Algorithms for Optimal Energy Management in the Smart Grid

by

Yu Wang

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 1, 2015

Keywords: Smart grid, microgrid, energy management, power scheduling, online algorithm, convex optimization, renewable energy

Approved by

Shiwen Mao, Chair, Associate Professor of Electrical and Computer Engineering
R. Mark Nelms, Professor of Electrical and Computer Engineering
Jitendra K Tugnait, Professor of Electrical and Computer Engineering
Ming Liao, Professor of Mathematics and Statistics

Abstract

Smart grid (SG) is regarded as the next generation power grid, which implements an innovative idea for a highly automated and integrated power system. The two-way energy and information flows in the SG, together with the smart devices, bring new perspectives to energy management and demand response. Meanwhile, innovative grid components, such as microgrid (MG) and electric vehicle, are emerging as new applications which bring many benefits as well as more chanllegens in SG. Therefore, we explore possible solutions to these chanllegening but interesting problems.

In this dissertation, we first present an introduction of the SG, and the research involved in different areas of SG. We then investigate an online algorithm for energy distribution in a SG environment. The proposed online algorithm are quite general, suitable for a wide range of utility, cost and pricing functions. And it is asymptotically optimal without any future information. Following this, we then propose a distributed online algorithm. Comparing to the previous one, it solves the online problem in a distributed manner and mitigates the user privacy issue by not sharing user utility functions. Both algorithms are evaluated with trace-driven simulations and shown to outperform a benchmark scheme.

We then propose a hierarchical power scheduling approach to optimally manage power trading, storage and distribution in a smart power grid with a Macrogrid and cooperative MGs. We develop online algorithms both for cooperative MGs and the Macrogrid. The proposed hierarchical power scheduling algorithms are evaluated with trace-driven simulations and are shown to outperform several existing schemes with considerable gains.

Also, we also introduce the simultaneous inference for power generation forecasting from renewable energy resources. We then apply it for solar intensity prediction using a real trace of weather data, where the performance is demonstrated over existing approaches.

Acknowledgments

I have received such immeasurable help from many people, and in such diverse ways, to support me during the development of this dissertation, and more generally, my academic experience. A few words mention here cannot adequately capture all my appreciation.

I would like to express my deepest gratitude to my dissertation advisor, Prof. Shiwen Mao, for his continuing and inspirational guidance, support and encouragement during my Ph.D. program. I would also like to sincerely thank my dissertation committee, Prof. R. Mark Nelms, Prof. Jitendra Tugnait, and Prof. Ming Liao for their valuable comments and time, who contribute their broad perspective in refining the ideas in this dissertation. I am also indebted to Prof. Peng Zeng for serving as the university reader, and reviewing my work.

I want to take this opportunity to recognize all my fellow colleagues in the Department of Electrical and Computer Engineering at Auburn University: Dr. Dongling Hu, Dr. Yingsong Huang, Dr. Yi Xu, Jing Ning, Zhifeng He, Zhefeng Jiang, Mingjie Feng, Xuyu Wang, Yu Wang, Ningkai Tang, Kefan Xiao, and Linjun Gao for the discussions, cooperation and assistance during these years. In addition, I am very grateful to Dr. Yihan Li for her hospitality and help in my study and life at Auburn.

Also, I want to express my thankfulness to my friends Will Abercrombie, Russel Johnson, Noeun Non and his family, who give me great friendships and share with me their great faithfulness. I am also thankful for my best friends, Siming Gu, Bowen Zhang, Hao Sun, Lu Lu, Bingyu Li, and Ning Hou. The time, happiness, memories, and friendships we share together will be a great treasure for the rest of my life.

Besides, I am also thankful for my girlfriend, Ms. Yiran Xu's help and cares. And I have to appreciate the supporting from the China Scholarship Council (CSC), which guarantees my life expenses during the whole time of my PhD programs.

Finally, I would like to extend my thanks with all my heart to my parents, and devote this work to them who give me continuous support and tremendous care. Without them, the achievements of this dissertation could not be possible.

This work was supported in part by the US National Science Foundation (NSF) under grant CNS-0953513, the NSF I/UCRC Broadband Wireless Access & Applications Center (BWAC) site at Auburn University, and the Wireless Engineering Research and Education Center (WEREC) at Auburn university. This work was also supported in part by the China Scholarship Council.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Smart Grid – The Future Power Grid

In 2003, when the Northeast Blackout happened in the United States, 50 million people were left without power up to 2 days [1]. Again, in 2012, when the "superstorm" Hurricane Sandy swept the Atlantic Ocean, 6 million people in 15 states and the District of Columbia were out of power for more than 2 days [2]. Even worse, in these events, people had no idea of what happened to the power grid, and had to wait in anxiety for many hours or even days. Admittedly, the major power grid infrastructure in the United States has lasted for more than 35 years, and is worse than that of many countries. The U.S. government now feels obliged to overhaul the old facility in power grids and to increase the reliability of power delivery. However, aging facility is not the only problem for the current power grid. It requires more efforts in many aspects to create a better power grid. Meanwhile, we have to face the challenges from depleting fossil fuels, global climate change, increasing power demand, etc. Fortunately, the advanced science and technology in many fields may help us to improve the existing power grid. Based on this, an innovative power grid is born – the smart grid.

Smart grid (SG), also called smart power grid or intelligent grid, is regarded as the next generation power grid. It is supposed to replace the current old, dirty, inefficient, and vulnerable power grid. With modern technologies in power system, control theory, communication system, and information theory, two-way flows of electricity and information will be enabled in SG to provide an advanced power system with higher energy efficiency and power delivery stability. Automated metering and monitoring will be realized in SG, based on a large number of smart meters and sensors installed throughout the grid, while

communication and networking technologies guarantee data collection and transmission in real time. And thus, SG will be able to respond quickly to blackouts or broken pieces inside the entire power grid, and then protect working circuits from being affected in the grid so that large area power outages can be avoided. Besides, SG will also support more distributed power generation of renewable energy, such as solar, wind, and geothermal energy, through which the power system capacity will be increased, and the reliance on the fossil fuel will be decreased. Consequently, *greenhouse gas* (GHG) emissions can be controlled.

More specifically, SG can be regarded as a large-scale and complicated power system that utilizes the advanced technologies in many fields to achieve a clean, efficient, reliable, and sustainable system. The intelligence penetrates into every component of the system from power generation to consumption by the customers. The realization of the ultimate SG requires incorporation of technologies in power system, information technology, communication, control theory, and computer science. The construction of SG needs support from government, society, utility companies, and end customers. It will be a large and complicated engineering project lasting for a few decades. We should be confident to accomplish this challenging task, because we have made so many great achievements, such as the space shuttle, spacecraft, and man-made satellites, which were unbelievable dreams several decades ago.

In the future SG, many new facilities and infrastructure will become common and indispensable, such as the distributed generation of renewable energy resources, smart meters and sensors, electric vehicles, and grid energy storage. By integrating these new components, the power grid becomes truly intelligent, efficient, and automatic. New SG components are deployed using the plug-and-play interfaces, which increases the flexibility, scalability, and security of SG. Smart meters and sensors can be embedded into SG directly through the configured interfaces as simple as connecting a laptop to the Internet. In this way, a huge and complex SG system can be decomposed into many small parts with different features.

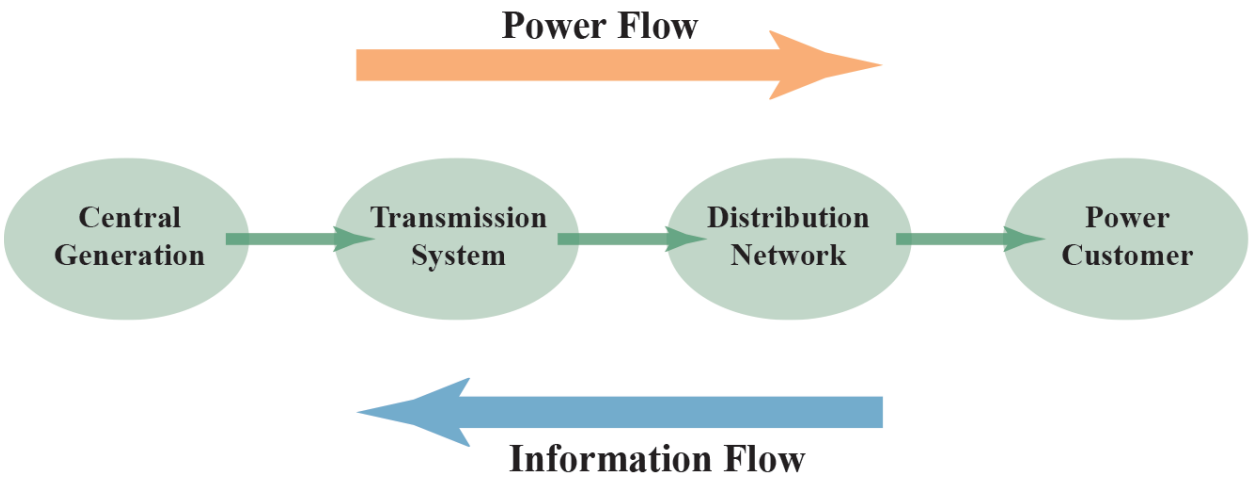

Figure 1.1: Traditional power grid.

For example, *Distributed generation* (DG) and *grid energy storage* (GES) are two new features in SG. DG makes it possible to incorporate more renewable energy generation, such as solar, wind, and tidal. GES is essential for optimal energy management, because it can not only store the extra energy, but also inject energy back to the grid when needed to avoid blackouts and reduce the cost.

Another important feature of SG is the two-way flows of electricity and information. In traditional power grids shown in Fig. 1.1, both electricity and information flow in a unidirectional fashion. Electric power is generated from a centralized generation plant, and then travels through the transmission system and distribution networks to power users. Utility company collects the information of user consumptions and grid status, while power users have no access to acquiring the grid or market information. However, in SG as shown in Fig. 1.2, two-way flows of electricity and information is supported, so that power customers are able to acquire the market information and the grid status, and sell energy back to the grid. In this way, exchanges of information and power become more flexible, and higher efficient power management is enabled for more reliable power distribution. For example, the utility company could lower the electricity price so that the load peak is reduced by power injection from end customers. Also, by periodic information communications, the control

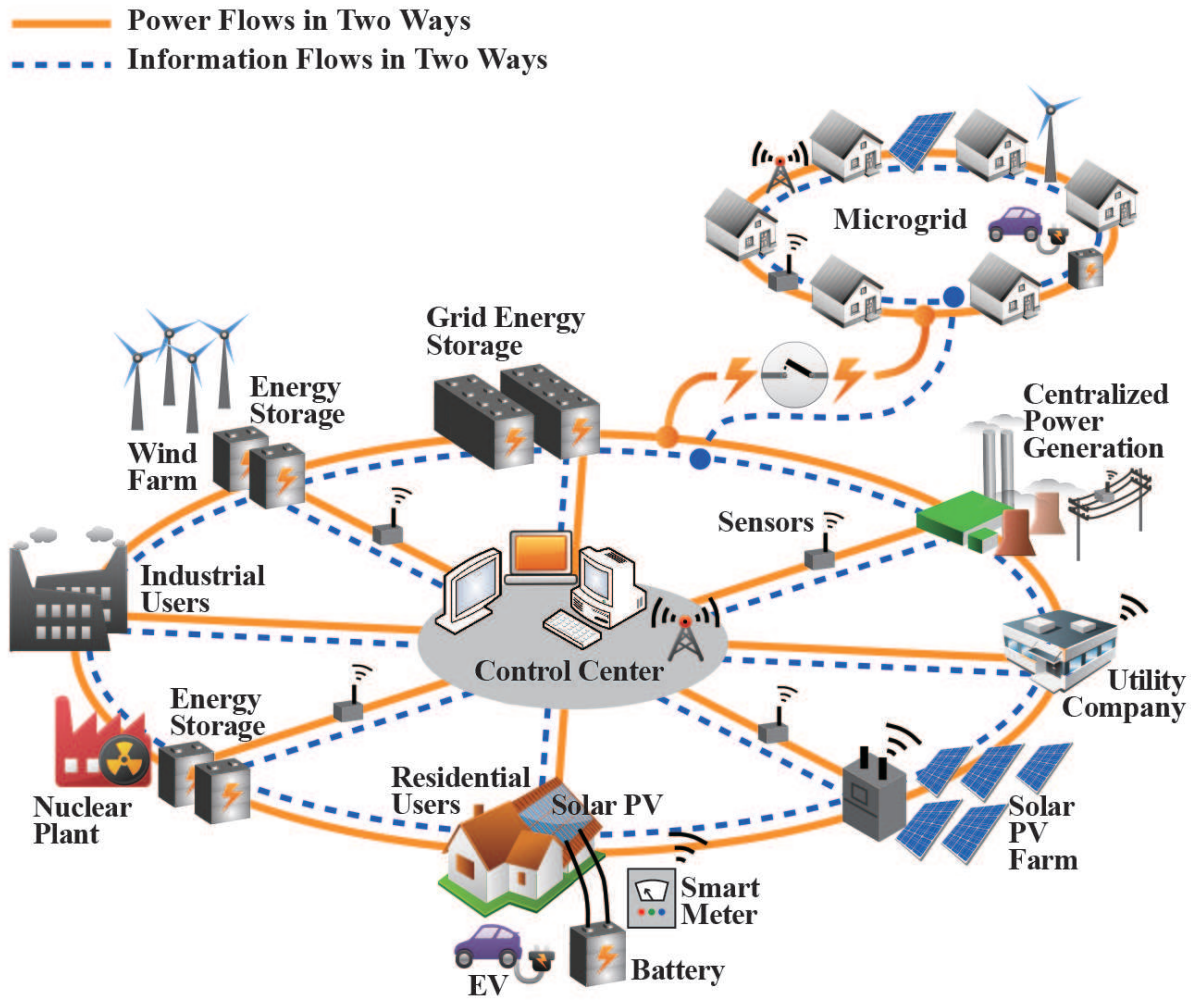

Figure 1.2: A vision of the future smart grid.

center monitors the grid in real time, and customers acquire updated price information in real time. In short, two-way flows of electricity and information are the foundation of the real time power control and many other SG applications.

Currently, there is no explicit definition of SG, because it is such a complex system that covers numerous subsystems, in which some research effort has just started. The research in SG is still in the infant stage. Therefore, it is not clear of the final shape of SG after several decades of development. But a road map could provide a main direction. Many countries or areas have proposed detailed SG road maps, such as the U.S. [3], European Union [4], Germany [5], China [6], and Japan [7]. According to these road maps, SGs are developed with

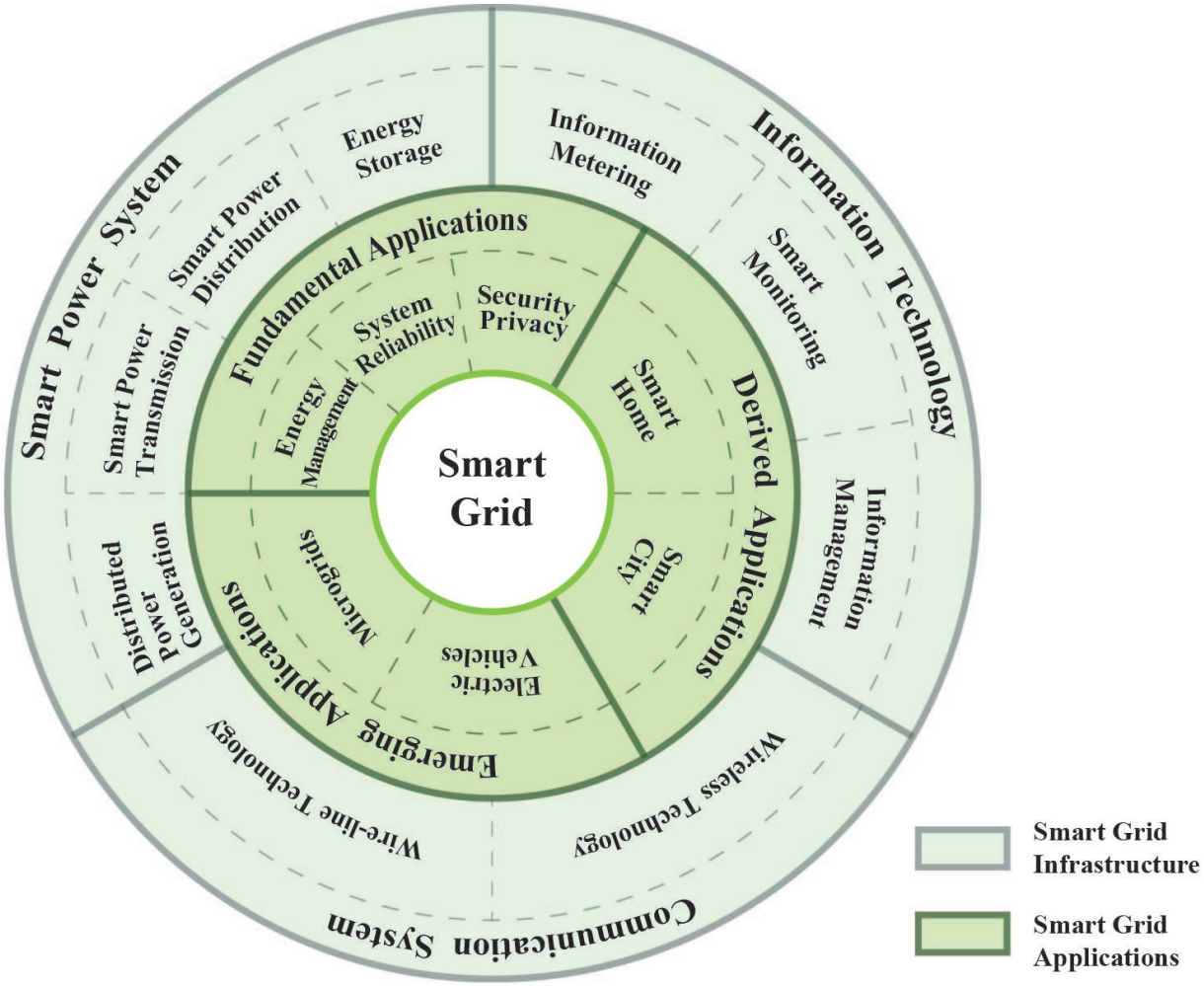

Figure 1.3: The structure of the smart grid.

different objectives, based on the conditions and policies of different countries. For example, in the United States, electric vehicles are supported by the government because the large population of vehicles consume more than 10 million barrels of petroleum products per day with enormous GHG emissions. In China, green energy generation finds its applications very well, especially in some cities with heavy environmental pollutions.

Similarly, given the broad scope of SG-related research, different researchers may focus on different topics and aspects of SG. Here, we introduce SG based on the hierarchical structure [8] as shown in Fig. 1.3, from SG infrastructure to SG applications.

- *Smart grid infrastructure.* SG infrastructure is the foundation of SG, including smart power system, information technology, and communication system:
  - *Smart power system* provides a reliable and intelligent power system which consists of power generation, power transmission, power distribution, and energy storage.
  - *Information technology* supports the advanced information metering, smart monitoring, and the corresponding information management.
  - *Communication system* builds on the advanced communication infrastructure and technologies.
- *Smart grid applications.* SG applications are further divided into fundamental applications and emerging applications:
  - *Fundamental applications* focus on the technologies of energy management, system reliability, security and privacy, featuring demand-side management for energy efficiency improvement, user utility maximization, and system protection.
  - *Emerging applications* introduce two new patterns in SG: *electric vehicle* (EV) and *microgrid* (MG), featuring energy management for large-scale support of EVs and DGs of renewable energy in MGs.
  - *Derived applications*: two examples are *smart home* and *smart city*, which are derived from SG, providing the impact of SG on human societies.

Based on this structure, we introduce SG in the remainder of this chapter. As an introduction, we focus on the fundamental concepts but avoid the complicated techniques. We also emphasize the contents relating to the following chapters, such as energy management in SG and MGs, which is the major scope of this dissertation. For other contents, we provide a brief introduction and provide readers with references for further reading.

1.2 Smart Grid Infrastructure

1.2.1 Smart Power System

The basic function of a power system is to generate power and deliver to end users through the transmission and distribution networks, in a reliable, efficient, and economic way. The traditional electricity power system shown in Fig. 1.1 mainly consists of three subsystems: generation system, transmission system, and distribution system. Electric power is produced in power plants from other forms of energy, such as fossil fuels, flowing water, and nuclear. Because of the economic and geographic factors, traditional power plants are located in places where large populations of people live. Then, the voltage of the generated power is stepped up through a transformer to a high level, reducing the power loss during transmission. Before distribution, the voltage is stepped down via a substation from transmission levels to distribution levels. Through another step down of the voltage when exiting the distribution system, electric power arrives at each house at a service level of voltage.

In SG, the main structure of power delivery system largely remains the same for several reasons. The current power system has been providing electricity services for more than 30 years in many developed countries. Our task is to make them intelligent instead of creating a new system from scratch. On the other hand, it will take several decades for the evolution from the current power system to a true SG. The major differences between the smart power system and the traditional power system are summarized in Fig. 1.4. Compared to the traditional power system, SG features with DG of renewable energy for power generation, flexible power transmission, DC-DC/AC-DC hybrid power distribution, and the new GES system. These new features make power generation and delivery in SG more flexible, reliable, secure, efficient, and sustainable.

Subsystems / Grid System	Generation	Transmission	Distribution	Storage
Traditional Power Grid	Centralized	Strict	AC-DC	Low Efficiency Small Capacity
Smart Grid	Distributed	Flexible	DC-DC &AC-DC	Grid Energy Storage

Figure 1.4: A comparison between the traditional power system and the smart power system.

Smart Power Generation

Electricity generation is a process of transforming other types of energy into electricity power, based on the theory of electromagnetic induction, discovered by Michael Faraday around the 1820s. From then, many sources of energy have been converted successfully into electricity generation including fossil fuels, hydro power, solar, wind, nuclear, etc. Currently, our power is mostly from five sources of energy: coal (40%), natural gas (23%), hydro power (17%), nuclear (11%) and oil (4%) [9]. Other sources of energy together only contribute 5% of the total power generation. According to the statistics in [10], the U.S. has similar generation percentages of coal (39%) and natural gas (27%). Meanwhile, the combustion of coal and natural gas are two major sources of the GHG emissions, which contributes to 73% and 24% respectively, in the U.S. It is recognized that by reducing the consumption of coal and natural gas, GHG emissions will be well controlled. A good way to achieve this goal is to incorporate more generation from clean and renewable energy resources. Also, fossil fuels on earth is getting depleted so that we have to find better ways to survive. It is even more urgent because of the fast increase in power demand caused by the fast growth of

the population and development of economy. Therefore, larger amounts of power generation from renewable energy resources is important in SG for sustainable development.

In SG, power generation has changed in many aspects, because of the support of two-way flows of energy and information. Compared to traditional power grids, power generation in SG is more flexible. In a traditional power grid, power is usually generated in a centralized way of a large amount; while in SG, power can be generated in a distributed way at a much smaller amount, because the renewable energy resources are normally *distributed energy resources* (DER), and power is often generated from small-scale power generators such as solar photovoltaic (PV) or small wind turbines (typically from 3 kW to 10 MW) [11]. To take advantage of the scattered DER, DG is applied as a solution in SG. Although power generation from each DG is very small, it is possible to gather the power from many DGs and manage them in a highly efficient way. The information and communication technologies applied in SG make it possible to accomplish this complicated task.

DG is normally described as the generation of electricity from small scale generators rather than from central generating plants, which can be connected to a power system in nearby places. The definitions of DG are different from many organizations such as Institute of Electrical and Electronics Engineers (IEEE), International Energy Agency (IEA), and International Council on Large Electricity Systems [12]. However, the main features of DG are similar in terms of location and generation capacity. The benefits of more DG penetrations are commonly recognized as follows.

- *Diversifying energy sources.* As DG is very flexible in gathering energy from a variety of sources, we can thus diversify our energy sources and reduce the reliance on the depleting fossil fuels. On the other hand, DG from DER will lead us to a new sustainable form of life.

- *Controlling GHG emissions.* Currently, our major power source is burning fossil fuels, which contributes to the largest part in GHG emission. More DG penetrations of

renewable energy will reduce the combustion of coals and natural gases, and thus limit GHG emission.

- *Improving power quality and reliability.* Distributed generation provides a reliable power support when some emergent events occur in the main power grid. In DG-based power grid, failures in different parts can be isolated effectively and immediately. As a result, the overall power quality in SG can be improved.

- *Increasing the flexibility of electricity market.* The electricity market can be more flexible with dynamic prices and demand response in SG of bidirectional power flows with participation of power customers.

On the other hand, DG penetrations, especially at a large scale, also bring many technical and economical issues. For example, power generation of DGs can be intermittent and fluctuating subject to the weather conditions for many DERs, such as solar and wind. It is also challenging to achieve the balance between power supply and demand [13]. This will no doubt increase the difficulties of power management and demand response, and it will be even harder given the real time control requirement. Besides, power transmitted from DG generators to the main grid requires new conversion circuits between different voltage levels, making power management much more complicated. Furthermore, economic cost needs to be considered for both utility companies and power customers [14]. More details of DG integration issues can be found in [15].

The future SG with large-scale DGs will be developed in three stages [16]: enlarging the scale of DG incorporation, building decentralized DG systems cooperating with the centralized generation, and generating most power from DG systems while limiting the generation from centralized generation. Nowadays, DG technologies have already been proposed and applied in many countries with a major installation of wind turbines and combined heat and power (CHP) plants, a system generating both heat and electricity simultaneously to

meet the basic demands temporarily for a small number of users, especially during the time isolated from the main power grid under extreme weather [17].

As the development of large scale penetration of DGs, control and power management in the highly distributed power system will be more difficult and complicated, with consideration of not only more flexible power flows, but customer demand and energy storage. These challenges lead to the concept of *virtual power plants* (VPP) [18], which aggregates power generation from many distributed generators and manages the power to meet the demand or save for future use. VPP provides a flexible way to manage distributed power generation of comparable capacity with conventional power plants [13]. However, VPP also requires a complex integration of optimization, management, and communication technologies. Most research works on VPP cover the topics on its structure and operation [19, 20]. VPP is also applied as an important power control approach for grid integration with EVs [21, 22]. In the future SG, VPP will play a more important role in power management.

As more DG units are deployed in the smart power system, power generation is being transformed from the centralized manner to the distributed manner.Although it brings about many difficult problems in several aspects, we are confident that with advanced scientific technologies and methodologies, large-scale DG penetrations can be achieved in the near future.

Power Transmission

In Thomas Edison's time, electric power was transmitted through a direct current (DC) system. It was soon replaced by the alternating current (AC) transmission system for power loss reduction during long distance transmission via high voltage level. Since then, the AC transmission system has been the major transmission manner until today. However, with the development of modern technologies, the DC transmission system has regained people's interests, especially in tomorrow's SG with a large number of DGs.

AC transmission has been a major choice over DC mainly because of its high voltage transmission for a lower power loss in long distance transmissions. However, the appearance of *high voltage direct current* (HVDC) system has changed the situation. Currently, the highest record of HVDC level in 2014 is ±800 kV DC in the XiangjiabaShanghai HVDC system in China [23]. And it is indicated in [24] that HVDC has many advantages over the high voltage AC system in flexibility, safety, and security. In some cases, HVDC is the best choice for economic consideration, such as long distance hydro-power transmission and offshore wind power transmission via submarine power lines. HVDC also has smaller footprints because of underground and submarine cables, which make HVDC a good solution for large-scale deployment of DGs in the future SG. To meet the requirements of SG in flexibility, accessibility and reliability, *flexible AC transmission systems* (FACTS) is an important technique to upgrade the current AC transmission system. In FACTS, fast DC/AC and AC/DC conversion is enabled to control power quality. The main technologies and prospects of HVDC and FACTS can be found in [24].

The transmission system in SG should also be incorporated with intelligence, in order to overcome the challenges from increasing load demand, market needs, environmental problems, and outdated low efficient components. As indicated in [25], the power transmission system in SG can be further divided as *smart control centers*, *smart transmission networks*, and *smart substations*. The future smart control centers will be capable of real-time monitoring, analysis, and control at a larger scale. In smart transmission networks, new facilities such as smart sensors will be installed and innovative technologies will be applied to achieve a high quality, reliable, and secure power transmission. Future smart substations will be highly automatic and self-healing with the support of new technologies, such as HVDC and FACTS. In sum, the entire power transmission grid in the future SG will be a digitalized system with the most advanced technologies from different areas to provide a reliable and sustainable power delivery system.

Power Distribution

The main function of the power distribution system is to deliver power efficiently and reliably to end users. In the future power distribution system, power delivery will be more flexible because more DGs will be integrated into SG. However, this will increase the complexity and difficulty for power control. On the other hand, the division of power transmission and distribution will be blurred in SG. For example, power generated from distributed renewable resources can be distributed directly to end users with DC/DC conversion. These features bring about new challenges to the power distribution system, and that is why we need a smart power distribution system.

