounting for temporal dependencies leads to a notable reduction in second-stage (real-time balancing) costs, particularly at lower confidence levels. For instance, at a 95% confidence level, spatio-temporal modeling achieves up to a 33.2% cost reduction. This improvement stems from the ability of the spatio-temporal PCHUS to exclude unrealistic RES fluctuation scenarios, effectively tightening the uncertainty set and yielding less conservative solutions. While the inclusion of temporal correlations slightly increases the computational burden due to the higher dimensionality of the uncertainty set, the trade-off is justifiable considering the achieved cost savings.

6.4.6 Impact Analysis of Various RES Clusters

HMGs typically incorporate high penetration of RES units. These RES units are often co-located within the same station or geographic area and experience identical environmental conditions (e.g.,

Table 6.4: Impact analysis of temporal correlations on solutions

Uncertainty Set		Total Cost (¥)	First-stage Cost (¥)	Second-stage Cost (¥)	CT (s)
Spatiality	PCHUS-99%	763888.2	573814.9	190073.3	32.6
	PCHUS-97%	725446.7	573445.4	152001.3	45.7
	PCHUS-95%	708120.5	573867.5	134253.0	42.1
Spatiality & Temporality	PCHUS-99%	750329.4	573032.6	177296.8	59.6
	PCHUS-97%	680564.6	573711.4	106853.2	49.0
	PCHUS-95%	663534.2	573827.1	89707.1	54.2

solar irradiance or wind speed), leading to highly synchronized output patterns. According to [162], such RES units exhibit strong mutual correlations and can be grouped into representative clusters to reduce modeling complexity and computational burden without sacrificing modeling accuracy. This section investigates the scalability of the proposed approach by evaluating its performance across increasing numbers of RES clusters. Each cluster aggregates multiple RES units with similar generation profiles. Table 6.5 presents the TC and CT for different numbers of clusters: 7 (35 units), 8 (40 units), and 9 (45 units). Here, ”OoM” denotes an ”out-of-memory” error after more than five days of computation.

The results indicate that as the number of RES clusters increases, the computational demand rises significantly, particularly for the FCHUS method. Specifically, FCHUS fails to find a feasible solution when scaled to 9 clusters (45 units) due to memory limitations, revealing its scalability issues. In contrast, the proposed PCHUS approach remains computationally efficient even at this larger scale. Within the PCHUS framework, the parallel version (PCHUS-Parallel) consistently outperforms the non-parallel version (PCHUS-No Parallel), achieving over five times faster computation at 9 clusters (13,783.4s vs. 77,312.6s) while maintaining the same solution quality. These findings demonstrate that the PCHUS formulation, especially when utilizing parallel processing, significantly enhances computational efficiency and ensures feasibility in large-scale scenarios. Moreover, the clustering strategy used in this study provides a practical and scalable solution for applying the proposed approach to real-world microgrid systems with many RES units. By grouping correlated units into manageable clusters, the method keeps computational complexity under control, making the proposed method well-suited for HMG applications.

Table 6.5: Impacts of RES cluster number on computational burden

Number of Clusters (Units)	FCHUS[160] TC (¥)	FCHUS[160] CT (s)	PCHUS-No Parallel TC (¥)	PCHUS-No Parallel CT (s)	PCHUS-Parallel TC (¥)	PCHUS-Parallel CT (s)
7 (35)	1616882.8	6053.5	1629196.5	1743.3	1629196.5	106.8
8 (40)	1804286.5	47987.9	1825066.2	3210.8	1825066.2	113.8
9 (45)	OoM	OoM	2050638.3	77312.6	2050638.3	13783.4

6.5 Summary

To manage the uncertainty of massive renewable energy sources (RESs) in a hybrid AC/DC microgrid, this paper introduces a data-driven pair convex hull uncertainty set (PCHUS) for modeling RES uncertainty with adjustable size. A quad-level energy management model aimed at minimizing solution conservativeness while ensuring robustness against RES fluctuations is also developed. To solve this quad-level model, a column-and-constraint generation algorithm is enhanced by incorporating a quasi-Taguchi orthogonal array testing and a parallel-capable solution architecture. Case study results show that the proposed adjustable size method can flexibly adjust the conservativeness of the energy management solution based on any given confidence level. Compared to the advanced full convex hull uncertainty set (FCHUS), the PCHUS solution closely approximates FCHUS with significantly fewer required vertices. Additionally, the tailored algorithm not only improves the ability to avoid local optimal traps but also significantly accelerates the solving process, especially when considering more than eight RES units.

Chapter 7

Data-Driven Uncertainty Set Optimization and Semi-Convex Relaxation for Microgrid Operation with Hydrogen Energy Storage and Electric Vehicles

Despite recent advancements, several critical research gaps impede the practical implementation of robust microgrid dispatch. First, prevailing uncertainty modeling techniques, such as the CHUS [15], [160], introduce a massive number of vertices that render the subsequent dispatch problem computationally intractable. These sets are also highly sensitive to data outliers, which can lead to overly conservative and economically inefficient solutions. Second, the inclusion of realistic complementary constraints for HMG components, such as BDCs, HESS units, and EVs, introduces significant non-convex constraints. These non-convexities transform the optimization into an NP-hard problem, forcing an unresolved trade-off between computationally intensive models that struggle to guarantee optimality and overly simplified relaxations that may yield infeasible solutions. Third, the existing nested C&CG algorithms for robust dispatch models suffer from a heavy computational burden that makes them intractable for the large-scale systems they are intended to manage, especially with a large number of BDCs, HESS units and EVs. Consequently, there is an urgent need for a computationally efficient, robust dispatch framework and algorithm that can simultaneously provide accurate representations of uncertainty, effectively handle non-convex constraints, and scale to practical, large-scale microgrid applications.

To address the aforementioned research gaps, this paper introduces a data-driven, probability-aided optimization method to characterize RES uncertainty, alongside a C&CG algorithm that integrates a semi-convex relaxation procedure for efficient model solution. The contributions of this work are threefold:

1) Data-Driven Tight Uncertainty Set Optimization: To reduce the number of vertices in CHUS, we formulate a data-driven tight uncertainty set optimization method by minimizing the volume of the uncertainty set. The formulated set not only approximates the CHUS as closely as possible but also enables the number of its vertices to be controllable. Furthermore, we construct a probability-driven DCUS (PDCUS) by leveraging the distribution information of the historical RES data. The constructed PDCUS addresses the issue of excessive conservatism in solutions caused by outliers, making the resulting scheme robust to anomalous data and more aligned with actual operating conditions.

2) Semi-Convex-Relaxation Procedure for Non-Convex Constraints: To address the non-convex complementary constraints in BDCs, HES units and EVs, we develop a semi-convex-relaxation procedure by iteratively identifying and adding the exact non-convex constraints to the solution process. Different from existing convex relaxation methods, our procedure can minimize the required number of non-convex constraints and thus reduce the computational burdens of solving the model while guaranteeing the accuracy of the solution.

3) Customized Solution Algorithm: To solve the robust dispatch model developed for HMG with the integration of HES units and massive EVs, we customize a modified C&CG algorithm embedded with the developed semi-convex relaxation procedure. Our customized algorithm can significantly accelerate the solution process while avoiding falling into the local optimal solution.

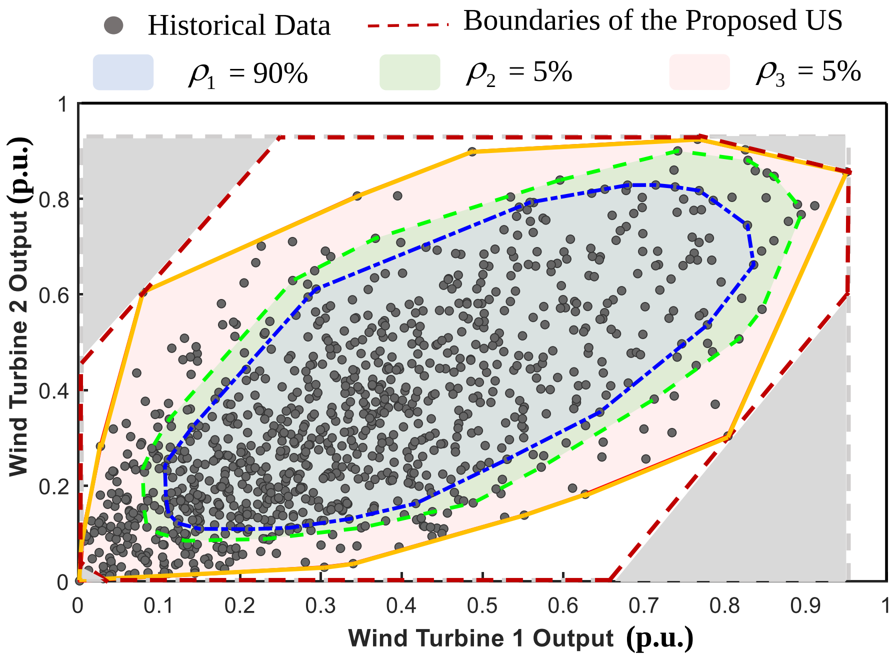


Figure 7.1: Probability-driven data-correlated uncertainty set construction.

7.1 Probability-Driven Data-Correlated Uncertainty Set Construction

The construction of the developed MIUS for two RES units is illustrated in Fig. 7.1. Unlike BbUS [161] and PUS [163], which lead to overly conservative solutions, our multi-interval approach introduces layered uncertainty ranges. Our structure allows for accurate modeling of scenario-dependent RES uncertainties, where different intervals represent distinct probabilistic solutions. By decoupling the whole uncertainty set into several interval-based sub-uncertainty sets rather than imposing uniform bounds, our multi-interval set can enhance flexibility and robustness, potentially reducing the conservatism of the solution compared to existing sets while ensuring robustness against RES uncertainty.

7.1.1 Existing Uncertainty Sets

BbUS, also referred to as interval uncertainty, is one of the most straightforward approaches. It characterizes the uncertain output of each RES unit, u_n, as varying independently within a defined hyperrectangular region. The formulation is given by:

$$\mathcal{V}^{\text{BbUS}}=\left\{u_n \left| \begin{array}{l} u_n = \hat{u}_n+\overline{\varepsilon}_n(\overline{u}_n - \hat{u}_n)+\underline{\varepsilon}_n(\underline{u}_n - \hat{u}_n) \\ \overline{\varepsilon}_n + \underline{\varepsilon}_n \leq 1, \overline{\varepsilon}_n, \underline{\varepsilon}_n \in [0,1], \forall n \end{array} \right. \right\} \tag{7.1}$$

Here, $\hat{u}_n$ is the nominal forecasted output for RES unit n. $\overline{u}_n$ and $\underline{u}_n$ are the maximum and minimum possible outputs, respectively. The parameters $\overline{\varepsilon}_n$ and $\underline{\varepsilon}_n$ represent the normalized extent of positive and negative deviation from the nominal forecast.

Similar to the BbUS, PUS incorporates an additional linear inequality, often termed an uncertainty budget Γ, which constrains the sum of normalized deviations across all RES units:

$$\mathcal{V}^{\text{PUS}}=\left\{u_n \left| \begin{array}{l} u_n = \hat{u}_n+\overline{\varepsilon}_n(\overline{u}_n - \hat{u}_n)+\underline{\varepsilon}_n(\underline{u}_n - \hat{u}_n) \\ \overline{\varepsilon}_n + \underline{\varepsilon}_n \leq 1, \overline{\varepsilon}_n, \underline{\varepsilon}_n \in [0,1], \forall n \\ \sum_{n=1}^{N} (\overline{\varepsilon}_n + \underline{\varepsilon}_n) \leq \Gamma \end{array} \right. \right\} \tag{7.2}$$

EUS characterizes the RES outputs as belonging to an ellipsoid, which is defined by its center, typically the vector of nominal forecasts $\hat{\boldsymbol{u}}$, and a shape matrix $\boldsymbol{\Sigma}$:

$$\mathcal{V}^{\text{EUS}}=\left\{\boldsymbol{u} \left| (\boldsymbol{u}-\hat{\boldsymbol{u}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{u}-\hat{\boldsymbol{u}}) \leq \Gamma \right. \right\} \tag{7.3}$$

A key advantage of the EUS is its ability to capture correlations between the outputs of different RES units through the off-diagonal elements of the shape matrix.

CHUS, also known as a scenario-based polytope, defines the uncertain region for RES outputs as the convex hull of a finite collection of discrete scenarios. Each scenario $\boldsymbol{u}_d$ represents a specific possible realization of the RES outputs. Any point within the CHUS can be expressed as a convex combination of these predefined scenarios:

$$\mathcal{V}^{\text{CHUS}}=\left\{u_n \left| \begin{array}{l} u_n = \hat{u}_n+\sum_{s\in\mathcal{S}^{\text{CHUS}}} \varepsilon_{ns}(u_{ns}-\hat{u}_n), \forall n \\ 0 \leq \varepsilon_{ns} \leq 1, \forall n, s \\ \sum_{s\in\mathcal{S}^{\text{CHUS}}} \varepsilon_{ns} = 1, \forall n \end{array} \right. \right\} \tag{7.4}$$

Here, $\mathcal{S}^{\text{CHUS}}$ is the set of discrete vertex indices, u_{ns} is the output of the n RES unit in the s-th vertex, and ε_{ns} is the weighting factor. Since the number of vertices in CHUS is excessively large, the robust counterpart remains computationally unmanageable.

7.1.2 Probability-Driven Data-Correlated Uncertainty Set

This section outlines the modelling for constructing a DCUS that has the minimum possible volume while still enclosing all historical data observations. This process is visually depicted in Fig. 7.1. The core idea is to optimize specific hyperplanes that form the DCUS boundaries. This optimization process can be viewed as maximizing the volume of the original BbUS that is cut off, represented by the shaded red area in Fig. 7.1.

A primary constraint is that every hyperplane defining the DCUS must align with a vertex from the BbUS, as the DCUS is constructed as a subset of the BbUS. The objective function for selecting these hyperplanes aims to maximize this removed volume. It is formulated by summing the volumes of the polyhedrons truncated from each BbUS vertex (indexed by k) ranging from 1 to 2^N:

$$\max_{\boldsymbol{x}_k^{\text{DCUS}}, \boldsymbol{a}_k^T, b_k, \delta_{kn}} \sum_{k=1}^{2^N} \frac{1}{N!} \prod_{n=1}^{N} \delta_{kn} \tag{7.5}$$

Here, δ_{kn} signifies the geometric distance between the k-th vertex $\boldsymbol{x}_k^{\text{BbUS}}$ of the BbUS and the corresponding j-th vertex $\boldsymbol{x}_j^{\text{DCUS}}$ of the DCUS along the n-th dimension, where N is the total number of dimensions/RES units. The index j for the DCUS vertex adjacent to the k-th BbUS vertex is determined by:

$$j = (k-1)N + n, \quad n = 1, 2, \ldots, N \tag{7.6}$$

The geometric distances δ_{kn} are defined such that the distance is zero in all dimensions except for the n-th dimension:

$$\delta_{kn} = \sigma_{kn}(x_{kn}^{\text{BbUS}} - x_{jn}^{\text{DCUS}}), \quad \forall n \tag{7.7}$$

$$x_{kp}^{\text{BbUS}} - x_{jp}^{\text{DCUS}} = 0, \quad \forall p = 1, 2, \ldots, n-1, n+1, \ldots, N \tag{7.8}$$

where x_{kn}^{BbUS} and x_{kn}^{DCUS} are the n-th components of vectors $\boldsymbol{x}_k^{\text{BbUS}}$ and $\boldsymbol{x}_k^{\text{DCUS}}$ respectively. σ_{kn} is a binary parameter ($+1$ or -1) indicating the orientation of $\boldsymbol{x}_k^{\text{BbUS}}$ relative to the BbUS boundary in the n-th dimension.

The model incorporates constraints reflecting the geometric positioning of the hyperplanes ($\boldsymbol{a}_k^T \boldsymbol{x} = b_k$), the vertices of both the BbUS $\mathcal{V}^{\text{BbUS}}$ and DCUS $\mathcal{V}^{\text{DCUS}}$, and the historical RES data points ($\boldsymbol{x}_h^{\text{Hist}}$ in set $\mathcal{H}$):

Vertices $\boldsymbol{x}_j^{\text{DCUS}}$ of the DCUS lie on the hyperplane:

$$\boldsymbol{a}_k^T \boldsymbol{x}_j^{\text{DCUS}} = b_k, \quad \forall \boldsymbol{x}_j^{\text{DCUS}} \in \mathcal{V}^{\text{DCUS}}, \forall k = 1, 2, \ldots, 2^N \tag{7.9}$$

Historical RES data are within the hyperplane boundaries:

$$\boldsymbol{a}_k^T \boldsymbol{x}_h^{\text{Hist}} \leq b_k, \quad \forall h \in \mathcal{H}, \forall k = 1, 2, \ldots, 2^N \tag{7.10}$$

Vertices $\boldsymbol{x}_k^{\text{BbUS}}$ of the BbUS are outside the region defined by the hyperplane for historical data:

$$\boldsymbol{a}_k^T \boldsymbol{x}_k^{\text{BbUS}} \geq b_k, \quad \forall \boldsymbol{x}_k^{\text{BbUS}} \in \mathcal{V}^{\text{BbUS}}, \forall k = 1, 2, \ldots, 2^N \tag{7.11}$$

Hyperplane coefficients are non-zero:

$$\boldsymbol{a}_k^T \neq \mathbf{0}, \quad b_k \neq 0, \quad \forall k = 1, 2, \ldots, 2^N \tag{7.12}$$

To address the computational burden for large-scale systems, two adjustments are made:

7.1.2.1 Decomposition

The optimization problem (7.5) can be separated for each hyperplane. It decomposes into 2^N smaller, independent nonlinear optimization problems that can be solved concurrently. The objective for each k-th vertex subproblem becomes:

$$\max_{\boldsymbol{x}_k^{\text{DCUS}}, \boldsymbol{a}_k^T, b_k, \delta_{kn}} \frac{1}{N!} \prod_{n=1}^{N} \delta_{kn} \tag{7.13}$$

Using logarithms simplifies this objective to:

$$\max_{\boldsymbol{x}_k^{\mathrm{DCUS}}, \boldsymbol{a}_k^T, b_k, \delta_{kn}} \sum\nolimits_{n=1}^{N} \ln \delta_{kn} \tag{7.14}$$

7.1.2.2 Constraint Simplification

While the DCUS must contain all historical data points, a brute-force check against every point by Eq. (7.10) is computationally intensive. To address this, the model simplifies the constraint by ensuring that the hyperplanes of the DCUS enclose the vertices $\boldsymbol{x}_s^{\mathrm{CHUS}}$ of the CHUS $\mathcal{V}^{\mathrm{CHUS}}$:

$$\boldsymbol{a}_k^T \boldsymbol{x}_s^{\mathrm{CHUS}} \leq b_k, \quad \forall s \in \mathcal{S}^{\mathrm{CHUS}}, \forall k = 1, 2, \ldots, 2^N \tag{7.15}$$

Proposition 9. *As long as all $\boldsymbol{x}_s^{\mathrm{CHUS}}$ satisfy the constraint $\boldsymbol{a}_k^T \boldsymbol{x}_s^{\mathrm{CHUS}} \leq b_k$, then all massive historical data points $\boldsymbol{x}_h^{\mathrm{Hist}}$ satisfy the constraint $\boldsymbol{a}_k^T \boldsymbol{x}_h^{\mathrm{Hist}} \leq b_k$.*

Proof. By definition of convex hull, every historical data point $\boldsymbol{x}_h^{\mathrm{Hist}}$ is either a vertex of the CHUS (i.e., $\boldsymbol{x}_h^{\mathrm{Hist}} = \boldsymbol{x}_s^{\mathrm{CHUS}}, \exists s$), or a convex combination of the vertices of the CHUS. For any historical data point $\boldsymbol{x}_h^{\mathrm{Hist}}$, it can be expressed as:

$$\boldsymbol{x}_h^{\mathrm{Hist}} = \sum\nolimits_{s \in \mathcal{S}^{\mathrm{CHUS}}} \varepsilon_s \boldsymbol{x}_s^{\mathrm{CHUS}}, \varepsilon_s \geq 0, \forall s \in \mathcal{S}^{\mathrm{CHUS}} \tag{7.16}$$

where $\sum_{s \in \mathcal{S}^{\mathrm{CHUS}}} \varepsilon_s = 1$, and $\mathcal{S}^{\mathrm{CHUS}}$ is the set of vertex indices of the CHUS. Given that the constraint $\boldsymbol{a}_k^T \boldsymbol{x}_s^{\mathrm{CHUS}} \leq \boldsymbol{b}_k$ holds for all vertices $\boldsymbol{x}_s^{\mathrm{CHUS}}$, we can apply this to any historical data point $\boldsymbol{a}_k^T \boldsymbol{x}_h^{\mathrm{Hist}} = \boldsymbol{a}_k^T \left(\sum_{s \in \mathcal{S}^{\mathrm{CHUS}}} \varepsilon_s \boldsymbol{x}_s^{\mathrm{CHUS}} \right) = \sum_{s \in \mathcal{S}^{\mathrm{CHUS}}} \varepsilon_s (\boldsymbol{a}_k^T \boldsymbol{x}_s^{\mathrm{CHUS}})$.

Since $\boldsymbol{a}_k^T \boldsymbol{x}_s^{\mathrm{CHUS}} \leq \boldsymbol{b}_k, \forall s \in \mathcal{S}^{\mathrm{CHUS}}, \varepsilon_s \geq 0$, we can obtain that $\boldsymbol{a}_k^T \boldsymbol{x}_h^{\mathrm{Hist}} = \sum_{s \in \mathcal{S}^{\mathrm{CHUS}}} \varepsilon_s (\boldsymbol{a}_k^T \boldsymbol{x}_s^{\mathrm{CHUS}}) \leq \sum_{s \in \mathcal{S}_{\mathrm{CHUS}}} \varepsilon_s \boldsymbol{b}_k$. We notice that $\boldsymbol{b}_k$ is a constant and $\sum_{s \in \mathcal{S}^{\mathrm{CHUS}}} \varepsilon_s = 1$, yielding $\boldsymbol{a}_k^T \boldsymbol{x}_h^{\mathrm{Hist}} \leq \boldsymbol{b}_k \sum_{s \in \mathcal{S}^{\mathrm{CHUS}}} \varepsilon_s = \boldsymbol{b}_k \cdot 1 = \boldsymbol{b}_k$. Thus, $\boldsymbol{a}_k^T \boldsymbol{x}_h^{\mathrm{Hist}} \leq \boldsymbol{b}_k$ for all historical data points $\boldsymbol{x}_h^{\mathrm{Hist}}$. □

As established in Proposition 9, checking Eq. (7.15) only at the CHUS vertices is sufficient to guarantee that Eq. (7.10) is satisfied for all historical data points. This substitution greatly enhances computational tractability, since the number of CHUS vertices is finite and typically much smaller than the number of historical data points.

Consequently, the hyperplane constraints defining the minimum volume DCUS are determined by solving the following nonlinear programming model for each vertex index k from 1 to 2^N:

$$\max_{\boldsymbol{x}_k^{Data}, \boldsymbol{a}_k^T, b_k, \delta_{kn}} \sum\nolimits_{n=1}^{N} \ln \delta_{kn} \tag{7.17}$$

$$\text{s.t.} \quad \text{Eqs. (7.7) and (7.8), (7.9), (7.11) and (7.12), (7.15)} \tag{7.18}$$

After finding the minimum volume of the uncertainty set through the above optimization, DCUS is expressed as follows:

$$\mathcal{V}^{\text{DCUS}} = \left\{ u_n \left| \begin{array}{l} u_n = \hat{u}_n + \sum_{s \in \mathcal{S}^{\text{DCUS}}} \varepsilon_{ns}(u_{ns} - \hat{u}_n), \forall n \\ 0 \leq \varepsilon_{ns} \leq 1, \forall n, s \\ \sum_{s \in \mathcal{S}^{\text{DCUS}}} \varepsilon_{ns} = 1, \forall n \end{array} \right. \right\} \tag{7.19}$$

Different from CHUS, where the size and shape are uncontrollable and are determined entirely by data distribution, the developed DCUS employs an optimization-driven approach that strategically positions hyperplanes to minimize the uncertainty volume while maintaining complete coverage of historical data. This optimization process essentially carves out unnecessary portions from the BbUS, thereby creating a more refined polytope that balances robustness with economic efficiency.

7.2 Semi-Convex-Relaxation for Complementarity Constraints in Hydrogen Energy Storage Units

7.2.1 Hybrid Microgrid Dispatch Framework and Model

Fig. 7.2 illustrates the typology of a hybrid AC/DC microgrid, which integrates both AC and DC subsystems to enhance energy efficiency and flexibility. In detail, the AC subsystem includes conventional and renewable energy sources such as wind turbines (WTs) and diesel engines (DEs), which supply power to AC loads through an AC bus. This AC bus is interconnected with a DC bus via power

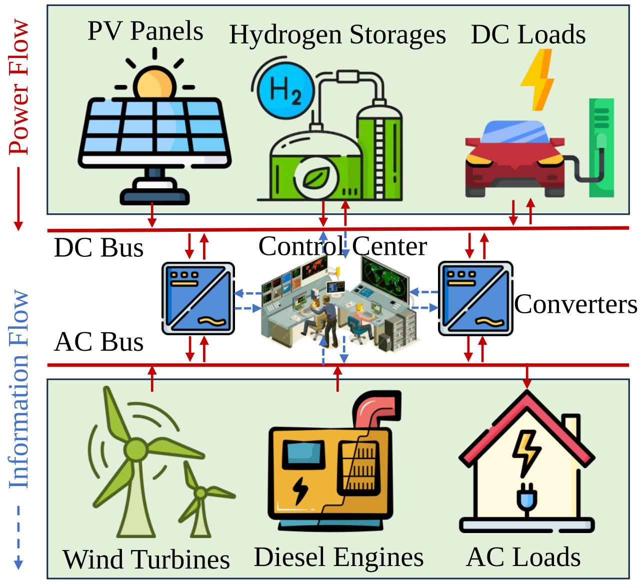


Figure 7.2: Hybrid AC/DC microgrid typology.

electronic converters that enable bidirectional power flow, facilitating efficient energy exchange between the two subsystems. On the DC side, the system incorporates PV panels and hydrogen energy storage (HES) units as primary energy sources, supplying power to various DC loads such as electric vehicles (EVs) that function as flexible loads and grid-supporting storage, adjusting charge and discharge cycles to lower costs and reduce grid stress. WTs and PV panels supply variable renewable energy, prioritized for their minimal operational costs. Hydrogen energy storage absorbs surplus renewables via electrolysis and releases energy during shortages, while EVs function as flexible loads, adjusting charge and discharge cycles to lower costs and reduce grid stress. A centralized control center manages both power and information flows across the entire microgrid, ensuring optimal energy distribution and enhanced system reliability, which minimizes the total operational cost of the microgrid, including DE cost C^{de}, HES degraded cost C^{hes}, renewable curtailment costC^{res}, and converter loss cost C^{bdc}.

$$\min C = C^{\text{de}} + C^{\text{hes}} + C^{\text{ev}} + C^{\text{res}} + C^{\text{bdc}} \tag{7.20}$$

$$C^{\text{de}} = \sum_{t=1}^{\mathcal{T}} \sum_{i=1}^{\mathcal{I}} (n_i^{\text{de}} U_{i,t}^{\text{de}} + f_i^{\text{de}} D_{i,t}^{\text{de}} + c_i^{\text{de}} P_{i,t}^{\text{de}}) \tag{7.21}$$

$$C^{\text{hes}} = \sum_{t=1}^{\mathcal{T}} \sum_{i=1}^{\mathcal{I}} c_i^{\text{hes}} (p_{i,t}^{\text{ES}+} + p_{i,t}^{\text{ES}-}) \tag{7.22}$$

$$C^{\mathrm{ev}}=\sum\nolimits_{t=1}^{\mathcal{T}}\sum\nolimits_{i=1}^{\mathcal{I}}c_i^{\mathrm{ev}}(p_{i,t}^{\mathrm{EV+}}+p_{i,t}^{\mathrm{EV-}}) \tag{7.23}$$

$$C^{\mathrm{res}}=\sum_{t=1}^{\mathcal{T}}\sum_{i=1}^{\mathcal{I}}(c_i^{\mathrm{w}}\hat{\rho}_{i,t}^{\mathrm{w}}+c_i^{\mathrm{pv}}\hat{\rho}_{i,t}^{\mathrm{pv}}) \tag{7.24}$$

$$C^{\mathrm{bdc}}=\sum\nolimits_{t=1}^{\mathcal{T}}\sum\nolimits_{l=1}^{\mathcal{L}}c_l^{\mathrm{loss}}(P_{i_lj_l,l,t}^{\mathrm{acdc}}+P_{i_1j_1,l,t}^{\mathrm{dcac}}) \tag{7.25}$$

s.t.

$$I_{i,t}^{\mathrm{de}}=I_{i,t-1}^{\mathrm{de}}+U_{i,t}^{\mathrm{de}}-D_{i,t}^{\mathrm{de}},\forall i,t \tag{7.26}$$

$$I_{i,t}^{\mathrm{de}}\geq\sum\nolimits_{\tau=1}^{T_i^{\mathrm{de,on}}}U_{i,t-\tau}^{\mathrm{de}},I_{i,t}^{\mathrm{de}}\leq 1-\sum\nolimits_{\tau=1}^{T_i^{\mathrm{de,off}}}D_{i,t-\tau}^{\mathrm{de}},\forall i,t \tag{7.27}$$

$$I_{i,t}^{\mathrm{de}}\underline{P}_i^{\mathrm{de}}\leq P_{i,t}^{\mathrm{de}}\leq I_{i,t}^{\mathrm{de}}\bar{P}_i^{\mathrm{de}},\forall i,t \tag{7.28}$$

$$I_{i,t}^{\mathrm{de}}\underline{Q}_i^{\mathrm{de}}\leq Q_{i,t}^{\mathrm{de}}\leq I_{i,t}^{\mathrm{de}}\bar{Q}_i^{\mathrm{de}},\forall i,t \tag{7.29}$$

$$-R_i^{\mathrm{P,de}}\leq P_{i,t}^{\mathrm{de}}-P_{i,t-1}^{\mathrm{de}}\leq R_i^{\mathrm{P,de}},\forall i,t \tag{7.30}$$

$$-R_i^{\mathrm{Q,de}}\leq Q_{i,t}^{\mathrm{de}}-Q_{i,t-1}^{\mathrm{de}}\leq R_i^{\mathrm{Q,de}},\forall i,t \tag{7.31}$$

$$0\leq\hat{\rho}_{i,t}^{\mathrm{w}}\leq\hat{P}_{i,t}^{\mathrm{w}},\forall i,t \tag{7.32}$$

$$V_{j,t}^{\mathrm{ac}}=V_{i,t}^{\mathrm{ac}}-\left(r_{ij}^{\mathrm{ac}}H_{ij,t}^{\mathrm{ac}}+x_{ij}^{\mathrm{ac}}G_{ij,t}^{\mathrm{ac}}\right)/V_{0,t}^{\mathrm{ac}},\forall ij,t \tag{7.33}$$

$$\underline{H}_{ij,t}^{\mathrm{ac}}\leq H_{ij,t}^{\mathrm{ac}}\leq\bar{H}_{ij,t}^{\mathrm{ac}},\forall ij,t \tag{7.34}$$

$$\underline{G}_{ij,t}^{\mathrm{ac}}\leq G_{ij,t}^{\mathrm{ac}}\leq\bar{G}_{ij,t}^{\mathrm{ac}},\forall ij,t \tag{7.35}$$

$$\underline{V}_{i,t}^{\mathrm{ac}}\leq V_{i,t}^{\mathrm{ac}}\leq\bar{V}_{i,t}^{\mathrm{ac}},\forall i,t \tag{7.36}$$

$$\begin{aligned}P_{i,t}^{\mathrm{de}}&-L_{i,t}^{\mathrm{ac}}+\sum\nolimits_{j_1\in\pi(i_1)}\sum\nolimits_{l\in\mathcal{L}}\left(P_{i_1j_1,l,t}^{\mathrm{ac}}-P_{i_1j_1,l,t}^{\mathrm{acdc}}\right)\\&=\sum\nolimits_{j\in\pi(i)}H_{ij,t}^{\mathrm{ac}}-\sum\nolimits_{k\in\delta(i)}H_{ki,t}^{\mathrm{ac}}-\left(\hat{P}_{i,t}^{\mathrm{w}}-\hat{\rho}_{i,t}^{\mathrm{w}}\right),\forall i,t\end{aligned} \tag{7.37}$$

$$Q_{i,t}^{\mathrm{de}}-Q_{i,t}^{\mathrm{ac}}=\sum\nolimits_{j\in\pi(i)}G_{ij,t}^{\mathrm{ac}}-\sum\nolimits_{k\in\delta(i)}G_{ki,t}^{\mathrm{ac}},\forall i,t \tag{7.38}$$

$$0\leq\hat{\rho}_{i,t}^{\mathrm{pv}}\leq\hat{P}_{i,t}^{\mathrm{pv}},\forall i,t \tag{7.39}$$

$$I_{i,t}^{\mathrm{ES+}}+I_{i,t}^{\mathrm{ES-}}\leq 1,\forall i,t \tag{7.40}$$

$$I_{i,t}^{\text{ES+}}\underline{P}_{i}^{\text{ES+}} \leq p_{i,t}^{\text{ES+}} \leq I_{i,t}^{\text{ES+}}\bar{P}_{i}^{\text{ES+}}, \forall i,t \tag{7.41}$$

$$I_{i,t}^{\text{ES}-}\underline{P}_{i}^{\text{ES}-} \leq p_{i,t}^{\text{ES}-} \leq I_{i,t}^{\text{ES}-}\bar{P}_{i}^{\text{ES}-}, \forall i,t \tag{7.42}$$

$$S_{i,1}^{\text{ES}} = S_{i,T_{\text{T}}}^{\text{ES}} = 0.7, \forall i,t \tag{7.43}$$

$$\underline{S}_{i}^{\text{ES}} \leq S_{i,t}^{\text{ES}} \leq \bar{S}_{i}^{\text{ES}}, \forall i,t \tag{7.44}$$

$$S_{i,t+1}^{\text{ES}} = S_{i,t}^{\text{ES}} + (p_{i,t}^{\text{ES+}}\eta_{i}^{\text{ES+}} - p_{i,t}^{\text{ES}-}/\eta_{i}^{\text{ES}-})/G_{i}^{\text{HES}}, \forall i,t \tag{7.45}$$

$$S_{i,t+1}^{\text{EV}} = (p_{i,t}^{\text{EV+}}\eta_{i}^{\text{EV+}} - p_{i,t}^{\text{EV}-}/\eta_{i}^{\text{EV}-})/G_{i}^{\text{EV}} + S_{i,t}^{\text{EV}} \tag{7.46}$$

$$\underline{S}_{i}^{\text{EV}} \leq S_{i,t}^{\text{EV}} \leq \bar{S}_{i}^{\text{EV}} \tag{7.47}$$

$$S_{i,t}^{\text{EV}} = S_{i}^{\text{EV,rem}}, \forall t = t_{i}^{\text{in}} \tag{7.48}$$

$$S_{i,t}^{\text{EV}} \geq S_{i}^{\text{EV,req}}, \forall t = t_{i}^{\text{out}} \tag{7.49}$$

$$0 \leq p_{i,t}^{\text{EV+}} \leq I_{i,t}^{\text{EV+}}\bar{P}_{i}^{\text{EV+}}, \forall t_{i}^{\text{in}} \leq t \leq t_{i}^{\text{out}} \tag{7.50}$$

$$0 \leq p_{i,t}^{\text{EV}-} \leq I_{i,t}^{\text{EV}-}\bar{P}_{i}^{\text{EV},-}, \forall t_{i}^{\text{in}} \leq t \leq t_{i}^{\text{out}} \tag{7.51}$$

$$I_{i,t}^{\text{EV+}} + I_{i,t}^{\text{EV}-} \leq 1, \forall t_{i}^{\text{in}} \leq t \leq t_{i}^{\text{out}} \tag{7.52}$$