In a smart power distribution system, the concept of power packet has attracted considerable interest recently [26, 27]. The authors in [26] present two systems for in-home power distribution, one of which is based on AC system, and the other is a DC power dispatching system. Power packets are used for DC dispatching. High frequency power switching technologies are used for power packetization. The DC-based power distribution can be a suitable paradigm for future power distribution systems, especially for in-home power distribution, because many in-home electric appliances are driven by DC power. Power control will be more efficient. An application of in-home DC distribution system is depicted in [28]. There is no doubt that DC system will be an important part of the smart power distribution system in the future SG.

Power Storage

Another major difference in the power system structure between traditional power grids and SG lies in power energy storage. Traditionally, limited primarily by cost and efficiency, energy storage cannot be widely deployed, although it has been regarded as an effective solution to many problems in the power system. In recent years, with development in technologies, especially physics, chemistry, biology, and material science, energy storage efficiency has been greatly improved and the cost of storage can be well controlled. More

and more *energy storage system* (ESS) are proposed, demonstrated, and deployed, which builds the foundation for the large-scale *grid energy storage* (GES) in the future power grid. As a new component, GES will not only change the structure of power grid, but also provide important supports for many features and applications in SG.

GES is a new feature and subsystem in SG, with power energy storage at the grid scale. It will be composed of a large number of ESS with different capacities of power storage. The ESS will use different storage mediums according to the power sources and the environment for energy storage. For example, in hot places, batteries may not be a good choice for energy storage because in these places, it costs more to keep batteries working in a suitable environment. Instead, *thermal energy storage* (TES) could be a good solution because it can store the heat from the sunshine efficiently with the help of *concentrating solar power* (CSP) [29].

Nowadays, many developed countries have invested on GES and expect for a prominent growth in the GES capacity. It is indicated in [30] that the energy storage capacity in the U.S. is expected to reach the level of 240 GW by 2030 from the current capacity of 24.6 GW [31]. Larger capacity of GES will bring about many benefits in SG:

- *Promoting renewable energy penetration.* The ESS with different storage mediums in GES increases the flexibility of energy storage, which helps to increase DG penetration. DGs of renewable energy will cover a large part of power generation in SG, however, the power supply from DGs cannot match the power demand very well for many reasons (see Section 1.2.1). With the support of GES, this problem can be solved. The extra power during off-peak times will be saved efficiently for usage in peak demand time. This also will increase the reliability of DGs.

- *Alleviating peak demand pressure and smoothing the grid load.* Due to low capacity of energy storage in current power systems, a certain amount of power is wasted everyday. On the other hand, the current power grid has to keep its generation capacity at a high level only to meet the peak power demand lasting only several hours in a day.

GES can alleviate the pressure of power demand in peak hours, reduce wasteful extra capacity, smooth the grid load by shifting power demand around, and also reduce GHG emissions.

- *Improving the performance of existing power system.* It is estimated by the U.S. Department of Energy (DOE) that over 60% transmission lines, power transformers, and circuit breakers are more than 35-years old. The energy storage provided by GES reduces the excess generation capacity and thus reduces line-congestion and line-loss in peak times, thus alleviating correspondingly the urgency of expanding the capability of the current power system facility.

- *Supporting the electrification of transportation.* EV is regarded as a good solution to reducing GHG mission. However, a high level of EV penetration will bring about a considerable charging load as well [32]. Although EV charging can be scheduled to shift the charging demand [33], it relies upon the GES capacity to achieve the demand response for a large number of EV fleets.

- *Increasing the overall grid resilience to extreme environmental conditions and emergencies.* Our current power grid is vulnerable to extreme weather conditions, such as hurricanes and tornadoes. The basic power supply during a disaster can be crucial. GES can provide this kind of power supply in areas out of power. A well-known example is the world's biggest battery of 40 MW power in Alaska, US, which is supposed to supply the power usage during blackouts in Fairbanks, Alaska, a so-called electrical island because of the extremely low temperature [34].

For a large GES capacity, technologies from many aspects are necessary. Currently, major technologies in GES include: battery energy storage system (BESS), flywheel, pumped hydro, compressed air energy storage (CAES), TES, superconducting magnetic energy storage (SMES), electrochemical capacitors (EC), etc. Flywheel and pumped hydro are two traditional deployments in U.S., in which pumped hydro provides more than 90% energy

storage capacity [31]. The research and development of battery technologies has already lasted for many years, and BESS has been developed and tested in many labs. Major batteries used for power storage are lithium-ion (Li-ion), sodium sulfur (NaS), and lead acid batteries, which have their advantages and disadvantages respectively. Please refer to [31,35] for details on batteries and other GES technologies.

1.2.2 Information Technology

In Section 1.2.1, we introduce the main structure of smart power system. In this subsection, we focus on information technologies that improve SG functions on metering, monitoring, and control throughout the entire SG. Information technology refers to the technologies used to process SG data such as data acquisition, data analysis, data optimization, data compression, data storage, etc. The improvement of information technology brings about many new features to SG, mainly on real-time metering and monitoring, automatic control, and self-healing. We first introduce information metering with smart meters, and smart monitoring with sensors and phasor measurement units. We then discuss information management in this section.

Information Metering

Information metering or smart metering is essential to information acquisition from end power users. The original concept of SG is actually based on *advanced metering infrastructure* (AMI) [36], which automatically collects information from all the metering equipments in power grids.

As part of AMI, *smart meters* are electric meters that record the power consumption information of end users and communicate with the control center. Smart meters support the two-way information flows with the control center. They not only record the power usage statistics and send to control center, but receive information from the control center as well. Power users are able to obtain the power grid status through smart meters. Another

role of the smart meter is the central controller at end users' places. Equipped with both capability of communication and computation, smart meters can communicate with the electrical appliances and the grid control center at the same time. In this way, power energy can be used more efficiently, and many customer-side services can be enabled. For example, smart meters can adjust the working schedules of electrical appliances according to the real-time price broadcast from the SG; while grid controller can shift the peak demand by adjusting power prices.

Smart Monitoring

Smart monitoring is another important function of SG that aims to provide real-time monitoring and measurement of the grid conditions covering everywhere in the grid. Two widely deployed monitoring and measuring schemes are *sensors* and *Phasor Measurement Unit* (PMU).

Sensors are devices that can transform other physical signals into electric signal. In the power grid, sensors can be used to detect failures, collapse, and malfunctions of electric components, and can be used to monitor the working environments such as temperature and humidity; alarms will be triggered under extreme conditions. A sensor network are composed of a group of sensors that can cover a large area. Wireless sensor network (WSN) in particular, is an organization of a large number of sensors that communicate with wireless technology. Given the characteristics of low cost and power saving, WSN has become a feasible deployment of real-time monitoring and precise sensing in a large-scale grid system [37], and a good solution to limited awareness in wide area power grids as described in [3].

PMUs are specially designed devices to measure the magnitude and phase angle of the electrical wave, which are represented by the phasor in mathematical form. The power waves information from different PMUs are synchronized precisely with the global positioning system (GPS). PMUs can be set to read in a short interval, from seconds to minutes depending

on the requirement of SG. Grid system controllers decide the substations for PMU installments so that each reading is expected to have a correct pattern. PMU readings from the selected places in the grid are then used to evaluate and analyze the health of the power grid. Further research in [38] indicates that PMUs can be used better by a delicately designed system. In short, large-scale deployment of PMUs will improve the reliability and avoid catastrophic failures in SG.

Information Management

An effective information management scheme takes time, efforts, and resources to develop and improve. However, a bad one will cost much more with low grid performance. A primary cause of the Northeast blackout of 2003 thath happened in the United States and Canada was a software bug in the alarm system, which led to an unawareness of a transmission line failure [1]. This exposes a big issue of ineffective information management and outdated control software. Fortunately, in the era of information, we could arm our power grid with advanced technologies in information management. On the other hand, large-scale deployment of smart meters, wireless sensors, and PMUs will generate a huge volume of data every several minutes in SG. Suppose the data can be collected successfully in SG, it is meaningless and even wasteful if the data cannot be processed on purpose and on time. This is one of the challenges of information management in SG. Information management in SG can be further divided into two aspects: *information modeling* and *information processing*.

Information data modeling is the basic of information management. It aims to provide a regulated, compatible, displayable, and robust data denotation for every application in SG. It should take into account many factors in SG, such as transmission time, processing complexity, and privacy protection. For example, a control center may receive many data packets at the same time, some of which are metering data and others sensing data. The computer at the control center should be able to analyze immediately the types of the data, and the meaning of each packet. The combination of each data packet needs to be regulated

for different uses, and the communication protocols need to be designed or modified for the best usage in the SG environment. Meanwhile, the data should be compatible in case of the need for more functions and applications in the future. And the computational complexity and privacy protection need to be well balanced for practical considerations. A well-defined structure may be used to simplify the modeling. A typical guide has been provided by the IEEE in [39].

After modeling, the information data need to be processed properly, and the corresponding operations need to be taken in a timely and precisely manner. Although the current microprocessors have amazing computational capability, it is still a challenge to process the vast data. Therefore, information analysis and optimization techniques are necessary. Information analysis extracts useful data from the raw data, and information optimization compresses and stores the data in a most effective way. For example, the sensors may be set to collect and transmit metering data every 5-min. Normally, the power grid status may remain the same in short periods. So data packets can be processed later based on the results of real-time information analysis. However, when some failures happen in SG, analyzing results may request sensors to generate reports every 15-s. Also, in SG, power lines are interconnected. According to the power system dynamics, the states on some lines can be inferred from the states of other lines, which saves the resources on monitoring devices, but requires more computations. The techniques of information compression and storage are also significant for timely optimization and resources usage. Big data techniques are a good solution to handling the large amount of data in SG. The authors in [40] list SG data management as an application of big data technology and discuss how big data can be applied in SG.

1.2.3 Communication System

The communication system supports two-way flows of information in SG, based on which information metering, monitoring, and management are applicable and energy management can be achieved. A variety of communication technologies can be applied in the SG environment to provide a fast, reliable, secure, and self-healing communication system covering the entire power grid. Fig. 1.5 shows an example of a communication network in SG, where both wireless and wireline communications find its applications. Different communication technologies are applied in different scenarios. In this subsection, we provide a brief introduction to the communication technologies used in SG, which includes *wireless communication* and *wireline communication*.

Wireless Communication

Wireless communication technologies have been developed and widely used for many years. Technologies such as Wi-Fi, 3G, and 4G LTE are used almost everyday in our life. Because of its development, wireless communication technologies can be used in plenty of applications in SG. Here, we introduce important wireless communication technologies that are strong candidates for SG, including wireless mesh networks (WMN), cellular communication systems, satellite communications, Wi-Fi, and Zigbee. For other communication technologies such as cognitive radio, microwave, free space optical communications, please refer to [41–44].

- *Wireless mesh network.* As a wireless network composed of nodes with a mesh topology, WMN has been regarded as the next-generation wireless networking paradigm [41]. WMN provides robust and reliable communications with self-organizing networking structure. Its feature of automatic connectivity is important for many applications in SG [42].

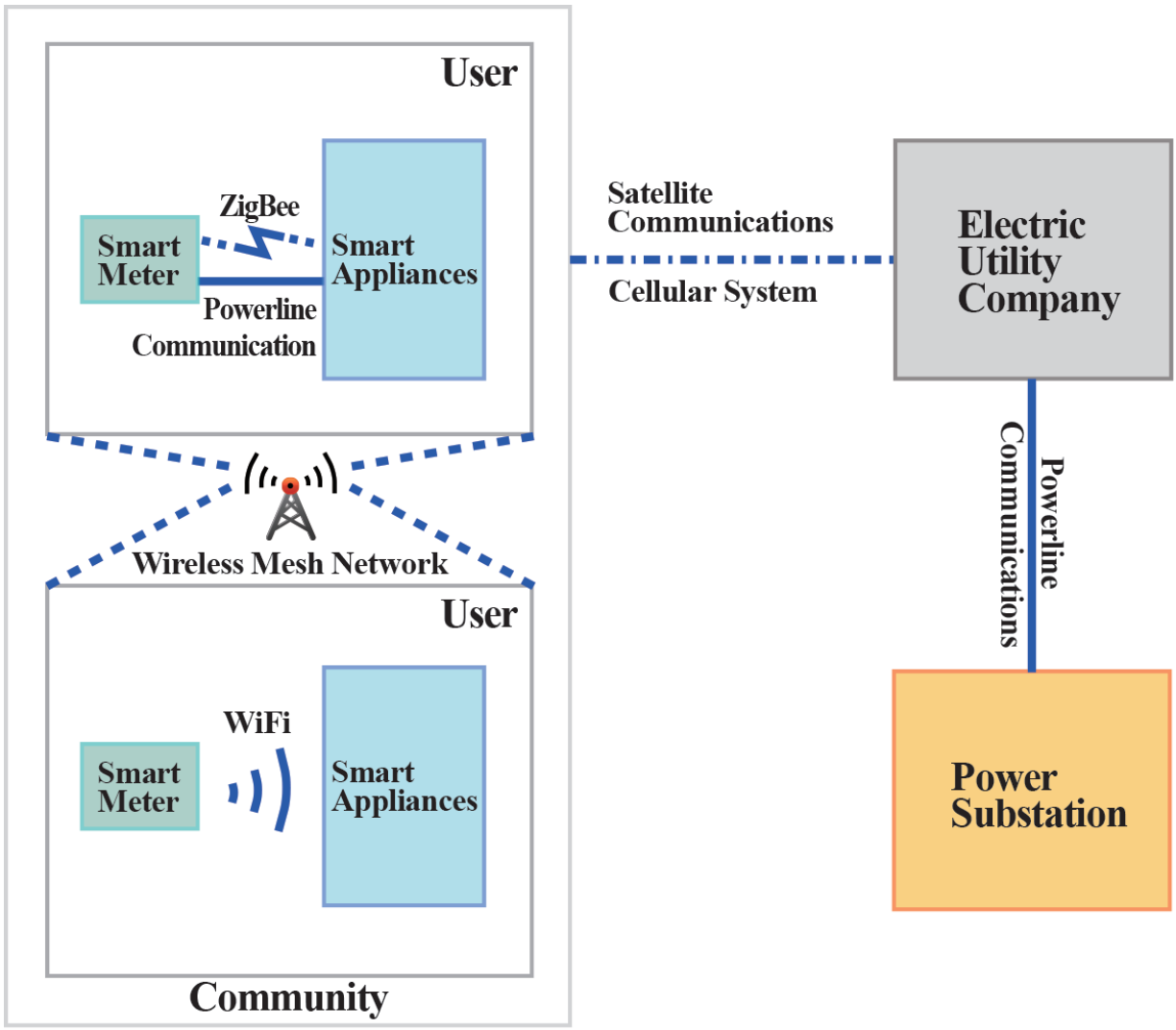

Figure 1.5: A typical communication network in SG.

- *Cellular communication system.* The wide application of cellular communication systems has proved its effectiveness and efficiency. As a mature system with many base stations deployed, it is convenient to be applied in SG covering a large area [43].

- *Zigbee and Wi-Fi.* Although Zigbee and Wi-Fi are different technologies with different standards and bandwidths, they share some common characteristics, especially in home networks. Featuring energy efficiency, long battery life, and high security, Zigbee is suitable for smart meters and have been developed to meet the needs of AMI [45], while Wi-Fi with higher rate and longer range, can be used for controlling electric

appliances in large apartments or buildings. And the ultra-low power Wi-Fi chips [45] of energy efficient features extend the lifetime toward conventional Wi-Fi. Zigbee is also a standard widely used in WSN [43], which has been widely deployed in environment monitoring and thus is a strong candidate for SG monitoring. For better services and further applications in SG, special routing infrastructure, security protection and quality-of-service are required [46].

- *Satellite communication.* Satellite communication offers a good option when other wireless signals are unavailable. And it provides GPS services for a global range, which are also important for PMUs. However, it is stated in [42] that, satellite communications can cause longer delays, and its channels and signals are easily affected by weather conditions. Besides, the high initial investment is another disadvantage. These drawbacks limit satellite communication applications in SG.

Wireline Communication

Two important wireline communication technologies are potential candidates for SG: *optical fiber communications* and *power line communications* (PLC).

- *Optical fiber communication.* Optical fiber communications have many advantages such as super high data rate, long-distance and electromagnetic interference immunity [42]. The current optical fiber infrastructure can be a good support for high speed communications in the future SG.

- *Power line communications.* PLC is a technology for transmission of data and electricity simultaneously through the power line [47]. PLC covers a large area where power line reaches, and is thus cost-effective because the power lines are already installed in most-part of the world. However, many drawbacks and technical problems have to be

identified, such as low capacity, large noise involvement, security, and lack of regulation [42]. We have to solve these problems before PLC can be widely applied in the future SG. The case of possible applications of PLC in SG can be found in [48].

1.3 Smart Grid Applications

In this section, we introduce the diverse applications in SG. If SG is viewed as an advanced computer, SG infrastructure is the hardware and SG applications are the software programs. More specifically, fundamental applications compose the operating system, based on which other applications provide the users with a variety of functions like advanced software programs. In the following, we begin with fundamental applications, and then discuss two important emerging applications, i.e., EV and microgrids, which are then followed by two interesting derived applications, i.e., smart home and smart city.

1.3.1 Fundamental Applications

As the foundation of SG applications, fundamental applications perform like the operating system in a computer. A well-developed and maintained operating system will effectively support more upper level applications. Therefore, fundamental applications are important in SG. Here, we category fundamental applications into three major classes: *energy management*, *system reliability* and *security and privacy*.

Energy Management

The feature of bidirectional flows of information and energy provides the basis for advanced energy management in SG. As stated in Section 1.2.2, information management is important for data acquisition and processing, which can be further used for energy management. Different from information management, energy management produces the results on many aspects that can be seen directly. For example, an effective demand-side management will smooth the grid load, which can be observed directly; management on environment

protection will result in a reduction of GHG emission. Therefore, smart energy management will fully reflect the intelligence of SG infrastructure. In SG, energy management is mainly focused on two aspects: *energy efficiency* and *environment protection*.

Energy Efficiency Even before the proposal of SG, energy efficiency had been regarded as an important issue. Power loss happens all the way from generation to consumption. Super high voltage power transmission is a practical method used to reduce power loss and thus increase energy efficiency. Considering the large amount of daily power consumption, even a 1% increase in energy efficiency will save considerably on energy resources. With optimized energy management, energy efficiency can be greatly improved. The research in energy efficiency improvement is mainly focused on *demand-side management* and *energy loss minimization*.

- *Demand-side management.* Demand-side management (DSM), also called demand response (DR), refers to the activities aiming to match the demand to supply and reduce the peak load or smooth the load profile. DSM covers a large portion in energy management and has attracted considerable research efforts. It is indicated in a recent survey on DR that the number of optimization models on DR programs was over 500 in 2013 [49].

  In so many different DR schemes, matching the demand to supply is the core idea. U.S. DOE defines DR as the activities to provide time-varying energy prices for end users according to the changing production costs and to reduce the peak demand by offering some incentives [50]. It motivates various works on dynamic pricing schemes such as real-time pricing (RTP), critical-peak pricing (CPP), time-of-use pricing (TOU), peak load pricing (PLP), and peak day rebates (PDR), etc. Applying the pricing schemes, smart meters at the power users can schedule the time for different power appliances. For example, based on the prediction of coming peak times, such as 7:00 a.m. in the morning and 6:00 p.m. in the evening, the control center in SG will update power

prices based on the predicted peak level, and smart meters will adjust accordingly the usage of the power appliances, such as to pause the washing machine for a few minutes.

By deploying the scheme in millions of houses, demand peak can be reduced and backup generations can be avoided, which will save the cost for both utility company and power customers, and increase the system reliability. User utility and energy provisioning cost are also considered in many existing works. We will provide a detailed discussion and present the models of user utility and energy provision cost in Chapters 2 and 3.

- *Energy loss minimization.* Energy loss minimization aims to reduce the power loss in the entire SG. In a power system with large penetrations of DGs, it is necessary to manage the power flows optimally [51]. For example, in some large power consuming industry plants, electricity is provided by designated sources, which is easier for operation but may not be the optimal solution in terms of power loss. More flexible power flows enabled in SG make it possible to distribute the optimal energy sources for energy consumers in terms of both cost and power loss, but requires more complicated energy management.

Environment Protection Environment protection is another important topic, which has been widely emphasized in SG. By employing specially designed management schemes, GHG emission can be limited and controlled. Incorporating more DGs of renewable energy is considered as a solution to reducing GHG emission. But the energy management targeting generation cost reduction or user utility maximization cannot directly guarantee the GHG emission reduction. This is mainly because renewable generation does not always have the lowest cost. Therefore, environment protection should be considered as a factor in energy management schemes [52, 53]. Besides, in the future SG, a large amount of batteries may be applied for energy storage and EVs. However, most of the current battery technologies are not fully environment friendly [35]. It can be a paradox if so many batteries in the future

EVs become hazardous to our environment. Therefore, new materials for battery are still in great need, and energy management on EVs are also necessary to extend the battery life.

System Reliability

System reliability has always been an important topic in power grid and also a key part in SG design. It is the ability for power grids to operate stably and reliably in each subsystem including generation, transmission, distribution, and storage. And it is commonly regarded as an important application in SG.

Traditionally, system reliability is interrupted by electric component failures, animals, severe weather, falling trees, and human factors [54]. By employing advanced technologies, many of these causes can be well controlled. The sensors distributed almost everywhere in SG can report severe conditions in a timely fashion [37]. Extreme weather conditions can be forecast several days ahead for preparation and the current material technologies improve the durability of power equipment under severe weather conditions. Human factors should be carefully considered in SG. The 2003 blackout failed to be stopped in time partly because of the slow reactions of engineers who were responsible for system control in emergency [1]. And it should be noted that even in a highly intelligent power system, human interference cannot be completely avoided.

Besides, some new interruption causes in SG need to be addressed carefully as well, such as the intermittence of DGs, fast conversions between AC/DC and DC/AC in a hybrid distribution system, the connection and disconnection of an MG (see Secttion 1.3.2). The work in [55] shows that a specially designed architecture can guarantee reliability in SG with DG penetrations. Interested readers are referred to [55] for more discussions of system reliability in SG.

Security and Privacy

Security and privacy issues are inevitable in almost all information communication systems. Thanks to the vast areas of information communications in SG, possible security and privacy attacks can be predicted in many aspects. Protections on security and privacy are especially important during the beginning of SG construction. User privacy, grid market, and control information are all sensitive areas and vulnerable to cyber attacks [3, 56]. The research of security and privacy is mainly focused on *smart metering* and *information monitoring*.

Security and Privacy in Smart Metering Smart meters, as a relative new type of device in SG, are ideal and attractive targets for malicious attackers [57]. By manipulating smart meters, the readings of power consumption can be tampered to gain economical benefits. This can happen for both sides. Opportunistic power consumers can reduce their power bills, while illegal companies may raise the power bills of their power customers.

On the other hand, the private information on power consumption of power users can expose some further information on their personal lives. With a long peirod monitoring of smart meters, hackers can thus analyze the habits ad daily routines of the power users, which may be used for illegal activities such as burglary. If not defended in advance, these cyber attacks are very difficult to track. Furthermore, the concerns on privacy may prevent some portions of people from installing smart meters, which will impede the development of SG. Fortunately, a great number of research works are focused on security and privacy for smart meters [58–60].

Security and Privacy in Information Monitoring It has been introduced in Section 1.2.2 that wireless sensors and PMUs will be widely deployed for real-time monitoring of the SG. Similar to smart meters, sensors and PMUs are potential targets of malicious attacks. False data injection attacks against measuring and monitoring data are stated as the

major form of cyber attacks [61]. These attacks are designed to attack SG information monitoring, it is thus important to defend these attacks to keep further information management working properly. Several approaches have been proposed in the literature [61–63].

1.3.2 Emerging Applications

Based on the basic infrastructure and fundamental applications, functional applications are made possible in SG. Recently, two emerging applications have attracted considerable interest: *electric vehicles* and *microgrids*. They are considered by many researchers as important components in the future SG, with their advantages in renewable energy generation incorporation and GHG emission control. Meanwhile, they are currently under development and thus have many technical and practical problems for large-scale applications. In this subsection, we discuss the applications and related issues on EVs and MGs.

Electric Vehicle

Driven by the environmental incentives and development in electric battery technologies, EVs are now available in the market and are gaining popularity. In the vehicle market, two types of EVs are now available: *Plug-in Hybrid Electric Vehicles* (PHEV) and *Plug-in Electric Vehicles* (PEV). PHEV can be driven by both fuel and battery, while PEV can only use power from its electric battery. In this dissertation, we use EV to represent both of them if not otherwise specified.

EVs are supported by the government in many countries [64] , especially in the U.S. It is projected by the U.S. DOE that the number of EVs will reach one million by the end of 2015. The Electric Power Research Institute (EPRI) projects that by 2050 EVs will comprise 62% of the entire U.S. vehicles under a moderate penetration scenario [33]. Large deployment of EVs will reduce a large portion of the GHG emission from traditional vehicles. However, it brings about many technical problems as well, such as charging infrastructure, extra charging load, and communication requirements. Most of these problems are related

to two basic concepts, *grid-to-vehicle* (G2V) and *vehicle-to-grid* (V2G). The former includes the impact and control of charging activities of EVs, while the latter denotes the effects and management of power injections from EVs back to the grid.

Grid-to-Vehicle EV batteries are designed to plug in for charging immediately or after a fixed start up. This means that they are common loads in the power grid. Although being convenient, EVs can generate a considerable and dynamic charging load to the grid. Especially when a large number of EVs are charging at the same time. For example, EVs for commutation are usually charged when the owners arrive at home after work, which is around 18:00 p.m. It is even worse if this is a peak load period for other power users. It has been estimated that under a 30% level of EV penetration in the U.S., the total charging load of EVs can reach 18% of the U.S. summer peak [32]. This will be a big impact and threat to the power grid. Therefore, a large number of emerging EVs cannot be deployed without optimized power management and scheduling of charging.

Fortunately, within SG, the charging of EVs can be controlled with specially designed schemes. Smart meters at the power users will play an important role in EV charging at home. Just like other electric appliances, EVs can communicate with smart meters. Also, EVs can be classified into different levels of urgency according to the scheduled uses with a basic power for occasional drivings. And thus the smart meter can schedule the charging of EVs according to the grid information from the grid control center. In this way, the peak load can be reduced by shifting a portion of demand to off-peak time. On the other hand, in public charging stations, a specific controller is required to coordinate the charging of multiple EVs. The controller needs to consider both customer satisfaction and grid stability. The optimal charging and coordinated charging schemes can be found in [32,65,66]. Furthermore, EV can be a good match of renewable energy, especially in DC-driven smart power system, because many DGs produce DC power directly [33].

Vehicle-to-Grid In SG, a large number of EVs also consist of a new means of energy storage and supply, when EVs are enabled to inject power back to the grid. V2G endows EVs with several new roles in SG.

- *Mobile energy storage system.* EVs can store electric power in batteries, which can serve as a small supply for the grid during peak periods. Incentives can be offered in the V2G market to encourage EV owners to charge in off-peak hours and discharge in peak hours. However, new management schemes on DR for EVs are needed to schedule the charging/discharging of EVs, and the V2G market also requires strict regulations. The mobility of EVs brings about new challenges to V2G management [67]. With distributed energy storage, EVs can also be used to provide ancillary services such as spinning reserves and frequency regulation [33].

- *Renewable energy storage system.* EVs driven directly by the renewable energy, such as solar vehicles, also serve as renewable energy storage in the future SG [68]. Also, some EV charging stations can be designed to use power from renewable energy. In this way, the intermittence of renewable energy generation can be mitigated with the storage from EVs.

- *Backup energy supply system.* With energy storage in batteries, EVs can serve as a backup and temporary supply for houses or communities during emergencies caused by natural disasters. Furthermore, the standards of batteries for EV uses are very strict. Therefore, some retired batteries from EVs will be still in good working conditions, which can be reused for ESS after some basic maintenance.