$$V_{j,t}^{\text{dc}} = V_{i,t}^{\text{dc}} - \left(r_{ij}^{\text{dc}}H_{ij,t}^{\text{dc}}\right)/V_{0,t}^{\text{dc}}, \forall ij,t \tag{7.53}$$

$$\underline{H}_{ij,t}^{\text{dc}} \leq H_{ij,t}^{\text{dc}} \leq \bar{H}_{ij,t}^{\text{dc}}, \forall ij,t \tag{7.54}$$

$$\underline{V}_{i,t}^{\text{dc}} \leq V_{i,t}^{\text{dc}} \leq \bar{V}_{i,t}^{\text{dc}}, \forall i,t \tag{7.55}$$

$$\begin{aligned} & p_{i,t}^{\text{ES}-} - p_{i,t}^{\text{ES+}} + p_{i,t}^{\text{EV}-} - p_{i,t}^{\text{EV+}} - L_{i,t}^{\text{dc}} \\ & + \sum\nolimits_{j_1 \in \pi(i_1)} \sum\nolimits_{l \in \mathcal{L}} \left(P_{i_1 j_1,l,t}^{\text{dc}} - P_{i_1 j_1,l,t}^{\text{dcac}}\right) \\ & = \sum\nolimits_{j \in \pi(i)} H_{ij,t}^{\text{dc}} - \sum\nolimits_{k \in \delta(i)} H_{ki,t}^{\text{dc}} - \left(\hat{P}_{i,t}^{\text{pv}} - \hat{\rho}_{i,t}^{\text{pv}}\right), \forall i,t \end{aligned} \tag{7.56}$$

$$0 \leq P_{i_1 j_1,l,t}^{\text{acdc}} \leq I_{l,t}^{\text{acdc}}\bar{P}_{i_1 j_1,l}^{\text{acdc}}, \forall l,t \tag{7.57}$$

$$0 \leq P_{i_1 j_1,l,t}^{\text{dcac}} \leq I_{l,t}^{\text{dcac}}\bar{P}_{i_1 j_1,l}^{\text{dcac}}, \forall l,t \tag{7.58}$$

$$I^{\text{acdc}}_{l,t} + I^{\text{dcac}}_{l,t} \leq 1, \forall l, t \tag{7.59}$$

$$P^{\text{ac}}_{i_1 j_1,l,t} = P^{\text{dcac}}_{i_1 j_1,l,t} \cdot \eta^{\text{dcac}}_{\text{BDC},l}(P^{\text{dcac}}_{i_1 j_1,l,t}), \forall l, t \tag{7.60}$$

$$P^{\text{dc}}_{i_1 j_1,l,t} = P^{\text{acdc}}_{i_1 j_1,l,t} \cdot \eta^{\text{acdc}}_{\text{BDC},l}(P^{\text{acdc}}_{i_1 j_1,l,t}), \forall l, t \tag{7.61}$$

$$P^{\text{ac}}_{i_1 j_1,l,t} = o_{0,l} + o_{1,l} P^{\text{dcac}}_{i_1 j_1,l,t}, \forall l, t \tag{7.62}$$

$$P^{\text{dc}}_{i_1 j_1,l,t} = o_{0,l} + o_{1,l} P^{\text{acdc}}_{i_1 j_1,l,t}, \forall l, t \tag{7.63}$$

Eq. (7.21) calculates the total cost of running the DEs, including fixed costs for being online, costs for starting up, and variable costs based on the amount of power produced. Eqs. (7.22) and (7.23) evaluate the degraded cost of HES units and EVs. Eq. (7.24) imposes a penalty for any available renewable energy that the microgrid fails to use. Eq. (7.25) accounts for the cost of energy that is lost as heat when power is converted between AC and DC forms by the bidirectional converters (BDCs). Eqs. (7.26)-(7.31) ensure the engines operate within their physical limits, including on/off status logic, minimum and maximum power output levels, and how quickly they can ramp their power output up and down. Eq. (7.32) ensures that any wind power curtailed does not exceed the total wind power available. Eqs. (7.33)-(7.36) model the electrical behavior of the AC grid. They include linearized power flow equations to calculate voltage drops and power flows, thermal limits to prevent power lines from overheating, and voltage limits to ensure power quality and equipment safety. Eqs. (7.37) and (7.38) ensure that at every bus in the AC grid, the power flowing in equals the power flowing out, including both active power and reactive power. Eq. (7.39) ensures that any solar power curtailed does not exceed the total solar power available. Eqs. (7.40)-(7.45) govern the operation of the HES unit. Eq. (7.40) ensures that the hydrogen storage unit cannot charge and discharge simultaneously. It is important to note that this non-convex complementarity constraint is physically fundamental to most energy storage technologies, such as battery energy storage systems (BESS) and pumped hydro storage. Eq. (7.41) limits the charging power within its minimum and maximum values. Eq. (7.42) limits the discharging power within its minimum and maximum values. Eq. (7.43) specifies that the initial and final SOC of the hydrogen storage unit is set to 70% of its capacity. Eq. (7.44) ensures that the SOC remains within its minimum and maximum limits. Eq.

(7.45) defines the SOC update equation based on charging and discharging, considering efficiency losses. Eqs. (7.46)-(7.52) model grid-interactive EVs. Eq. (7.46) defines the SOC update equation based on charging and discharging. Eq. (7.47) ensures SOC remains within its limits. Eq. (7.48) sets the SOC at the arrival time. Eq. (7.49) ensures SOC meets the required level before departure. Eqs. (7.50)-(7.51) bound the charging and discharging power within their limits. Eq. (7.52) ensures the EV is not charging and discharging simultaneously. Eqs. (7.53)-(7.56) model the DC portion of the grid, including linearized power flow, line limits, voltage limits, and the crucial DC power balance that connects all DC components (PV, HES, EVs, DC loads). Eqs. (7.57)-(7.63) model the power converters that link the AC and DC subgrids. Eqs. (7.57) and (7.58) set the power flow limits for the converter. Eq. (7.59) is the non-convex complementarity constraint preventing the converter from simultaneously sending power from AC-to-DC and DC-to-AC. Eqs. (7.60)-(7.63) model the energy efficiency of the BDC using a simplified but high-accuracy linear approximation of the true non-linear efficiency curve to make the problem easier to solve computationally.

7.2.2 Closed-Loop Semi-Convex-Relaxation Procedure

The matrix form of the deterministic microgrid dispatch model is rewritten as follows:

$$\min_{\boldsymbol{x}=(\boldsymbol{x}_1,\boldsymbol{x}_2),\boldsymbol{y}} \boldsymbol{a}^T\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{y} \tag{7.64}$$

$$\text{s.t.} \quad \mathrm{Cx_1 \leq c, x_1 \in \{0,1\}} \tag{7.65}$$

$$\boldsymbol{D}_1\boldsymbol{x}_1 + \boldsymbol{D}_2\boldsymbol{x}_2 + \boldsymbol{E}\boldsymbol{y} \leq \hat{\boldsymbol{u}}, \boldsymbol{x}_2 \in \{0,1\} \tag{7.66}$$

$$\boldsymbol{x}_2' + \boldsymbol{x}_2'' \leq 1, \boldsymbol{x}_2 = (\boldsymbol{x}_2', \boldsymbol{x}_2''), \boldsymbol{x}_2', \boldsymbol{x}_2'' \in \{0,1\} \tag{7.67}$$

Here, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{C}$, $\boldsymbol{c}$, $\boldsymbol{D}_1$, $\boldsymbol{D}_2$ and $\boldsymbol{E}$ are coefficient matrices. $\boldsymbol{x}$ and $\boldsymbol{y}$ represent binary and continuous variables for the deterministic microgrid dispatch model (7.20)-(7.63). In detail, $\boldsymbol{x}_1$ represents binary commitment decisions for DE units, while $\boldsymbol{x}_2$ denotes binary variables for the energy store/release states of HES units and EVs, and the inversion and rectification states of BDCs. Note that, the mutual exclusion constraint (7.67) prevents simultaneous charging and discharging actions for each HES unit and EV, and avoids simultaneous inverting and rectifying operations for each BDC in any timeslot, by introducing $\boldsymbol{x}_2'$ and $\boldsymbol{x}_2''$.

Then, the robust microgrid dispatch model under our PDCUS is formulated as follows:

$$\min_{\boldsymbol{x}} \boldsymbol{a}^T\boldsymbol{x} + \mathbb{E}_r[\max_{\boldsymbol{u}_r} \min_{\boldsymbol{y}_r} \boldsymbol{b}^T\boldsymbol{y}_r] \tag{7.68}$$

s.t.

$$\boldsymbol{C}\boldsymbol{x}_1 \leq \boldsymbol{c}, \boldsymbol{x}_1 \in \{0,1\} \tag{7.69}$$

$$\boldsymbol{u}_r \in \mathcal{V}_r^{\text{DCUS}} \subseteq \mathcal{V}^{\text{DCUS}} \tag{7.70}$$

$$\boldsymbol{D}_1\boldsymbol{x}_1 + \boldsymbol{D}_2\boldsymbol{x}_2 + \boldsymbol{E}\boldsymbol{y}_r \leq \boldsymbol{u}_r, \boldsymbol{x}_2 \in \{0,1\} \tag{7.71}$$

$$\boldsymbol{x}_2' + \boldsymbol{x}_2'' \leq 1, \boldsymbol{x}_2 = (\boldsymbol{x}_2', \boldsymbol{x}_2''), \boldsymbol{x}_2', \boldsymbol{x}_2'' \in \{0,1\} \tag{7.72}$$

Here, $\boldsymbol{u}_r$ denotes uncertainty realizations within the PDCUS, and $\boldsymbol{y}_r$ represents adaptive corrective actions that can be taken in response to RES uncertainties.

Note that, the constraint (7.67) in the third level of the robust dispatch model presents non-convexity, significantly increasing computational complexity and affecting the solving approaches. Currently, the state-of-the-art NCC&G algorithms can only solve models with a small scale, and also suffer from a heavy computational burden. This is verified by the subsequent Table 7.2.

To make the model solvable within the minimal computational burden, we design a closed-loop semi-convex-relaxation (CS) procedure to minimize the required number of non-convex constraints (7.67) during the solution process. The procedure is illustrated in **Algorithm 10**. Specifically, we first relax the binary variable $\boldsymbol{x}_2$ to a continuous variable in [0,1], which produces the following formulation, denoted as the semi-convex-relaxation model:

$$(7.68) - (7.70) \tag{7.73}$$

$$\boldsymbol{D}_1\boldsymbol{x}_1 + \boldsymbol{D}_2\boldsymbol{x}_2 + \boldsymbol{E}\boldsymbol{y}_r \leq \boldsymbol{u}_r, \boldsymbol{x}_2 \in [0,1] \tag{7.74}$$

$$\boldsymbol{x}_2' + \boldsymbol{x}_2'' \leq 1, \boldsymbol{x}_2 = (\boldsymbol{x}_2', \boldsymbol{x}_2''), \boldsymbol{x}_2', \boldsymbol{x}_2'' \in [0,1] \tag{7.75}$$

This relaxation can simplify complex, non-convex problems into more solvable forms. However, the solution of such a relaxed model might yield results where, for instance, an HES unit is partially charging and partially discharging simultaneously, which is physically impossible. Then, the simultaneous action metric (SAM) is defined to quantify the violation level of mutual exclusion constraints as follows:

$$\mathrm{SAM}_{i,t}^{\mathrm{HES}} = p_{i,t}^{\mathrm{HES}+} p_{i,t}^{\mathrm{HES}-}, \forall i, t \tag{7.76}$$

$$\mathrm{SAM}_{i,t}^{\mathrm{EV}} = p_{i,t}^{\mathrm{EV}+} p_{i,t}^{\mathrm{EV}-}, \forall i, t \tag{7.77}$$

$$\mathrm{SAM}_{l,t}^{\mathrm{BDC}} = P_{i_1 j_1,l,t}^{\mathrm{acdc}} P_{i_1 j_1,l,t}^{\mathrm{dcac}}, \forall l, t \tag{7.78}$$

SAM provides a numerical value to represent the magnitude of the solution's infeasibility. A non-zero SAM value indicates that one or more mutual exclusion constraints have been violated in the solution of the relaxed model. For $\exists i', l', t', \mathrm{SAM}_{i',t'}^{\mathrm{ES}} \neq 0, \mathrm{SAM}_{i',t'}^{\mathrm{EV}} \neq 0, \mathrm{SAM}_{l',t'}^{\mathrm{acdc}} \neq 0$, the corresponding mutual exclusion constraint is added to ensure the accuracy of the solution, as follows:

$$I_{i',t'}^{\mathrm{ES}+}, I_{i',t'}^{\mathrm{ES}-} \in [0, 1] \Rightarrow I_{i',t'}^{\mathrm{ES}+}, I_{i',t'}^{\mathrm{ES}-} \in \{0, 1\} \tag{7.79}$$

$$I_{i',t'}^{\mathrm{EV}+}, I_{i',t'}^{\mathrm{EV}-} \in [0, 1] \Rightarrow I_{i',t'}^{\mathrm{EV}+}, I_{i',t'}^{\mathrm{EV}-} \in \{0, 1\} \tag{7.80}$$

$$I_{l',t'}^{\mathrm{acdc}}, I_{l',t'}^{\mathrm{dcac}} \in [0, 1] \Rightarrow I_{l',t'}^{\mathrm{acdc}}, I_{l',t'}^{\mathrm{dcac}} \in \{0, 1\} \tag{7.81}$$

It can be seen that the SAM serves as an indicator within an iterative algorithm to guide the reintroduction of necessary non-convex constraints, ensuring that the final dispatch solution is both optimal and adheres to the practical constraints of the microgrid components. The goal of this procedure is to achieve a zero gap of simultaneous action across all cases, indicating that the method maintains solution precision.

Algorithm 10 Closed-Loop Semi-Convex-Relaxation Procedure

Initialize the binary variable set

$\mathcal{I}^{\text{B}} = \left\{ I^{\text{ES+}}_{i',t'}, I^{\text{ES}-}_{i',t'}, I^{\text{EV+}}_{i',t'}, I^{\text{EV}-}_{i',t'}, I^{\text{acdc}}_{l',t'}, I^{\text{dcac}}_{l',t'} \right\} = \emptyset$

Solve robust dispatch with semi-relaxed model (7.73)-(7.75)

Calculate the value of $\text{SAM}^{\text{ES}}_{i,t}$, $\text{SAM}^{\text{EV}}_{i,t}$, and $\text{SAM}^{\text{BDC}}_{l,t}$

while $\exists i', l', t', \text{SAM}^{\text{ES}}_{i',t'} \neq 0, \text{SAM}^{\text{EV}}_{i',t'} \neq 0, \text{SAM}^{\text{BDC}}_{l',t'} \neq 0$ **do**

  Add all binary variables in (7.79)-(7.81) to the binary set $\mathcal{I}^{\text{B}}$

  Form complementary constraints:

$$I^{\text{ES+}}_{i',t'} + I^{\text{ES}-}_{i',t'} \leq 1, \forall I^{\text{ES+}}_{i',t'}, I^{\text{ES}-}_{i',t'} \in \mathcal{I}^{\text{B}} \tag{7.82}$$

$$I^{\text{EV+}}_{i',t'} + I^{\text{EV}-}_{i',t'} \leq 1, \forall I^{\text{EV+}}_{i',t'}, I^{\text{EV}-}_{i',t'} \in \mathcal{I}^{\text{B}} \tag{7.83}$$

$$I^{\text{acdc}}_{l',t'} + I^{\text{dcac}}_{l',t'} \leq 1, \forall I^{\text{acdc}}_{l',t'}, I^{\text{dcac}}_{l',t'} \in \mathcal{I}^{\text{B}} \tag{7.84}$$

  Solve model (7.73)-(7.75) with updated constraints (7.82)-(7.84)

  Calculate the value of $\text{SAM}^{\text{ES}}_{i',t'}$, $\text{SAM}^{\text{EV}}_{i',t'}$, and $\text{SAM}^{\text{BDC}}_{l',t'}$

end while

7.3 Customized Algorithm Framework Embedded with the Developed Semi-Convex-Relaxation

To effectively solve the tri-level robust dispatch model with non-convex complementary constraints, we design a customized C&CG algorithm embedded with the developed CS procedure. The algorithm decomposes the whole model into an MILP-based master problem and a CS-based slave problem.

7.3.1 MILP-based Master Problem

Specifically, the MILP-based master problem at q-th iteration is formulated as follows:

$$\min_{\boldsymbol{x},\Theta} \boldsymbol{a}^T \boldsymbol{x} + \Theta \tag{7.85}$$

$$s.t. \quad \boldsymbol{C}\boldsymbol{x}_1 \leq \boldsymbol{c}, \boldsymbol{x}_1 \in \{0,1\}, \Theta \geq 0, \tag{7.86}$$

$$\boldsymbol{D}_1\boldsymbol{x}_1 + \boldsymbol{D}_2\boldsymbol{x}_2 + \boldsymbol{E}\boldsymbol{y}_r \leq \hat{\boldsymbol{u}}_r, \boldsymbol{x}_2 \in \{0,1\} \tag{7.87}$$

$$\dot{\boldsymbol{x}}_2 + \ddot{\boldsymbol{x}}_2 \leq 1, \boldsymbol{x}_2 = (\dot{\boldsymbol{x}}_2, \ddot{\boldsymbol{x}}_2), \dot{\boldsymbol{x}}_2, \ddot{\boldsymbol{x}}_2 \in \{0,1\} \tag{7.88}$$

$$\Theta \geq \mathbb{E}_r \boldsymbol{b}^T \boldsymbol{y}_r^v, v = 1, 2, 3, .., q-1 \tag{7.89}$$

$$\boldsymbol{D}_1 \boldsymbol{x}_1 + \boldsymbol{D}_2 \boldsymbol{x}_2^v + \boldsymbol{E} \boldsymbol{y}_r^v \leq \boldsymbol{u}_r^v, \boldsymbol{x}_2^v \in \{0, 1\}, \forall v \tag{7.90}$$

$$\dot{\boldsymbol{x}}_2^v + \ddot{\boldsymbol{x}}_2^v \leq 1, \boldsymbol{x}_2^v = (\dot{\boldsymbol{x}}_2^v, \ddot{\boldsymbol{x}}_2^v), \dot{\boldsymbol{x}}_2^v, \ddot{\boldsymbol{x}}_2^v \in \{0, 1\}, \forall v \tag{7.91}$$

Here, $\boldsymbol{y}_r^v$ and $\boldsymbol{x}_2^v$ denote the continuous and binary variables of the third-level model with the constraint set of the v-th cuts. For each iteration, a series of optimality cuts and feasibility cuts (7.89)-(7.91) are generated and added to the master problem. The master problem is resolved with these new constraints, leading to a new set of $\boldsymbol{x}_1$ decisions and an updated Θ. By adding these constraints iteratively, the master problem gains more information about the consequences of its decisions under uncertainty, effectively learning to make more robust choices.

7.3.2 CS-based Slave Problem

Given the first-stage decisions $\boldsymbol{x}_1^{*q}$ obtained from the master problem, the slave problem is responsible for finding the worst-case scenario of RES output and the corresponding optimal operational decisions. Its primary goal is to determine the maximum possible unbalanced operation cost that could occur given $\boldsymbol{x}_1^*$. This information is then fed back to the master problem to generate new cuts. Specifically, the slave problem with semi-convex relaxation at the q-th iteration is expressed as follows:

$$\mathbb{E}_r[\max_{\boldsymbol{u}_r^q} \min_{\boldsymbol{y}_r^q} \boldsymbol{b}^T \boldsymbol{y}_r^q] = \sum\nolimits_r \max_{\boldsymbol{u}_r^q} \min_{\boldsymbol{y}_r^q} \rho_r \boldsymbol{b}^T \boldsymbol{y}_r^q \tag{7.92}$$

$$\boldsymbol{u}_r^q \in \mathcal{V}_r^{\text{DCUS}} \subseteq \mathcal{V}^{\text{DCUS}} \tag{7.93}$$

$$\boldsymbol{D}_1 \boldsymbol{x}_1^{*q} + \boldsymbol{D}_2 \boldsymbol{x}_2^q + \boldsymbol{E} \boldsymbol{y}_r^q \leq \boldsymbol{u}_r^q, \boldsymbol{x}_2^q \in \{0, 1\}, \forall q \tag{7.94}$$

$$\dot{\boldsymbol{x}}_2^q + \ddot{\boldsymbol{x}}_2^q \leq 1, \boldsymbol{x}_2^q = (\dot{\boldsymbol{x}}_2^q, \ddot{\boldsymbol{x}}_2^q), \dot{\boldsymbol{x}}_2^q, \ddot{\boldsymbol{x}}_2^q \in \{0, 1\}, \forall q \tag{7.95}$$

where ρ_r is the probability for the r-th DCUS. For each r, the model (7.92)-(7.95) can be expressed as follows:

$$\max_{\boldsymbol{u}_r^q} \min_{\boldsymbol{y}_r^q} \boldsymbol{b}^T \boldsymbol{y}_r^q \tag{7.96}$$

$$\boldsymbol{u}_r^q \in \mathcal{V}_r^{\text{DCUS}} \tag{7.97}$$

$$(7.94) - (7.95) \tag{7.98}$$

The presence of non-convex constraints (7.98) in the bi-level slave model (7.96)-(7.98) means the typical block coordinate descent algorithm cannot be used to solve it. This is because the algorithm only works when all constraints in the slave model are convex. Thus, a modified block coordinate descent algorithm embedded with the semi-convex-relaxation procedure is tailored to solve the employed slave model with non-convex constraints, which can ensure the accuracy of the dispatch solution while accelerating the solution process. First, the inner-level model of the slave problem with the relaxation technique (**ISP-R**) at the q'-th iteration is formulated as follows:

$$\min_{\boldsymbol{u}_r^{q,q'},\boldsymbol{y}_r^{q,q'}} \boldsymbol{b}^T \boldsymbol{y}_r^{q,q'} \tag{7.99}$$

$$\boldsymbol{D}_1 \boldsymbol{x}_1^{*q} + \boldsymbol{D}_2 \boldsymbol{x}_2^{q,q'} + \boldsymbol{E} \boldsymbol{y}_r^{q,q'} \leq \boldsymbol{u}_r^{q,q'}, \boldsymbol{x}_2^{q,q'} \in [0,1] \tag{7.100}$$

$$\dot{\boldsymbol{x}}_2^{q,q'} + \ddot{\boldsymbol{x}}_2^{q,q'} \leq 1, \boldsymbol{x}_2^{q,q'} = (\dot{\boldsymbol{x}}_2^{q,q'}, \ddot{\boldsymbol{x}}_2^{q,q'}), \dot{\boldsymbol{x}}_2^{q,q'}, \ddot{\boldsymbol{x}}_2^{q,q'} \in [0,1] \tag{7.101}$$

$$\boldsymbol{u}_r^{q,q'} = \boldsymbol{u}_r^{q,q'-1} : \boldsymbol{\pi}^{q,q'} \tag{7.102}$$

Here, the variable $\boldsymbol{u}_r^{q,q'}$ is fixed as the solution $\boldsymbol{u}_r^{q,q'-1}$ at the $q'-1$ iteration, and $\boldsymbol{\pi}^{q,q'}$ is the marginal value of the constraint (7.102), which represents the rate of change in the optimal value of the objective function when the corresponding constraint is slightly relaxed. Then, the closed-loop semi-convex-relaxation procedure in **Algorithm** 10 is employed to identify the binary variable set $\mathcal{I}^{\text{B}}$ to form complementary constraints and ensure the solution accuracy. If any of the values of $\text{SAM}_{i',t'}^{\text{ES}}$, $\text{SAM}_{i',t'}^{\text{EV}}$, and $\text{SAM}_{l',t'}^{\text{BDC}}$ does not equal to zero, the inner-level model of the slave problem with the semi-convex-relaxation technique (**ISP-S**) is developed by adding the constraints associated with the binary variables in the set $\mathcal{I}^{\text{B}}$:

$$(7.99) - (7.101) \tag{7.103}$$

$$\boldsymbol{D}_1 \boldsymbol{x}_1^{*} + \boldsymbol{D}_2 \boldsymbol{x}_2^{q,q'} + \boldsymbol{E} \boldsymbol{y}_r^{q,q'} \leq \boldsymbol{u}_r^{q,q'}, \boldsymbol{x}_2^{q,q'} \in \mathcal{I}^{\text{B}} \in \{0,1\} \tag{7.104}$$

$$\dot{\boldsymbol{x}}_2^{q,q'} + \ddot{\boldsymbol{x}}_2^{q,q'} \leq 1, \dot{\boldsymbol{x}}_2^{q,q'}, \ddot{\boldsymbol{x}}_2^{q,q'} \in \mathcal{I}^{\mathrm{B}} \in \{0,1\} \tag{7.105}$$

After solving the ISP-S model, we extract the marginal value $\boldsymbol{\pi}^{q,q'}$ by fixing the value of binary variables $\boldsymbol{x}_2^{*q,q'}$, $\dot{\boldsymbol{x}}_2^{*q,q'}$ and $\ddot{\boldsymbol{x}}_2^{*q,q'}$, and solving the following convex inner-level model of the slave problem (**CISP**):

$$(7.99) - (7.101) \tag{7.106}$$

$$\boldsymbol{D}_1\boldsymbol{x}_1^* + \boldsymbol{D}_2\boldsymbol{x}_2^{*q,q'} + \boldsymbol{E}\boldsymbol{y}_r^{q,q'} \leq \boldsymbol{u}_r^{q,q'} \tag{7.107}$$

$$\boldsymbol{u}_r^{q,q'} = \boldsymbol{u}_r^{q,q'-1} : \boldsymbol{\pi}^{q,q'} \tag{7.108}$$

With the solution for the inner-level model of the slave problem, the outer-level model of the slave problem (**OSP**) at the q'-th iteration is formulated as follows:

$$R_r^{q'} = \max_{\boldsymbol{u}_r^{q,q'}} \boldsymbol{b}^T\boldsymbol{y}_r^{*q,q'} + (\boldsymbol{\pi}^{*q,q'})^T(\boldsymbol{u}_r^{q,q'} - \boldsymbol{u}_r^{*q,q'-1}) \tag{7.109}$$

$$s.t. \quad \boldsymbol{u}_r^{q,q'} \in \mathcal{V}_r^{\mathrm{DCUS}} \tag{7.110}$$

Based on the block coordinate descent theory [164], this optimization model is formulated by the first-order Taylor series expansion of the optimal value $\boldsymbol{\pi}^{*q,q'}$ in **CISP**.

7.3.3 Framework and Flowchart

The framework and the flowchart of the customized C&CG algorithm are illustrated in Fig. 7.3 and **Algorithm** 11. Our customized C&CG algorithm embedded with the semi-convex relaxation procedure provides an efficient method to solve the tri-level robust microgrid dispatch problem under uncertainty and non-convex constraints. The algorithm begins by solving a MILP-based master problem that determines the first-stage decisions (e.g., unit commitments). Then, a CS-based slave problem is employed to iteratively identify the worst-case renewable energy scenarios and corresponding optimal dispatch strategies through a semi-convex relaxation process. The key innovation lies in dynamically identifying and reintroducing only those non-convex complementary constraints (such as mutual exclusivity of charging/discharging in HES and EVs) that are violated in the relaxed

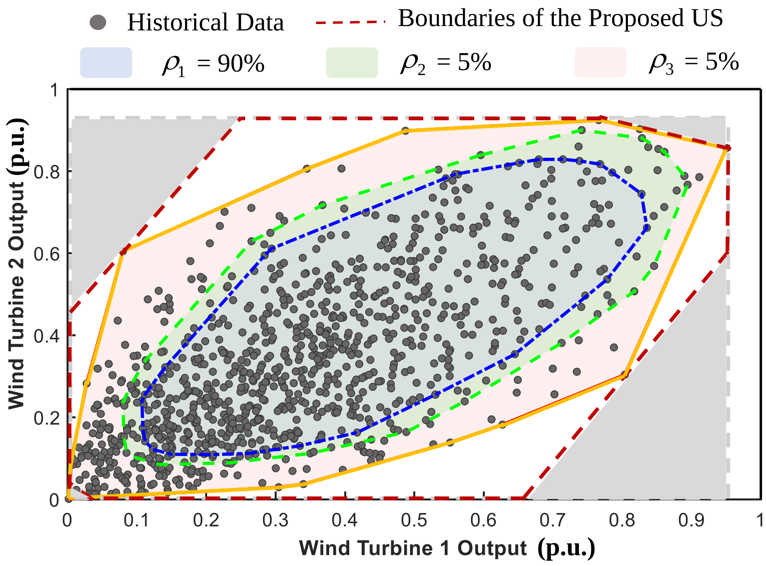


Figure 7.3: Customized C&CG algorithm with semi-convex relaxation procedure.

solution. The SAM is used to detect these violations, and their corresponding binary constraints are reactivated, ensuring solution feasibility and accuracy without solving the full non-convex model from the start.

Handling the non-convex complementary constraints for discrete resources, such as HES and EVs, within a robust optimization framework typically requires a nested algorithm, where a full C&CG process is embedded within the slave problem. This approach creates a significant computational burden, making it impractical for microgrids with many discrete resources. In comparison, our semi-convex relaxation embedded C&CG algorithm significantly reduces computation time by avoiding unnecessary binary constraints unless they are truly violated. On the other hand, unlike fully convex relaxation methods, which relax all non-convex constraints and are likely to result in infeasible and inaccurate solutions, our method guarantees exact satisfaction of non-convex complementary constraints, leading to zero violation gaps. Thus, the proposed approach achieves a balance between computational efficiency and solution accuracy, solving large-scale robust dispatch problems more effectively.

Algorithm 11 Customized C&CG Algorithm with the Closed-Loop Semi-Convex Relaxation Procedure

Input: Algorithm Gap $\varepsilon_{\mathrm{O}} = \varepsilon_{\mathrm{I}} = 10^{-5}$.
Initialize: $UB_{\mathrm{O}} \leftarrow +\infty$, $LB_{\mathrm{O}} \leftarrow -\infty$, $q \leftarrow 1$
while $\left|UB_{\mathrm{O}} - LB_{\mathrm{O}}\right| / UB_{\mathrm{O}} > \varepsilon_{\mathrm{O}}$ **do**
    Compute $\boldsymbol{x}^{*q}$, Θ^{*q} by solving master problem (7.85)-(7.91)
    Update $LB_{\mathrm{O}} = \boldsymbol{a}^T\boldsymbol{x}^{*q} + \Theta^{*q}$
    Initialize $q' \leftarrow 1$ and $\boldsymbol{u}_{z'} \leftarrow \hat{\boldsymbol{u}}$ (predicted value)
    Initialize: $UB_{\mathrm{I}} \leftarrow +\infty$, $LB_{\mathrm{I}} \leftarrow -\infty$
    while $\left|UB_{\mathrm{I}} - LB_{\mathrm{I}}\right| / UB_{\mathrm{I}} > \varepsilon_{\mathrm{I}}$ **do**
        Solve ISP-R (7.99)-(7.102) $\rightarrow \boldsymbol{b}^T\boldsymbol{y}_r^{*q,q'}$ and $\boldsymbol{\pi}^{*q,q'}$
        if $\exists i', l', t', \mathrm{SAM}_{i',t'}^{\mathrm{ES}} \neq 0, \mathrm{SAM}_{i',t'}^{\mathrm{EV}} \neq 0, \mathrm{SAM}_{l',t'}^{\mathrm{acdc}} \neq 0$ **then**
            Identify binary variable set $\mathcal{I}^{\mathrm{B}}$ by **Algorithm 10**
            Solve ISP-S (7.103)-(7.105)
            Solve ISP-S (7.106)-(7.108) $\rightarrow \boldsymbol{b}^T\boldsymbol{y}_r^{*q,q'}$ and $\boldsymbol{\pi}^{*q,q'}$
            Update the value of $\boldsymbol{b}^T\boldsymbol{y}_r^{*q,q'}$ and $\boldsymbol{\pi}^{*q,q'}$
        end if
        Update $LB_{\mathrm{I}} = \mathbb{E}_r[\boldsymbol{b}^T\boldsymbol{y}_r^{*q,q'}]$
        Solve OSP (7.109)-(7.110)
        Update $UB_{\mathrm{I}} = \mathbb{E}_r[R_r^{q'}]$ and $q' \leftarrow q' + 1$
    end while
    Update $UB_{\mathrm{O}} = \boldsymbol{a}^T\boldsymbol{x}^{*q} + LB_{\mathrm{I}}$, and $q \leftarrow q + 1$
end while
Return optimal solution $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$

7.4 Case Studies

To verify the effectiveness and superiority of our developed uncertainty set and solution algorithm, we conducted numerical case studies on the IEEE 66-bus hybrid AC/DC microgrid system in [17], and detailed data are provided in [165]. All experiments are carried out in MATLAB 2023b and GAMS 24.4.

7.4.1 Comparisons of Various Uncertainty Sets

To validate the effectiveness of the proposed uncertainty sets, we compare them against several widely adopted models, including the BbUS [161], the PUS [163], and the CHUS [95]. The results are summarized in Table 7.1. Here, robustness is quantified using the robustness rate (RR), defined as the proportion of solutions remaining feasible for 10,000 out-of-sample test scenarios. It can be seen that, BbUS results in the highest total cost of ¥1,142,644.2 due to its overly conservative

Table 7.1: Comparisons with different uncertainty set (Unit:¥)

Uncertainty Set (US)		TC	GC	RC	LC	EVC	HESC	RR
BbUS [161]		1142644.2	822094.8	81581.9	503.2	547.6	2158.1	100%
PUS [163]	$\Gamma = 0.75$	972488.5	824648.0	85303.0	448.8	570.4	1984.4	100%
	$\Gamma = 0.50$	944168.5	824431.4	81958.6	447.0	549.6	1860.2	100%
	$\Gamma = 0.25$	893454.5	818712.6	66137.4	506.3	586.9	2223.9	100%
CHUS [95]		836189.0	809812.2	22657.6	719.2	607.5	2392.5	100%
Our Developed DCUS		836500.9	810069.4	22793.6	706.2	594.7	2337.1	100%
Our Developed PDCUS		825496.0	807853.9	14422.7	732.4	566.0	1921.0	99.46%

TC: total cost; GC: generation cost; RC: renewable curtailment cost; LC: loss cost on BDCs; UBC: unbalanced cost; HESC: HESS degraded cost; EVC: EV degraded cost.

treatment of uncertainty, assuming simultaneous worst-case deviations of all renewables, which ignores realistic correlations and forces reliance on expensive dispatchable generation. PUS addresses this by introducing a budget parameter Γ to limit conservatism. As Γ is reduced from 0.75 to 0.25, the total cost drops significantly from ¥972,488.5 to ¥893,454.5, and RC is also reduced. However, the tuning of Γ is subjective and may not provide sufficient robustness against real-world uncertainties, making the solution less reliable for practical operations.

In contrast, CHUS constructs uncertainty sets directly from historical scenarios, inherently capturing correlations among renewable energy sources and offering a less conservative yet robust solution. With a total cost of ¥836,189.0 and an RC of ¥22,657.6, CHUS performs well but suffers from a heavy computational burden due to the extensive number of vertices required. Our developed DCUS closely approximates CHUS in performance (total cost: ¥836,500.9), while significantly reducing the number of vertices and computational complexity. However, DCUS remains sensitive to historical outliers, which can expand the uncertainty set excessively. To overcome this, we propose the PDCUS, which introduces probabilistic layering to distinguish between likely and unlikely scenarios. PDCUS achieves the lowest total cost (¥825,496.0) and the lowest RC (¥14,422.7), while also minimizing EV and HES degradation costs. These results demonstrate that PDCUS offers a more intelligent, robust, and cost-efficient approach for modeling renewable uncertainty in practical microgrid applications.