Microgrids

Microgrid is another emerging paradigm in SG, which is a small power grid composed of localized medium or low level power generation, energy storage, and loads. Because of flexible DGs, MG is considered as one of the most important feature application in SG. In

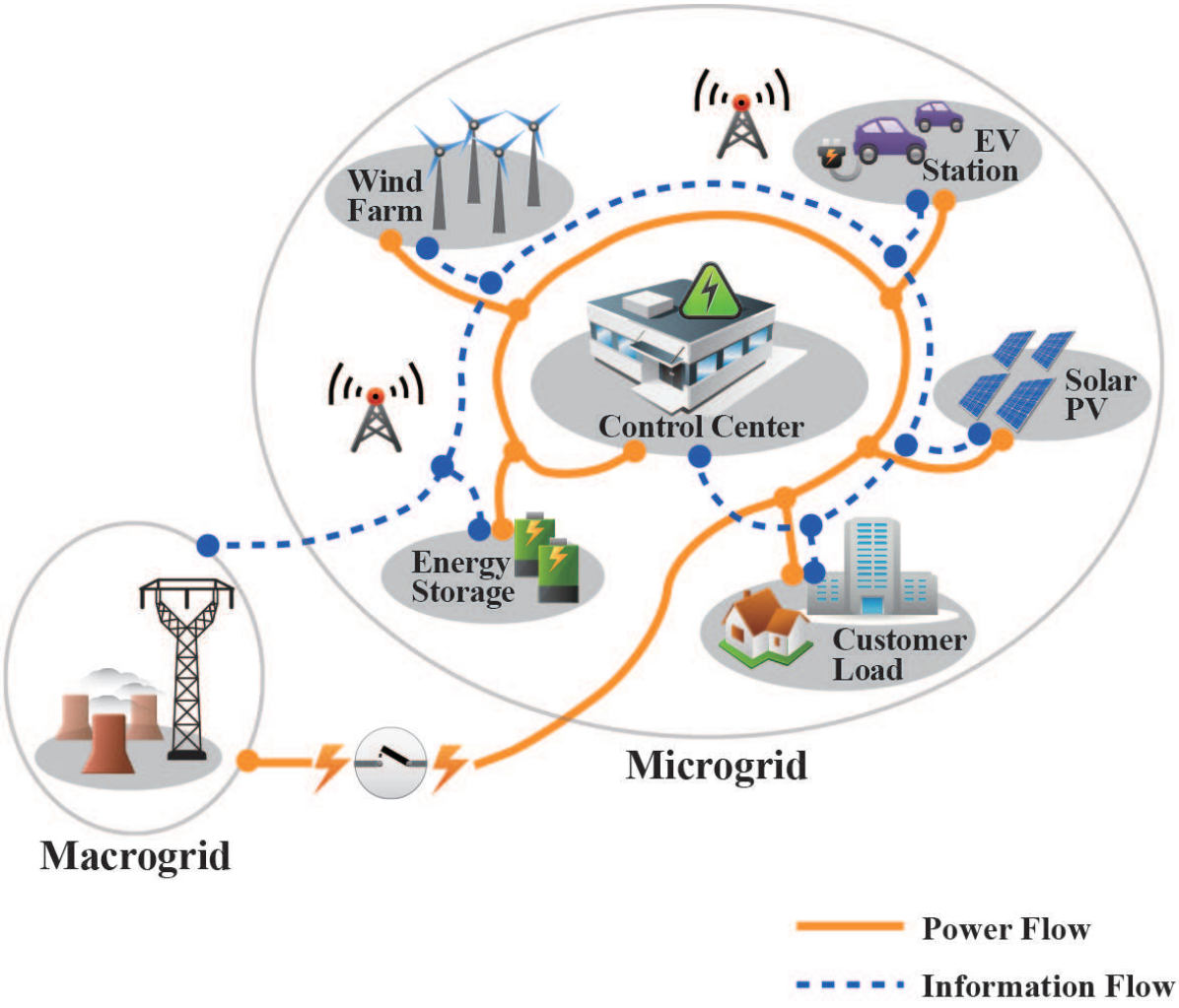

Figure 1.6: An example of a microgrid.

the *connected* mode, the MG is connected to the macrogrid, which is a main power grid with a large amount of centralized generation and loads; the connection is through the point of common coupling (PCC), which can also be disconnected for an *islanded* operation, when the MG operates as a small but independent power system, supporting the local load with its own local power generation. A typical example of an MG is shown in Fig. 1.6. MG can be viewed as an integral smart power system with more flexibility and higher intelligence, such as multiple DERs, two-way power and information flows, energy storage, etc. New MGs are expected to be integrated into SG through plug-and-play mode, which requires dedicated designs of the connection interfaces to match different levels of voltages [8].

When the concept was proposed initially in [69], the utilization of DER was emphasized in MG. Much renewable energy is usually generated as low level DC power, and MG is supposed to integrate them from multiple sources of DER. In the connected mode, the macrogrid absorbs the integrated power from MGs when renewable power generation is abundant. The macrogrid injects the power to MG when the generation from DER is not sufficient. On the other hand, the islanded mode provides reliability and flexibility to both macrogrid and MG. When emergencies or blackouts happen in the macrogrid, MG can be disconnected autonomously for safety and reliability. A type of MGs with CHP system are constructed mainly for natural disasters or extreme weathers. After Hurricane Sandy in 2012, CHP-driven MG projects has been started in several islands in the northeastern parts of the U.S. In this way, the organization of SG becomes much more flexible and energy efficiency can be further improved. Therefore, it is not surprising that MG is considered as the most important application and cornerstone in SG. Many interesting and featured MG projects and demonstrations are introduced in [70].

Energy Management and Control of MGs In engineering, a beautiful idea often incurs more difficulties. This rule works for most parts in SG, while MG is not an exception. But as stated, the core feature of two-way flows of energy and information in SG make it possible for complex power control and management, which is essential for MGs as well [71]. Because of the several features, power management in MGs is very complicated, and is different in the connected mode and islanded mode.

- *Power management for connected MG.* In the connected mode, the major role of MG is an integrator of generations from multiple DERs. Thus, the power from distributed energy sources need to be optimally managed through the MG control center (MGCC). The intermittence of renewable energy generation needs to be considered, and thus, the scheduling of charging and discharging for energy storage devices is also required. The interconnection with macrogrid requires the management for both sides. When the generation in MG is low, power flows from the macrogrid to MG, making MG

as a load in the macrogrid. Reversely, MG is an extra supply. Therefore, if the energy distribution is controlled wisely, the MG can help shape the peak load profile in the macrogrid. Other factors such as power customer's satisfaction and EV charging will increase the difficulty of power management for connected MGs. Popular control schemes and algorithms for energy management in connected MGs are convex optimization, nonlinear programming, stochastic optimization, machine learning, and game theory [72–85]. We also have a brief review of these techniques in Chapter 2.

- *Power management for islanded MG.* The power management for MG in the islanded mode is very different from that of the connected mode. In the islanded operation mode, MG operates as an independent/isolated power system, supplying its own loads with its own generation. It is stated in [86] that frequency control and real power injections caused by distributed DER generators require different management schemes. The energy storage, including the batteries in EVs, becomes essential to balance the intermittence of renewable energy generation. However, in the islanded mode for emergency, some constraints can be relaxed, such as customer satisfaction.

1.3.3 Derived Applications

Different from the fundamental and emerging applications that mainly focus on SG, derived applications are those based on or driven by SG. They bring us to a new smart era. We briefly introduce *smart home* and *smart city* in the following.

Smart Home

Customer participation is an important feature in SG. The enabling component is smart meter, which is also the bridge between *smart home* (SH) and SG. Smart meters no longer perform only as data collector for utility companies. They play quite different roles for both customers and the grid. The importance of smart meters for SG has been emphasized in Section 1.2.2. On the customer side, smart meter serves as a controller in SH. In the future

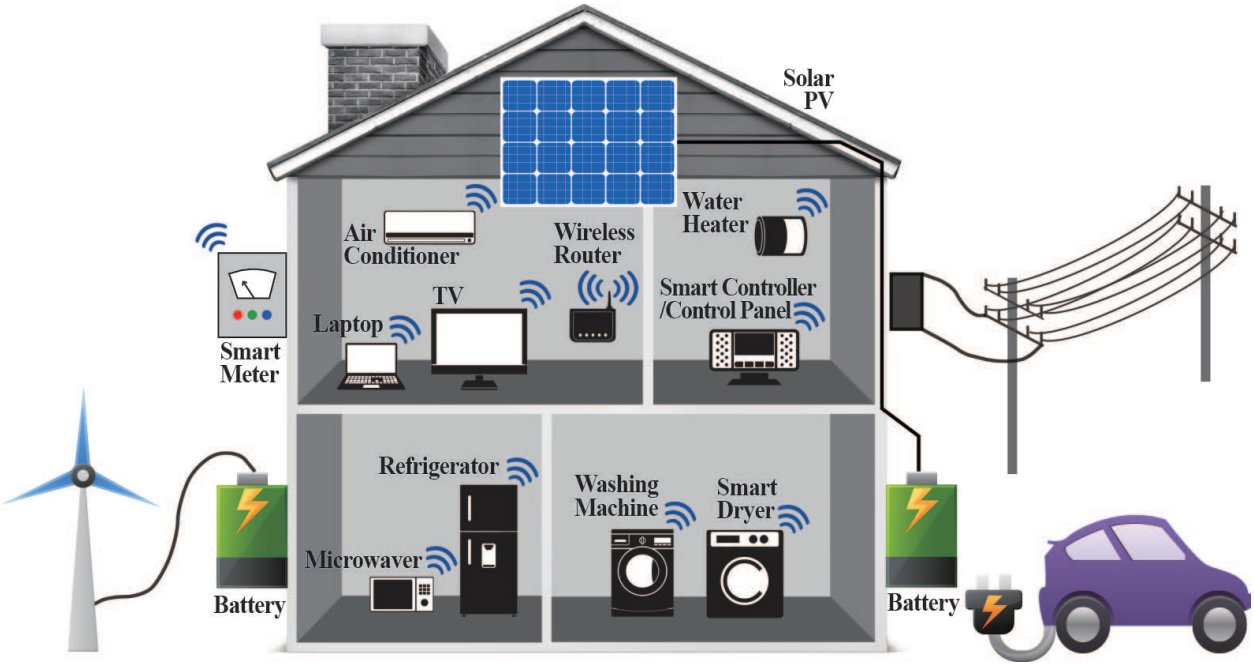

Figure 1.7: A vision of the future smart home.

SH, the appliances are equipped with communication capabilities, and are controlled by the smart meter through the in-home networking system. New smart appliances are deployed with plug-and-play scheme with specific interfaces. All the information can be displayed through a controller on a displayable control panel. The SH control provides users many optional functions according to users' preferences, such as energy saving, money saving, and low carbon.

A vision of future SH is depicted in Fig. 1.7. With small renewable generations and energy storage equipment, future SH operates like a small power system or a small connected MG. Also, the hybrid DC/AC distribution system will be realized and enhanced in the future SH to use both DC power and AC power [28]. In short, the future SH will be a highly integrated system featuring high automation, customer preferences, low carbon and energy efficiency, which will bring us much convenience, health, relax, and sustainability. An interesting SH solution proposed by ZTE featuring high security can be found in [87].

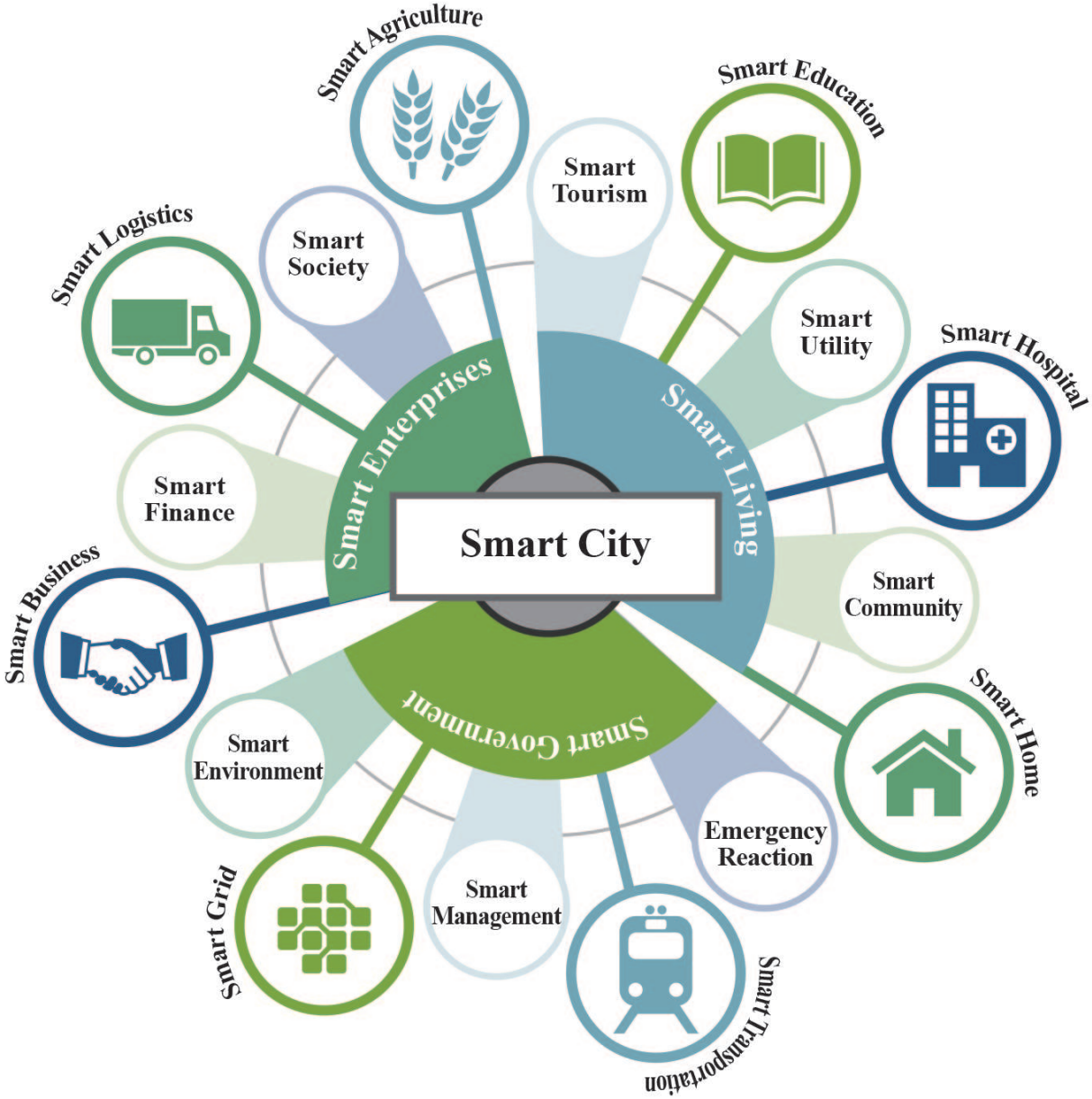

Figure 1.8: The coverage of smart city.

Smart City

With most advanced technologies and innovations, a city is considered as a representative of the civilization in all eras. Currently, we are experiencing the revolution of information technology, and *smart city* (SC) will be a product of that. Different from SH, SC includes more elements and components, which can be categorized into *smart government*, *smart*

enterprises, and *smart living*, as shown in Fig. 1.8. Similar to SG, information management and security protection are highly important problems. The realization of SC requires the support and cooperation of many aspects, from the government to each home. Thus this will be a long process and will be penetrated into the city gradually. Although the idea of SC has just been proposed, we are confident that our cities are approaching the ultimate SC in the future.

1.4 Overview of the Dissertation

In this dissertation, we focus on energy management in SG environment. As stated in Section 1.3.1, energy management in SG is very important in the SG. Therefore, we study the new characteristics of SG environment, and investigate the optimal power distribution schemes in Chapters 2 and 3. We also perform the research on energy management in MGs, especially the cooperative MGs in Chapter 4. Besides, we explore the forecasting on the power generation from renewable energy in Chapter 5.

In Chapter 2, we investigate an online algorithm for electricity energy distribution in a smart grid environment. We first present a formulation that captures the key design factors such as user's utility and cost, grid load smoothing, dynamic pricing, and energy provisioning cost. The problem is shown to be convex and can be solved with an offline algorithm if future user and grid related information are known a priori. We then develop an online algorithm that only requires past and present information about users and the grid, and prove that the online solution is asymptotically optimal. The proposed energy distribution framework and the online algorithm are quite general, suitable for a wide range of utility, cost and pricing functions. It is evaluated with trace-driven simulations and shown to outperform a benchmark scheme.

We propose a distributed online algorithm for electricity distribution in Chapter 3. We first present a formulation that captures the key design factors such as user's utility, grid load smoothing, and energy provisioning cost. The problem is shown to be convex and can be

solved with a centralized online algorithm that only requires present information about users and the grid in our prior work. In this chapter, we develop a distributed online algorithm that decomposes and solves the online problem in a distributed manner, and prove that the distributed online solution is asymptotically optimal. The proposed algorithm is also practical and mitigates the user privacy issue by not sharing user utility functions. It is evaluated with trace-driven simulations and shown to outperform a benchmark scheme.

In Chapter 4, we investigate a hierarchical power scheduling approach to optimally manage power trading, storage and distribution in a smart power grid with a Macrogrid and cooperative MGs. We first formulate the problem as a convex optimization problem and then decompose it into a two-tier formulation: the first-tier problem jointly considers user utility, transmission cost, and grid load variance, while the second-tier problem minimizes the power generation and transmission cost and exploits distributed storage in the MGs. We develop an effective online algorithm to solve the first-tier problem and prove its asymptotic optimality, as well as a distributed optimal algorithm for solving the second-tier problem. The proposed hierarchical power scheduling algorithms are evaluated with trace-driven simulations and are shown to outperform several existing schemes with considerable gains.

Energy crisis and environmental problems are forcing us to incorporate more renewable energy in the new Smart Grid, which also provides better power management. Forecast on renewable power generation, from sources such as solar and wind, is crucial for better energy management. However, the current forecast methods lack a comprehensive understanding of the natural processes, and are thus limited in precise prediction. In Chapter 5, we introduce the simultaneous inference to analyze the solar generation and weather data. We first introduce a local linear model for nonlinear time series, and present the construction of the simultaneous confidence bands of the time-varying coefficients, which provide more information on the dynamic properties of the model. We then apply the simultaneous inference for solar generation analysis using a real trace of weather data.

We conclude the dissertation and present the future work in Chapter 6.

Chapter 2

Centralized Online Algorithm for Optimal Energy Distribution in the Smart Grid

2.1 Introduction

A smart grid is an electrical grid that is enhanced with communications and networking, computing, and signal processing technologies [44]. Unlike the traditional power grid that is strictly hierarchical, the smart grid is characterized by the two-way flows of electricity and real-time information, which offers tremendous benefits and flexibility to both users and energy providers. With full-duplex information flows, configuration of the grid devices can be customized for timely response to the grid status. For example, energy storage systems can cooperate with distributed renewable energy resources (DRERs) to balance the supply and demand, and users can adapt their demand for energy according to the market price fluctuations [81].

The two-way energy and information flows, along with the smart devices, also bring new perspectives to energy management and demand response in the smart grid. Demand side management is one of the most important problems in smart grid research, which aims to match electricity demand to supply for enhanced energy efficiency and demand profile while considering user utility, cost and price [44]. Researchers have been focusing on peak shift or peak reduction for reducing the grid deployment and operational cost [88,89], as well as on reducing user or energy provider's cost [90, 91]. In particular, some prior works aim to achieve a single objective, such as to improve the users' utility or reduce the cost of the energy provider [92], while others jointly consider both the user and energy provider costs, to increase the users' utility as much as possible while keeping the energy provider's cost at a relatively lower level [93]. Given the wide range of smart grid models and the challenge in characterizing the electricity demand and supply processes and the utility, cost, pricing

functions, a general model that can accommodate various application scenarios would be highly desirable. Furthermore, it is important to jointly consider the utilities and costs of the key components of the system to achieve optimized performance for the overall smart grid system.

In this chapter, we consider real-time energy distribution in a smart grid system. As shown in Fig. 2.1, the distribution control center (DCC) collects real-time information from the three key components, i.e., the users, the grid, and the energy provider, makes decisions on, e.g., electricity distribution, and then sends the decisions back to the key components to control their operations. The smart meters at the user side will be responsible for the information exchange with the DCC and for enforcing the electricity schedule received from the DCC. The information flows will be carried through a communications network infrastructure, such as a wireless network or a powerline communication system [44].

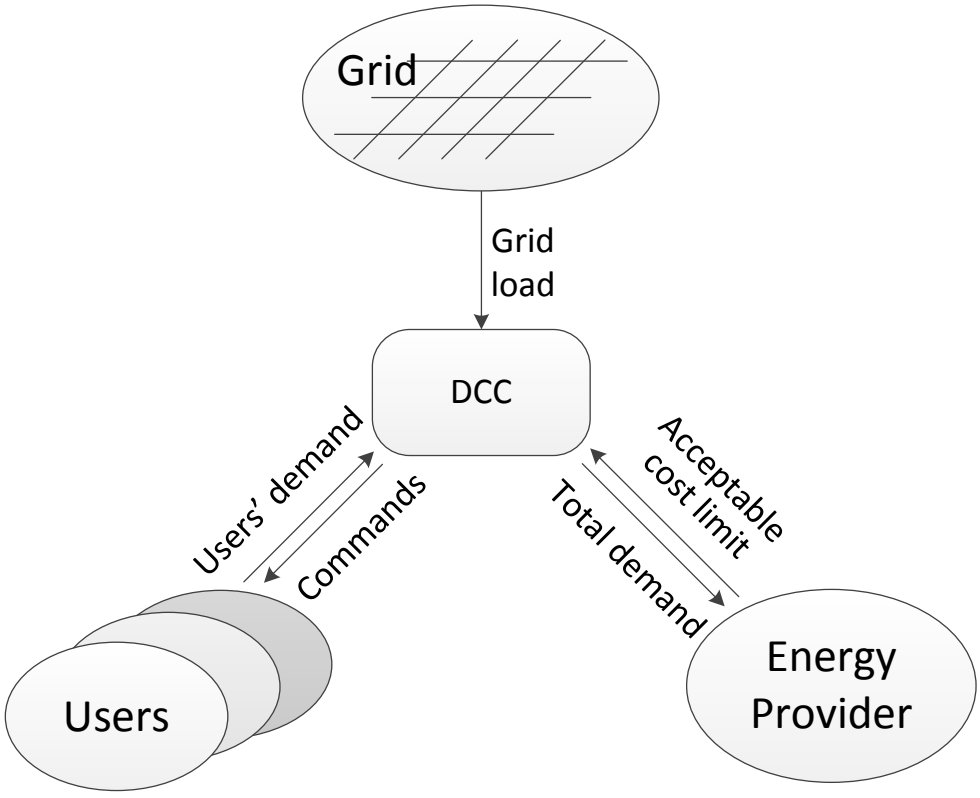

Figure 2.1: Illustration of the key elements and interactions in the smart grid.

For optimizing the performance of such a complex network system, the utilities and costs of the three key components, i.e., the users, the grid, and the energy provider, should be jointly considered. In this chapter, we take a holistic approach, to incorporate the key design factors including user's utility and cost, grid load smoothing, dynamic pricing, and energy provisioning cost in a problem formulation. To solve the real-time energy distribution problem, we first present an offline algorithm that can produce optimal solutions but assuming that the future user and grid information are known in advance. Based on the offline algorithm, we then develop an online algorithm that does not require any future information. As the name suggests, an online algorithm operates in an online setting, where the complete input is not known a priori [94]. It is very useful for solving problems with uncertainties [95]. We find the online algorithm particularly suitable in addressing the lack of accurate mathematical models and the lack of future information for electricity demand and supply in this problem. We also prove that the online algorithm converges to the optimal offline algorithm almost surely.

The proposed framework is quite general. It does not require any specific models for the electricity demand and supply processes, and only have some mild assumptions on the utility, cost, and price functions (e.g., convex and differentiable). The proposed algorithm can thus be applied to many different scenarios. The online algorithm also does not require any future information, making it easy to be implemented in a real smart grid system. It is also asymptotically optimal, a highly desirable property. Since there is no need for communications among the users, their privacy can be easily protected. The proposed algorithm is evaluated with trace-driven simulation using energy consumption traces recorded in the field. It outperforms a benchmark scheme that assumes global information.

The remainder of this chapter is organized as follows. We present the system model and problem formulation in Section 2.2. The offline algorithm is introduced in Section 2.3, and the online algorithm is developed and analyzed in Section 2.4. The communications protocol for supporting the online algorithm is discussed in Section 2.5. A practical online algorithm

is given in Section 2.6. We present the simulation studies in Section 2.7 and review related work in Section 2.8. Section 2.9 concludes this chapter with a discussion future work.

2.2 Problem Statement

2.2.1 System Model

Network Structure

We consider a power distribution system in a smart grid environment where one energy provider supports the power usage of N users. The users could be residential, commercial and industrial energy consumers. Each user has a *smart meter* that manages the schedule of electrical devices [44]. We envisage that the smart meters could be a controller of electrical appliances in a house and are connected to the DCC of the energy provider through a communication network. At each time cycle, the smart meters update user information to, and receive control information from the DCC, while the DCC decides the power distribution among the users based on the real-time system information such as grid load, user demand and provider's cost. The DCC manages the entire system as a whole to achieve an optimum distribution scheme that balances the users' utility, supply cost of the energy provider, and the variance of the grid.

Here, the time cycles or slots indexed by $t \in \{1, 2, \cdots\}$ could be, e.g., 1 hour, 0.5 hour, 15 minutes and even shorter, according to the updating period of the smart meters and the size of the smart grid. Usually, the DCC takes a one-day operation cycle based on the daily periodical nature of electricity usage. Note that this is not a requirement for the model but a practical scenario in most cases, which will be applied in the performance evaluation section. Let $\mathbb{N} = \{1, 2, \cdots, N\}$ be the set of users. We denote the power consumption of user i at time t as $p_i(t)$. At each time slot, user i's minimum demand $p_{i,min}(t)$ should be guaranteed, i.e.,

$$p_i(t) \geq p_{i,min}(t), \forall\, i \in \mathbb{N}, t. \tag{2.1}$$

Table 2.1: Notation Table for Chapter 2

Symbol	*Description*
$\mathbb{N}$	set of electricity users in the system
$\mathbb{P}$	set of power demand or consumption for users in a time slot
$\mathbb{C}$	set of maximum cost for the energy provider at any time t
$\mathbb{U}$	set of user utility functions
N	number of users in the system
T	total number of time slots in offline problem
$\mathbf{P}$	power usage by the N users from time $1:T$, offline
$\vec{P}_i$	power usage by user i from time $1:T$, offline
$\vec{P}(t)$	power usage by the N users at time t, offline
$P_i(t)$	power usage by user i at time t, offline
$\mathbf{P}^*$	optimal solution of the offline problem
$\vec{P}_i^*$	optimal power distribution for user i from time $1:T$, offline
$\vec{P}^*(t)$	optimal power distribution for N users at time t, offline
$P_i^*(t)$	optimal power distribution for user i at time t, offline
$\eta(t)$	Lagrange multipliers associated with the offline problem
$\gamma_i(t)$	Lagrange multipliers variable associated with the offline problem
$\vec{p}(t)$	power usage by the N users at time t, online
$p_i(t)$	power usage by user i at time t, online
p_i	power usage by user i at a fixed time, online
$\hat{\vec{p}}(t)$	asymptotically convergent vector in the online problem
$\hat{p}_i(t)$	asymptotically convergent variable in the online problem
$\mathbf{p}^*$	optimal solution of the online problem
$\vec{p}^*(\hat{\vec{p}}, c(t))$	optimal power distribution for N users at time t, online
$\vec{p}^*(t)$	short term for $\vec{p}^*(\hat{\vec{p}}, c(t))$
$p_i^*(\hat{\vec{p}}, c(t))$	optimal power distribution for user i at time t, online
$p_i^*(t)$	short term for $p_i^*(\hat{\vec{p}}, c(t))$
$p_{i,min}(t)$	minimum power demand for user i at time t
$\omega_i(t)$	the flexibility of user i at time t
$L(t)$	grid load at time t
$c(t)$	maximum cost for the energy provider at time t
$U(\cdot)$	user utility function
$C(\cdot)$	cost function of energy provider
$f(\cdot)$	price function
$\Psi(\cdot)$	optimal objective value of the offline problem
$\Phi(\cdot)$	sum of online Lagrange dual function for t from $1:T$
ρ	modified parameter of p

Besides, we assume that the users are rational, which means that at each time slot, power demand of each user has an upper bound, i.e., $p_i(t) \leq p_{i,max}(t)$. This will not become a constraint in our problem, because we aim to satisfy the user demand as much as possible under other constraints. However, this assumption together with (2.1) guarantees a closed set $\mathbb{P}$ which includes all the possible value of power demanded and used, that is, $p_i(t) \in \mathbb{P}$.

User Utility Function

We assume independent users with their own preferences of power usage. For example, each user could have its own time schedule for using different electrical appliances. Also, the user demand may vary as weather changes. Usually the power consumption is larger in a hot summer day than that in a mild day in the spring. Besides, different users may have different reactions to different price schemes [91]. Therefore, it is difficult to characterize user preference with a precise mathematical model. In prior work, user preference is usually represented by a *utility function* [90]. Similarly, we use function $U(p_i(t), \omega_i(t))$ to represent user i's satisfaction on power consumption. We assume $U(\cdot, \cdot)$ to be a strictly increasing, concave function of the allocated power $p_i(t)$; its form could be general. One example is the widely used quadratic utility function [90, 91, 93]. For each user i, the other parameter $\omega_i(t)$ of the utility function indicates the user's flexibility at time t. A larger $\omega_i(t)$ means higher flexibility. $\omega_i(t)$ could be different for users or vary over time. Its values are sent to the DCC at each updating cycle by the smart meter.