Additionally, all conventional models (BbUS, PUS, CHUS) and DCUS achieve a perfect 100% RR, ensuring absolute feasibility under the tested conditions. PDCUS attains an RR of 99.46%, re-

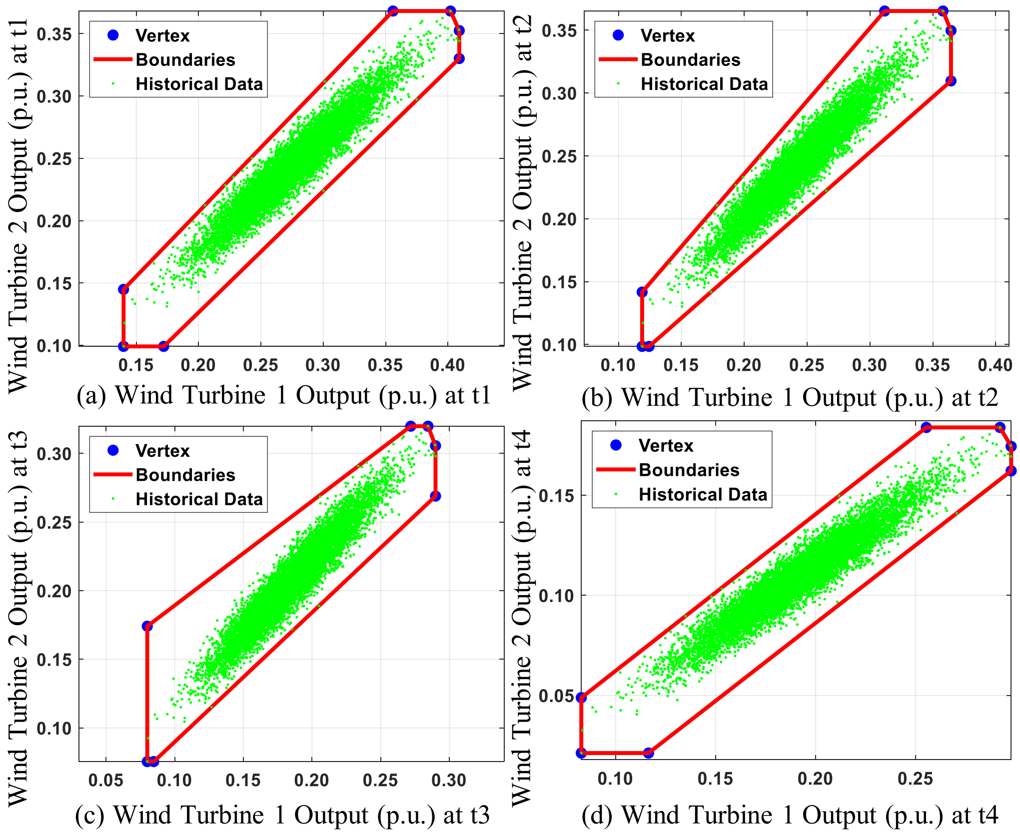
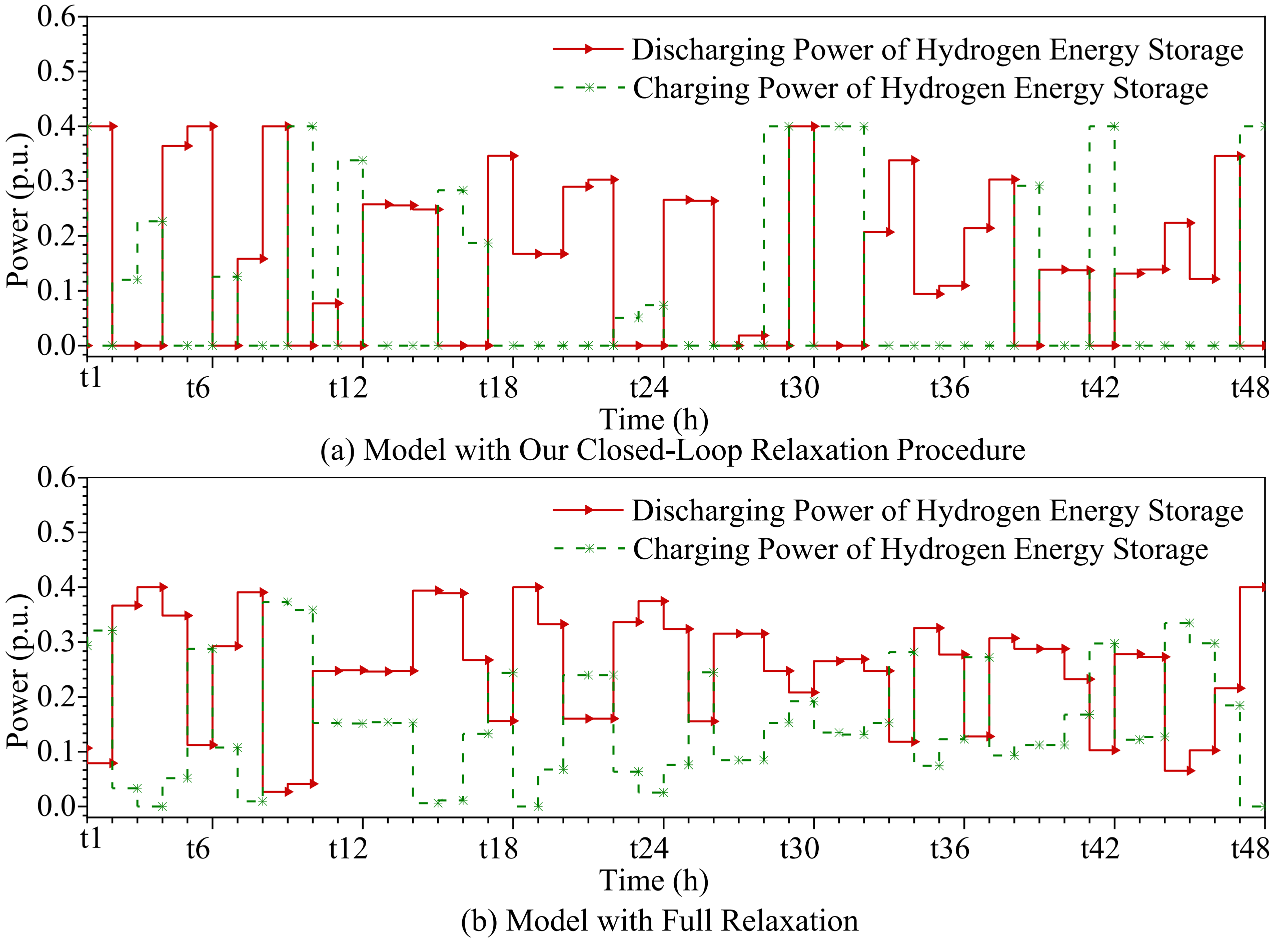


Figure 7.4: Data-correlated uncertainty set construction.

maining feasible in 9,946 of the 10,000 cases and failing in only 54 extreme, low-probability scenarios. This marginal reduction in RR is an intentional design choice, i.e., PDCUS prioritizes economic performance by not hedging against the most unlikely 0.54% of events. This choice can make a good trade-off between operational efficiency and reliability of the solutions.

Fig. 7.4 consists of four scatter plots, (a) through (d), each illustrating the constructed DCUS for the outputs of two wind turbines at four different time intervals: t1-t4. It can be observed that, this output data shows a strong positive correlation between the outputs of the two turbines; when one turbine's output is high, the other's tends to be high as well. Our DCUS is designed to be the smallest possible set that still contains all historical observations, which helps in reducing the over-conservatism found in other uncertainty models like the BbUS and creating a more refined model that balances robustness with economic efficiency.



(a) Model with Our Closed-Loop Relaxation Procedure

(b) Model with Full Relaxation

Figure 7.5: Discharging and charging power of hydrogen energy storage.

7.4.2 Verifications of the Closed-Loop Relaxation Procedure

To verify the effectiveness of the developed closed-loop relaxation procedure, we compare the constraint satisfaction results of the model with our procedure with the existing models of the full relaxation under the condition of 10 EVs. Here, we take HESS as an example for illustration, and the charging and discharging results are shown in Fig. 7.5. In the model using full relaxation, it is evident that simultaneous charging and discharging occurs during several time intervals, which violates the physical operational constraints of HESS units. This leads to infeasible or unrealistic dispatch results in practice. To quantify this violation, we introduced the maximum simultaneous action metric (MSAM) as defined in (7.111). The MSAM captures the highest level of mutual exclusivity violation across all time intervals and all units of HESSs, EVs and BDCs. A non-zero MSAM indicates simultaneous charging and discharging, which is not physically implementable.

$$\text{MSAM} = \max_{\phi\in\Phi} \left\| \begin{matrix} p^{\phi+}_{i_1,t_1} p^{\phi-}_{i_1,t_1} & \cdots & p^{\phi+}_{i_1,|\mathcal{T}|} p^{\phi-}_{i_1,|\mathcal{T}|} \\ \vdots & \ddots & \vdots \\ p^{\phi+}_{|\mathcal{I}|,t_1} p^{\phi-}_{|\mathcal{I}|,t_1} & \cdots & p^{\phi+}_{|\mathcal{I}|,|\mathcal{T}|} p^{\phi-}_{|\mathcal{I}|,|\mathcal{T}|} \end{matrix} \right\|_\infty \tag{7.111}$$

As shown in Table 7.2, the full relaxation model exhibits MSAM values ranging between 0.7 and 2.0 under different EV penetration levels, confirming the presence of infeasible behaviors. In contrast, our proposed closed-loop relaxation procedure consistently maintains MSAM $= 0.0$ across all test cases, ensuring that the final solution is physically valid and adheres to the complementary constraints. These results validate the effectiveness of the closed-loop semi-convex relaxation procedure in the proposed framework. It dynamically identifies and enforces only the necessary non-convex constraints during the optimization process, thereby reducing problem complexity without sacrificing solution feasibility. This approach ensures zero violation gap, as required for real-world microgrid operation, while significantly accelerating the solution process compared to traditional methods.

7.4.3 Comparisons of Various Solution Algorithms

NC&CG algorithms [41], [166] and full relaxed algorithms [17], [23], [37] are the two advanced and most widely employed algorithms for effectively handling the robust dispatch model with non-convex constraints. To verify the superiority of our customized C&CG algorithm, we compare the results obtained by our algorithm with those obtained by these existing algorithms, as presented in Table 7.2. To evaluate the solution quality of different algorithms, we conduct out-of-sample robustness tests. We define a metric, the robustness failure gap (RFG), to quantify the comparative performance. The RFG is defined as the difference in the number of scenarios that fail the robustness test between a competing algorithm and our proposed method. The formula is given by:

$$\text{RFG} = |S_{\text{NC\&CG/FullRelax}}| - |S_{\text{Ours}}|$$

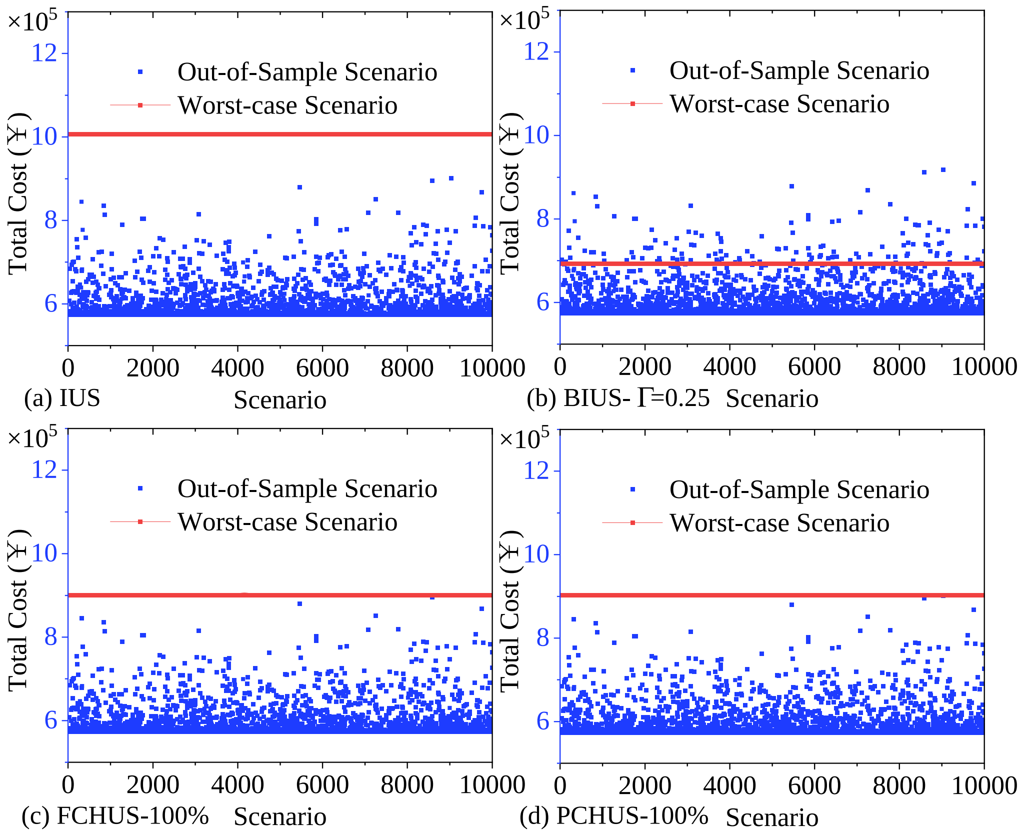


Figure 7.6: Out-of-sample performance of various algorithms.

where $|S_{\text{Ours}}|$ is the number of failed scenarios for our algorithm, and $|S_{\text{NC\&CG}}|$, $|S_{\text{FullRelax}}|$ is the number for competing algorithms (e.g., NC&CG algorithms or full relaxed algorithms).

The results from Table 7.2 reveal the superior performance of the proposed customized C&CG algorithm compared to both the established NC&CG and full relaxed algorithms, particularly as the number of EVs increases. Specifically, the proposed algorithm consistently achieves a MSAM of zero, indicating that it completely avoids the physically impossible scenario of simultaneous charging and discharging, a significant advantage over the full relaxed algorithm which shows violations. While the NC&CG algorithm also maintains a zero MSAM, it suffers from a drastic increase in CT, becoming impractically slow and even failing to solve within seven days for 200 EVs.

In contrast, the proposed algorithm demonstrates remarkable computational efficiency, solving the 200-EV case in just 1057.0 seconds, a fraction of the time required by NC&CG at smaller scales. Furthermore, the out-of-sample tests, quantified by the RFG, highlight the superiority of the proposed method in seeking a worst-case fluctuation of renewable uncertainty. The proposed algorithm records

Table 7.2: Verifications of our proposed semi-convex-relaxation method

EV No.	Methods	TC (¥)	MSAM	CT (s)	RFG
	NC&CG Algorithm	832626.6	0.0	11999.6	5
10	Full Relaxed Algorithm	832423.4	1.4	138.8	17
	Our Algorithm	834225.2	0.0	642.5	0
	NC&CG Algorithm	833655.8	0.0	18756.1	1
20	Full Relaxed Algorithm	831705.7	1.4	60.0	16
	Our Algorithm	833759.8	0.0	483.4	0
	NC&CG Algorithm	830371.2	0	89598.8	249
50	Full Relaxed Algorithm	829680.4	2.0	401.1	10
	Our Algorithm	831857.5	0	1648.3	0
	NC&CG Algorithm	826673.5	0	204075.2	257
100	Full Relaxed Algorithm	827547.5	1.4	455.6	12
	Our Algorithm	829812.2	0	4509.3	0
	NC&CG Algorithm	–	–	>7-d	–
200	Full Relaxed Algorithm	823944.0	0.7	785.6	9
	Our Algorithm	825496.0	0.0	2463.5	0

an RFG of zero in all scenarios, meaning it successfully navigated all out-of-sample tests without failure. This stands in stark contrast to the NC&CG and full relaxed algorithms, which exhibited numerous failures, with the NC&CG algorithm failing in as many as 257 scenarios for the 100-EV case. This demonstrates that the customized algorithm not only provides accurate solutions and avoids locally optimal solutions, but also significantly accelerates the solution process.

7.4.4 Impacts of EV Number on Model Solution

To evaluate the EV number on the proposed model solution, the results are summarized in Table 7.3 and Table 7.4. The analysis reveals that increasing the penetration of V2G-enabled EVs yields significant economic and operational benefits for the microgrid. As the EV fleet expands from 50 to 300 vehicles, the total operating cost progressively decreases, driven by the fleet's ability to perform peak shaving and valley filling. This strategic charging and discharging reduces reliance on expensive generators and drastically cuts renewable energy curtailment by up to 69%. In contrast, the scenario with only unidirectional (G2V) charging highlights the substantial opportunity cost of lacking bidirectional capability. The G2V-only system consistently incurs higher total costs at all penetration levels, a gap that widens as more EVs are added, quantifying the direct economic value of V2G technology. This system is less effective at integrating renewables and demonstrates lower operational resilience, exhibiting a more brittle response to stress. While it manages with incentive-based controls at lower numbers, it experiences a sudden and severe onset of involuntary load shedding at high penetration, indicating a less graceful failure mode compared to the V2G system. This comparative analysis underscores that V2G is not merely an economic tool but a vital asset for enhancing grid resilience, justifying the investment in bidirectional technology and the sophisticated control algorithms required to unlock its full potential.

Table 7.3: Impacts of EV number on solutions

EV Number	Gap	TC (¥)	GC (¥)	RC (¥)	LC (¥)	Computing Time (s)
50	0	831857.5	811248.9	18120.1	649.5	1648.3
100	0	829812.2	810300.3	16793.3	678.2	4509.3
200	0	825496.0	807853.9	14422.6	732.4	2463.5
300	0	822189.6	808413.4	10965.7	710.6	2767.9

Table 7.4: Impacts of EV number on solutions without V2G

EV Number	Gap	TC (¥)	GC (¥)	RC (¥)	LC (¥)	Computing Time (s)
50	0	833634.2	812189.0	19065.8	617.3	110.4
100	0	832202.9	812181.5	17687.8	612.2	118.6
200	0	829416.4	811930.2	15109.2	632.1	180.2
300	0	826148.0	811834.4	11990.9	622.3	450.1

7.5 Summary

This paper proposes a robust dispatch framework to address the complex challenges of hybrid AC/DC microgrids integrating renewable energy sources (RES), hydrogen energy storage systems (HESS), and electric vehicles (EVs). To reduce the number of vertices and over-conservatism in traditional uncertainty sets, a probability-driven data-correlated uncertainty set (PDCUS) is developed, capturing temporal correlations and scenario probabilities from historical RES data. This approach ensures robustness while avoiding the computational burden of convex hull uncertainty sets. Additionally, to tackle the inherent non-convex complementary constraints introduced by HESS and EV operations, a closed-loop semi-convex relaxation procedure is developed. By iteratively identifying only the violated non-convex constraints, this method ensures computational efficiency while maintaining physical feasibility, a critical requirement for large-scale real-time microgrid operations. Finally, a customized column-and-constraint generation (C&CG) algorithm integrates this relaxation to solve the resulting tri-level optimization problem effectively. Case studies on the IEEE 66-bus hybrid microgrid system demonstrate that the proposed PDCUS significantly reduces total operating costs compared to existing uncertainty sets. The proposed algorithm solves large-scale problems with 200 EVs orders of magnitude faster than state-of-the-art methods while guaranteeing physical feasibility with zero simultaneous action metric and superior out-of-sample robustness with zero robustness failure gap.

Table 7.5: Notation for HMG operation models

Indices and Sets	
i, j	Index for buses in the microgrid
$l, \mathcal{L}$	Index and set for BDCs
$t, \mathcal{T}$	Index and set for time periods in the horizon
Parameters	
$n_i^{\mathrm{de}}/f_i^{\mathrm{de}}/c_i^{\mathrm{de}}$	Startup/shutdown/variable cost coefficient of DE
$c_i^{\mathrm{w}}/c_i^{\mathrm{pv}}$	Penalty cost coefficient for WT/PV curtailment
c_i^{loss}	Cost coefficient for power losses in BDC
$\bar{P}_i^{\mathrm{de}}/\underline{P}_i^{\mathrm{de}}$	Upper/lower active power limits of DE
$\bar{Q}_i^{\mathrm{de}}/\underline{Q}_i^{\mathrm{de}}$	Upper/lower reactive power limits of DE
$R_i^{\mathrm{P,de}}/R_i^{\mathrm{Q,de}}$	Ramp-rate limit for active/reactive power of DE
$\hat{P}_{i,t}^{\mathrm{w}}/\hat{P}_{i,t}^{\mathrm{pv}}$	Forecasted power output of WT/PV unit
$\underline{V}_{i,t}^{\mathrm{ac}}/\bar{V}_{i,t}^{\mathrm{ac}}$	Upper/lower voltage limits at AC bus
$\underline{V}_{i,t}^{\mathrm{dc}}/\bar{V}_{i,t}^{\mathrm{dc}}$	Upper/lower voltage limits at DC bus
$\underline{H}_{ij,t}^{\mathrm{ac}}/\bar{H}_{ij,t}^{\mathrm{ac}}$	Upper/lower active power flow limits on AC line
$\underline{G}_{ij,t}^{\mathrm{ac}}/\bar{G}_{ij,t}^{\mathrm{ac}}$	Upper/lower reactive power flow limits on AC line
$L_{i,t}^{\mathrm{ac}}/Q_{i,t}^{\mathrm{ac}}$	Active/reactive load at AC bus
$r_{ij}^{\mathrm{ac}}/x_{ij}^{\mathrm{ac}}$	Resistance/reactance of AC line
$\bar{S}_i^{\mathrm{ES}}/\underline{S}_i^{\mathrm{ES}}$	Upper/lower state of charge (SOC) limits of HES
$\bar{S}_i^{\mathrm{EV}}/\underline{S}_i^{\mathrm{EV}}$	Upper/lower SOC limits of EV
$S_i^{\mathrm{EV,rem}}/S_i^{\mathrm{EV,req}}$	Remaining SOC at arrival and required SOC at departure for EV
$t_i^{\mathrm{in}}/t_i^{\mathrm{out}}$	Arrival and departure times for EV
$\bar{P}_i^{\mathrm{ES+}}/\underline{P}_i^{\mathrm{ES+}}$	Upper/lower charging power limits of HES unit
$\bar{P}_i^{\mathrm{ES-}}/\underline{P}_i^{\mathrm{ES-}}$	Upper/lower discharging power limits of HES unit
$\eta_i^{\mathrm{ES+}}/\eta_i^{\mathrm{ES-}}$	Charging and discharging efficiency of HES unit
$G_i^{\mathrm{HES}}/G_i^{\mathrm{EV}}$	Capacity of HES/EV unit
$\bar{P}_i^{\mathrm{EV+}}/\bar{P}_i^{\mathrm{EV-}}$	Maximum charging/discharging power limits of EV
$\eta_i^{\mathrm{EV+}}/\eta_i^{\mathrm{EV-}}$	Charging and discharging efficiency of EV
$\underline{H}_{ij,t}^{\mathrm{dc}}/\bar{H}_{ij,t}^{\mathrm{dc}}$	Upper/lower power flow limits on DC line
$L_{i,t}^{\mathrm{dc}}$	Active load at DC bus
r_{ij}^{dc}	Resistance of DC line
$\eta_{\mathrm{BDC},l}^{\mathrm{acdc}}(\cdot)/\eta_{\mathrm{BDC},l}^{\mathrm{dcac}}(\cdot)$	Efficiency functions for BDC
$o_{0,l}/o_{1,l}$	Coefficients for the linear approximation of BDC efficiency
$c_i^{\mathrm{hes}}/c_i^{\mathrm{ev}}$	Coefficients of HES/EV degraded cost
Variables	
$U_{i,t}^{\mathrm{de}}/D_{i,t}^{\mathrm{de}}$	Binary variables for startup/shutdown of DE
$I_{i,t}^{\mathrm{de}}$	Binary variable for operational status of DE
$P_{i,t}^{\mathrm{de}}/Q_{i,t}^{\mathrm{de}}$	Active/reactive power output of DE
$P_{i_lj_l,l,t}^{\mathrm{acdc}}/P_{i_1j_1,l,t}^{\mathrm{dcac}}$	Power flow from AC to DC / DC to AC on BDC l
$\hat{\rho}_{i,t}^{\mathrm{w}}/\hat{\rho}_{i,t}^{\mathrm{pv}}$	Power curtailment of WT/PV unit
$V_{i,t}^{\mathrm{ac}}/V_{i,t}^{\mathrm{dc}}$	Voltage magnitude at AC/DC bus
$H_{ij,t}^{\mathrm{ac}}/G_{ij,t}^{\mathrm{ac}}$	Active and reactive power flow on AC line
$S_{i,t}^{\mathrm{ES}}/S_{i,t}^{\mathrm{EV}}$	SOC of HES/EV unit
$p_{i,t}^{\mathrm{ES+}}/p_{i,t}^{\mathrm{ES-}}$	Charging and discharging power of HES
$I_{i,t}^{\mathrm{ES+}}/I_{i,t}^{\mathrm{ES-}}$	Binary variables for charging/discharging of HES
$p_{i,t}^{\mathrm{EV+}}/p_{i,t}^{\mathrm{EV-}}$	Charging and discharging power of EV
$I_{i,t}^{\mathrm{EV+}}/I_{i,t}^{\mathrm{EV-}}$	Binary variables for charging/discharging of EV
$\bar{P}_{i_1j_1,l}^{\mathrm{acdc}}/\bar{P}_{i_1j_1,l}^{\mathrm{dcac}}$	Maximum power flow limits for BDC
$I_{l,t}^{\mathrm{acdc}}/I_{l,t}^{\mathrm{dcac}}$	Binary variables for AC-DC and DC-AC operating modes of BDC

Chapter 8

Conclusions and Future Perspectives

8.1 Conclusions

The global energy landscape is undergoing a profound transformation towards a decentralized and decarbonized paradigm, with hybrid AC/DC microgrids (HMGs) emerging as a cornerstone technology for integrating renewable energy sources (RES), energy storage, and flexible demand. However, unlocking the full potential of HMGs requires overcoming a tripartite of interconnected challenges: the non-convex modeling of bi-directional converters (BDCs), which complicates optimization; the limited observability of end-user load behavior, which hinders effective demand-side management; and the inherent uncertainty of renewable generation, which threatens system reliability and economic efficiency. This thesis has addressed these critical challenges by developing a unified framework that bridges the gap between granular, appliance-level smart meter analytics and system-level, uncertainty-aware optimization. Through a series of five core contributions, this work establishes both the theoretical foundations and practical tools necessary to enhance the flexibility, robustness, and economic viability of future renewable-penetrated power systems.

The principal contributions of this research are synthesized as follows:

1) First, this thesis developed a novel framework for the data-driven convex modeling of bi-directional converters. Conventional BDC models introduce significant non-convexities into optimization problems through their non-linear efficiency characteristics and the mutually exclusive nature of their rectifying and inverting operations. These non-convexities, often represented by bilinear terms or integer variables, render large-scale economic dispatch problems computationally intractable, particularly within robust or stochastic optimization frameworks that depend on the strong duality theorem and require convex inner-level subproblems. This research systematically overcomes this fundamental barrier through two key innovations. A least squares approximation (LSA) method was proposed to transform the complex, trigonometric BDC efficiency function into a computationally efficient convex form without introducing auxiliary binary variables. More critically, a set of Karush-Kuhn-Tucker (KKT)-based sufficient conditions was rigorously derived to guarantee non-simultaneous rectification and inversion, thereby obviating the need for non-convex complementary constraints entirely under common operating scenarios. An iterative procedure was also designed to selectively add back the non-convex constraints only when these conditions are not met, minimizing the computational burden. Case studies on systems up to 66 nodes validated that this convex modeling approach reduces solution time by more than two orders of magnitude while preserving high fidelity in economic dispatch outcomes, thus providing an essential enabling technology for the advanced robust optimization methods developed subsequently in this work.

2) Second, a heterogeneous multi-task learning architecture for high-fidelity load identification was introduced to address the challenge of limited load observability. Accurate appliance-level load disaggregation from low-frequency smart meter data is a prerequisite for unlocking demand-side flexibility. However, existing Non-Intrusive Load Monitoring (NILM) methods often struggle due to their reliance on single-task models or homogeneous expert networks that are ill-suited to capture the diverse electrical signatures of different appliances. This thesis proposed a heterogeneous multi-gate mixture-of-experts (HMMOE) framework that integrates a diverse ensemble of neural network architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention-based models, as specialized ”experts”. The heterogeneity of this design is a direct and necessary response to the multifaceted nature of appliance signatures; for instance, the sharp, short-duration power draws of a kettle are best captured by CNNs, while the long-term cyclical patterns of a refrigerator are better suited to RNNs. A multi-gate mechanism dynamically allocates weights to each expert for the dual tasks of state classification and power estimation, while a novel dynamic multi-objective optimization strategy adaptively balances these competing learning objectives. Extensive validation on public datasets (UK-DALE, REDD) and a real-world Hong Kong dataset, with a challenging 15-minute sampling interval, demonstrated significant improvements in F1 score, classification accuracy, and mean absolute error over state-of-the-art NILM methods.

Third, this research extended the NILM framework to a multimodal architecture for robust demand response evaluation, bridging the gap between passive analytics and active system control. Real-world appliance usage is heavily influenced by contextual factors such as weather and occupancy, which are ignored by conventional NILM models that rely solely on electrical data, leading to inaccurate demand response assessments. Building upon the HMMOE architecture, this work developed a multimodal, dual-task framework that integrates environmental data, including temperature, humidity, and solar irradiance, with smart meter signals to create a more holistic view of energy consumption drivers. A key innovation was the derivation of an analytical, closed-form solution for computing the optimal weights to balance the classification and regression tasks, which eliminates heuristic trial-and-error and ensures stable, interpretable training. The high-fidelity outputs of this multimodal NILM model were then embedded directly into a robust energy management framework for a building-level HMG, enabling the flexible and informed scheduling of household appliances. This integration provides a direct link between appliance-level data intelligence and system-level operational decision-making, resulting in more accurate and reliable demand response evaluations.

Fourth, a suite of advanced data-driven uncertainty sets for renewable energy with adjustable conservativeness was formulated. Robust optimization is a powerful tool for managing RES uncertainty, but its effectiveness is highly dependent on the construction of the uncertainty set. Conventional sets, such as box, polyhedral, or standard convex hull formulations, often suffer from over-conservativeness, sensitivity to data outliers, and a high computational burden, which can lead to economically inefficient dispatch solutions. To overcome these limitations, this thesis introduced three novel uncertainty set formulations: the pair convex hull uncertainty set (PCHUS), which reduces dimensionality and computational complexity by constructing hulls from pairwise data relationships; the intersected convex hull uncertainty set (ICHUS), which achieves tighter feasible regions by intersecting multiple PCHUSs; and the probability-driven data-correlated uncertainty

Table 8.1: Comparison of different uncertainty set methods

Method	Distribution Assumption	Data Requirement	Conservativeness Control	Computational Scalability	Applicability to HMGs
CHUS	None	Historical samples	Fixed	Low	Limited
PCHUS	None	Historical samples	Confidence level	High	High
ICHUS	None	Historical samples	Intersection depth	High	High
PDCUS	None	Historical samples	Probability-driven	Medium–High	High

set (PDCUS), which leverages probabilistic confidence levels to mitigate the impact of outliers and construct adaptive, data-centric sets. Case studies confirmed that these advanced sets significantly improve out-of-sample performance, reduce renewable curtailment, and lower total operational costs compared to traditional methods, offering operators a flexible framework to balance robustness and economic efficiency. As summarized in Table 8.1, the proposed uncertainty sets provide a favorable trade-off between modeling flexibility and computational scalability, which is particularly important for hybrid AC/DC microgrids with bi-directional converters.

Fifth, to ensure the practical applicability of the advanced theoretical models, this thesis developed customized solution algorithms for large-scale robust microgrid operation. The multi-level robust optimization problems arising from the integration of novel uncertainty sets and complex components like hydrogen energy storage systems (HESS) and electric vehicles (EVs) are notoriously difficult to solve with standard approaches. This work designed a series of tailored algorithmic frameworks to ensure computational tractability and scalability. These include a semi-convex relaxation method to efficiently handle the non-convex complementarity constraints inherent in HESS and EV operation, a customized column-and-constraint generation (C&CG) algorithm enhanced with a checking loop procedure and block coordinate descent to improve convergence, and a quad-level decomposition strategy supported by a novel Taguchi Orthogonal Array Testing (TOAT)-based initialization method for efficient exploration of high-dimensional uncertainty spaces. Comparative studies demonstrated that these customized algorithms achieve faster convergence, higher-quality solutions, and superior scalability, effectively translating the theoretical advancements of this thesis into practical tools for real-world HMG operation.

In summation, the primary significance of this thesis lies in the integration of these five contributions into a unified, multi-layered framework. It establishes a novel and powerful synergy between data-driven smart meter analytics and uncertainty-aware system optimization. This holistic approach enables the intelligent harnessing of demand-side flexibility, informed by high-fidelity, real-world data, to cost-effectively mitigate the uncertainty of renewable generation at the system level. By jointly addressing the fundamental challenges of converter modeling, load observability, and renewable uncertainty, this work provides a comprehensive set of theoretical foundations and practical tools to advance the flexible, robust, and economically efficient operation of the hybrid microgrids that will be integral to future low-carbon energy systems.

8.2 Future Perspectives

While this thesis provides a comprehensive framework for the flexible economic dispatch of HMGs, its findings also illuminate several promising avenues for future research. The work presented herein serves as a foundation upon which more dynamic, adaptive, integrated, and intelligent energy systems can be built. The following research directions extend the contributions of this thesis toward these long-term goals.

1) From Dispatch to Real-Time Control: Integrating Dynamic Modeling and Hardware-in-the-Loop Validation. The scope of this thesis is centered on economic dispatch, which operates on a timescale of minutes to hours and relies on steady-state models for power system components. This approach effectively optimizes energy scheduling but intentionally excludes the faster electromagnetic and electromechanical dynamics (on the order of milliseconds to seconds) that are critical for maintaining system stability. A crucial next step is to bridge the gap between high-level economic dispatch and low-level real-time control. This would involve developing detailed dynamic models for power electronic interfaces, such as BDCs and inverters, and integrating these models into a hierarchical control structure. Advanced control strategies, particularly model predictive control (MPC), could be explored to manage these fast dynamics in a way that respects the economic objectives determined by the dispatch layer, thus creating a cohesive multi-timescale control framework. To validate and de-risk such real-time controllers before field deployment, power-hardware-in-the-loop (PHIL) simulation is an indispensable methodology. This would involve testing the developed control algorithms on a real-time digital simulator that models the HMG, while interfacing with physical hardware components. This research direction would advance the findings of this thesis from the realm of planning and scheduling toward practical implementation and operational stability assurance.

2) Towards Lifelong Learning Systems: Online, Adaptive, and Privacy-Preserving NILM. The HMMOE models for NILM developed in this thesis are trained offline on static datasets. In real-world applications, however, households continuously acquire new appliances and user behavior patterns evolve, which would cause the performance of a static model to degrade over time. Future work should therefore focus on developing online and continual learning frameworks for NILM. Such frameworks would enable the HMMOE model to adapt to new appliance signatures and changing consumption patterns in a lifelong learning fashion without requiring periodic, computationally expensive retraining from scratch. Furthermore, the data-intensive nature of NILM raises significant user privacy concerns, which represent a major barrier to real-world adoption.

Federated learning (FL) offers a powerful, privacy-by-design solution to this challenge. A future FL-based HMMOE framework could be developed where local models are trained on individual household data, keeping sensitive raw data private and on-premise, while only anonymized model parameter updates are shared with a central server for aggregation. This approach would allow for the collaborative creation of a powerful, generalized NILM model that benefits from diverse data sources without compromising user privacy. It also naturally handles the data heterogeneity that exists across different households, which is a known challenge in FL applications. This path transforms the NILM framework from a research prototype into a scalable and socially acceptable technology.

3) AI-Driven Decision Making. The robust optimization framework central to this thesis is a powerful, model-based approach that provides deterministic guarantees on performance. However, its efficacy depends on the availability of an accurate system model, and it can become computationally intensive as system complexity and the number of uncertainties grow. A paradigm-shifting future direction is to explore model-free approaches using deep reinforcement learning (DRL) for HMG energy management. A DRL agent could learn an optimal control policy through direct trial-and-error interaction with the microgrid environment, without requiring an explicit mathematical model of the system's dynamics. DRL is particularly adept at solving complex sequential decision-making problems in dynamic and uncertain environments. A DRL-based energy management system could potentially learn more sophisticated and adaptive control strategies than those derived from traditional optimization, automatically adapting to unforeseen events, real-time electricity price fluctuations, and evolving load patterns. This research path represents a move towards a truly autonomous, self-learning intelligent grid, aligning with the long-term vision of AI-driven power systems.