Energy Provisioning Cost

For energy providers, when demand is in the normal level, the generation cost increases only slowly as the demand grows. However, it will cost much more when the load peak is approaching the grid capacity, because the provider has to transmit more power from the outside or backup batteries to avoid a blackout. Therefore, we use an increasing and strictly convex function to approximate the *cost function* for energy provisioning. Similar to [91, 93],

we choose a quadratic function to model the provider's cost.

$$C(L(t)) = a \cdot L^2(t) + b \cdot L(t) + c, \tag{2.2}$$

where $a > 0$ and $b, c \geq 0$ are pre-selected for the power grid and $L(t) = \sum_{i \in \mathbb{N}} p_i(t)$ denotes the grid load, i.e., the total power consumption for time slot t. From the provider's perspective, we assume that it aims to meet the user demand under an acceptable cost constraint $c(t)$ at time t, which shall not be exceeded.

$$C(L(t)) \leq c(t), \forall\, t \in \{1, 2, \cdots, T\}. \tag{2.3}$$

We call $c(t)$ *budget* in the rest of this chapter. Without loss of generality, we assume $c(t)$ to be an ergodic process, which is taken from a set $\mathbb{C}$, i.e., $c(t) \in \mathbb{C}$.

Price Model

Dynamic pricing like real-time pricing (RTP), critical peak pricing (CPP) and time of use pricing (TUP) [96] could be incorporated in the smart grid environment. However, real electricity market is still dominated by simple pricing schemes. In this chapter, we use a simple price model that can characterize most real electricity markets, especially for residential usage. As shown in [89, 97], without dynamic price demand, the price load curve has the shape of a *hockey stick*; it remains flat over a long range of grid load and then grows upward steeply as demand approaches the grid capacity. Let $f(\cdot)$ be the *price function* and $f(L(t))$ the price at time t. Therefore, we assume $f(\cdot)$ to be a twice-differentiable increasing convex function that maps the total load to a price. Similar to the utility function $U(\cdot)$, the price function $f(\cdot)$ could have a general form as well.

2.2.2 Problem Formulation

As mentioned in Section 2.1, we aim to minimize the load variance in the grid while maximizing user satisfaction. Large load variance is undesirable for grid operation. It brings about uncertainties that affect not only user satisfaction but also the stability of the power system. Furthermore, the energy provisioning cost should be bounded and users' necessary power needs should be guaranteed.

We first consider an offline scenario where the DCC distributes the power to users during time $t = 1, 2, \cdots, T$, and all the information on users' flexibility $\omega_i(t)$ and provider's budget $c(t)$ are assumed to be known in advance. Let $P_i(t)$ denote the power usage for user i at time t, for $t \in \{1, 2, \cdots, T\}$. In this chapter, we use upper case P in the *offline problem* (see Section 2.3), where all the necessary constraints are known a priori. In the corresponding *online problem*, which will be examined in Section 2.4, we use lower case p for the corresponding variables. A vector with subscript i is used to denote a time sequence, e.g., $\vec{P}_i$ for the power usage by user i for $t \in \{1, 2, \cdots, T\}$. The offline problem can be formulated as follows.

$$\text{max: } \sum_{t=1}^{T} \sum_{i \in \mathbb{N}} \left[U(P_i(t), \omega_i(t)) - f\left(\sum_{i \in \mathbb{N}} P_i(t) \right) P_i(t) \right] - \frac{\alpha T}{2} \text{Var}\left(\sum_{i \in \mathbb{N}} \vec{P}_i \right) \tag{2.4}$$

subject to:

$$P_i(t) \geq P_{i,min}(t), \forall\, i \in \mathbb{N}, t \in \{1, 2, \cdots, T\} \tag{2.5}$$

$$C\left(\sum_{i \in \mathbb{N}} P_i(t) \right) \leq c(t), \forall\, t \in \{1, 2, \cdots, T\}, \tag{2.6}$$

where

$$\text{Var}\left(\sum_{i \in \mathbb{N}} \vec{P}_i \right) = \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i \in \mathbb{N}} P_i(t) - \frac{1}{T} \sum_{k=1}^{T} \sum_{i \in \mathbb{N}} P_i(k) \right)^2.$$

The objective function (2.4) consists of two parts. The first part represents users' satisfaction and preference as the difference between user utility and cost. The second part

represents the load variance of the grid. These two parts are integrated with a parameter $\alpha > 0$, allowing a trade-off between the two. Constraint (2.5) indicates the minimum user demand should be guaranteed, while constraint (2.6) represents the cost upper bound for the energy provider. In section 2.3, we present an algorithm that can solve this offline problem and explain how we can move from offline to online. In Section 2.4, we present an algorithm to solve the corresponding online problem that does not require any a priori user/grid information, and show that the online algorithm is asymptotically optimal.

2.3 Offline Algorithm

In the offline problem (2.4), the user power consumption $P_i(t)$'s are independent. Hence the variance term can be rewritten as $\text{Var}(\sum_{i\in\mathbb{N}} \vec{P_i}) = \sum_{i\in\mathbb{N}} \text{Var}(\vec{P_i})$ and the *price function* $f\left(\sum_{i\in\mathbb{N}} P_i(t)\right)$ is same for each user, which means

$$\sum_{i\in\mathbb{N}} f\left(\sum_{i\in\mathbb{N}} P_i(t)\right) P_i(t) = f\left(\sum_{i\in\mathbb{N}} P_i(t)\right) \sum_{i\in\mathbb{N}} P_i(t)$$

. Therefore, we could depart the first term of (2.4) and rewrite the price term and variance term respectively. Then the problem can be reformulated as follows (termed Prob-OFF).

$$\begin{aligned} \text{max: } & \Psi(\mathbf{P}) = \sum_{t=1}^{T}\sum_{i\in\mathbb{N}} U(P_i(t),\omega_i(t)) - \sum_{t=1}^{T} f\left(\sum_{i\in\mathbb{N}} P_i(t)\right) \sum_{i\in\mathbb{N}} P_i(t) - \frac{\alpha T}{2}\sum_{i\in\mathbb{N}} \text{Var}\left(\vec{P_i}\right) \quad (2.7) \\ \text{subject to: } & (2.5) - (2.6), \end{aligned}$$

where $\mathbf{P}$ is an $N \times T$ matrix that denotes the power allocated for each user i at time $t \in \{1, 2, \cdots, T\}$ and $\text{Var}(\vec{P_i}) = \frac{1}{T}\sum_{t=1}^{T}\left(P_i(t) - \frac{1}{T}\sum_{k=1}^{T} P_i(k)\right)^2$.

In Prob-OFF, $U(\cdot)$ is concave and $C(\cdot)$ is convex. Since the price function $f(\cdot)$ is convex, $f(\sum_{i\in\mathbb{N}} P_i(t)) \sum_{i\in\mathbb{N}} P_i(t)$ is also convex. We only need to show the convexity of $\text{Var}(\vec{P_i})$ to establish a convex optimization problem. The convexity of $\text{Var}(\vec{P_i})$ can be easily proved by its definition.

Lemma 2.1. *Prob-OFF is a convex optimization problem and has a unique solution.*

The complete proof of Lemma 2.1 is presented in Appendix A.3. As Lemma 2.1 holds, we can carefully choose $P_{i,min}(t)$ to meet Slater's condition [98], and thus the KKT conditions [98] are sufficient and necessary for the optimality of Prob-OFF. Let $\mathbf{P}^*$ be an optimal solution to Prob-OFF. Let $\eta(t)$ and $\gamma_i(t)$ be the Lagrange multipliers and variables, respectively, for $i \in \mathbb{N}$ and $t \in \{1, 2, \cdots, T\}$. We have

$$\begin{cases} U'(P_i^*(t), \omega_i(t)) - h\left(\sum_{i\in\mathbb{N}} P_i^*(t)\right) - \alpha\left(P_i^*(t) - \bar{P}_i^*\right) - \\ \qquad\qquad \eta(t) C'\left(\sum_{i\in\mathbb{N}} P_i^*(t)\right)/c(t) + \gamma_i(t) = 0 \\ \eta(t)\left(C\left(\sum_{i\in\mathbb{N}} P_i^*(t)\right)/c(t) - 1\right) = 0 \\ \gamma_i(t)\left(P_i^*(t) - P_{i,min}(t)\right) = 0 \\ \eta(t), \gamma_i(t) \geq 0, \forall\, i \in \mathbb{N},\ t \in \{1, 2, \cdots, T\}, \end{cases} \tag{2.8}$$

where

$$h\left(\sum_{i\in\mathbb{N}} P_i^*(t)\right) = f'\left(\sum_{i\in\mathbb{N}} P_i^*(t)\right)\sum_{i\in\mathbb{N}} P_i^*(t) + f\left(\sum_{i\in\mathbb{N}} P_i^*(t)\right)$$

and

$$\bar{P}_i^* = \frac{1}{T}\sum_{k=1}^{T} P_i^*(k). \tag{2.9}$$

From the above equations, we can solve for $\eta(t)$ as

$$\eta(t) = \frac{\alpha\left(\bar{P}_i^* - P_i^*(t)\right) + U'(P_i^*(t), \omega_i(t)) - h(\sum_{i\in\mathbb{N}} P_i^*(t)) + \gamma_i(t)}{C'(\sum_{i\in\mathbb{N}} P_i^*(t))/c(t)} \tag{2.10}$$

Therefore, to achieve optimality, there is an identical $\eta(t)$ for all users in a time slot t. The optimal solution guarantees that the right-hand-side (RHS) of (2.10) has the same value for all users. Furthermore, we observe that only the $\bar{P}_i^*$ term requires information from other time slots. This implies that if $\bar{P}_i^*$ could be accurately estimated, the optimal energy

distribution $\mathbf{P}^*$ could be determined using only information in the current time slot, such as $c(t)$ and $P_{i,min}(t)$. This is essential, because in the offline scenario, our assumption that future information are known a priori is not a possible case in the real smart grid. Based on this observation, we are able to present an *online algorithm* for the energy distribution problem in the next section which requires no future information.

2.4 Online Algorithm

In this section, we present an online algorithm for energy distribution, and prove that the online solution is asymptotically convergent to the offline optimal solution, i.e., *asymptotically optimal*. The online energy distribution algorithm consists of the following three steps, denoted as Algorithm 2.1:

Algorithm 2.1: Online Energy Distribution Algorithm

Step 1: For each $i \in \mathbb{N}$, initialize $\hat{p}_i(0) \in \mathbb{P}$.

Step 2: In each time slot t, the DCC solves the following convex optimization problem (termed Prob-ON).

$$\text{max:} \quad \sum_{i\in\mathbb{N}} U(p_i(t),\omega_i(t)) - f\left(\sum_{i\in\mathbb{N}} p_i(t)\right)\sum_{i\in\mathbb{N}} p_i(t) - \frac{\alpha}{2}\sum_{i\in\mathbb{N}}(p_i(t)-\hat{p}_i(t-1))^2 \quad (2.11)$$

$$\text{subject to:} \quad p_i(t) \geq p_{i,min}(t), \forall\, i \in \mathbb{N} \quad (2.12)$$

$$C\left(\sum_{i\in\mathbb{N}} p_i(t)\right) \leq c(t), \forall\, t. \quad (2.13)$$

Let $\vec{p}^*(t)$ denote the solution to Prob-ON, where each element $p_i^*(t)$ represents the optimal power allocation to user i.

Step 3: Update $\hat{p}_i(t)$ for all $i \in \mathbb{N}$ as follows.

$$\hat{p}_i(t) = \hat{p}_i(t-1) + \frac{\alpha}{t+\alpha}\cdot(p_i^*(t) - \hat{p}_i(t-1)). \quad (2.14)$$

$\vec{p}^*(t)$ is indeed the short term of $\vec{p}^*(\hat{\vec{p}}, c(t))$. For brevity, we use $\vec{p}^*(t)$ instead in the chapter when it is clear in context. Comparing to (2.7), the variance term is approximated by $\sum_{i\in\mathbb{N}}(p_i(t) - \hat{p}_i(t-1))^2$ in (2.11). In Prob-ON, (2.14) can be viewed as a stochastic approximation updating equation, if the budget of the energy provider, $c(t)$, is viewed as a stationary stochastic process. This interpretation can be justified because $c(t)$ is assumed to be ergodic, and thus is stationary.

Similar to Prob-OFF, problem Prob-ON is also a convex optimization problem satisfying Slater's condition. Its KKT conditions with KKT multipliers $\lambda(t)$ and KKT variables $\nu_i(t)$, for $i \in \mathbb{N}$, are as follows.

$$\begin{cases} U'(p_i^*(t), \omega_i(t)) - h\left(\sum_{i\in\mathbb{N}} p_i^*(t)\right) - \alpha\left(p_i^*(t) - \hat{p}_i(t-1)\right) \\ \qquad\qquad -\lambda(t)C'\left(\sum_{i\in\mathbb{N}} p_i^*(t)\right)/c(t) + \nu_i(t) = 0 \\ \lambda(t)\left(C\left(\sum_{i\in\mathbb{N}} p_i^*(t)\right)/c(t) - 1\right) = 0 \\ \nu_i(t)\left(p_i^*(t) - p_{i,min}(t)\right) = 0 \\ \lambda(t), \nu_i(t) \geq 0, \ \ \forall\, i, t. \end{cases} \tag{2.15}$$

In the remainder of this section, we firstly prove that $\hat{p}_i(t)$ approaches a limit for t goes to infinity and then we show that $\hat{p}_i(t)$ converges to the mean of the power allocated to each user $i \in \mathbb{N}$ over time, as given in (2.9).

We begin with the definition the function $g(\hat{\vec{p}}, c(t))$:

$$\begin{aligned} g(\hat{\vec{p}}, c(t)) \ = \ & \sum_{i\in\mathbb{N}} U(p_i^*(\hat{\vec{p}}, c(t)), \omega_i(t)) - \\ & f\left(\sum_{i\in\mathbb{N}} p_i^*(\hat{\vec{p}}, c(t))\right) \sum_{i\in\mathbb{N}} p_i^*(\hat{\vec{p}}, c(t)) - \\ & \frac{\alpha}{2}\sum_{i\in\mathbb{N}} (p_i^*(\hat{\vec{p}}, c(t)) - \hat{p}_i)^2. \end{aligned} \tag{2.16}$$

Note that the optimized function $g(\hat{\vec{p}}, c(t))$ share the same form with (2.11), but with a different meaning. Here we regard the optimizer $\vec{p}^*(\hat{\vec{p}}, c(t))$ and the optimized objective $g(\hat{\vec{p}}, c(t))$ as stochastic processes. We need to show the process $\hat{\vec{p}}(t)$ converges almost surely, for given stationary stochastic process $c(t)$. We have the following immediate properties of $\vec{p}^*(\hat{\vec{p}}, c(t))$ and $g(\hat{\vec{p}}, c(t))$.

Property 2.1. *Continuity of $\vec{p}^*(\hat{\vec{p}}, c(t))$ and $g(\hat{\vec{p}}, c(t))$.*

For any $c(t) \in \mathbb{C}$, we have

i) $\vec{p}^*(\hat{\vec{p}}, c(t))$ and $g(\hat{\vec{p}}, c(t))$ are continuous functions of $\hat{\vec{p}}$;

ii) $E[\vec{p}^*(\hat{\vec{p}}, c(t))]$, $E[g(\hat{\vec{p}}, c(t))]$ are continuous functions of $\hat{\vec{p}}$.

Property 2.2. *Differentiability of $g(\hat{\vec{p}}, c(t))$ and $E[g(\hat{\vec{p}}, c(t))]$.*

For any $c(t) \in \mathbb{C}$ and each $i \in \mathbb{N}$, we have

i) $\nabla_{\hat{p}_i} g(\hat{\vec{p}}, c(t)) = \alpha(p_i^*(\hat{\vec{p}}, c(t)) - \hat{p}_i)$;

ii) $\nabla_{\hat{p}_i} E[g(\hat{\vec{p}}, c(t))] = \alpha(E[p_i^*(\hat{\vec{p}}, c(t))] - \hat{p}_i)$.

With Properties 2.1 and 2.2, we are able to show the following result, which is an important step to the proof of the convergence of process $\hat{\vec{p}}$. We next show the convergence of $\hat{p}_i(t)$ stated in the following Lemma 2.2 and Lemma 2.3. The complete proofs of Properties 2.1 and 2.2, Lemmas 2.2 and 2.3 are shown in Appendix A.

Lemma 2.2. *The solution of the following fixed point equation is unique*

$$E[\vec{p}^*(\hat{\vec{p}}, c(t))] = \hat{\vec{p}}. \tag{2.17}$$

Lemma 2.3. *$\hat{p}_i(t)$ converges almost surely to the unique solution $\hat{\vec{p}}$ of the fixed point equation $E[\vec{p}^*(\hat{\vec{p}}, c(t))] = \hat{\vec{p}}$.*

Based on the convergence of $\hat{p}_i(t)$, we are ready to prove the asymptotic optimality of the online algorithm, which indicates that for a sufficiently long time period, the time averaged difference between the online and offline objective values will become negligible. The results are shown in the following lemma and theorem.

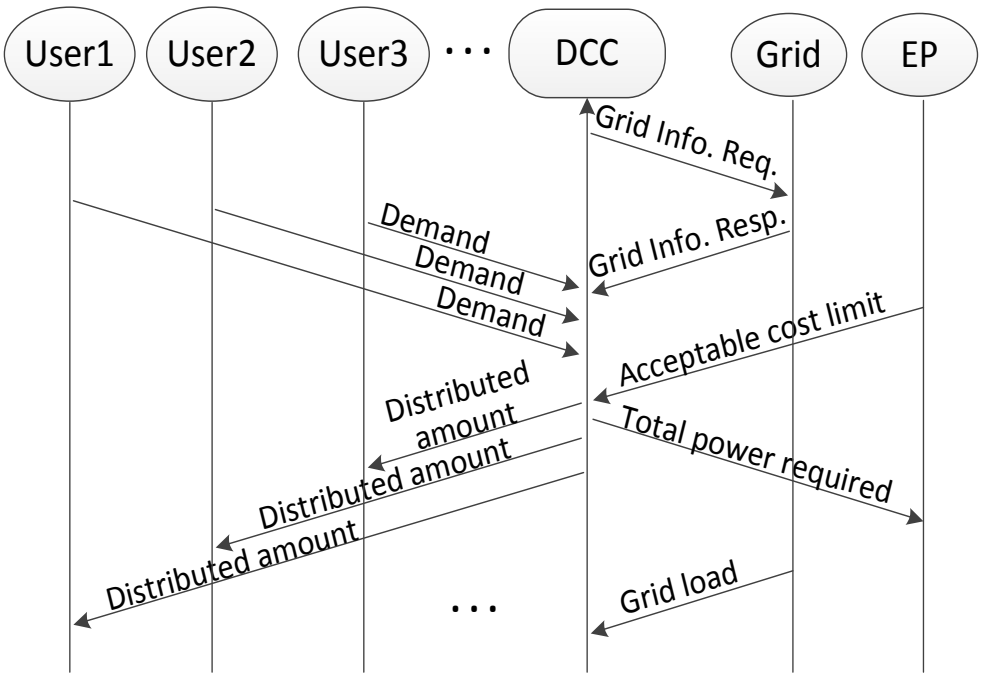

Figure 2.2: Information flows in the power distribution network.

Lemma 2.4. *The following limit exits and converges for* $i \in \mathbb{N}$*:*

$$\lim_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} p_i^*(t) - \hat{p}_i(T)\right) = 0.$$

Theorem 2.1. *The online optimal solution converges asymptotically and almost surely to the offline optimal solution.*

Lemma 2.4 and Theorem 2.1 guarantee the asymptotic convergence of the online solutions to the offline solutions. See Appendix A.6 and A.7 for the proof of Lemma 2.4 and Theorem 2.1 respectively.

2.5 Communication Network Protocol

Information exchange is an important element of the emerging smart grid. Communications between smart meters and the DCC are essential for both control and distribution. Algorithm 2.1 is also based on such information exchanges. As more advances are made in smart grid, there is a compelling need for network architectures, standards, and protocols for

communications in smart grid. We hereby introduces a basic protocol for communications network support in the smart grid, which is simple but sufficient to support the real time online power distribution algorithm.

In Algorithm 2.1, the users' basic demand for power and the maximum acceptable cost of the energy provider (EP) should be updated in every decision period at the DCC for real time execution, because these are the constraints and are sometimes unpredictable. As we try to smooth the total power consumption of all users in the system, grid stability is another objective. We have four entities in the system: the DCC is the core and Users, EP and Grid are also important participants. At the beginning of each time slot, users send their demands to the DCC through their smart meters, while EP informs the DCC its acceptable cost limit. The DCC also collects other information from the grid, such as the actual grid load. Then the DCC executes the online power distribution algorithm using the updated information. It sends the allocated amounts to the users and the total usage or demand to the EP. Moreover, DCC is able to send other control information to the EP or users for regulation, accounting, emergency response and alerts, etc.

Fig. 2.2 illustrates the information flows in the network system. At each updating slot, the DCC sends a grid information request to the grid, which returns relevant real-time grid parameters such as load condition and capacity. Meanwhile, users send their basic power demands to the DCC to request power for the time period. Also, the EP sends its cost limit to the DCC to get their energy provisioning cost controlled within an acceptable range. After the DCC have gathered these necessary information, it applies Algorithm 2.1 and then sends the results to the EP and users, so that the EP could supply the corresponding amount of power to the users. Finally, the grid will update the actual grid load to the DCC for grid inspection and control. Note that there will be no information exchange among the users, so that their privacy (e.g., electricity usage habit) could be protected. Note that the update interval are at the order of hour or tens of minutes. Given the data rate of existing

wireless networks, such exchange of control information only takes a negligible fraction of the interval.

2.6 Practical Online Algorithm

In the smart grid communication network discussed in Sectoin 2.5, we notice that the DCC has to communicate with all the Users, the Grid and the EP. It is a large burden when the smart grid becomes larger. This only brings more users, but will increase the time for both calculation and communication. With modern network infrastructure and protocol, time for information exchange could be well controlled. However, in complicated practical situations, the *utility function*, the *cost function* and the *price function* may have different realizations, some of which are very complex. This will no doubt bring much difficulty for the DCC to solve the Prob-ON. In some cases, the KKT conditions (see (2.15)) are very difficult to solve especially in short intervals. Therefore,in this section, we present a practical online algorithm (termed Algorithm 2.2) for energy distribution in smart grid, motivated by Algorithm 1 stated in Section 2.4.

At each time slot t, $p_i(t)$ for all $i \in \mathbb{N}$ is the distribution power to user i. In the above practical algorithm, the derivative could be replaced by the difference equation, when the analytic function form of the function is difficult to be acquired. For example, use $\frac{C(\sum_{i\in\mathbb{N}} p_i(t)) - C(\sum_{i\in\mathbb{N}} p_i(t-1))}{\sum_{i\in\mathbb{N}}[p_i(t) - p_i(t-1)]}$ instead of $C'(\sum_{i\in\mathbb{N}} p_i(t))$, when the *cost function* cannot be formulated. From the practical algorithm, we see clearly the allocation process in evaluating λ_i in Step 2, which is a natural expression from (2.10). In this way, DCC distributes the energy uniformly while not giving a user too much. So we would expect a more smoothy allocation.

Algorithm 2.2: Practical Online Energy Distribution Algorithm

---

Step 1: For each $i \in \mathbb{N}$, initialize $\hat{p}_i(0) \in \mathbb{P}$.

In each time slot t, do the next two steps:

Step 2: For each $i \in \mathbb{N}$, initialize $p_i(t) = p_{i,min}(t)$. Let the set $\mathbb{S} = \{1, 2, ..., N\}$. Then take the following loop:

while $\mathbb{S} \neq \emptyset$

For each $i \in \mathbb{S}$, take

$$\lambda_i = \frac{\alpha(\hat{p}_i(t-1) - p_i(t)) - h(\sum_{i\in\mathbb{N}} p_i(t)) + U'(p_i(t), \omega_i(t))}{(1/c(t))C'(\sum_{i\in\mathbb{N}} p_i(t))};$$

if $\max_{i\in\mathbb{S}}\lambda_i < 0$, then $\mathbb{S} = \emptyset$;

else $j = \text{argmax}_{i\in\mathbb{S}}\lambda_i$;

$p_j(t) = p_j(t) + step$;

if $p_j(t) > \max\{p : p \in \mathbb{P}\}$ or $C(\sum_{i\in\mathbb{N}} p_i(t)) > c(t)$

then $p_j(t) = p_j(t) - step$;

delete j from $\mathbb{S}$;

end

end

end

Step 3: In each time slot t, update $\hat{p}_i(t)$ for all $i \in \mathbb{N}$ as follows:

$$\hat{p}_i(t) = \hat{p}_i(t-1) + \frac{\alpha}{t+\alpha} \cdot (p_i(t) - \hat{p}_i(t-1)).$$

The parameter *step* controls the incremental precision and the running number(and time) of the loop in Step 2. When the updating interval is short, it is safe to set *step* very small, which leads to a longer running time and vice versa. The complexity of the practical algorithm is roughly proportional to $N \cdot \max(p)/step$, i.e., the number of users times the maximum distributed energy over the increment. *step* also decides the error between the practical solution and the theoretical solution to KKT conditions. So *step* is an important parameter in the practical online algorithm. The DCC could choose *step* according to the

length of updating periods and the number of users. With controlled *step*, the DCC could support a large number of users. Although the good power distribution from the practical algorithm is not the optimal one, it is more practical as it can be used with complicated functions and its running time and precision could be controlled.

2.7 Performance Evaluation

2.7.1 Simulation Configuration

In this section, we evaluate the proposed online algorithm with trace-driven simulations. The simulation data and parameters are acquired from the traces of power consumption in the Southern California Edison (SCE) area recorded in 2011 [99]. We first study the performance of the proposed algorithm on convergence, grid load variance and peak reduction. We then compare the online algorithm with an existing scheme under different numbers of users.

Consider a power distribution system in a small area with $N = 20$ users and 15 minutes updating periods. Note that a quarter is a practical set which allows DCC to have sufficient time to coordinate all the users so that the system could support more users and that in most cases, 15 minutes is short enough to show the users' change of demand. We will show results within a 24-hour time pattern for an evaluation of the daily operations. We choose users' utility function from a function set $\mathbb{U}$ in which the functions are generated as widely used quadratic expressions (see [90, 91, 93]), with $\omega_i(t) \in (0, 1)$ randomly selected.

$$U(p_i(t), \omega_i(t)) = \begin{cases} \omega_i(t)p_i(t) - \frac{1}{8}p_i(t)^2, & \text{if } 0 \le p_i(t) \le 4\omega_i(t) \\ 4\omega_i(t), & \text{if } p_i(t) \ge 4\omega_i(t). \end{cases} \tag{2.18}$$

We also assume the basic user demand $p_{i,min}(t)$ and the initial value $p_i(0)$ are selected from the set of $\mathbb{P} = [0.5, 3]$, for all i. The parameters in the energy provisioning cost function (2.2) are set as $a = 0.05$, $b = c = 0$, and $c(t)$ is selected randomly from the set $\mathbb{C} = [1, 20]$ for each time slot. These parameters are carefully determined after studying the characteristics

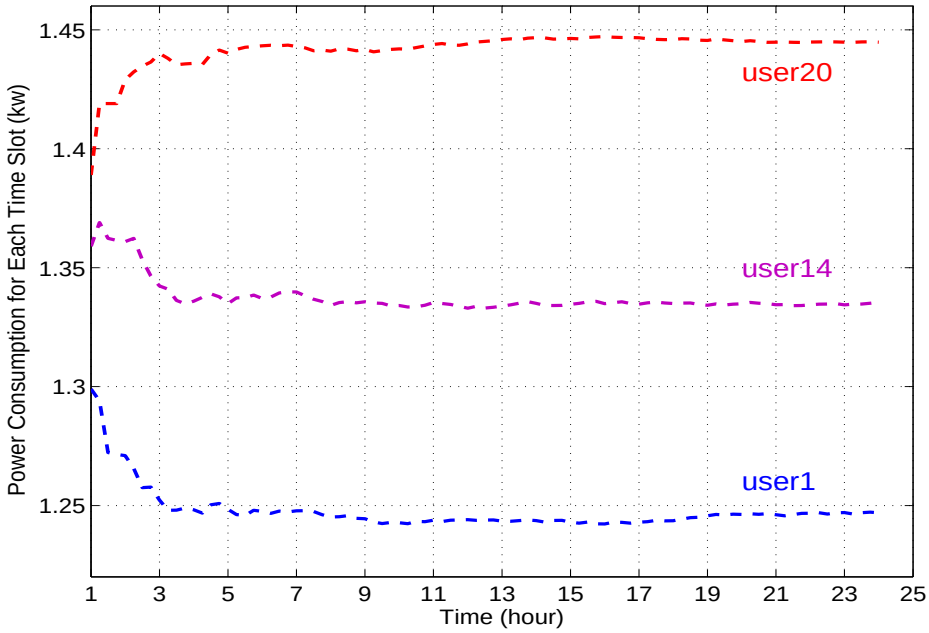

Figure 2.3: Convergence of $\hat{p}_i(t)$ for different users ($\alpha = 1$).

of the SCE trace. In addition, we choose the price function as

$$f(L(t)) = 0.047 \cdot L(t)^2 - 0.38 \cdot L(t) + 27.67. \tag{2.19}$$

It is a quadratic function and also a twice-differentiable increasing convex function as discussed in Section 2.2.1. This model is formulated from the predicted and actual prices from the SCE trace [100]. We simulate two scenarios with α set as 1 and 0.01, respectively, to examine how it affects the result.