Chapter 9

References

[1] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, “Microgrids”, *IEEE Power Energy Mag.*, vol. 5, no. 4, pp. 78–94, 2007.

[2] G. S. Chawda, W. Su, and M. Wang, “A comprehensive review of high-frequency ac microgrids for distribution systems”, *IEEE Trans. Smart Grid*, vol. 15, no. 6, pp. 5415–5428, 2024.

[3] J. I. D. Cisneros-Saldana, S. Samal, M. M. Begovic, and S. R. Samantaray, “On protection schemes for ac microgrids: Challenges and opportunities”, *IEEE Trans. Ind. Appl.*, vol. 60, no. 3, pp. 4843–4854, 2024.

[4] B. Patnaik, M. Mishra, R. C. Bansal, and R. K. Jena, “Ac microgrid protection–a review: Current and future prospective”, *Appl. Energy*, vol. 271, p. 115 210, 2020.

[5] B. Tan, Z. Liang, C. Chung, H. Tan, H. Wang, and H. Yang, “Risk-based data-driven energy management for integrated electrical and hydrogen microgrids with improved hydrogen vehicle charging prediction”, *Appl. Energy*, vol. 398, p. 126 410, 2025.

[6] F. Alasali, H. Mustafa, A. S. Saidi, *et al.*, “The recent development of protection coordination schemes based on inverse of ac microgrid: A review”, *IET Gener. Transm. Distrib.*, vol. 18, no. 1, pp. 1–23, 2024.

[7] F. S. Al-Ismail, “A critical review on dc microgrids voltage control and power management”, *IEEE Access*, vol. 12, pp. 30 345–30 361, 2024.

[8] K. Quizhpe, P. Arévalo, D. Ochoa-Correa, and E. Villa-Ávila, “Optimizing microgrid planning for renewable integration in power systems: A comprehensive review”, *Electronics*, vol. 13, no. 18, p. 3620, 2024.

[9] M. S. Alam, M. A. Hossain, M. Shafiullah, *et al.*, “Renewable energy integration with dc microgrids: Challenges and opportunities”, *Electr. Power Syst. Res.*, vol. 234, p. 110 548, 2024.

[10] D. A. Memon, K. H. Ali, A. A. Memon, *et al.*, “Comparative analysis and implementation of dc microgrid systems versus ac microgrid performance”, *Front. Energy Res.*, vol. 12, p. 1 370 547, 2024.

[11] V. Patel, V. K. Giri, and A. Kumar, “Efficient power management strategies for ac/dc microgrids with multiple voltage buses for sustainable renewable energy integration”, *Energy Informatics*, vol. 7, no. 1, p. 97, 2024.

[12] A. Kumar, A. Maulik, and K. Chinmaya, “A decentralized energy management scheme for a dc microgrid with correlated uncertainties and integrated demand response”, *Electr. Power Syst. Res.*, vol. 238, p. 111 093, 2025.

[13] H. Binqadhi, W. M. Hamanah, M. Shafiullah, M. S. Alam, M. M. AlMuhaini, and M. A. Abido, “A comprehensive survey on advancement and challenges of dc microgrid protection”, *Sustainability*, vol. 16, no. 14, p. 6008, 2024.

[14] A. E. Shivadasan and P. K. Shanmugam, “Advanced fault detection methodologies and communication protocols for dc micro grid-a technical review”, *Results Eng.*, p. 104 398, 2025.

[15] Z. Liang, X. Yin, C. Y. Chung, S. K. Rayeem, X. Chen, and H. Yang, “Managing massive res integration in hybrid microgrids: A data-driven quad-level approach with adjustable conservativeness”, *IEEE Trans. Ind. Inform.*, vol. 21, no. 10, pp. 7698–7709, 2025.

[16] Z. Liang, C. Chung, Q. Wang, H. Chen, H. Yang, and C. Wu, “Fortifying renewable-dominant hybrid microgrids: A bi-directional converter based interconnection planning approach”, *Engineering*, vol. 51, pp. 130–143, 2025.

[17] Z. Liang, C. Chung, W. Zhang, Q. Wang, W. Lin, and C. Wang, “Enabling high-efficiency economic dispatch of hybrid AC/DC networked microgrids: Steady-state convex bi-directional converter models”, *IEEE Trans. Smart Grid*, vol. 16, no. 1, pp. 45–61, 2025.

[18] A. S. Dahane and R. B. Sharma, “Hybrid ac-dc microgrid coordinated control strategies: A systematic review and future prospect”, *Renew. Energ. Focus*, vol. 49, p. 100 553, 2024.

[19] Z. Liang, C. Chung, H. Dong, S. He, X. Yin, and J. Chen, “Energy management in microgrids: Sufficient conditions for convex bi-directional converter operation model”, in *2024 IEEE PES 16th Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2024, pp. 1–5.

[20] A. Mahmoudian, R. Garmabdari, F. Bai, J. M. Guerrero, M. Mousavizade, and J. Lu, “Adaptive power-sharing strategy in hybrid ac/dc microgrid for enhancing voltage and frequency regulation”, *Int. J. Electr. Power Energy Syst.*, vol. 156, p. 109 696, 2024.

[21] X. Fu and Y. Zhou, “Collaborative optimization of PV greenhouses and clean energy systems in rural areas”, *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 642–656, 2023.

[22] B. Papari, C. S. Edrington, I. Bhattacharya, and G. Radman, “Effective energy management of hybrid AC–DC microgrids with storage devices”, *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 193–203, 2019.

[23] Z. Li, Y. Xu, S. Fang, X. Zheng, and X. Feng, “Robust coordination of a hybrid AC/DC multi-energy ship microgrid with flexible voyage and thermal loads”, *IEEE Trans. on Smart Grid*, vol. 11, no. 4, pp. 2782–2793, 2020.

[24] H. Qiu, W. Gu, X. Xu, *et al.*, “A historical-correlation-driven robust optimization approach for microgrid dispatch”, *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1135–1148, 2021.

[25] J.-C. Fernández-Pérez, F. M. Echavarren Cerezo, and L. Rouco Rodríguez, “Linear power flow algorithm with losses for multi-terminal VSC AC/DC power systems”, *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 1739–1749, 2022.

[26] Y. Fu, Z. Zhang, Z. Li, and Y. Mi, “Energy management for hybrid AC/DC distribution system with microgrid clusters using non-cooperative game theory and robust optimization”, *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1510–1525, 2020.

[27] J.-S. Lai, L. Leslie, J. Ferrell, and T. Nergaard, “Characterization of HV-IGBT for high-power inverter applications”, in *Fourtieth IAS Annual Meeting. Conference Record of the 2005 Industry Applications Conference, 2005.*, vol. 1, 2005, 377–382 Vol. 1.

[28] P. Li, P. Han, and S. He, “Double-uncertainty optimal operation of hybrid AC/DC microgrids with high proportion of intermittent energy sources”, *J. Mod. Power Syst. Clean Energy*, vol. 5, no. 6, pp. 838–849, 2017.

[29] K. Thirugnanam, M. S. El Moursi, V. Khadkikar, H. H. Zeineldin, and M. A. Hosani, “Energy management strategy of a reconfigurable grid-tied hybrid AC/DC microgrid for commercial building applications”, *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 1720–1738, 2022.

[30] B. Wei, X. Han, P. Wang, H. Yu, W. Li, and L. Guo, “Temporally coordinated energy management for AC/DC hybrid microgrid considering dynamic conversion efficiency of bidirectional AC/DC converter”, *IEEE Access*, vol. 8, pp. 70 878–70 889, 2020.

[31] T. Zhao, X. Liu, P. Wang, and F. Blaabjerg, “More efficient energy management for networked hybrid AC/DC microgrids with multivariable nonlinear conversion losses”, *IEEE Syst. J.*, vol. 17, no. 2, pp. 3212–3223, 2023.

[32] J.-W. Lim, J. Hassan, and M. Kim, “Bidirectional soft switching push–pull resonant converter over wide range of battery voltages”, *IEEE Trans. Power Electron.*, vol. 36, no. 11, pp. 12 251–12 267, 2021.

[33] K. Mozaffari, M. Amirabadi, and Y. Deshpande, “A single-phase inverter/rectifier topology with suppressed double-frequency ripple”, *IEEE Trans. Power Electron.*, vol. 33, no. 11, pp. 9282–9295, 2018.

[34] Z. Liang, H. Chen, X. Wang, S. Chen, and C. Zhang, "Risk-based uncertainty set optimization method for energy management of hybrid AC/DC microgrids with uncertain renewable generation", *IEEE Trans. on Smart Grid*, vol. 11, no. 2, pp. 1526–1542, 2020.

[35] H. Qiu, W. Gu, Y. Xu, and B. Zhao, "Multi-time-scale rolling optimal dispatch for AC/DC hybrid microgrids with day-ahead distributionally robust scheduling", *IEEE Trans. Sustainable Energy*, vol. 10, no. 4, pp. 1653–1663, 2019.

[36] H. Qiu and H. B. Gooi, "Identifying differential scheduling plans for microgrid operations under diverse uncertainties", *IEEE Trans. Sustainable Energy*, vol. 14, no. 1, pp. 309–324, 2023.

[37] Q. Xu, T. Zhao, Y. Xu, Z. Xu, P. Wang, and F. Blaabjerg, "A distributed and robust energy management system for networked hybrid AC/DC microgrids", *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3496–3508, 2020.

[38] Z. Li, L. Wu, Y. Xu, and X. Zheng, "Stochastic-weighted robust optimization based bilayer operation of a multi-energy building microgrid considering practical thermal loads and battery degradation", *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 668–682, 2022.

[39] X. Wang, Z. Li, M. Shahidehpour, and C. Jiang, "Robust line hardening strategies for improving the resilience of distribution systems with variable renewable resources", *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 386–395, 2019.

[40] H. Qiu, L. Wang, W. Gu, *et al.*, "Multistage scheduling of regional power grids against sequential outage and power uncertainties", *IEEE Trans. on Smart Grid*, vol. 13, no. 6, pp. 4624–4637, 2022.

[41] X. Zheng, M. E. Khodayar, J. Wang, M. Yue, and A. Zhou, "Distributionally robust multistage dispatch with discrete recourse of energy storage systems", *IEEE Trans. Power Syst.*, vol. 39, no. 6, pp. 6960–6973, 2024.

[42] EMSD, *Hong kong energy end-use data 2023*, `https://www.emsd.gov.hk/filemanager/en/content_762/HKEEUD2023.pdf`, Accessed 2023.

[43] R. Gopinath, M. Kumar, C. Prakash Chandra Joshua, and K. Srinivas, "Energy management using non-intrusive load monitoring techniques – State-of-the-art and future research directions", *Sust. Cities Soc.*, vol. 62, p. 102 411, Nov. 2020, ISSN: 22106707.

[44] J. Liang, S. K. Ng, G. Kendall, and J. W. Cheng, "Load signature study—part i: Basic concept, structure, and methodology", *IEEE Trans. Power Deliv.*, vol. 25, no. 2, pp. 551–560, 2010.

[45] P. A. Schirmer and I. Mporas, "Non-intrusive load monitoring: A review", *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 769–784, 2023.

[46] D. Egarter and W. Elmenreich, "Load disaggregation with metaheuristic optimization", 2015.

[47] S. Zeinal-Kheiri, A. M. Shotorbani, and B. Mohammadi-Ivatloo, "Residential load disaggregation considering state transitions", *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 743–753, 2020.

[48] M. Z. A. Bhotto, S. Makonin, and I. V. Bajić, "Load disaggregation based on aided linear integer programming", *IEEE Trans. Circuits Syst. II-Express Briefs*, vol. 64, no. 7, pp. 792–796, 2017.

[49] F. M. Wittmann, J. C. López, and M. J. Rider, "Nonintrusive load monitoring algorithm using mixed-integer linear programming", *IEEE Trans. Consum. Electron.*, vol. 64, no. 2, pp. 180–187, 2018.

[50] C. Lu, L. Ma, T. Xu, G. Ding, C. Wu, and X. Jiang, "Non-intrusive load monitoring method based on improved differential evolution algorithm", in *Proc. 11th Int. Conf. Meas. Technol. Mechatronics Autom.*, 2019, pp. 279–283.

[51] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An extensible approach for non-intrusive load disaggregation with smart meter data", *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3362–3372, 2018.

[52] P. Kumar and A. R. Abhyankar, "A time efficient factorial hidden markov model-based approach for non-intrusive load monitoring", *IEEE Trans. Smart Grid*, vol. 14, no. 5, pp. 3627–3639, 2023.

[53] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, “Non-intrusive load monitoring by using active and reactive power in additive factorial hidden markov models”, *Appl. Energy*, vol. 208, pp. 1590–1607, 2017.

[54] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao, and F. J. Luo, “A hierarchical hidden markov model framework for home appliance modeling”, *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3079–3090, 2018.

[55] Z. Wu, C. Wang, H. Zhang, W. Peng, and W. Liu, “A time-efficient factorial hidden semi-markov model for non-intrusive load monitoring”, *Electr. Power Syst. Res.*, vol. 199, p. 107 372, 2021.

[56] D. García-Pérez, D. Pérez-López, I. Díaz-Blanco, A. González-Muñiz, M. Domínguez-González, and A. A. Cuadrado Vega, “Fully-convolutional denoising auto-encoders for nilm in large non-residential buildings”, *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2722–2731, 2021.

[57] P. A. Schirmer and I. Mporas, “Double fourier integral analysis based convolutional neural network regression for high-frequency energy disaggregation”, *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 3, pp. 439–449, 2022.

[58] J. Yuan, R. Jin, L. Wang, and T. Wang, “A nonintrusive load identification method based on dual-branch attention gru fusion network”, *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–10, 2024.

[59] H. Hwang and S. Kang, “Nonintrusive load monitoring using an lstm with feedback structure”, *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[60] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, “Short-term residential load forecasting based on lstm recurrent neural network”, *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.

[61] Y. Zhang, H. Wu, Q. Ma, Q. Yang, and Y. Wang, “A learnable image-based load signature construction approach in nilm for appliances identification”, *IEEE Trans. Smart Grid*, vol. 14, no. 5, pp. 3841–3849, 2023.

[62] K. Chen, Y. Zhang, Q. Wang, J. Hu, H. Fan, and J. He, "Scale- and context-aware convolutional non-intrusive load monitoring", *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 2362–2373, 2020.

[63] Y. Liu, J. Qiu, and J. Ma, "Samnet: Toward latency-free non-intrusive load monitoring via multi-task deep learning", *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2412–2424, 2022.

[64] S. Dash and N. C. Sahoo, "A multi-task deep learning approach for non-intrusive load monitoring of multiple appliances", *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 3337–3340, 2024.

[65] R. Xie, C. Li, X. Zhou, H. Chen, and Z. Dong, "Differentially private federated learning for multitask objective recognition", *IEEE Trans. Ind. Inform.*, vol. 20, no. 5, pp. 7269–7281, 2024.

[66] Y. Zhou, G. Karamanolakis, V. Soto, *et al.*, "Mergeme: Model merging techniques for homogeneous and heterogeneous moes", *arXiv preprint arXiv:2502.00997*, 2025.

[67] A. Wang, X. Sun, R. Xie, *et al.*, "Hmoe: Heterogeneous mixture of experts for language modeling", *arXiv preprint arXiv:2408.10681*, 2024.

[68] G. Jawahar, S. Mukherjee, X. Liu, *et al.*, "Automoe: Heterogeneous mixture-of-experts with adaptive computation for efficient neural machine translation", *arXiv preprint arXiv:2210.07535*, 2022.

[69] W. Cai, J. Jiang, F. Wang, J. Tang, S. Kim, and J. Huang, "A survey on mixture of experts in large language models", *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 7, pp. 3896–3915, 2025.

[70] B. Papari, C. S. Edrington, I. Bhattacharya, and G. Radman, "Effective energy management of hybrid AC–DC microgrids with storage devices", *IEEE Trans. on Smart Grid*, vol. 10, no. 1, pp. 193–203, 2019.

[71] X. Fu, C. Zhang, Y. Xu, Y. Zhang, and H. Sun, "Statistical machine learning for power flow analysis considering the influence of weather factors on photovoltaic power generation", *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–15, 2024.

[72] H. Yang, W. Zhang, C. Y. Chung, Z. Wang, W. Qiu, and Z. Liang, “Ac false data injection attack based on robust tensor principle component analysis”, *IEEE Trans. Ind. Inf.*, vol. 20, no. 8, pp. 9887–9897, 2024.

[73] X. Liu, Y. Wang, X. Liu, T. Zhao, and P. Wang, “More efficient energy management for hybrid AC/DC microgrids with hard constrained neural network-based conversion loss surrogate models”, *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 2534–2552, 2024.

[74] T. Zhao, X. Pan, S. Yao, C. Ju, and L. Li, “Strategic bidding of hybrid AC/DC microgrid embedded energy hubs: A two-stage chance constrained stochastic programming approach”, *IEEE Trans. Sustainable Energy*, vol. 11, no. 1, pp. 116–125, 2020.

[75] P. Sharma, H. D. Mathur, P. Mishra, and R. C. Bansal, “A critical and comparative review of energy management strategies for microgrids”, *Appl. Energy*, vol. 327, p. 120 028, 2022.

[76] Z.-P. Yuan, P. Li, Z.-L. Li, and J. Xia, “Data-driven risk-adjusted robust energy management for microgrids integrating demand response aggregator and renewable energies”, *IEEE Trans. on Smart Grid*, vol. 14, no. 1, pp. 365–377, 2023.

[77] F. Valencia, J. Collado, D. Sáez, and L. G. Marín, “Robust energy management system for a microgrid based on a fuzzy prediction interval model”, *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1486–1494, 2016.

[78] Y. Xiang, J. Liu, and Y. Liu, “Robust energy management of microgrid with uncertain renewable generation and load”, *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 1034–1043, 2016.

[79] B. Wang, C. Zhang, C. Li, P. Li, Z. Y. Dong, and J. Lu, “Hybrid interval-robust adaptive battery energy storage system dispatch with soc interval management for unbalanced microgrids”, *IEEE Trans. Sustainable Energy*, vol. 13, no. 1, pp. 44–55, 2022.