2.7.2 Algorithm Performance

We first study the convergence of $\hat{p}_i(t)$. Earlier discussions in Section 2.4 show that $\hat{p}_i(t)$ is convergent. Fig. 2.3 illustrates that for $\alpha = 1$, one day is sufficient for $p_i^*(t)$ to converge to steady state values. In Fig. 2.4, it takes more time to converge. In the online problem Prob-ON, α is not only a parameter integrating different objectives, but also an important coefficient affecting the convergence of the algorithm. In the online updating

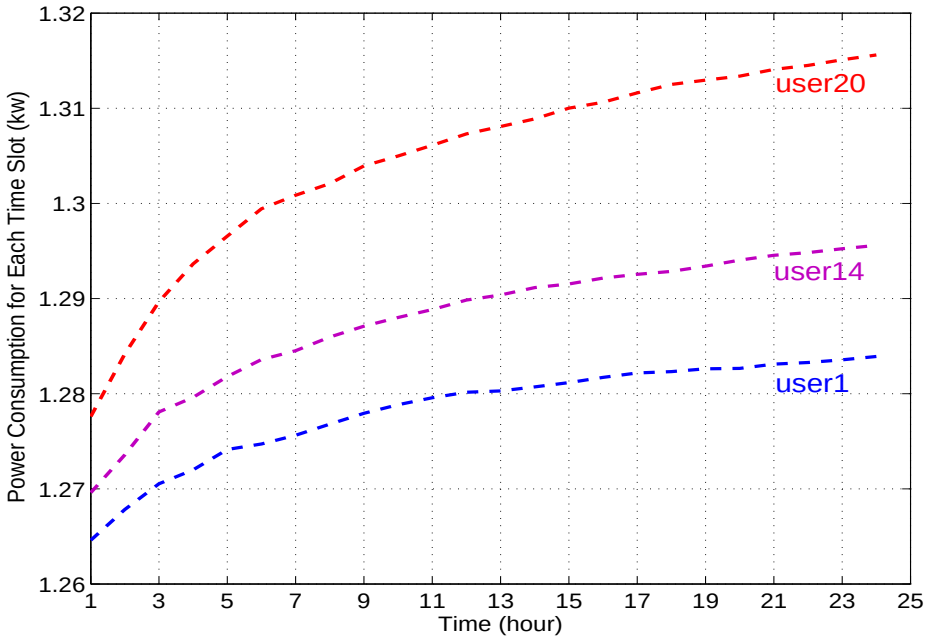

Figure 2.4: Convergence of $\hat{p}_i(t)$ for different users ($\alpha = 0.01$).

equation (2.14), it is clear that a large α will cause relatively a large disturbance, especially at the very beginning. However, a large α will also lead to fast convergence, and vice versa, as shown in Figs. 2.3 and 2.4. Besides, α also affects the impact of the variance (or, smoothness) on the overall objective value (2.11). It shapes the grid load curve to some degree, as we will see in Section 2.7.3.

Lemma 2.4 states that $\hat{p}_i(t)$ will converge to the time averaged $p_i^*(t)$ if we run the simulation sufficiently long. For a larger α, the convergence will be faster, shown in Fig. 2.5, where we find that $\hat{p}_i(t)$ fluctuates uniformly along the $p_i^*(t)$ curve for different users. For a smaller α, the convergence could be very slow. Fig. 2.6 demonstrates the slow convergence when $\alpha = 0.01$. However, the convergence of $\hat{p}_i(t)$ is proved to be true as $T \to \infty$ (see the proof of Lemma 2.4). In Fig. 2.6, it can be seen that $\hat{p}_i(t)$ is still approaching $p_i^*(t)$, although slowly. Therefore, the value of α should be carefully chosen to trade-off between convergence and other objectives.

More importantly, our main objective is to develop an optimal online algorithm to reduce the variance of the grid load and to balance electricity demand and supply. In Fig. 2.7, we

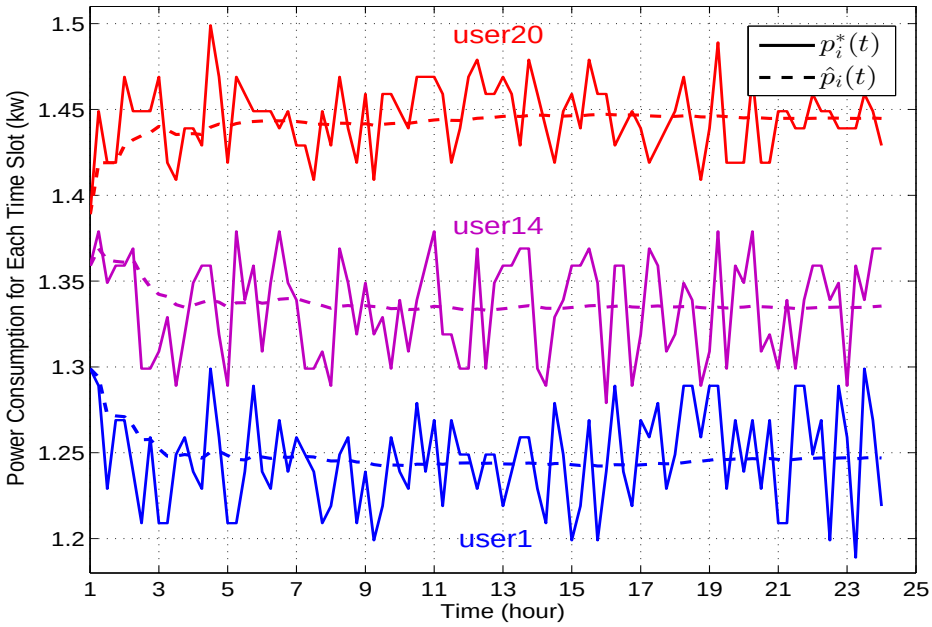

Figure 2.5: Online power distribution $p_i^*(t)$ and $\hat{p}_i(t)$ for different users when $\alpha = 1$.

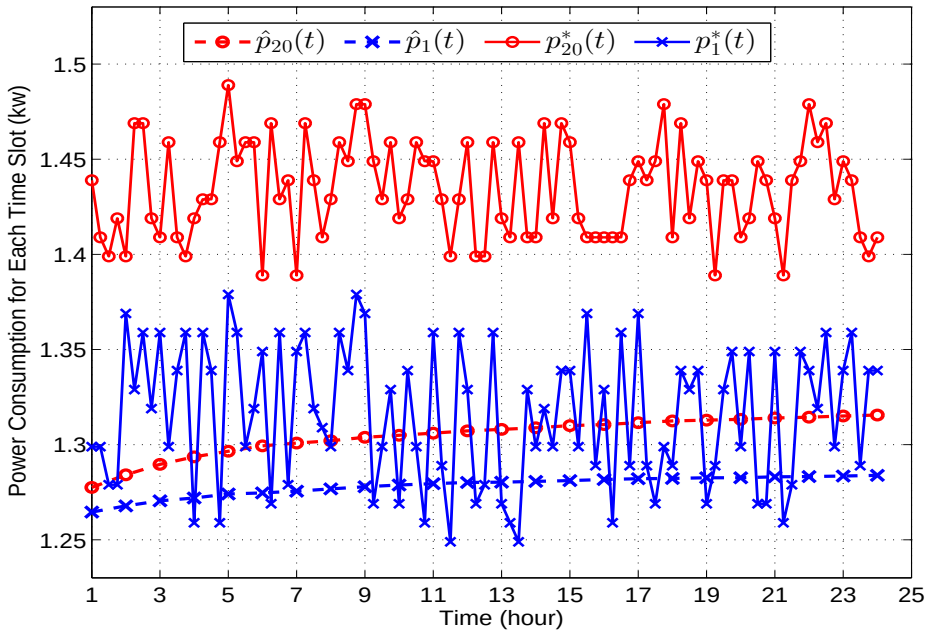

Figure 2.6: Online power distribution $p_i^*(t)$ and $\hat{p}_i(t)$ for different users when $\alpha = 0.01$.

plot the total power consumption achieved with the online algorithm and the actual load. The real power usage is the summation of 20 independent users' consumption generated by the average real load in the SCE trace on a hot day (i.e., Sept. 1, 2011) [99]. The constraints

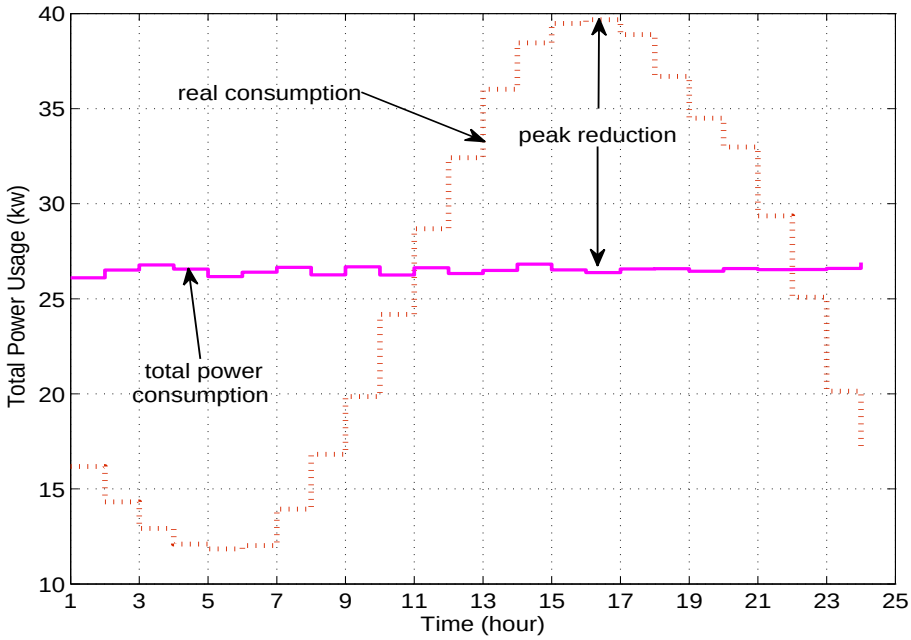

Figure 2.7: Real power usage and total power usage by the online algorithm when $\alpha = 1$.

are derived from the real load in the 2011 SCE trace. For better presentation, we only plot the result of the online algorithm with $\alpha = 1$. The results with $\alpha = 0.01$ will be shown in Section 2.7.3.

In Fig. 2.7, we find that the online algorithm achieves a well smoothed grid load. Interestingly, although the power usage of each user varies over time (as shown in Fig. 2.5), the total power usage is effectively smoothed out by the online algorithm. This result demonstrates the effectiveness of variance detection and reduction of the online algorithm. Although the controlled curve lies slightly above the average level of the real load, it reduces the cost of energy provisioning by achieving a considerable peak reduction, which is about 35% in this scenario with only 20 users.

2.7.3 Comparison with a Benchmark

We next compare the online algorithm with the Optimal Real-time Pricing Algorithm (ORPA) presented in [93] as a Benchmark. Comparing to prior work, this one formulates a similar but simpler problem to our problem. It adopts a real-time pricing strategy to

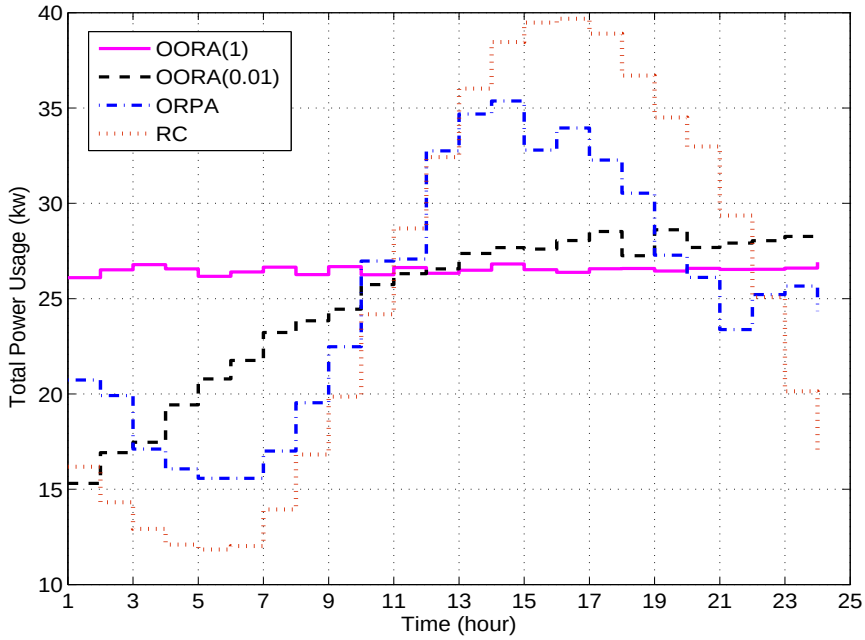

Figure 2.8: Total power consumption for OORA(1), OORA(0.01), ORPA and RC.

maximize social welfare of the smart grid, as

$$\max \sum_{i \in \mathbb{N}} \left(U(p_i(t), \omega_i(t)) - C\left(\sum_{i \in \mathbb{N}} p_i(t) \right) \right), \tag{2.20}$$

for $t \in \{1, 2, \cdots, T\}$ and for all independent user i. As we can see, (2.20) is similar to but simpler than (2.11). With the same parameters as in the online algorithm, this is also a convex optimization problem. We can solve (2.20) with a centralized interior-point method as discussed in [93].

Firstly, we show the total power consumption of different algorithms in Fig. 2.8. From the aspect of smoothness, we could see clearly that the online optimal real-time energy distribution algorithm with $\alpha = 1$ (termed OORA(1)) achieves the best performance. The figure also shows that the online algorithm with $\alpha = 0.01$ (termed OORA(0.01)) also outperforms the benchmark ORPA. All the three algorithms achieve smoother total loads than the real consumption (RC). The peak reductions over RC are 35% for OORA(1), 28% for

OORA(0.01), and 12.5% for ORPA. Therefore, OORA(1) achieves the largest peak reduction, while OORA(0.01) still outperforms ORPA with considerable gains.

Next, we plot the variance of the total load in Fig. 2.9 for different system settings. These results are consistent with that in Fig. 2.8. We find that OORA(1) achieves the minimum variance for all the cases simulated, while OORA(0.01) still outperforms ORPA with a much smaller variance. This is because variance is explicitly incorporated into the objective function in the online problem formulation, while ORPA is designed mainly to maximize the social welfare as in (2.20) and cannot guarantee a smooth total grid load.

Finally, we provide a more detailed comparison of the three schemes in Table 2.2, where the simulation results of several individual performance measures are listed for networks of 200, 500, and 1000 users. Note that the price function is different for different network sizes, which is a function of the total load. As defined in (2.21), $\overline{V}$, $\overline{U}$, $\overline{F}$, and $\overline{PK}$ denote the averages across users of the total power variance, users' utility, users' cost, and the peak of the total load, respectively, while c is the total energy provisioning cost for the entire period.

$$\begin{cases} \overline{V} = \frac{1}{N}\sum_{i\in\mathbb{N}} \mathrm{Var}(p_i^*(t)) \\ \overline{U} = \frac{1}{N}\sum_{t=1}^{T}\sum_{i\in\mathbb{N}} U(p_i^*(t), \omega_i(t)) \\ \overline{F} = \frac{1}{N}\sum_{t=1}^{T} f(\sum_{i\in\mathbb{N}} p_i^*(t))(\sum_{i\in\mathbb{N}} p_i^*(t)) \\ \overline{PK} = \frac{1}{N}\mathrm{max}_{t\in[1:T]}\sum_{i\in\mathbb{N}} p_i^*(t) \\ c = \sum_{t=1}^{T} C(\sum_{i\in\mathbb{N}} p_i^*(t)). \end{cases} \tag{2.21}$$

For $\overline{V}$, the best performer is OORA(1), which is consistent with the earlier results. Also, the variance is increasing as the user number grows. For $\overline{F}$, we observe a relatively stable number of the averaged cost on daily electricity consumption for each user. In the first three algorithms, $\overline{F}$ is almost the same while RC always has the largest number because in reality where the RC curve was recorded, supply was always matched to the user demand. This is confirmed by the results of users' utility $\overline{U}$: as users could use electricity freely, they should

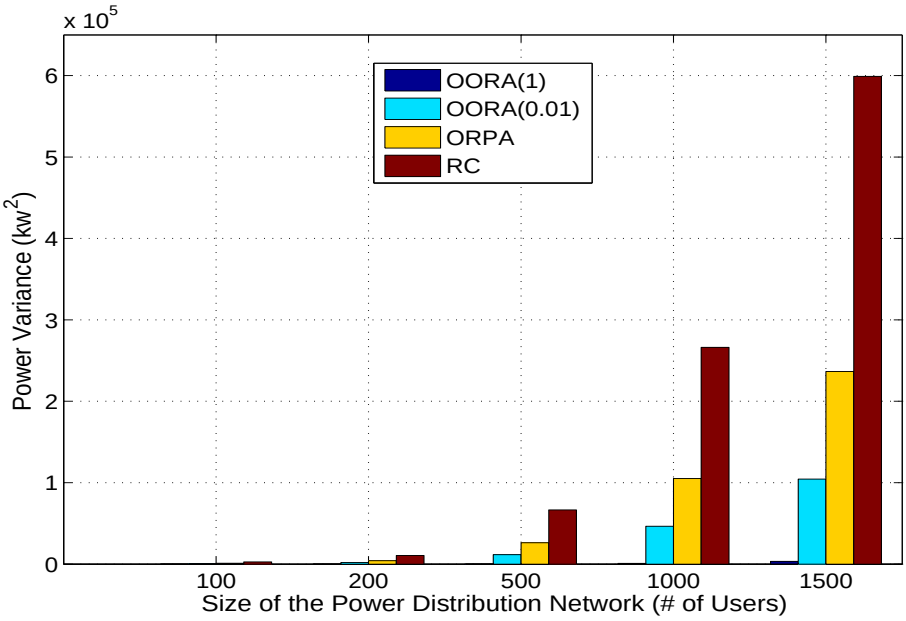

Figure 2.9: Total power variance by OORA(1), OORA(0.01), ORPA and RC.

Table 2.2: Simulation Results of Individual Performance Measures for Different Algorithms

Algorithm	N	$\overline{V}$	$\overline{U}$	$\overline{F}$	c ($\times 10^3$)	$\overline{PK}$
OORA(1)	200	**0.02**	3.52	3.41	1.69	**1.35**
OORA(0.01)	200	9.3	3.59	3.27	1.61	1.46
ORPA	200	21.5	3.56	3.43	1.54	1.79
RC	200	53.5	3.86	3.65	1.86	2.07
OORA(1)	500	**0.05**	3.53	3.31	10.5	**1.37**
OORA(0.01)	500	23.2	3.63	3.42	10.1	1.51
ORPA	500	52.6	3.54	3.28	9.54	1.83
RC	500	113	3.88	3.61	14.0	2.27
OORA(1)	1000	**0.10**	3.51	3.25	42.2	**1.41**
OORA(0.01)	1000	44.1	3.59	3.30	40.2	1.59
ORPA	1000	105	3.54	3.25	38.1	1.93
RC	1000	266	3.87	4.23	54.1	2.58

have the highest satisfaction level. Observing $\overline{U}$ and $\overline{F}$, we see that a higher satisfaction level is achieved with a higher cost. Moreover, it is interesting to see that utility $\overline{U}$ of OORA(1), OORA(0.01), ORPA are almost the same for different numbers of users, with OORA(0.01) being slightly better. This is because, as in ORPA, the utility is incorporated in the objective function of OORA. When α is small, the first two terms in (2.11) will have larger weights.

For energy provisioning cost c, ORPA exhibits its advantage by including c in the objective function. Also, if we take $\overline{U} - c$, ORPA is also the best performer, which could be expected from its objective function (2.20). However, this advantage becomes insignificant when the variance $\overline{V}$ and the peak $\overline{PK}$ are considered. OORA has unique advantages on variance control and peak reduction. It is also worth noting that OORA is an online algorithm that requires minimal exchange of control/state information within the grid, while the ORPA results are obtained with a centralized solver assuming accurate global information.

2.8 Related Work

Smart grid is characterized by the two-way flows of electricity and information and is envisioned to replace the existing power grid in the future [8, 101]. A comprehensive review on smart grid technologies and research can be found in [44], where the research on smart grid is classified into three major areas: infrastructure, management and protection.

In the three areas, demand side management or demand response has been attracting considerable research efforts [81, 84, 88, 90–93, 102, 103]. Researchers work mainly on demand profile shaping, user utility maximization and cost reduction. For example, machine learning is used in [90] to develop a learning algorithm for energy costs reduction and energy usage smoothing, while [92] aims to achieve a balance between user's cost and waiting time. In [93], the authors propose an optimal real-time pricing algorithm to maximize the social welfare, considering user utility maximization and energy provider cost minimization. In [102], the authors formulate a Stackelberg game between utility companies and end-users aiming to maximize the revenue of each utility company and the payoff of each user. In [103], the authors discuss the architecture of home machine-to-machine (M2M) networks for energy management, which is an important component in the smart grid. In these works, convex programming, machine learning and game theory are mostly used.

On the other hand, online algorithms [94] are widely used in wireless communications and networking, where precise channel and network information are hard to obtain. Recent

research on solving wireless networking problems using online algorithms can be found in [95, 104–106]. In [104], two online algorithms are developed from the optimal offline algorithms to maximize the amount of unit-length packets scheduled in a packet-switching mechanism. The authors of [105] address the energy-efficient uplink scheduling problem in a multiuser wireless system. With an online algorithm, an optimal scheduling is achieved without prior knowledge on arrival and channel statistics. In [106], online algorithm is applied to overcome the dynamic nature of the time-varying channels in wireless networks and then the throughput of the single-transmitter is maximized by optimal power assignment. In [95], online algorithm is used for multi-user video streaming in a wireless system so that user's perceived video quality and its variations are jointly considered for a maximization with almost no statistical information about the congested channels.

Our work is inspired by the online algorithm works, which demonstrate the high potential of online algorithms for solving optimization problems with relatively limited information. In power systems, it is possible to use online algorithms to detect and control the grid load variance in real time. Motived by this observation, we propose an energy distribution online algorithm to achieve utility maximization and load smoothing. We consider the key design factors from users, energy provider and load variance in the problem formulation. The proposed online algorithm is quite effective as shown in Section 2.7.

2.9 Conclusion

In this chapter, we present a study of optimal real-time energy distribution in smart grid. With a formulation that captures the key design factors of the system, we first present an offline algorithm that can solve the problem with optimal solutions. We then develop an online algorithm that requires no future information about users and the grid. We also show that the online solution converges to the offline optimal solution asymptotically and almost surely. The proposed online algorithm is evaluated with trace-driven simulations and is shown to outperform an existing benchmark scheme.

Chapter 3

Distributed Online Algorithm for Optimal Energy Distribution in Smart Grid

3.1 Introduction

According to the National Institute of Standard and Technology (NIST) standard [107], the Smart Grid model includes seven domains: Customer, Market, Service Provider, Operations, Bulk Generation, Transmission and Distribution. Each domain functions differently, interactively and cooperatively. However, in some cases, different domains may share some actors and applications. For instance, the Distribution and Customer domains probably both contain smart meters. On the other hand, an integrated utility may have actors in many domains: a distribution system operator could have applications in both Operation and Market domains [107].

In this chapter, we consider real-time energy distribution in a certain area with the Smart Grid system. As shown in Fig. 3.1, the system considered in this chapter includes three large domains: Customer, Power Grid Operator (PGO), and Energy Distributor (ED). The Customer domain here is similar to the one in the NIST model, which represents power users including resident, industrial and others. The PGO performs as Market, Service Provider and Operations do in the seven-domain model. The ED includes the Generation, Transmission and Distribution domains. It generates power to meet local demand and stores excessive power. It also transmits power from outside when there is not enough local generation and storage. This way, we simplify the seven domains to three large domains or utilities. The smart meter (SM) in the Customer domain is responsible for information exchange with the PGO and for scheduling the electrical appliances on the user side. The information flows are carried through a communications network infrastructure, such as a wireless network or a powerline communication system [39, 44, 108]. With both energy and communication

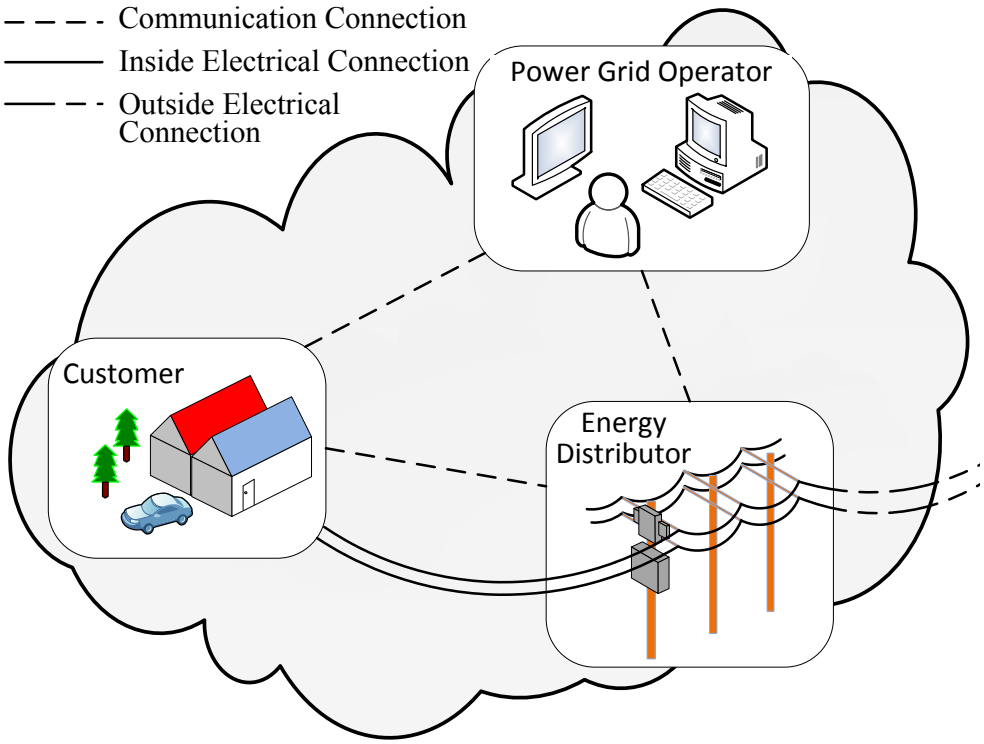

Figure 3.1: Illustration of the key domains in the smart grid of NIST standard.

connections among the domains, the PGO can exchange information with the Customer and ED and thus it controls the energy operation of the entire area.

Demand side management is one of the most important problems in smart grid research, which aims to match electricity demand to supply for enhanced energy efficiency and demand profile while considering user utility, cost and price [44]. Researchers have been focusing on peak shift or peak reduction for reducing grid deployment and operational cost [88, 89], as well as user or energy provider's cost [90, 91]. In particular, some prior works have jointly considered both user and energy provider costs, to maximize users' utility while keeping energy provider's cost at a lower level [93]. Furthermore, privacy is also emphasized in demand side management in practice. Some researches investigate the privacy problem in the smart grid from many aspects and show that an individual's daily life can even be reconstructed with collection of data on power usage [109, 110].

Given the wide range of smart grid models and the challenge in characterizing the electricity demand and supply processes and the utility/cost/pricing functions, a general model

that can accommodate various application scenarios would be highly desirable. Furthermore, it is important to jointly consider the utilities and costs of the key components of the system to achieve optimized performance for the overall smart grid system. For optimizing the performance of such a complex power system, the utilities and costs of the three key components, i.e., Customer, PGO and ED, should be jointly considered.

In this chapter, we take a holistic approach to incorporate the key design factors including Customer's utility, grid load smoothing, and energy provisioning cost in a problem formulation. To solve the real-time energy distribution problem, we first introduce a centralized offline solution and then a centralized online algorithm from Chapter 2 and our prior publication [74], which is variance sensitive without requiring any future information of the system. Furthermore, we propose a distributed online algorithm, which firstly decomposes the master problem into several subproblems and then solves them locally at each user and the PGO with the online approach. We also investigate a communications protocol to facilitate the information exchange for the iterative distributed online algorithm, which can be built on existing or emerging smart grid communication standards [39, 108].

The proposed framework is quite general. It does not require any specific models for the electricity demand and supply processes, and only has some mild assumptions on the utility and cost functions (e.g., convex and differentiable). The proposed algorithm can thus be applied to many different scenarios. It inherits the advantages of online algorithms that requires no future information for a convergent solution, and the advantages of distributed algorithms, which solves the problem in a distributed manner with local information. Although user power usages are still exchanged with the PGO, the distributed online algorithm mitigates the privacy problem since it does not require disclosure of user's utility function and its parameters. The proposed algorithm is easy to be implemented in a real smart grid system. The distributed computation allows scalability for handling large systems. The distributed online algorithm inherits the variance sensitive nature from the online algorithm, while converging to the offline optimal solution almost surely, a highly desirable property.

The proposed algorithm is evaluated with trace-driven simulation using energy consumption traces recorded in the field. It outperforms a benchmark scheme that is also distributed online but with no control for grid load smoothing.

The remainder of this chapter is organized as follows. We present the system model in Section 3.2. The problem formulation with both centralized offline and online solutions are introduced in Section 3.3. The distributed online algorithm is developed and analyzed in Section 3.4. The communications protocol is discussed in Section 3.5. We present the simulation studies in Section 3.6 and review related work in Section 3.7. Section 3.8 concludes this chapter.

3.2 System Model

3.2.1 Network Structure

We consider a power distribution system in the smart grid environment where the PGO supports the power usage of all users in the Customer domain. The users could be residential, commercial and industrial energy consumers. Each user deploys an SM to monitor and control the energy consumption of the electrical appliances [44]. All SMs are connected to the PGO through the information infrastructure such as a wireless or wireline local area network. During each distribution time cycle, SMs and PGO exchange status and control information to maximize users' utility, to minimize the PGO's generating cost, and to smooth the total power variance. The ED then transmits and distributes electricity to the users accordingly.

The relevant time period for the operation is divided into T time slots, indexed by $t \in \mathbb{T} = \{1, 2, \cdots, T\}$ and $\mathbb{T}$ is the set of all the time slots. Usually, the operation time period is a one-day cycle based on the daily periodic nature of electricity usage, while the time slot duration could be 1 hour, 0.5 hour, or 15 minutes, etc., according to users' power demand pattern/timescale in consideration of varying demand in different time of the day, as well as the amount of users in an area in consideration of communications cost.

Table 3.1: Notation Table for Chapter 3

Symbol	*Description*
$\mathbb{N}$	set of electricity users in the system
$\mathbb{P}$	set of feasible powers for users in a time slot
$\mathbb{U}$	set of user utility functions
N	number of users in the system
T	total number of time slots in offline problem
$\mathbf{P}$	power usage by the N users from time $1:T$, offline
$\vec{P_i}$	power usage by user i from time $1:T$, offline
$P_i(t)$	power usage by user i at time t, offline
$P_i^*(t)$	optimal power distribution for user i at time t, offline
$\vec{p}(t)$	power usage by the N users at time t, online
$p_i(t)$	power usage by user i at time t, online
p_i	power usage by user i at a fixed time, online
$p_i^*(t)$	short term for $p_i^*(\hat{\vec{p}}, c(t))$
$\lambda(t)$	Lagrange multipliers associated with the online problem
$\lambda^*(t)$	Lagrange multipliers associated with the online optimal solution
$p_{i,min}(t)$	minimum power demand for user i at time t
$\omega_i(t)$	flexibility level of user i at time t
$g(t)$	total power that need to be generated at time t
$g_{min}(t)$	minimum total power that need to be generated at time t
$g_{max}(t)$	maximum total power that need to be generated at time t
$c(t)$	maximum cost for the energy provider at time t
$L^T(\cdot)$	Lagrange function for the offline problem
$S_i^T(\cdot)$	distributed subproblem for users, offline
$R^T(\cdot)$	distributed subproblem for the PGO, offline
$D^T(\cdot)$	dual problem for the offline problem
$\lambda^T(t)$	Lagrange multipliers associated with the offline problem
$L(\cdot)$	Lagrange function for the online problem
$S_i(\cdot)$	distributed subproblem for users, online
$R(\cdot)$	distributed subproblem for the PGO, online
$D(\cdot)$	dual problem for the online problem
$\lambda_t(k)$	k-th update for $\lambda(t)$
$g_t(k)$	k-th solution to the distributed subproblem of the PGO
$p_{i,t}^*(k)$	k-th solution to the distributed subproblem of user i
δ	step-size for updating $\lambda_t(k)$
$U(\cdot)$	user utility function
$C(\cdot)$	cost function of energy provider
$Var(\cdot)$	variance function

We denote the power consumption of user i at time slot t as $p_i(t)$ and denote the set of all users as $\mathbb{N} = \{1, 2, \cdots, N\}$. We also define a set $\mathbb{P}$ of energy consumption at each time slot t for each user as

$$\mathbb{P} = [p_{i,min}(t), p_{i,max}(t)], \text{ for all } t \in \mathbb{T}, i \in \mathbb{N}, \tag{3.1}$$

where $p_{i,min}(t)$ is the minimum power demand and $p_{i,max}(t)$ the maximum power demand of user i at time t, as the users are assumed to be rational. That is, $\mathbb{P}$ includes all the possible value of power requested and consumed, that is, $p_i(t) \in \mathbb{P}$, for all i and t. It is noted that $\mathbb{P}$ is defined to be a nonnegative set, because even today, few users are able to generate enough power for themselves in a short time.

3.2.2 User Utility Function

We assume that each user behaves independently in the power grid. They have their own preferences and time schedules for using different electrical appliances. For instance, users may set their air conditioner at different temperatures and different users may use their washer and dryer at different times of the day. Also, the user demand may vary as weather condition changes. Usually the power consumption is larger in a hot summer (or a cold winter) day than that in a mild day in the spring (or autumn). Furthermore, different users may have different reactions to different pricing schemes [91]. Therefore, it is non-trivial to characterize the diverse user preference with a precise mathematical model.

In prior work, user preference is usually represented by a *utility function* [90]. Similarly, we adopt a function $U(p_i(t), \omega_i(t))$ to represent user i's satisfaction on power consumption in this chapter. Here $U(\cdot)$ is a general, strictly increasing, concave function of the allocated power $p_i(t)$, although the quadratic utility function is also popular in the literature [90,91,93]. The other parameter $\omega_i(t)$ of the utility function indicates user i's level of flexibility at time t. It is a "sorting" parameter for users and thus, can be normalized to be within the interval

$[0, 1]$ [111]. A larger $\omega_i(t)$ indicates a higher level of flexibility or level of power consumption. For example, a user with $\omega_i(t)$ close to 1 will probably consume more power than others. Different users can have different $\omega_i(t)$, and $\omega_i(t)$ can vary over time.

In a centralized scheduling scheme, the PGO will require the $\omega_i(t)$'s from all users in every updating interval. The user utility function and preference are private information, which can be used possibly to reconstruct many aspects of users' daily life and infringe their privacy [109, 110]. Information about a user's utility function and its parameters should be protected. To this end, a distributed algorithm that does not require exchanging privacy information would be appealing.

3.2.3 Energy Provisioning Cost Function

For ED, when demand is in the normal level, the generation cost increases only slowly as the demand grows. However, it will cost much more when the load peak is approaching the grid capacity, because PGO has to ask ED to transmit more power from outside to avoid a blackout, which incurs considerable power loss on the transmission line. Therefore, we could use a general increasing and strictly convex function to approximate the *cost function* for energy provisioning.

Similar to [91, 93], we choose a quadratic function to model the ED's cost, as

$$C(g(t)) = a \cdot g^2(t) + b \cdot g(t) + c, \tag{3.2}$$

where $a > 0$ and $b, c \geq 0$ are pre-selected for the power grid and $g(t)$ denotes the total amount of electricity generated by the ED at time slot t. ED has to provide sufficient power for users while reducing its cost.

In addition, we assume a maximum generating capacity $g_{max}(t)$ for ED at time slot t. Thus, we have the following constraint for $g(t)$:

$$\sum_{i \in \mathbb{N}} p_i(t) \leq g(t) \leq g_{max}(t), \text{ for all } t \in \mathbb{T}. \tag{3.3}$$

The constraint indicates that $g(t) \in \mathbb{G} = [g_{min}(t), g_{max}(t)]$, where $g_{min}(t) = \sum_{i \in \mathbb{N}} p_i(t)$ and $\mathbb{G}$ is a closed positive set. Because the cost function $C(\cdot)$ is strictly convex and increasing, $C(\cdot)$ is reversible so that the energy provisioning cost $C(g(t))$ is also bounded in a closed set, i.e., $C(g(t)) \in \mathbb{C}$ for all t. In other words, by adjusting the amount of power generation, the ED can control its provisioning cost.

3.3 Problem Formulation and Centralized Solutions

In this section, we summarize the problem formulation and the centralized offline and online algorithms presented in Chapter 2, for the sake of completeness. The proposed distributed online algorithm will be presented in Section 3.4 and evaluated in Section 3.6.

3.3.1 Problem Formulation

We take into account three core parts in the smart grid environment: Customer, ED and PGO in the model. Under certain constraints, we aim to achieve the triple goals of (i) maximizing users' utility, (ii) minimizing ED's cost, and (iii) smoothing the total power load of the grid.

We first consider an offline scenario where the PGO has global information on users' flexibility $\omega_i(t)$ and ED's total generated power $g(t)$ for the entire period (i.e., future information is known). Let $P_i(t)$ denote the power usage for user i at time t, for $t \in \mathbb{T}$. We use upper case P in the *offline problem*. In the corresponding *online problem*, which will be examined in Section 3.3.3, we use lower case p for the corresponding variables. A vector with subscript i is used to denote a time sequence, e.g., $\vec{P}_i$ for the power usage by user

i for $t \in \mathbb{T}$. The offline problem (termed Prob-OFF) can be formulated as follows. For $P_i(t) \in \mathbb{P}, g(t) \in \mathbb{G}$, for all $i \in \mathbb{N}, t \in \mathbb{T}$, we have the offline problem Prob-OFF as

$$\text{maximize: } \Theta(\vec{P}_1, \cdots, \vec{P}_N) = \sum_{t\in\mathbb{T}} \left[\sum_{i\in\mathbb{N}} U(P_i(t), \omega_i(t)) - C(g(t)) \right] - \frac{\alpha T}{2} \text{Var}\left(\sum_{i\in\mathbb{N}} \vec{P}_i \right) \tag{3.4}$$

$$\text{subject to: } \sum_{i\in\mathbb{N}} P_i(t) \leq g(t), \text{for all } t \in \mathbb{T}, \tag{3.5}$$

where $\text{Var}(\cdot)$ is the variance function defined as

$$\text{Var}\left(\sum_{i\in\mathbb{N}} \vec{P}_i \right) = \frac{1}{T} \sum_{t\in\mathbb{T}} \left(\sum_{i\in\mathbb{N}} P_i(t) - \frac{1}{T} \sum_{k\in\mathbb{T}} \sum_{i\in\mathbb{N}} P_i(k) \right)^2 .$$

The objective function (3.4) consists of three parts. The first part represents users' satisfaction and preference. The second part represents ED's cost for energy provisioning. The third part represents the load variance of the grid. It is integrated with a parameter $\alpha > 0$, to enable a trade-off between the grid and users' benefits. All the users' demand and generating power should be included in the set $\mathbb{P}$ and $\mathbb{G}$ as we have discussed in Sections 3.2.1 and 3.2.3.

3.3.2 Centralized Offline Algorithm

In Problem Prob-OFF (3.4), the user power consumption $P_i(t)$'s are independent. Hence the grid load variance term can be rewritten as $\text{Var}\left(\sum_{i\in\mathbb{N}} \vec{P}_i\right) = \sum_{i\in\mathbb{N}} \text{Var}\left(\vec{P}_i\right)$. It can be verified that Prob-OFF is a convex optimization problem because function $U(\cdot)$ is concave and $C(\cdot)$ and $\text{Var}(\cdot)$ are both convex. Also due to convexity of the variance function $\text{Var}(\cdot)$ we can show that Prob-OFF has a unique solution. If we carefully define sets $\mathbb{P}$ and $\mathbb{G}$, the Slater's condition can be satisfied as well, which indicates that the KKT conditions are sufficient and necessary for the optimality of Prob-OFF [98]. By solving the KKT conditions, we can derive the optimal energy allocation for each of the users at each time slot.

In Prob-OFF, all information are assumed to be known a priori. Because of this, its solution is optimal. However, since it requires future information for computing the grid load variance (i.e., the third part in (3.4)), we cannot solve the KKT conditions at each time slot in practice.

3.3.3 Centralized Online Algorithm

We now present the online algorithm for energy distribution, and show the main result that the online solution is asymptotically convergent to the offline optimal solution, i.e., *asymptotically optimal*. The online energy distribution algorithm consists of the following three steps.

Algorithm 3.1: Centralized Online Algorithm

Step 1: For each $i \in \mathbb{N}$, initialize $\hat{p}_i(0) \in \mathbb{P}$.

Step 2: In each time slot t, the PGO solves the following convex optimization problem (termed Prob-ON). For $p_i(t) \in \mathbb{P}, g(t) \in \mathbb{G}$, for all $i \in \mathbb{N}$,

$$\begin{aligned} \text{maximize:} \quad & \sum_{i \in \mathbb{N}} U(p_i(t), \omega_i(t)) - C(g(t)) - \\ & \frac{\alpha}{2} \sum_{i \in \mathbb{N}} (p_i(t) - \hat{p}_i(t-1))^2 \qquad (3.6) \\ \text{subject to:} \quad & \sum_{i \in \mathbb{N}} p_i(t) \leq g(t), \text{for all } t \in \mathbb{T}. \qquad (3.7) \end{aligned}$$

Let $\vec{p}^*(t)$ denote the solution to Prob-ON, where each element $p_i^*(t)$ represents the optimal power allocation to user i.

Step 3: Update $\hat{p}_i(t)$ for all $i \in \mathbb{N}$ as follows and go to Step 2.

$$\hat{p}_i(t) = \hat{p}_i(t-1) + \frac{\alpha}{t+\alpha} \cdot (p_i^*(t) - \hat{p}_i(t-1)). \qquad (3.8)$$

Comparing to (3.4), the variance term is approximated by $\sum_{i\in\mathbb{N}}(p_i-\hat{p}_i(t-1))^2$ in (3.6). Similar to problem Prob-OFF, Prob-ON is also a convex optimization problem satisfying Slater's condition. The KKT conditions can be derived as follows.

$$\begin{cases} U'(p_i^*(t),\omega_i(t)) - \alpha\left(p_i^*(t) - \hat{p}_i(t-1)\right) - \lambda^*(t) = 0 \\ -C'(g(t)) + \lambda^*(t) = 0 \\ \lambda^*(t)\left(\sum_{i\in\mathbb{N}} p_i^*(t)/g(t) - 1\right) = 0 \\ \lambda^*(t) \geq 0, \quad \forall\, t. \end{cases} \tag{3.9}$$

where $\lambda^*(t)$ is the Lagrange multiplier. In (3.9), only information for time slot t is needed to solve the equations. This allows us to solve the problem in each time slot without needing any future information. The following theorem states that the offline solution converges to the optimal Prob-OFF solution, which is obtained assuming all future information is available.

Theorem 3.1. *The centralized online optimal solution converges asymptotically and almost surely to the centralized offline optimal solution.*

Although the formulation in Chapter 2 is slightly different with our problem in this chapter, the conditions of the theorem are still satisfied in our model. Therefore, the theorem still holds true. It presents a strong result, based on which we could solve Prob-ON instead of Prob-OFF but with an equally good result.

However, Prob-ON is still solved in a centralized manner, which means that at each time slot, PGO still requires the accurate utility functions of all users with their preference parameters $\omega_i(t)$, which are important user privacy information. It will be appealing to develop a distributed algorithm that can preserve user privacy, but still achieve the optimal performance. The distributed online algorithm will also provide scalability and have low control and communication overhead.

3.4 Distributed Online Algorithm

In this section, we firstly decompose problem Prob-OFF in a distributed manner so that the PGO and every user can solve the subproblems independently without requiring global information. We then present a distributed offline algorithm for the decomposed problem. Finally, we show that the distributed offline problem can also be solved with an online approach, and the distributed online solution is asymptotically convergent to that of the centralized offline problem. Therefore we can eliminate the need to share users' utility functions and their parameters.

3.4.1 Decomposition and Distributed Offline Algorithm

Firstly, the offline objective function (3.4) can be rewritten as

$$\Theta = \sum_{t=1}^{T} \left[\sum_{i \in \mathbb{N}} (U(P_i(t), \omega_i(t)) - \frac{\alpha}{2} \left(P_i(t) - \frac{1}{T} \sum_{k=1}^{T} P_i(k)^2 \right) \right) - C(g(t)) \right], \tag{3.10}$$

where the first two terms are functions of $P_i(t)$ and $\omega_i(t)$ (i.e., information available at user i) and the third term is a function of the total load $g(t)$ (i.e., information available at the PGO). However, we cannot decompose the problem in this simple way, because constraint (3.5) involves both user information $P_i(t)$ and PGO information $g(t)$. Note that the superscript (e.g., $L^T(\cdot)$ or $\lambda^T(\cdot)$) indicates the functions and Lagrange multiplier of the distributed offline problem. The superscript is removed for the corresponding functions and Lagrange multiplier of the distributed online problem.

To decompose the problem, we first derive the Lagrangian for Prob-OFF as

$$
\begin{aligned}
& L^T(\vec{P}(t), g(t), \lambda^T(t)) \\
= & \sum_{t=1}^{T} \left[\sum_{i \in \mathbb{N}} \left(U(P_i(t), \omega_i(t)) - \frac{\alpha}{2} \left(P_i(t) - \frac{1}{T} \sum_{k=1}^{T} P_i(k) \right)^2 \right) \right. \\
& \left. - C(g(t)) - \lambda^T(t) \left(\sum_{i \in \mathbb{N}} P_i(t) - g(t) \right) \right] \\
= & \sum_{t=1}^{T} \left[\sum_{i \in \mathbb{N}} \left(U(P_i(t), \omega_i(t)) - \frac{\alpha}{2} \left(P_i(t) - \frac{1}{T} \sum_{k=1}^{T} P_i(k) \right)^2 \right. \right. \\
& \left. \left. - \lambda^T(t) P_i(t) \right) \right] + \sum_{t=1}^{T} \left[\lambda^T(t) g(t) - C(g(t)) \right], \qquad (3.11)
\end{aligned}
$$

where $\lambda^T(t)$ is the Lagrange multiplier. In (3.11), functions of $P_i(t)$ and $g(t)$ are decoupled. For each $P_i(t) \in \mathbb{P}$, define

$$
\begin{aligned}
S_i^T(\lambda^T(t)) = \max \left\{ \sum_{t=1}^{T} \left[U(P_i(t), \omega_i(t)) - \right. \right. \qquad (3.12) \\
\left. \left. \frac{\alpha}{2} \left(P_i(t) - \frac{1}{T} \sum_{k=1}^{T} P_i(k) \right)^2 - \lambda^T(t) P_i(t) \right] \right\}.
\end{aligned}
$$

For $g(t) \in \mathbb{G}$, define

$$
R^T(\lambda^T(t)) = \max \left\{ \sum_{t=1}^{T} \left[\lambda^T(t) g(t) - C(g(t)) \right] \right\}. \qquad (3.13)
$$

We can reformulate problem Prob-OFF to the Lagrange dual problem as follows [98].

$$
\begin{aligned}
\text{minimize: } & D^T(\lambda^T(t)) \qquad (3.14) \\
\text{subject to: } & \lambda^T(t) \geq 0,
\end{aligned}
$$

where

$$D^T(\lambda^T(t)) = \max\left\{L^T(\vec{P}(t), g(t), \lambda^T(t))\right\}$$
$$= \sum_{i\in\mathbb{N}} S_i^T(\lambda^T(t)) + R^T(\lambda^T(t)). \tag{3.15}$$

This way, problem Prob-OFF is decomposed into two parts: (i) the first one is an optimization problem $S_i^T(\lambda^T(t))$ defined in (3.12) for each user to solve, and (ii) the other one is also an optimization problem $R^T(\lambda^T(t))$ defined in (3.14) for the PGO to solve. Since they are both concave and have linear constraints, *strong duality* holds for careful selections of $P_i(t)$ and $g(t)$, which guarantees the zero gap between Prob-OFF and the dual problem $D^T(\lambda^T(t))$.

3.4.2 Distributed Online Subproblem

Although we can apply several methods from convex optimization to solve the problems of (3.12) and (3.14) in a distributed way, such an approach is still not practical because the offline problem and solution require future information to be known a prior. We next develop an online distributed algorithm to further eliminate such need for future information.

Observe that in (3.11), the only term that needs future information other than that at time t is $\frac{1}{T}\sum_{k=1}^{T} P_i(k)$, i.e., the average of $P_i(t)$ over $\mathbb{T}$, which is also a term in subproblem (3.12) for users. Therefore, if the average of $P_i(t)$ can be revealed with accumulated historic information, we will be able to solve (3.14) in an online manner. Similar to the idea of transforming problem Prob-OFF into Prob-ON, we use $\hat{p}_i(t-1)$ to approximate the average in the distributed online algorithm and show that the solution obtained this way is still asymptotically optimal.

We first present the distributed online subproblems by rewriting the distributed offline optimization problems for users and PGO according to (3.12), (3.13), (3.14) and (3.15). At

each time slot t, for each user i, define

$$S_i(\lambda(t)) = \max\left\{U(p_i(t), \omega_i(t)) - \frac{\alpha}{2}(p_i(t) - \hat{p}_i(t-1))^2 - \lambda(t)p_i(t)\right\}. \tag{3.16}$$

$$R(\lambda(t)) = \max\left\{\lambda(t)g(t) - C(g(t))\right\}. \tag{3.17}$$

And the objective function for $\lambda(t)$ is

$$\begin{aligned} &\text{minimize: } D(\lambda(t)) \\ &\text{subject to: } \lambda(t) \geq 0, \end{aligned} \tag{3.18}$$

where

$$D(\lambda(t)) = \sum_{i\in\mathbb{N}} S_i(\lambda(t)) + R(\lambda(t)). \tag{3.19}$$

This way, we derive the distributed online subproblems for users and the PGO to solve. Note that the dual decomposition is only able to decompose the online problem and we still need to show that the distributed online problem is optimal and convergent. The following theorem shows that the distributed online subproblems can be solved and the solutions are asymptotically optimal.

Theorem 3.2. *The optimal solution to the distributed online subproblems converges asymptotically and almost surely to the offline optimal solution.*

The proof of Theorem 3.2 is shown in Appendix B.1. It clarifies the relationship between problems Prob-OFF, Prob-ON and the distributed online subproblems. Actually, we can also achieve the distributed online decomposition from Prob-ON by dual decomposition as we did for Prob-OFF. Theorem 3.2 also presents an effective means of solving the online distribution problem in a practical manner. We next present the distributed online algorithm.

3.4.3 Distributed Online Algorithm

Following Theorem 3.2, we can solve the dual problem (3.18) to acquire the optimal online solution. Because of constraint (3.7), $S_i(\lambda(t))$ and $R(\lambda(t))$ are coupled by the Lagrange multiplier $\lambda(t)$; $\lambda(t)$ is associated with both the user utility maximization problem (3.16) and the ED cost minimization problem (3.17). As the dual variable, it is also a key parameter for solving the dual problem.

In our case, the dual function $D(\lambda(t))$ is differentiable. So we can apply the following gradient method to acquire the dual variable $\lambda(t)$ at each time slot t [112].

$$\lambda_t(k+1) = \left[\lambda_t(k) - \delta\left(g_t(k) - \sum_{i\in\mathbb{N}} p^*_{i,t}(k)\right)\right]^+, \tag{3.20}$$

where δ is the step-size; $[\cdot]^+$ is the projection onto the nonnegative orthant; $\lambda_t(k)$ is the k-th update of $\lambda(t)$; $g_t(k)$ and $p^*_{i,t}(k)$ are the solutions to (3.16) and (3.17), respectively.

At each time slot t, this method requires that PGO and the users exchange $\lambda_t(k)$ and $p^*_{i,t}(k)$ for a number of times to obtain the convergent $\lambda(t)$, the power that will be generated $g(t)$, and the energy $p_i(t)$ allocated to each user i. We then present the distributed online algorithm, Algorithm 2, to solve the dual problem (3.18) as well as problem Prob-ON. The algorithm consists of two parts:

- a three-step Algorithm 2.a for all users;
- a three-step Algorithm 2.b executed by the PGO.

Algorithm 3.2.a: Distributed Online Algorithm for Users

Step 1: For each user $i \in \mathbb{N}$, initialize $\hat{p}_i(0) \in \mathbb{P}$.

Step 2: In time slot t, the SM of each user does the following:

1) Receives the updated $\lambda_t(k)$ from the PGO;

2) Solves problem (3.16) for user utility maximization;

3) Transmits the solution $p^*_{i,t}(k)$ to the PGO for energy demand;

4) Repeats 1) to 3) until $|\lambda_t(k+1) - \lambda_t(k)| < \epsilon$, where $\epsilon > 0$.

Step 3: Update $\hat{p}_i(t)$ for all $i \in \mathbb{N}$ as (3.8).

Algorithm 3.2.b: Distributed Online Algorithm for the PGO

Step 1: For each $i \in \mathbb{N}$, initialize $p^*_{i,t}(0) \in \mathbb{P}$. Choose an arbitrary $\lambda_t(0) \geq 0$.

Step 2: In each time slot t, the PGO does the following:

1) Solves problem (3.17) to obtain $g_t(k)$;

2) Receives $p^*_{i,t}(k)$ from all the users;

3) Updates the value of $\lambda_t(k)$ using (3.20) and broadcasts it to the users;

4) Repeats 1) to 3) until $|\lambda_t(k+1) - \lambda_t(k)| < \epsilon$, where $\epsilon > 0$.

Step 3: Sends $g(t)$ to ED for energy generation for time slot t and distributes $p^*_i(t)$ to user i, for all $i \in \mathbb{N}$.

Note that for each time t, we have a terminating condition that $|\lambda_t(k+1) - \lambda_t(k)| < \epsilon$ for the inner loop, where ϵ is a positive real number small enough to indicate the convergence of $\lambda_t(k)$. A smaller ϵ will produce a more precise $\lambda(t)$. But the computation will also take more time. The other factor affecting the convergence of $\lambda_t(k)$ is the step-size δ in (3.20). For the gradient method, a small δ guarantees the convergence of $\lambda_t(k)$ but may require more iterations. In fact, the terminating condition could be rewritten as

$$\left| \delta \left(g_t(k) - \sum_{i \in \mathbb{N}} p^*_{i,t}(k) \right) \right| < \epsilon.$$

Therefore, δ and ϵ should be carefully selected for Algorithm 2 to achieve fast convergence within one time slot. This is especially important for large scale systems with a large population of users in the Customer domain, where the information exchanged increases fast for more users. However, we conjecture that the communications will not be a big issue

under today's advanced wired and wireless communication infrastructure. We will evaluate the effect of δ on the convergence of $\lambda_t(k)$ in Section 3.6.2.

In Algorithm 2, we see an interaction between users and the PGO realized by the dual variable $\lambda(t)$. It not only is the necessary parameter to solve both (3.16) and (3.17), but also connects users and the PGO decisions. The PGO has no information about user utilities, while $\lambda(t)$ instead conveys information from users to the PGO. By updating $\lambda(t)$ as in (3.20), the new value contains new information from both users and the PGO. Thus, by using Algorithm 2, the online problem can be solved in a distributed fashion with comparable optimality to the centralized online algorithm. Furthermore, from Theorems 3.1 and 3.2, the distribution solution from Algorithm 2 will also converge asymptotically to the offline optimal solution.

It is worth noting that no information on user utility and preference parameter is transmitted between the users and the PGO. Consider practical data communication networks for the smart grid, less transmitted data brings about higher security, reliability and sufficiency. This also helps simplify the communication protocol designs for the grid. Furthermore, the computational load is offloaded from the PGO to the SMs at each user's site; the computation at the PGO is greatly simplified, leading to resource and time savings so that a larger number of users can be supported. In conclusion, the distributed online Algorithm 2 could be useful in practice.

3.5 Communication Network Protocol

Information exchange is an important element of the emerging smart grid. Communications between SMs and the PGO are essential for both control and distribution [39,108]. The distributed online algorithm is also based on such information exchanges. As more advances are made in smart grid, there is a compelling need for network architectures, standards, and protocols for communications in smart grid. We hereby introduces a basic protocol for communications network support in the smart grid for the proposed distributed online algorithm,

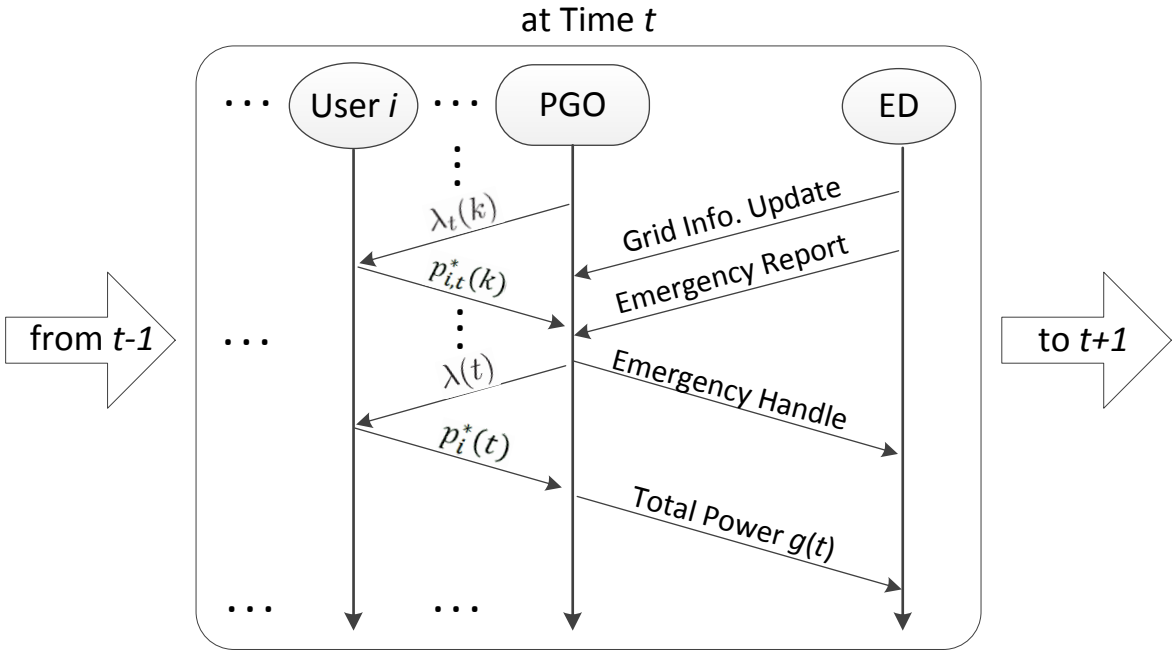

Figure 3.2: Information flows in the smart grid of NIST standard.

which is simple but sufficient to support the real time online power distribution algorithm and can be built upon existing or future smart grid communication standards [39, 108].