[80] W. Lin, Z. Yang, J. Yu, K. Xie, X. Wang, and W. Li, “Tie-line security region considering time coupling”, *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1274–1284, 2021.

[81] Y. Guo, K. Baker, E. Dall’Anese, Z. Hu, and T. H. Summers, “Data-based distributionally robust stochastic optimal power flow—part i: Methodologies”, *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1483–1492, 2019.

[82] R. Zhu, H. Wei, and X. Bai, “Wasserstein metric based distributionally robust approximate framework for unit commitment”, *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2991–3001, 2019.

[83] C. Zhao and Y. Guan, “Data-driven stochastic unit commitment for integrating wind generation”, *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2587–2596, 2016.

[84] H. Qiu and H. B. Gooi, “Identifying differential scheduling plans for microgrid operations under diverse uncertainties”, *IEEE Trans. Sustainable Energy*, vol. 14, no. 1, pp. 309–324, 2023.

[85] C. Zhang, Y. Xu, Z. Y. Dong, and L. F. Yang, “Multitimescale coordinated adaptive robust operation for industrial multienergy microgrids with load allocation”, *IEEE Trans. Ind. Inform.*, vol. 16, no. 5, pp. 3051–3063, 2020.

[86] Z. Li, Y. Xu, S. Fang, X. Zheng, and X. Feng, “Robust coordination of a hybrid AC/DC multi-energy ship microgrid with flexible voyage and thermal loads”, *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 2782–2793, 2020.

[87] Z. Zhao, J. Guo, X. Luo, *et al.*, “Distributed robust model predictive control-based energy management strategy for islanded multi-microgrids considering uncertainty”, *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2107–2120, 2022.

[88] Z. Liang, H. Chen, X. Wang, S. Chen, and C. Zhang, “Risk-based uncertainty set optimization method for energy management of hybrid AC/DC microgrids with uncertain renewable generation”, *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1526–1542, 2020.

[89] M. Fan, Z. Li, T. Ding, *et al.*, “Uncertainty evaluation algorithm in power system dynamic analysis with correlated renewable energy sources”, *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5602–5611, 2021.

[90] H. Qiu, H. Long, W. Gu, and G. Pan, “Recourse-cost constrained robust optimization for microgrid dispatch with correlated uncertainties”, *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2266–2278, 2021.

[91] M. A. El-Meligy, A. M. El-Sherbeeny, and A. Anvari-Moghaddam, “Transmission expansion planning considering resistance variations of overhead lines using minimum-volume covering ellipsoid”, *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 1916–1926, 2022.

[92] M. A. El-Meligy, A. M. El-Sherbeeny, and A. Anvari-Moghaddam, “Transmission expansion planning considering resistance variations of overhead lines using minimum-volume covering ellipsoid”, *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 1916–1926, 2022.

[93] P. Li, X. Guan, J. Wu, and X. Zhou, “Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling”, *IEEE Trans. Sustainable Energy*, vol. 6, no. 4, pp. 1594–1605, 2015.

[94] X. Li, Y. Liu, L. Guo, X. Li, and C. Wang, “Data-driven based uncertainty set modeling method for microgrid robust optimization with correlated wind power”, *CSEE J. Power Energy Syst.*, vol. 9, no. 2, pp. 420–432, 2022.

[95] A. Velloso, A. Street, D. Pozo, J. M. Arroyo, and N. G. Cobos, “Two-stage robust unit commitment for co-optimized electricity markets: An adaptive data-driven approach for scenario-based uncertainty sets”, *IEEE Trans. Sustainable Energy*, vol. 11, no. 2, pp. 958–969, 2020.

[96] L. Zeng, H.-D. Chiang, and T. Wang, “A pairwise convex hull approach for effective representation of uncertainty for system analysis and its application to power grids”, *IEEE Trans. Circuits Syst. Express Briefs*, vol. 68, no. 7, pp. 2498–2502, 2021.

[97] S. Chen, Z. Wei, G. Sun, W. Wei, and D. Wang, “Convex hull based robust security region for electricity-gas integrated energy systems”, *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1740–1748, 2019.

[98] J. Zhong, Y. Li, Y. Cao, *et al.*, “Robust coordinated optimization with adaptive uncertainty set for a multi-energy microgrid”, *IEEE Trans. Sustainable Energy*, vol. 14, no. 1, pp. 111–124, 2023.

[99] H. Qiu, W. Gu, X. Xu, *et al.*, “A historical-correlation-driven robust optimization approach for microgrid dispatch”, *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1135–1148, 2021.

[100] B. Zhou, J. Fang, X. Ai, *et al.*, "Partial-Dimensional correlation-aided convex-hull uncertainty set for robust unit commitment", *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2434–2446, 2023.

[101] Q. Xu, T. Zhao, Y. Xu, Z. Xu, P. Wang, and F. Blaabjerg, "A distributed and robust energy management system for networked hybrid AC/DC microgrids", *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3496–3508, 2020.

[102] M. R. Ebrahimi and N. Amjady, "Adaptive robust optimization framework for day-ahead microgrid scheduling", *Int. J. Electr. Power Energy Syst.*, vol. 107, pp. 213–223, 2019.

[103] B. Tan, S. Chen, Z. Liang, X. Zheng, Y. Zhu, and H. Chen, "An iteration-free hierarchical method for the energy management of multiple-microgrid systems with renewable energy sources and electric vehicles", *Appl. Energy*, vol. 356, p. 122 380, 2024.

[104] H. Qiu, W. Gu, Y. Xu, and B. Zhao, "Multi-time-scale rolling optimal dispatch for AC/DC hybrid microgrids with day-ahead distributionally robust scheduling", *IEEE Trans. Sustainable Energy*, vol. 10, no. 4, pp. 1653–1663, 2019.

[105] B. Zhao, H. Qiu, R. Qin, X. Zhang, W. Gu, and C. Wang, "Robust optimal dispatch of AC/DC hybrid microgrids considering generation and load uncertainties and energy storage loss", *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 5945–5957, 2018.

[106] P. T. An, N. D. Hoang, and N. K. Linh, "An efficient improvement of gift wrapping algorithm for computing the convex hull of a finite set of points in R n", *Numer. Algorithms*, vol. 85, no. 4, pp. 1499–1518, 2020.

[107] N. K. Linh, P. T. An, and T. Van Hoai, "A fast and efficient algorithm for determining the connected orthogonal convex hulls", *Appl. Math. Comput.*, vol. 429, p. 127 183, 2022.

[108] D. Bertsekas, *Convex optimization theory*. Athena Scientific, 2009, vol. 1.

[109] H. Chen, Y. Zhu, and Z. Liang, "Energy scheduling of virtual power plant: A data-driven enclosing polyhedron method", *IEEE Trans. Ind. Inform.*, vol. Early Access, pp. 1–1, 2024.

[110] L. Che, X. Zhang, M. Shahidehpour, A. Alabdulwahab, and A. Abusorrah, "Optimal interconnection planning of community microgrids with renewable energy sources", *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1054–1063, 2017.

[111] A. A. Hamad, M. E. Nassar, E. F. El-Saadany, and M. M. A. Salama, "Optimal configuration of isolated hybrid AC/DC microgrids", *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2789–2798, 2019.

[112] H. Qiu, W. Gu, and F. You, "Bilayer distributed optimization for robust microgrid dispatch with coupled individual-collective profits", *IEEE Trans. on Sustain. Energy*, vol. 12, no. 3, pp. 1525–1538, 2021.

[113] K. Thirugnanam, S. K. Kerk, C. Yuen, N. Liu, and M. Zhang, "Energy management for renewable microgrid in reducing diesel generators usage with multiple types of battery", *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6772–6786, 2018.

[114] P. Li, P. Han, S. He, and X. Wang, "Double-uncertainty optimal operation of hybrid AC/DC microgrids with high proportion of intermittent energy sources", *J. Mod. Power Syst. Clean Ener*, vol. 5, no. 6, pp. 838–849, 2017.

[115] B. Zhao, H. Qiu, R. Qin, X. Zhang, W. Gu, and C. Wang, "Robust optimal dispatch of AC/DC hybrid microgrids considering generation and load uncertainties and energy storage loss", *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 5945–5957, 2018.

[116] L. Guo, R. Hou, Y. Liu, C. Wang, and H. Lu, "A novel typical day selection method for the robust planning of stand-alone wind-photovoltaic-diesel-battery microgrid", *Appl. Energy*, vol. 263, p. 114 606, 2020.

[117] C. Zhang, Q. Liu, B. Zhou, *et al.*, "A central limit theorem-based method for dc and ac power flow analysis under interval uncertainty of renewable power generation", *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 563–575, 2023.

[118] Y. Yu, G.-P. Liu, Y. Huang, and P. Shi, "Optimal cooperative secondary control for islanded DC microgrids via a fully actuated approach", *IEEE-CAA J. Automatica Sin.*, vol. 11, no. 2, pp. 405–417, 2024.

[119] F. Blaabjerg, R. Teodorescu, M. Liserre, and A. Timbus, “Overview of control and grid synchronization for distributed power generation systems”, *IEEE Trans. Ind. Electron.*, vol. 53, no. 5, pp. 1398–1409, 2006.

[120] A. Bidram and A. Davoudi, “Hierarchical structure of microgrids control system”, *IEEE Trans. on Smart Grid*, vol. 3, no. 4, pp. 1963–1976, 2012.

[121] M. Dabbaghjamanesh, A. Kavousi-Fard, S. Mehraeen, J. Zhang, and Z. Y. Dong, “Sensitivity analysis of renewable energy integration on stochastic energy management of automated reconfigurable hybrid AC–DC microgrid considering dlr security constraint”, *IEEE Trans. Ind. Inf.*, vol. 16, no. 1, pp. 120–131, 2020.

[122] H.-G. Yeh, D. F. Gayme, and S. H. Low, “Adaptive VAR control for distribution circuits with photovoltaic generators”, *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1656–1663, 2012.

[123] C. Zhang, Y. Xu, Y. Wang, Z. Y. Dong, and R. Zhang, “Three-stage hierarchically-coordinated voltage/var control based on PV inverters considering distribution network voltage stability”, *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 868–881, 2022.

[124] G. Notton, M.-L. Nivet, C. Voyant, *et al.*, “Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting”, *Renewable Sustainable Energy Rev.*, vol. 87, pp. 96–105, 2018.

[125] A. Zakaria, F. B. Ismail, M. H. Lipu, and M. A. Hannan, “Uncertainty models for stochastic optimization in renewable energy applications”, *Renew. Energy*, vol. 145, pp. 1543–1571, 2020.

[126] K. Antoniadou-Plytaria, D. Steen, O. Carlson, B. Mohandes, M. A. F. Ghazvini, *et al.*, “Scenario-based stochastic optimization for energy and flexibility dispatch of a microgrid”, *IEEE Trans. on Smart Grid*, vol. 13, no. 5, pp. 3328–3341, 2022.

[127] Z. Zhuo, E. Du, N. Zhang, C. Kang, Q. Xia, and Z. Wang, “Incorporating massive scenarios in transmission expansion planning with high renewable energy penetration”, *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1061–1074, 2020.

[128] Z. Liang, H. Chen, S. Chen, Z. Lin, and C. Kang, "Probability-driven transmission expansion planning with high-penetration renewable power generation: A case study in northwestern China", *Appl. Energy*, vol. 255, p. 113 610, 2019.

[129] Z. Liang, H. Chen, S. Chen, Y. Wang, C. Zhang, and C. Kang, "Robust transmission expansion planning based on adaptive uncertainty set optimization under high-penetration wind power generation", *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 2798–2814, 2021.

[130] B. Wang, C. Zhang, C. Li, G. Yang, and Z. Y. Dong, "Transactive energy sharing in a microgrid via an enhanced distributed adaptive robust optimization approach", *IEEE Trans. on Smart Grid*, vol. 13, no. 3, pp. 2279–2293, 2022.

[131] K. Antoniadou-Plytaria, D. Steen, L. A. Tuan, O. Carlson, B. Mohandes, and M. A. F. Ghazvini, "Scenario-based stochastic optimization for energy and flexibility dispatch of a microgrid", *IEEE Trans. on Smart Grid*, vol. 13, no. 5, pp. 3328–3341, 2022.

[132] Z. Liu, L. Wang, and L. Ma, "A transactive energy framework for coordinated energy management of networked microgrids with distributionally robust optimization", *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 395–404, 2020.

[133] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts", in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 1930–1939.

[134] J. Chen, X. Wang, X. Zhang, and W. Zhang, "Temporal and spectral feature learning with two-stream convolutional neural networks for appliance recognition in nilm", *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 762–772, 2022.

[135] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need", *Advances in neural information processing systems*, vol. 30, 2017.

[136] H. Yang, Z. Wang, and R. C. Qiu, "Data domain adaptation for voltage stability evaluation considering topology changes", *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2834–2844, 2023.

[137] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis, “Context aware energy disaggregation using adaptive bidirectional lstm models”, *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3054–3067, 2020.

[138] Q. Mercier, F. Poirion, and J.-A. Désidéri, “A stochastic multiple gradient descent algorithm”, *Eur. J. Oper. Res.*, vol. 271, no. 3, pp. 808–817, 2018.

[139] R. M. Freund and P. Grigas, “New analysis and results for the frank–wolfe method”, *Math. Program.*, vol. 155, no. 1, pp. 199–230, 2016.

[140] J. E. Martínez-Legaz, “On weierstrass extreme value theorem”, *Optim. Lett.*, vol. 8, no. 1, pp. 391–393, 2014.

[141] C. Xiong, Z. Cai, S. Liu, J. Luo, and G. Tu, “An improved sequence-to-point learning for non-intrusive load monitoring based on discrete wavelet transform”, *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–16, 2023.

[142] R. S. Mollel, L. Stankovic, and V. Stankovic, “Using explainability tools to inform nilm algorithm performance: A decision tree approach”, in *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2022, pp. 368–372.

[143] Z. Yan, P. Hao, M. Nardello, D. Brunelli, and H. Wen, “A generalizable load recognition method in nilm based on transferable random forest”, *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025.

[144] J. Kelly and W. Knottenbelt, “The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes”, *Sci. Data*, vol. 2, no. 1, pp. 1–14, 2015.

[145] J. Z. Kolter and M. J. Johnson, “REDD: A public data set for energy disaggregation research”, in *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA*, Citeseer, vol. 25, 2011, pp. 59–62.

[146] *Hong kong observatory*, `https://www.hko.gov.hk/sc/index.html`, Accessed 2025.

[147] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting", *International journal of forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[148] EVbox, *Tesla model 3 charging specifications*. [Online]. Available: https://evbox.com/en/electric-cars/tesla/tesla-model-3.

[149] *Data and supplement materials for hybrid microgrid systems with nilm*, 2025. [Online]. Available: https://1drv.ms/f/c/a67e41e18eb3d70e/Ejdgv_jpY8FJuyytrHmMnvEBFCNMIW8K_BVWgUpBPRGXRg?e=fxnnag.

[150] N. Yang, Y. Wang, Y. Zhang, and D. Yuan, "Non-intrusive load monitoring for energy management in smart grids incorporating evs", *IEEE Trans. Consum. Electron.*, vol. 71, no. 1, pp. 1696–1706, 2025.

[151] H. Çimen, N. Çetinkaya, J. C. Vasquez, and J. M. Guerrero, "A microgrid energy management system based on non-intrusive load monitoring via multitask learning", *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 977–987, 2021.

[152] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance", in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, vol. 30, 1934, pp. 178–191.

[153] K. Bruninx and E. Delarue, "A statistical description of the error on wind power forecasts for probabilistic reserve sizing", *IEEE Trans. Sustainable Energy*, vol. 5, no. 3, pp. 995–1002, 2014.

[154] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls", *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, 1996.

[155] R.-P. Liu, S. Lei, C. Peng, W. Sun, and Y. Hou, "Data-based resilience enhancement strategies for electric-gas systems against sequential extreme weather events", *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5383–5395, 2020.

[156] Z. Liang, *Data and Supplement Materials of Pair-Convex HMG*. 2024. [Online]. Available: https://1drv.ms/f/s!Ag7Xs47hQX6mhPQgFZ8y24V_OvsPfA?e=CCGKQc.

[157] H. Qiu, W. Gu, and F. You, "Bilayer distributed optimization for robust microgrid dispatch with coupled individual-collective profits", *IEEE Trans. Sustain. Energy*, vol. 12, no. 3, pp. 1525–1538, 2021.

[158] *Data and supplement materials for hybrid microgrid systems*, 2024. [Online]. Available: `https://1drv.ms/f/s!Ag7Xs47hQX6mhPdRaf0ix6H7tTyolg?e=VrMjma`.

[159] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks", *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, 2021.

[160] J. Zhong, Y. Li, Y. Cao, *et al.*, "Robust coordinated optimization with adaptive uncertainty set for a multi-energy microgrid", *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 111–124, 2023.

[161] Z. Liang, H. Chen, X. Wang, S. Chen, and C. Zhang, "Risk-based uncertainty set optimization method for energy management of hybrid AC/DC microgrids with uncertain renewable generation", *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1526–1542, 2020.

[162] Z. Zhuo, N. Zhang, J. Yang, *et al.*, "Transmission expansion planning test system for ac/dc hybrid grid with high variable renewable energy penetration", *IEEE Trans on Power Systems*, vol. 35, no. 4, pp. 2597–2608, 2020.

[163] C. Li, Z. Ren, and W. Li, "Cvar-wasserstein metric based power system scheduling considering frequency support of hydrogen production stations", *IEEE Trans. on Industry Applications*, 2024.

[164] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion", *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[165] *Data and Supplement Materials of HMG Uncertainty*. [Online]. Available: `https://1drv.ms/f/s!Ag7Xs47hQX6mhPQgFZ8y24V_OvsPfA?e=CCGKQc`.

[166] Z. Wen, X. Zhang, G. Wang, Y. Li, J. Qiu, and F. Wen, "Data-driven stochastic-robust planning for resilient hydrogen-electricity system with progressive hedging decoupling", *IEEE Trans. Sustain. Energy*, vol. 16, no. 3, pp. 1545–1561, 2025.