In the distributed online energy distribution algorithm, the users and the PGO needs to exchange $\lambda_t(k)$ and $p^*_{i,t}(k)$ several times at each time t, to achieve a satisfactory $p^*_i(t)$ for users and $g(t)$ for the ED. And the ED should update periodically the grid information to the PGO and the emergency report should be timely. The PGO also collects other information from the ED, such as the actual grid load. After the distributed online algorithm is executed, the users obtains their own power consumption and the PGO sends the total energy usage to the ED. Moreover, the PGO is able to send other control information to the EP or users for, e.g., regulation, accounting, emergency response and alerts, etc.

Fig. 3.2 illustrates the information flows in the network system, where we have three large entities in the system: the PGO is the core controller and Users and the ED are also important participants. Fig. 3.2 illustrates the communications at time t. For other time slots, the communications protocols are almost the same. We take user i as example, because other users have similar interactions with the PGO.

With Algorithm 2, at each updating slot, the PGO solves the subproblem, receives $p^*_{i,t}(k)$ from the users and updates $\lambda_t(k)$ to the users; the users receive the updated $\lambda_t(k+1)$, use it to solve the distributed optimization problem for users and update the new solution to the PGO. The iteration process terminates when the terminating condition is satisfied. Then the PGO will inform the ED to transmit the power request and distribute to the users. Meanwhile, the ED updates the power grid information to the PGO and sends alarms when emergency events happen. The PGO returns corresponding commands for the ED to execute.

3.6 Performance Evaluation

3.6.1 Simulation Configuration

In this section, we evaluate the proposed distributed online algorithm (denoted as DOA in this section) with trace-driven simulations. The simulation data and parameters are acquired from the recorded power consumption in the Southern California Edison (SCE) area in 2011 [99]. We first study the performance of DOA on convergence comparing to the centralized online algorithm (termed COA in this section) described in Section 3.3.2. We then compare the distribution solutions between DOA and COA, as well as with an existing scheme as benchmark.

Consider a power distribution system in a small area with $N = 20$ users and 15-minute updating periods. For COA, the 15-minute interval is sufficient to obtain the required user information and execute the centralized optimization algorithm. The 15-minute interval is also short enough to show the users' change of demand, although with DOA, shorter time slots are also practical. We will show results within a 24-hour time pattern for an evaluation of the daily operations.

We choose users' utility function from a function set $\mathbb{U}$ in which the functions are generated as widely used quadratic expression (see [90, 91]) with $\omega_i(t) \in (0, 1)$ randomly

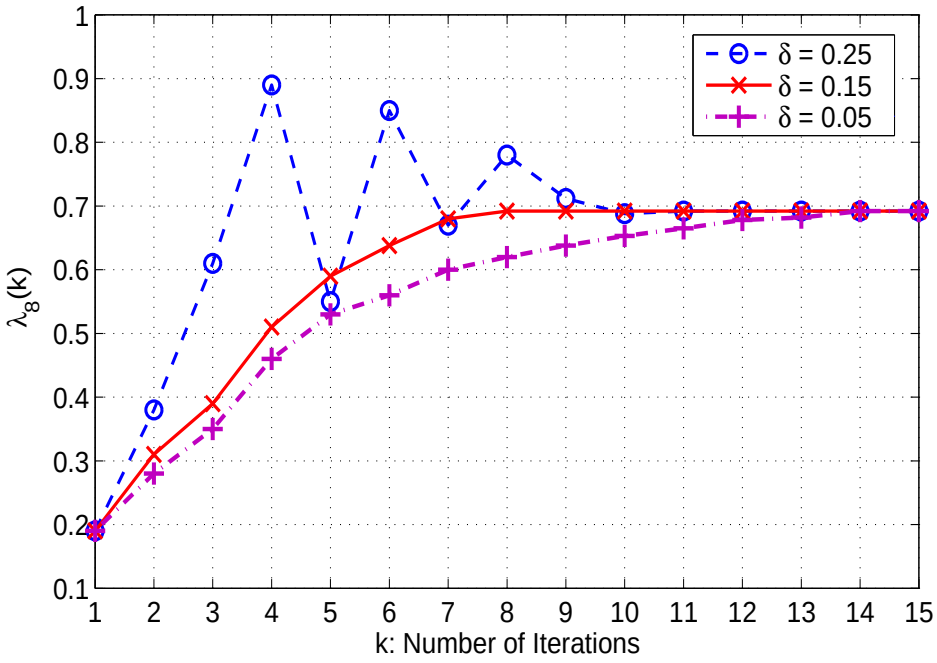

Figure 3.3: Convergence of $\lambda_8(k)$ when $\epsilon = 0.2$.

selected.

$$U(p_i(t), \omega_i(t)) = \begin{cases} \omega_i(t)p_i(t) - \dfrac{1}{8}p_i(t)^2, \\ \qquad\qquad \text{if } 0 \le p_i(t) \le 4\omega_i(t) \\ 4\omega_i(t), \quad \text{if } p_i(t) \ge 4\omega_i(t). \end{cases} \tag{3.21}$$

We also assume user's energy demand $p_i(t)$ is selected from the set of $\mathbb{P} = [1.0, 3.0]$, for all i. The maximum generating power $g_{max}(t)$ is set to the maximum total power demand of all the users, that is $g_{max}(t) = \sum_{i\in\mathbb{N}} p_{i,max}(t)$, which implies that the generating power is equal to the power demand. The initial value of $\lambda(t)$ in Algorithm 2 is picked randomly from the set $(0, 1)$ and the termination condition ϵ is chosen as 0.2. The parameters in the energy provisioning cost function (3.2) are set as $a = 0.05$, and $b = c = 0$. These parameters are carefully determined after studying the characteristics of the SCE trace. For parameter α in the updating function (3.8), we take $\alpha = 1$ in the following simulations. In 2, we have shown that $\alpha = 1$ is a proper value for fast convergence.

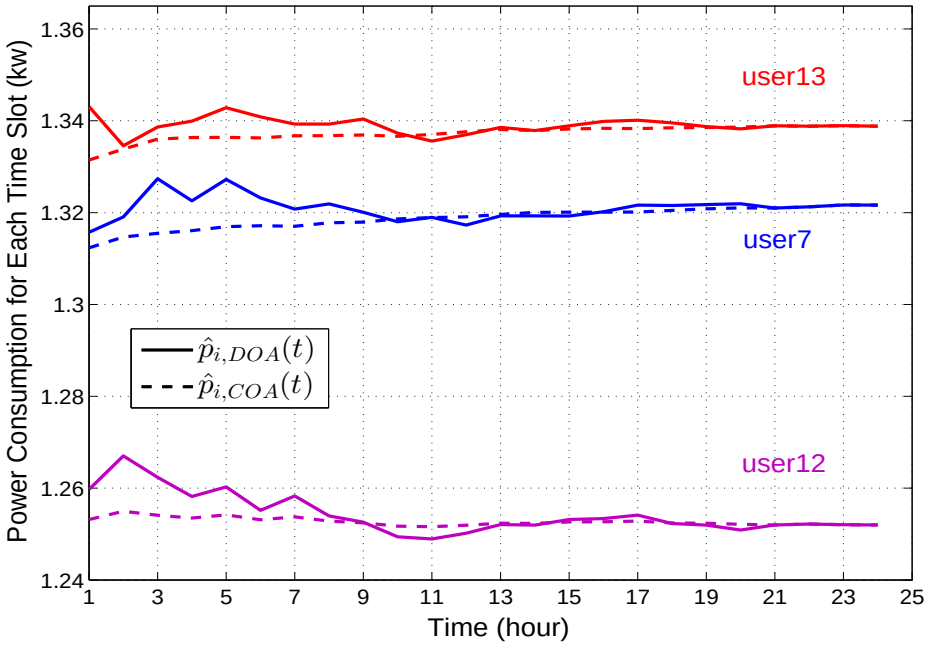

Figure 3.4: Convergence of $\hat{p}_i(t)$ for DOA and COA for users of different levels of flexibility.

3.6.2 DOA Performance Evaluation

As shown in Section 3.4, DOA is based on the convergence of $\lambda_t(k)$. We first show the convergence of $\lambda_t(k)$ in a time slot t. The gradient method applied in Algorithm 2 (see the updating function (3.20) for $\lambda_t(k)$) requires that the positive step-size δ be sufficiently small to guarantee the convergence of $\lambda_t(k)$. However, small δ may slow down the convergence. For a fixed ϵ, which indicates the same tolerance for the convergent $\lambda_t(k)$, Fig. 3.3 illustrates the evolution of $\lambda_8(k)$ as a function of k for the same user at the eighth time slot with different step-sizes δ. It is observed that the series of $\lambda_8(k)$ with larger δ of 0.25 has large perturbation than the other two series of smaller δ. Also, the the series of $\lambda_8(k)$ with the smallest δ of 0.05 has the slowest speed of convergence. Although the $\lambda_8(k)$ with δ of 0.25 converges faster than the one of 0.05, it is slower than the one of 0.15. This implies that increasing δ cannot guarantee faster convergence of $\lambda_t(k)$, because a larger δ may make $\lambda_t(k)$ not convergent. In practice, a proper δ is important for convergence and thus the efficiency of DOA. It can be decided after several simple experiments. From Fig. 3.3, we also observe

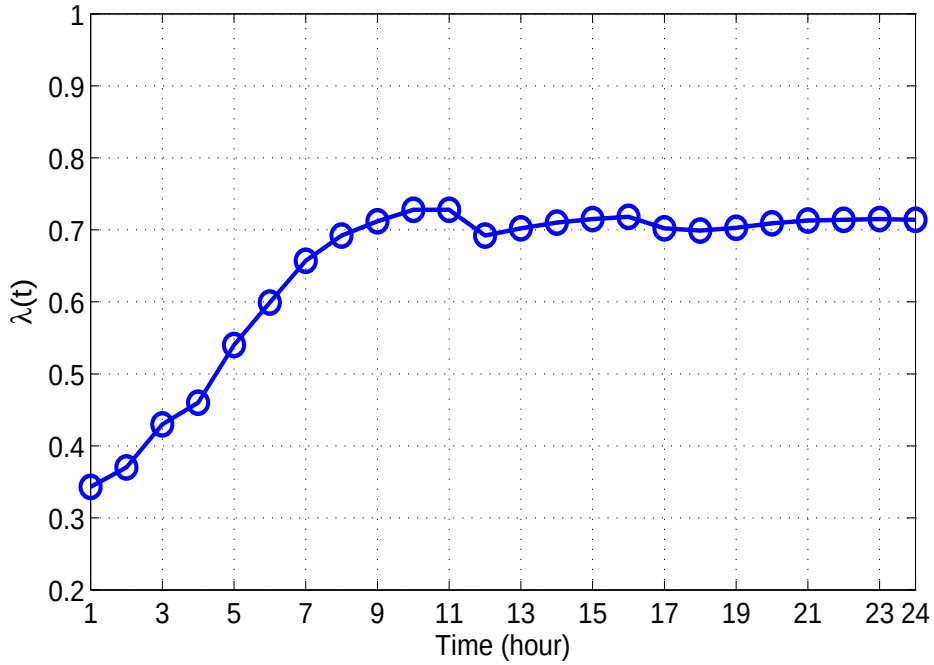

Figure 3.5: Evolution of $\lambda(t)$ for a 24 hours period.

a fast convergence of $\lambda_t(k)$ in about 10 times of information exchanges. We set δ to 0.15 in all the following simulations.

We then show the convergence of $\hat{p}_i(t)$ from both COA and DOA. $\hat{p}_i(t)$ is a key variable in the online algorithm. Its convergence indicates that the gap between the online and offline solutions becomes zero (see the updating function (3.8)). In Fig. 3.4, $\hat{p}_{i,COA}(t)$ and $\hat{p}_{i,DOA}(t)$ for three users are both convergent. For COA, we see a fairly fast convergence with a very short transient period. For DOA, it shows slower convergence with larger variance before stable values are achieved. This is because comparing to COA, DOA has another iteration function brought about by (3.20) for updating $\lambda_t(k)$. The initial value $\lambda_t(0)$ is set randomly, so it requires extra time for the convergence of $\hat{p}_{i,DOA}(t)$. Also in Fig. 3.4, we find the coincidence of two curves for the several last time slots. This can be explained by Theorem 3.2, which indicates that DOA and COA deliver identical solutions. It can be also observed in Fig. 3.4 that both algorithms achieve convergence for users of different levels of consumption, where user 13 has a $\omega(t)$ larger than that of users 7 and 12.

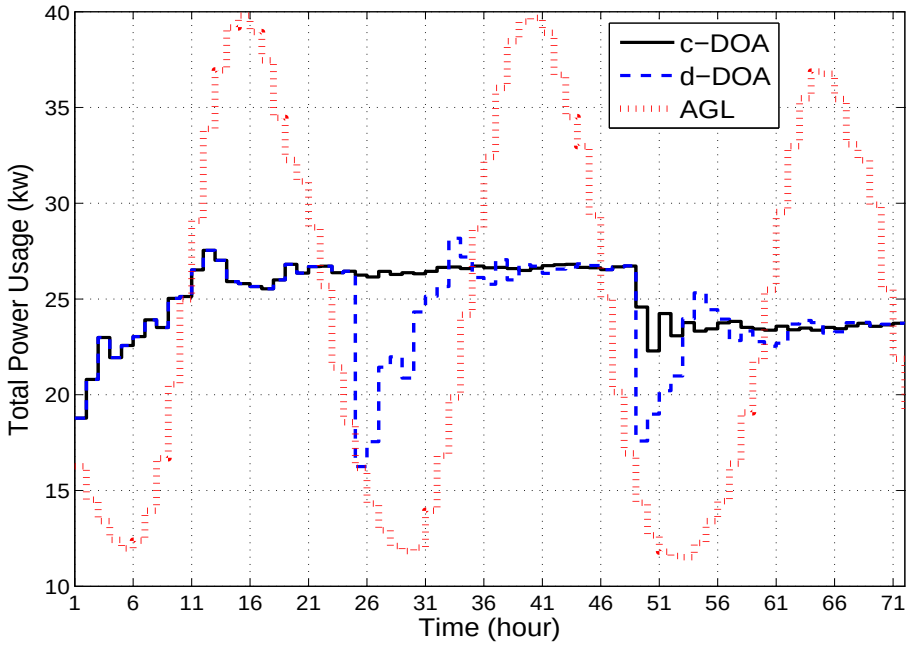

Figure 3.6: The actual grid load (AGL) and total power consumption by d-DOA and c-DOA of three consecutive days.

It is common that the power usage of users may not have large perturbations within one time slot. As discussed, $\lambda(t)$ is related to the users and the power grid. It is natural to assume that the $\lambda(t)$'s of consecutive time slots are correlated. If such a correlation could be revealed, DOA can be further improved. Therefore, we plot the variable $\lambda(t)$ in Fig. 3.5. We observe a convergent trend of $\lambda(t)$ for the 24-hour period. However, it is not clear whether it is convergent or not at this time. Because the initial value of $\lambda_t(k)$ is selected randomly, we can confirm our assumption that $\lambda(t)$ and $\lambda(t+1)$ are highly correlated. Thus, set $\lambda_t(0)$ as $\lambda(t-1)$ would reduce the iteration steps and speed up convergence in time slot t.

Furthermore, the power consumption of users and the grid load are usually closely related for consecutive days. Therefore, we can use the final results/parameters from the previous day as a starting point for the present day, which leads to a better performance. We plot the grid load of three consecutive days by applying DOA separately on *daily* basis (d-DOA) and by applying DOA *consecutively* (c-DOA), as discussed, in Fig. 3.6. For the first two days, the grid loads are almost the same. We find c-DOA achieves an obviously

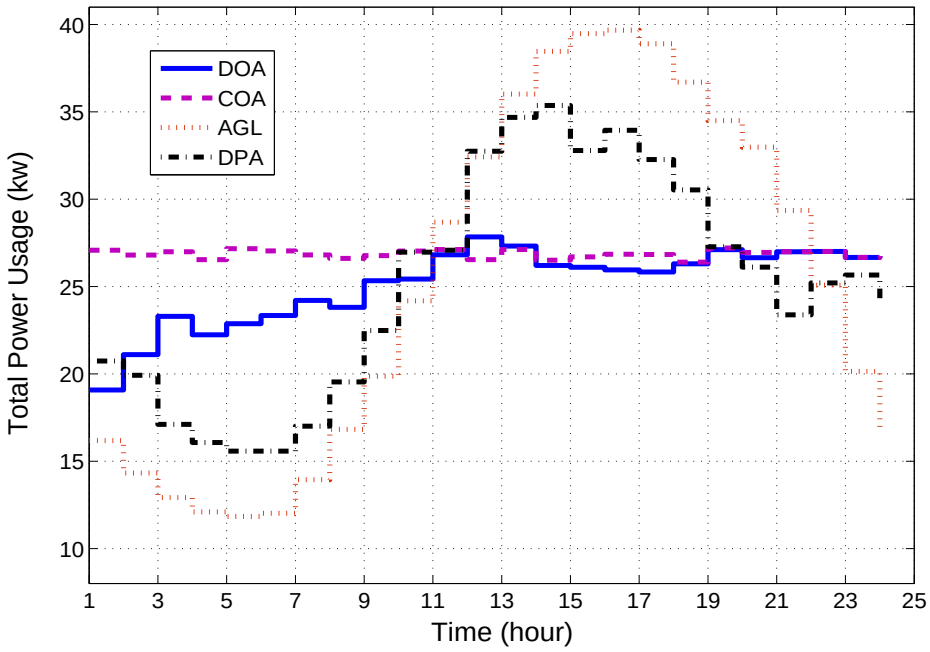

Figure 3.7: The AGL and total power consumptions achieved by DOA, COA and DPA for a hot day.

better convergence performance over d-DOA in Day Two because the initial values of Day Two are set to the final values of Day One. Although the third day has a lower grid load, c-DOA still achieves a better convergence and smoothness performance over d-DOA because the initial values for d-DOA are randomly chosen. This way, we can enhance the proposed algorithm to achieve fast convergence and reduce communication requirements. In the remaining simulations, the enhanced DOA algorithm is used whenever possible.

3.6.3 Comparison with Other Algorithms

One important benefit of DOA is the variance control it offers, which is inherited from COA. In Figs. 3.8 and 3.7, we plot the AGL and total power consumption by DOA, COA and a state-of-the-art algorithm proposed in [93], which is a dynamic pricing algorithm (DPA) based on utility maximization. DPA considers both users and the ED as we do in our chapter, but it has no consideration on the load variance. The actual grid load in Fig. 3.7 is the summation of 20 independent users' consumptions generated by the average real load

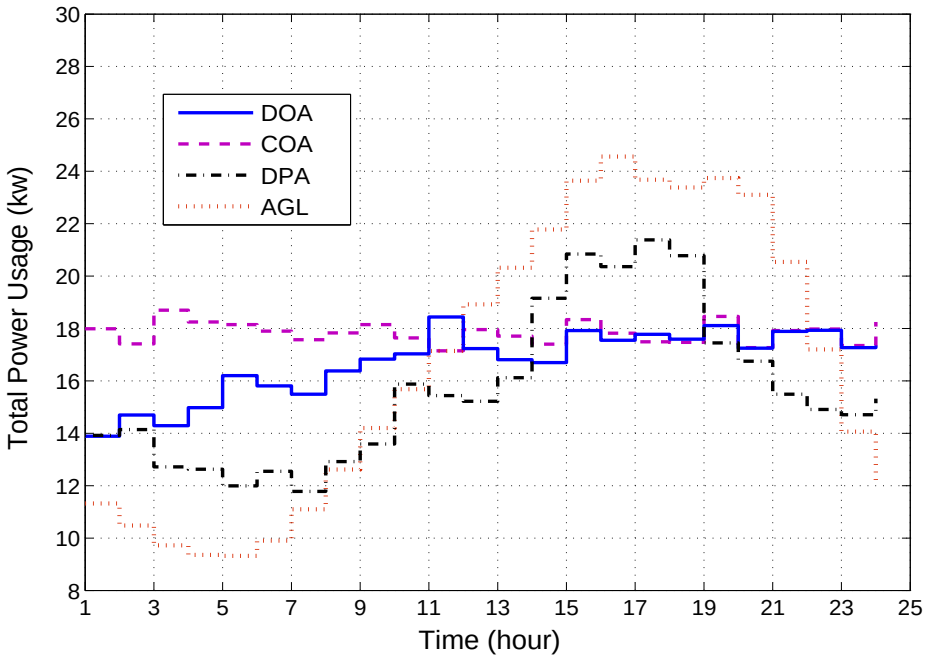

Figure 3.8: The AGL and total power consumptions achieved by DOA, COA and DPA for an average day.

in the SCE trace on a hot day (i.e., Sep. 1, 2011) [99]; while the AGL in Fig. 3.8 is based on a typical day in the same SCE trace when the grid load is the average case (i.e., Oct. 5, 2011) [99]. We show these two figures to have a direct comparison of our energy distribution algorithms.

From both Figs. 3.7 and 3.8, we first observe that DOA needs several time slots to converge to COA. This is caused by the effect of ϵ, as discussed before. On the other hand, we also observe a larger gap between the DOA and COA curves for the hot day in Fig. 3.7 than that for a typical day in Fig. 3.8. This is because the typical day has a much lower peak demand. This confirms that under average condition, DOA has very good performance, which is close to COA.

On the other hand, peak reduction is another objective of our algorithm. Peak refers to the highest point of the grid load curve and for different curves, the amount of peak reduction is represented by the normalized percentage, which is calculated as the ratio of the difference of the peak between the actual load curve and the controlled load curve, and the peak of

Table 3.2: Simulation Results of Several Performance Metrics for DOA, COA, DPA and AGL

System Size N	Algorithm	$\overline{V}$	$\overline{U}$	$c(T)$	$\overline{PK}$
200	DOA	3.3	3.49	1.62	1.39
	COA	0.02	3.52	1.69	1.35
	DPA	24.5	3.66	1.74	1.61
	AGL	53.5	3.86	1.86	1.97
500	DOA	9.2	3.51	9.54	1.41
	COA	0.05	3.54	10.1	1.37
	DPA	62.6	3.64	10.5	1.73
	AGL	113	3.88	14.0	2.27
1000	DOA	18.1	3.47	38.1	1.55
	COA	0.10	3.53	40.2	1.41
	DPA	125	3.69	42.2	1.99
	AGL	266	3.88	54.1	2.63

actual curve. We have three controlled curves here: COA, DOA and DPA. And the peak reduction percentages for COA, DOA and DPA are 29.8%, 31.7% and 10.9%, respectively, for the hot day, and 23.9%, 23.9% and 12.9%, respectively, for the typical day. We can see that DOA achieves almost the same peak reduction as COA in both cases, which are superior than DPA. Note that both DOA and COA have better performance on peak reduction of the hot day over the typical day; while DPA has the opposite result. This is because it does not consider variance reduction.

Finally, we compare several performance metrics for the three schemes (i.e., DOA, COA and DPA) together with the actual trace results (i.e., the AGL based on the worse condition in the hot day) in Table 3.2. These metrics are usually used as optimization objectives in prior work (see Section 3.7). As defined in (3.22), $\overline{V}$, $\overline{U}$ and $\overline{PK}$ denote the averages (across all users) of the grid load variance, users' utility, users' cost, and the peak of the total grid load, respectively, while $c(T)$ is the total energy provisioning cost for the entire period. The

simulation results are listed for systems with 200, 500 and 1000 users.

$$\begin{cases} \overline{V} = \frac{1}{N}\sum_{i\in\mathbb{N}} \mathrm{Var}(\vec{p}^*) \\ \overline{U} = \frac{1}{N}\sum_{t=1}^{T}\sum_{i\in\mathbb{N}} U(p_i^*(t), \omega_i(t)) \\ \overline{PK} = \frac{1}{N}\mathrm{max}_{t\in[1:T]}\sum_{i\in\mathbb{N}} p_i^*(t) \\ c(T) = \sum_{t=1}^{T} C(\sum_{i\in\mathbb{N}} p_i^*(t))/1000. \end{cases} \tag{3.22}$$

For $\overline{V}$, the best performer is COA, which is closely followed by DOA. This is consistent with the curves in Fig. 3.7. For $\overline{U}$, we observe a slightly better performance for DPA without the variance control. For energy provisioning cost $c(T)$, the three algorithms all yield similar results, because they all include the function $C(\cdot)$ as a part of objective function. For the peak $\overline{PK}$ we see the same result as in Fig. 3.7, with COA achieving the best and DOA following COA tightly.

Overall, the distributed online algorithm proposed in this chapter achieves better results than DPA. Although COA is slightly better than DOA, its centralized manner in energy distribution limits its usage in practice for large scale systems. It also has the disadvantage of requiring user's privacy information. DOA successfully mitigates these problems with the distributed approach. In summary, DOA is a practical method with a highly competitive performance comparing to the optimum, especially on variance control and peak reduction, for online energy distribution in the smart grid.

3.7 Related Work

Smart grid, characterized with the two-way flows of electricity and information, is envisioned to replace the existing power grid in the future [8, 101]. A comprehensive review on smart grid technologies and research can be found in [44], where major topics on smart grid is discussed in three areas: infrastructure, management and protection.

Within the three areas, demand side management or demand response has been attracting considerable research efforts [81–83, 88, 90–92, 113, 114]. Researchers work mainly on demand profile shaping, user utility maximization and cost reduction. For example, machine learning is used in [90] to develop a learning algorithm for energy costs reduction and energy usage smoothing, while [92] aims to balance the users' cost and waiting time. A constrained multi-objective optimization problem is formulated in [113] to minimize energy consumption cost and to maximize a certain utility among a group of users. Lyapunov optimization is adopted in [81–83] to stabilize the energy storage and user utility while reducing the operation cost of a microgrid. Lyapunov optimization is also used in [114] to optimally schedule the usage of all the energy resources in the system and minimize the long-term time averaged expected total cost of supporting all users load demand. In these works, convex programming, machine learning and game theory are mostly used. In some other works, online algorithms [94], which are widely used in wireless communications and networking, is also utilized [74, 93]. In [93], the authors propose an dynamic pricing algorithm based on utility maximization in a distributed way. Ref. [74] presents a centralized online algorithm that achieves the optimal energy distribution and variance control without any future information.

Furthermore, for practical considerations, user's privacy is emphasized more and more by many authors [109, 110]. In [109], the authors examine privacy in smart grid from definition to different concerns in detail. In [110], the author studies how high resolution user electricity information can be used to reconstruct a user's daily life and preference.

Our work is inspired by considering the above two aspects for the energy distribution in smart grid. In power systems, it is possible to use online algorithms to detect and control the grid load variance in real time. Also the online algorithm can be decomposed into subproblems for users to solve locally. Motived by this two observations, we propose an energy distribution distributed online algorithm to achieve utility maximization, load smoothing and

privacy protection. The proposed distributed online algorithm is quite effective as shown in Section 3.6.

3.8 Conclusion

In this chapter, we presented a study of optimal distributed online energy distribution in the smart grid. With a formulation that captures the key design factors of the system, we extend our prior work of a centralized online algorithm, by decomposing the problem into many subproblems that can be solved in a distributed manner, thus protecting users' privacy and achieving scalability. We also show that the distributed online solution converges to the optimal offline solution asymptotically. The proposed distributed online algorithm is evaluated with trace-driven simulations and outperforms a benchmark scheme.

Chapter 4

Hierarchical Power Management for the Macrogrid and Cooperative Microgrids

4.1 Introduction

The decentralized generation at most renewable energy sources and the supporting technologies such as photovoltaics and micro-turbines, have driven the demand for a new distributed power grid system, the *Microgrid* (MG) [69]. Unlike traditional centralized power generation, the MG features distributed generation (DG) to support local users. DG is the basis of distributed energy resource (DER) systems, which is usually comprised of small power units, such as micro-turbines (25∼100 KW) and small photovoltaic panels (1∼10 KW). An MG can operate either in the *island* mode, where the local demand is supported with the MG's own DG and power storage, or the *grid-connected* mode, where the MG can acquire energy from, and/or contribute extra power to the Macrogrid [71]. MG is regarded as an important paradigm for the next generation power grid, the Smart Grid (SG) [44, 115, 116]. SG technologies, such as smart metering, communications and distributed control, will speed up the integration of MGs, and thus the penetration of DGs.

Over the past decade, MGs are built, experimented and tested around the world [70]. In a single MG, research works cover several main topics, including interface or coupling between an MG and the Macrogrid, DER dispatching and power support, and energy management [71, 72, 117–119].

Although more works are focused on the optimization and control of a single MG [120–123], the problem of *cooperation* among MGs and the Macrogrid has attracted considerable interest recently. With such cooperation, MGs and the Macrogrid will each gain tremendous benefits, such as reduced power loss, lower operational cost, and load peak reduction [124–130]. The obvious advantages stem from exploiting the *temporal, spatial, and technological*

diversities in a multiple MG system. For instance, an MG supporting a business area will have a very different temporal demand profile from that of an MG supporting a residential area; the DGs in geographically distributed MGs can also have different generation levels at same time of the day; and different DGs are affected by weather differently: an MG with a photovoltaic array may suffer low generation during a storm, while a neighboring micro-turbine based MG, caught in the same storm, may generate a large amount of power exceeding its own demand. As in wireless communications systems, exploiting such diversity through MG cooperation could bring about more efficient power generation and distribution.

The power grid is currently under a transition from traditional centralized distribution to decentralized distribution. In practice, the DG in MGs are usually not able to generate power stably and constantly. On the other hand, MGs can provide surplus power to the Macrogrid. Therefore, it is important to incorporate all the key factors in a holistic manner, i.e., the generation cost, power generation and transmission losses, load smoothing, distributed storage, and the utility of power users. A control strategy would be highly desired that considers all the key factors for both the Macrogrid and MGs.

In this chapter, we consider a power grid consisting of the Macrogrid and several cooperative MGs. The goal is to exploit *MG diversity gain* to optimize both the MG performance and user satisfaction. With cooperation, an MG is able to share its excess power with other MGs nearby or with the Macrogrid. Due to limited storage capacity, the MG can sell its extra power to other MGs suffering power shortage. Alternatively, the MG could buy power from other MGs as well when its DG suffers low generation, such that the power loss and cost can both be reduced compared to buying power directly from the Macrogrid. On the other hand, the Macrogrid could provide more storage capacity for the MGs, while the extra power from the MGs will in turn reduce the need of traditional power generation in the Macrogrid. Grid load smoothness of the Macrogrid could be achieved if the power flows from/to the MGs are optimally managed and scheduled.

In particular, under some mild assumptions, we firstly formulate the cooperative MG problem as a convex optimization problem by capturing the key factors in a grid system, i.e., operation cost, power generation and transmission losses, user utility, distributed storage, and grid load smoothing. We then decompose the original problem into a two-tier power control problem. The first-tier control is for the Macrogrid, aiming to maximize user utility, minimize power transmission cost from/to the Macrogrid, and smooth the grid load of the Macrogrid. The second-tier control is for each MG, aiming to minimize the cost of the MGs for power generation and transmission, while guaranteeing the power demand of MG users. It balances the power level with the Macrogrid and makes energy trading and storage decisions within the MG network.

The power flow between MGs and the Macrogrid is on one side the power injected from outside of the MG network for MGs, and on the other side a special load for the Macrogrid, which is positive as usual if the power is transmitted to the MGs out from the Macrogrid, and is negative reversely. This way, the two-tier controls are well integrated. For the first-tier problem, we develop an effective online algorithm that does not require any future information and is proven to be asymptotically optimal; for the second-tier problem, we develop a distributed algorithm for optimal solutions. The performance of the proposed hierarchical power scheduling scheme is validated with trace-driven simulations, where fast convergence and superior performance over several comparison schemes are observed.

The remainder of this chapter is organized as follows. We present the system model and problem formulation in Section 4.2. We develop the asymptotically optimal online algorithm for the Macrogrid in Section 4.3, a distributed algorithm for cooperative MGs in Section 4.4, and present these algorithms in Section 4.5. Performance evaluation is presented in Section 4.6. Section 4.8 concludes this chapter. The notation used in the rest of this chapter is summarized in Table 4.1.

Table 4.1: Notation Table for Chapter 4

Symbol	*Description*
$\mathbb{N}$	set of electricity users in Macrogrid
$\mathbb{M}$	set of Microgrids
$\mathbb{T}$	set of time slots from 1 to T
N	total number of users in Macrogrid
M	total number of MGs
$U(\cdot)$	user utility function
$\omega_i(t)$	flexibility level of user i at time t
α	coupling parameters in Prob-MAMG
$C_m(\cdot)$	transmission cost between Macrogrid and MG m
$Var(\cdot)$	variance of Macrogrid load
$G(\cdot)$	generation cost in Macrogrid
$d_j(t)$	power usage by user i at time t
$p_m(t)$	power between Macrogrid and MG m
$p_{m,max}(t)$	maximum power allowed between Macrogrid and MG m
$l(t)$	grid load in Macrogrid at time t
$I(\cdot), R(\cdot)$	indicator functions
$b_{max}(t)$	maximum generation cost in Macrogrid at time t
σ_m	transmission loss ratio between Macrogrid and MG m
σ_{km}	transmission loss ratio between MG k and MG m
$G_k(\cdot)$	generation cost in MG k
$C_{km}(\cdot)$	transmission cost between MG k and m
$p_{km}(t)$	power transmitted from MG k to m
$g_{k,max}(t)$	maximum generation in MG k at time t
$C_{m,max}(t)$	maximum power received in MG m from other MGs
$C_{m,min}(t)$	minimum power received in MG m from other MGs
$s_m^*(t)$	power to be stored in MG m at time t
$s_m(t)$	the storage level in MG m at time t
ξ	storage loss coefficient in time t in MGs
$\lambda_m(t)$	Dual multiplier of the upper bound in Prob-MG1
δ	step-size for updating $\lambda_{m,t}(k)$
ϵ	terminating condition for updating $\lambda_{m,t}(k)$
$\beta_m(t)$	Dual multiplier of the lower bound in Prob-MG1
τ	step-size for updating $\beta_{m,t}(k)$
ε	terminating condition for updating $\beta_{m,t}(k)$
$\tilde{(\cdot)}$	the offline optimal solution of $(\cdot)$
$(\cdot)^*$	the online optimal solution of $(\cdot)$
$\hat{(\cdot)}$	the iterative replacement of $(\cdot)$
$\bar{(\cdot)}$	the average of $(\cdot)$

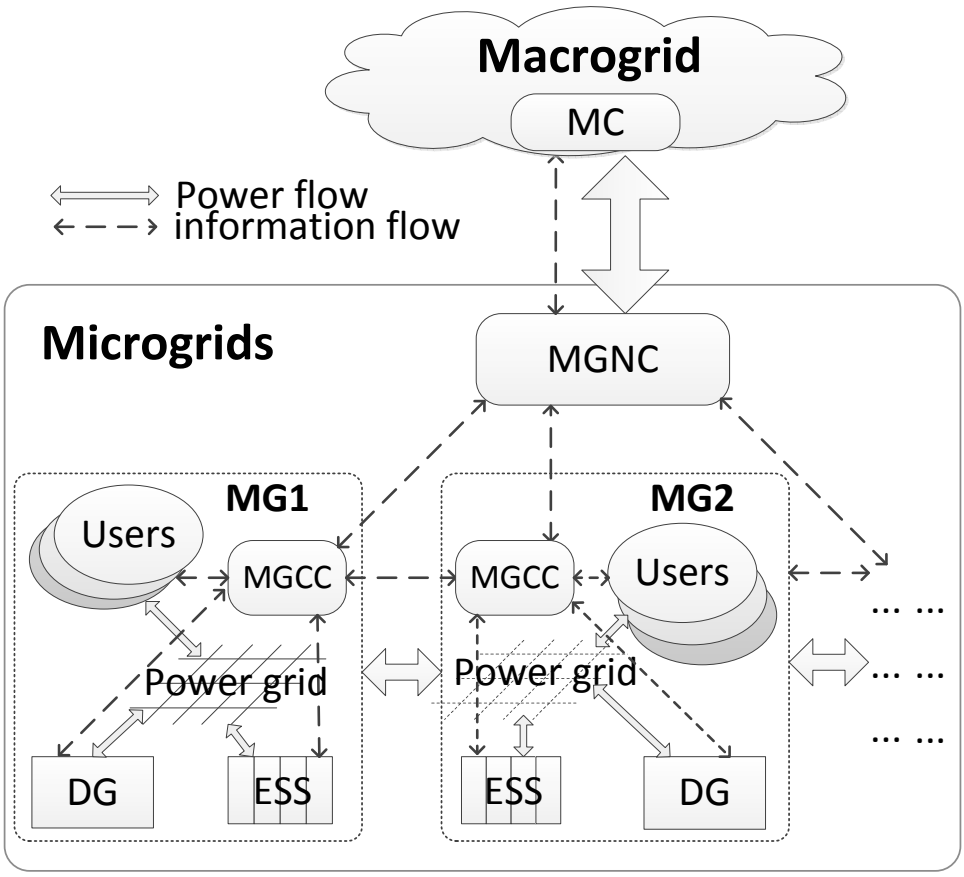

Figure 4.1: Illustration of the power grid network.

4.2 Problem Statement

4.2.1 System Model

We consider a power grid system with one Macrogrid and many MGs as shown in Fig. 4.1. The Macrogrid supports its own set of power users mostly with the traditional power generation sources. The Macrogrid controller (MC) collects information from the smart meters at the Macrogrid users to optimally distribute power to the users, and from the MG control center (MGCC) to trade power with the MGs.

As shown in Fig. 4.1, each MG consists of one or more DGs, an energy storage system (ESS), a smart infrastructure (such as smart meters and communication links), a set of users, and an MGCC. The energy users demand power from a designated MG. The independent DGs generate power to support the demand inside the MG. The ESS stores the extra power

and discharges to satisfy the excess demand exceeding the DG generation. It usually consists of many batteries as well as some PHEVs. Furthermore, for excessive needs from its users, each MG will request and buy energy from other MGs or from the Macrogrid. The MGCC in each MG controls the power distribution of the entire MG. Note that in an MG of the SG environment, information flow is essential for the control and cooperation. Both energy and information flows are enabled among the MGs and the Macrogrid. The MGCC acquires user demand from the smart meters on the user side through a wired or wireless communication network. The MGCC also decides to sell or store excessive power based on the grid information. In this way, the energy generated among the MGs could be used efficiently, to minimize the power from the Macrogrid and to support the Macrogrid needs when possible.

While MC and MGCC are the core of power scheduling inside the Macrogrid and MGs, respectively, the MG network controller (MGNC) works both as a controller of the MG network and a bridge between the Macrogrid and the MGs. It coordinates the information exchange and power transmissions between the Macrogrid and the cooperative MGs.

4.2.2 Problem Formulation

We assume a time slotted system with $\mathbb{T} = \{1, 2, \cdots, T\}$ time slots. We denote the set of independent power users in the Macrogrid as $\mathbb{N} = \{1, 2, \cdots, N\}$. Each user $i \in \mathbb{N}$ demands power $d_i(t)$ at time t. Let $U(d_i(t), \omega_i(t))$ be the utility function for user i, which indicates the users' overall satisfactory level, and is a concave and strictly increasing function of $d_i(t)$. The parameter $\omega_i(t) \in (0, 1)$ denotes user i's level of flexibility, while a larger number closer to 1 (0) indicating a higher (lower) level of flexibility. The function $G(\cdot)$ indicates the generation cost in the Macrogrid, which is strictly convex and increasing. In practice, a quadratic function is used; see our previous work [74] for more details on the utility function and generation cost function.

We use $\mathbb{M} = \{1, 2, \cdots, M\}$ to denote the set of all MGs. Each MG m has N_m users with total demand $d_m(t)$ at time t. Unlike the Macrogrid, the power supply in MGs may

be unstable in many cases, and thus it provides more flexibility to consider users' demand as a whole. Let $p_m(t)$ be the power load in the Macrogrid, transmitted to or received from MG $m \in \mathbb{M}$ at time t. It is positive when power is transmitted from the Macrogrid to MG m, and is negative for the reverse direction. We define $\sigma_{0m} \in (0,1)$ [127] as the ratio of transmission loss from the Macrogrid to MG m, and $\sigma_{m0} \in (0,1)$ as from the MG m to the Macrogrid. And to simplify the expression, we define $\sigma_m = \sigma_{0m}$ if $p_m(t) > 0$, and $\sigma_m = \sigma_{m0}$ otherwise. Thus, when $p_m(t)$ is positive, the power received in MG m is $p_m(t)\sigma_m$; when $p_m(t)$ is negative, the generation in MG m is $p_m(t)/\sigma_m$. Let $p_{km}(t)$ denote the power received in MG m from MG k at time t, and $p_{mm}(t)$ be the power generated and used in MG M by itself. Similarly, $\sigma_{km} \in (0,1)$ denotes the ratio of transmission loss between MG k and MG m. Note that although σ_{0m} can be same as σ_{m0}, and σ_{km} can be same as σ_{mk} in some cases, the reciprocity of transmission loss ratios is not an essential requirement in our model. Therefore, $\sum_{m\in\mathbb{M}} p_{km}(t)/\sigma_{km}$ is the total power generated in MG k for MGs, and $\sum_{k\in\mathbb{M}} p_{km}(t)$ is the total power in MG m received from all the MGs.

We use a general convex function $C_m(\cdot)$ to represent the transmission cost between the Macrogrid and MG m ,because it costs more for the same amount of loss of power as the total power loss increases to a higher level. Similarly, we use convex functions $G_k(\cdot)$ and $C_{km}(\cdot)$ to denote the power generation cost in MG k and the power transmission cost between MG k and MG m, respectively. Without loss of generality, we assume the utility functions, the transmission cost functions and the generation cost functions all have the same unit (e.g., dollar).

Jointly considering user utility, power transmission cost, and the load variance in the Macrogrid, and power generation cost, power transmission cost in the MG network, we

formulate the power scheduling problem Prob-MAMG as follows.

$$\begin{aligned}
\text{max: } & \sum_{t=1}^{T}\sum_{i\in\mathbb{N}} U(d_i(t),\omega_i(t)) - \frac{\alpha T}{2}\text{Var}(\vec{l_T}) - \\
& \sum_{t=1}^{T}\left(\sum_{m\in\mathbb{M}} C_m(p_m(t)) + \sum_{k\in\mathbb{M}} G_k\left(\sum_{m\in\mathbb{M}}\frac{p_{km}(t)}{\sigma_{km}} - \right.\right. \\
& \left.\left. p_k(t)I(\sigma_k)\right) + \sum_{m\in\mathbb{M}}\sum_{k\in\mathbb{M}} C_{km}(p_{km}(t))\right) \qquad (4.1)
\end{aligned}$$

$$\text{s.t. } d_i(t) \geq d_{i,min}(t), \forall\, i \in \mathbb{N}, t \in \mathbb{T} \qquad (4.2)$$

$$G(l(t)) \leq b_{max}(t), \forall\, t \in \mathbb{T}, \qquad (4.3)$$

$$|p_m(t)| \leq p_{m,max}(t), \forall\, m \in \mathbb{M}, t \in \mathbb{T} \qquad (4.4)$$

$$\sum_{k\in\mathbb{M}} p_{km}(t) + p_m(t)R(\sigma_m) - s'_m(t) = d_m(t), \quad \forall\, m \in \mathbb{M}, t \in \mathbb{T} \qquad (4.5)$$

$$\sum_{m\in\mathbb{M}}\frac{p_{km}(t)}{\sigma_{km}} - p_k(t)I(\sigma_k) \leq g_{k,max}(t), \forall\, k \in \mathbb{M}, t \in \mathbb{T}, \qquad (4.6)$$

where α is the weight to trade-off the dual objectives, and the variance function $\text{Var}(\cdot)$ is defined as

$$\text{Var}(\vec{l_T}) = \frac{1}{T}\sum_{t=1}^{T}\left(l(t) - \frac{1}{T}\sum_{k=1}^{T} l(k)\right)^2, \qquad (4.7)$$

where each element of vector $\vec{l_T}$ is the load of the Macrogrid at time t computed as $l(t) = \sum_{i\in\mathbb{N}} d_i(t) + \sum_{m\in\mathbb{M}} p_m(t)$. The indicator function $I(\sigma_k)$ is defined as $I(\sigma_k) = 1/\sigma_k$ if $p_k(t) < 0$, and $I(\sigma_k) = 0$ otherwise; the indicator function $R(\sigma_m)$ is defined as $R(\sigma_m) = \sigma_m$ if $p_k(t) > 0$, and $R(\sigma_m) = 1/\sigma_m$ otherwise. Moreover, $d_{i,min}(t)$ in constraint (4.2) is the minimum demand of user i, $b_{max}(t)$ in constraint (4.3) is the generation cost limit for the energy provider, and $p_{m,max}(t)$ in (4.4) is the maximum amount of transmission allowed in one time slot. In constraint (4.6), $g_{k,max}(t)$ is the maximum generation in MG k at time t, and $s_m(t)$ is the

power storage level of MG m at time t and $s'_m(t)$ is the amount of power to be stored in the time slot, computed as

$$s'_m(t) = s_m(t) - \xi \cdot s_m(t-1), \tag{4.8}$$

$$s_{m,min} \leq s_m(t) \leq s_{m,max}, \forall\ m \in \mathbb{M}, . \tag{4.9}$$

where ξ is the storage loss ratio in the ESS, and $s_{m,min}$ ($s_{m,max}$) is the lower (upper) bound on the storage capacity.

In Prob-MAMG, all the functions and the constraints are convex, which means it is a convex optimization problem. But it cannot be solved unless all the constraints from the Macrogrid and MGs are known a priori for the entire time period $\mathbb{T}$. Even with all these necessary information, it is very difficult to solve such a complex problem in practice. Note that in Prob-MAMG, the Macrogrid and MG m is coupled by the power flow $p_m(t)$. As discussed, the Macrogrid usually generates much more power than the MGs, and thus $p_m(t)$ can be seen as a special load in the Macrogrid. Therefore, we can decompose the Prob-MAMG into two tiers. The first-tier related to the Macrogrid solves for power distribution for users, i.e., $d_i(t)$ and power exchanged with the MGs, i.e., $p_m(t)$. The second-tier for the MGs matches $p_m(t)$ in MG m and solves for $p_{km}(t)$, i.e., the power transmissions among the cooperative MGs.

The first-tier problem for the Macrogrid Prob-MA1 is formulated as follows.

$$\text{max: } \sum_{t=1}^{T}\left(\sum_{i\in\mathbb{N}} U(d_i(t),\omega_i(t)) - \sum_{m\in\mathbb{M}} C_m(p_m(t))\right) - \frac{\alpha T}{2}\text{Var}(\vec{l_T}) \tag{4.10}$$

$$\text{s.t. } (4.2) \sim (4.4),$$

The second-tier problem for the MGs Prob-MG1 is as follows.

$$\begin{aligned} \text{min:} \quad & \sum_{k\in\mathbb{M}} \left(G_k \left(\sum_{m\in\mathbb{M}} \frac{p_{km}(t)}{\sigma_{km}} - p_k^*(t) I(\sigma_k) \right) \right. \\ & \left. + \sum_{m\in\mathbb{M}} C_{km}(p_{km}(t)) \right) \qquad (4.11) \\ \text{s.t.} \quad & (4.5), (4.6) \text{ and } (4.9), \end{aligned}$$

where $p_k^*(t)$ is part of the solution to Prob-MA1 (see Section 4.3). Now, it is clear that in Prob-MA1, user utility, load variance and power transmission cost to/from the Macrogrid are optimized, and Prob-MG1 aims to minimize the power generation cost and transmission cost among the cooperative MGs. In Section 4.3, we reformulate Prob-MA1 and develop an online algorithm that is asymptotically optimal. In Section 4.4, we solve problem Prob-MG1 with a distributed algorithm for optimal solutions.

4.3 Online Power Distribution in the Macrogrid

4.3.1 Reformulation and Optimal Offline Solution

In Prob-MA1, all the power users and MGs are independent. Thus we can reformulate Prob-MA1 by replacing the grid load variance term with $\text{Var}(\vec{l}_T) = \sum_{i\in\mathbb{N}} \text{Var}(\vec{d}_{i,T}) + \sum_{m\in\mathbb{M}} \text{Var}(\vec{p}_{m,T})$. We thus obtain Prob-MA2 as follows.

$$\begin{aligned} \text{max:} \quad & \mathbf{F}(\mathbf{d},\mathbf{p}) = \sum_{t=1}^{T} \left(\sum_{i\in\mathbb{N}} U(d_i(t),\omega_i(t)) - \sum_{m\in\mathbb{M}} C_m(p_m(t)) \right) \\ & - \frac{\alpha T}{2} \left(\sum_{i\in\mathbb{N}} \text{Var}(\vec{d}_{i,T}) + \sum_{m\in\mathbb{M}} \text{Var}(\vec{p}_{m,T}) \right) \qquad (4.12) \\ \text{s.t.} \quad & (4.2) \sim (4.4), \end{aligned}$$

where

$$\mathrm{Var}(\vec{d}_{i,T}) = \frac{1}{T}\sum_{t=1}^{T}\left(d_i(t) - \frac{1}{T}\sum_{k=1}^{T} d_i(k)\right)^2,$$

$$\mathrm{Var}(\vec{p}_{m,T}) = \frac{1}{T}\sum_{t=1}^{T}\left(p_m(t) - \frac{1}{T}\sum_{k=1}^{T} p_m(k)\right)^2.$$

In Prob-MA2, the utility function $U(\cdot)$ is concave, and the transmission cost function $C_m(\cdot)$ and the variance function $\mathrm{Var}(\cdot)$ are both convex. Therefore, Prob-MA2 is a convex optimization problem with a convex set of the constraints. Furthermore, Prob-MA2 has a unique solution since $U(\cdot)$ is strictly increasing. Thus, we can select the constraints $d_{i,min}(t)$ and $p_{m,max}(t)$ so that Prob-MA2 is feasible and the Slater's condition is satisfied, and obtain the optimal solution by solving the KKT conditions [98], as

$$\begin{cases} I_{\mathbb{N}}\left(U'(\tilde{d}_i(t),\omega_i(t)) - \alpha(\tilde{d}_i(t) - \bar{\tilde{d}}_{i,T}) + \tilde{\nu}_i(t)\right) + \\ I_{\mathbb{M}}\left(-C'_m(\tilde{p}_m(t)) - \alpha(\tilde{p}_m(t) - \bar{\tilde{p}}_{m,T}) + \tilde{\gamma}_m(t) - \right. \\ \left.\tilde{\rho}_m(t)\right) - \tilde{\mu}(t)G'(\tilde{l}(t))/b_{max}(t) = 0 \\ \tilde{\mu}(t)\left(G(\tilde{l}(t))/b_{max}(t) - 1\right) = 0 \\ \tilde{\nu}_i(t)\left(\tilde{d}_i(t) - d_{i,min}(t)\right) = 0 \\ \tilde{\gamma}_m(t)\left(\tilde{p}_m(t) + p_{m,max}(t)\right) = 0 \\ \tilde{\rho}_m(t)\left(\tilde{p}_m(t) - p_{m,max}(t)\right) = 0 \\ \tilde{\mu}(t), \tilde{\nu}_i(t), \tilde{\gamma}_m(t), \tilde{\rho}_m(t) \geq 0, \forall\, i \in \mathbb{N},\ m \in \mathbb{M},\ t \in \mathbb{T}, \end{cases} \tag{4.13}$$

where $\tilde{d}_i(t)$ and $\tilde{p}_m(t)$ are the optimal points; the indicator $I_{\mathbb{N}} = 1$ for the users in $\mathbb{N}$, and $I_{\mathbb{N}} = 0$ otherwise; the indicator $I_{\mathbb{M}} = 1$ for the MGs in $\mathbb{M}$, and $I_{\mathbb{M}} = 0$ otherwise; and

$$\bar{\tilde{d}}_{i,T} = \frac{1}{T}\sum_{k=1}^{T}\tilde{d}_i(k), \quad \bar{\tilde{p}}_{m,T} = \frac{1}{T}\sum_{k=1}^{T}\tilde{p}_m(k). \tag{4.14}$$

From the gradient condition of the above KKT conditions, we derive the Lagrange multiplier $\mu(t)$ as

$$\tilde{\mu}(t) = \Big(I_{\mathbb{N}}\left(U'(\tilde{d}_i(t), \omega_i(t)) - \alpha(\tilde{d}_i(t) - \bar{\tilde{d}}_{i,T}) + \tilde{\nu}_i(t)\right) + I_{\mathbb{M}}\left(-C'_m(\tilde{p}_m(t)) - \alpha(\tilde{p}_m(t) - \bar{\tilde{p}}_{m,T}) + \tilde{\gamma}_m(t) - \tilde{\rho}_m(t)\right)\Big) / (G'(\tilde{l}(t))/b_{max}(t)). \tag{4.15}$$

Thus, the optimal solution to Prob-MA2 can be found by solving its KKT conditions (4.13). However, it is indicated in (4.15) that solving the KKT conditions requires the information on $\bar{\tilde{d}}_{i,T}$ and $\bar{\tilde{p}}_{m,T}$, which are the average of the user demand $\tilde{d}_i(t)$ and the exchanged power with MG m $\tilde{p}_m(t)$ for the entire time period T, respectively. To derive the optimal solution to Prob-MA2, the constraints $d_{i,min}(t)$, $p_{m,max}(t)$, and $b_{max}(t)$ over the entire time window $\mathbb{T}$ are also needed. This is an *offline optimal solution*, which may not be practical in some cases.

4.3.2 Online Power Distribution in the Macrogrid

We next present an online algorithm for Prob-MA2 in this section. It can be seen that in (4.15), $\bar{\tilde{d}}_{i,T}$ and $\bar{\tilde{p}}_{m,T}$ are the only two terms requiring future information, while these time averages can be approximated by properly defined updating equations. Motivated by this observation, we first present an approximation problem that can be solved without future information, and then prove that its solution is convergent to the optimal offline solution to the original problem Prob-MA2.

Specifically, we replace the average terms $\frac{1}{T}\sum_{k=1}^{T} d_i(k)$ and $\frac{1}{T}\sum_{k=1}^{T} p_m(k)$ in (4.12) by two new terms $\hat{d}_i(t)$ and $\hat{p}_m(t)$, respectively, and remove the time sum notation so that the problem can be solved at time t. We thus obtain a new problem Prob-MA3 at time t as

follows.

$$\begin{aligned} \text{max:} \quad & \sum_{i\in\mathbb{N}} U(d_i(t),\omega_i(t)) - \sum_{m\in\mathbb{M}} C_m(p_m(t)) - \frac{\alpha}{2}\sum_{i\in\mathbb{N}}(d_i(t)- \\ & \hat{d}_i(t-1))^2 - \frac{\alpha}{2}\sum_{m\in\mathbb{M}}(p_m(t)-\hat{p}_m(t-1))^2 \qquad (4.16) \\ \text{s.t.} \quad & (4.2)\sim(4.4), \end{aligned}$$

where $\hat{d}_i(t)$ and $\hat{p}_m(t)$ are updated at each time slot t as

$$\begin{cases} \hat{d}_i(t) = \hat{d}_i(t-1) + \frac{\alpha}{t+\alpha}\cdot(d_i^*(t) - \hat{d}_i(t-1)) \\ \hat{p}_m(t) = \hat{p}_m(t-1) + \frac{\alpha}{t+\alpha}\cdot(p_m^*(t) - \hat{p}_m(t-1)), \end{cases} \qquad (4.17)$$

where $d_i^*(t)$ and $p_m^*(t)$ denote the solutions to Prob-MA3. This way, we decompose the problem over a time window $\mathbb{T}$ into many problems to be solved by the MC at each time t without requiring any future information. Because the updating equations in (4.17) only use the solutions to Prob-MA3 in the previous time slot, we use $\hat{d}_i(t)$ and $\hat{p}_m(t)$ to approximate the average terms $\bar{d}_{i,T}^*$ and $\bar{p}_{m,T}^*$, respectively. The following lemma and theorem state that $\hat{d}_i(t)$ and $\hat{p}_m(t)$ are convergent, and the online solutions are convergent to the offline solutions. The complete proofs of Lemma 4.1 and Theorem 4.1 are presented in Appendix C.

Lemma 4.1. *The updating terms in Prob-MA3, i.e., $\hat{d}_i(t)$ and $\hat{p}_m(t)$, are convergent to the time averages of its solution $\bar{d}_{i,T}^*$ and $\bar{p}_{m,T}^*$, respectively, when T is sufficiently large. That is, for $i\in\mathbb{N}$ and $m\in\mathbb{M}$, we have*

$$\lim_{T\to\infty}\hat{d}_i(T) = \lim_{T\to\infty}\bar{d}_{i,T}^* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} d_i^*(t) \qquad (4.18)$$

$$\lim_{T\to\infty}\hat{p}_m(T) = \lim_{T\to\infty}\bar{p}_{m,T}^* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} p_m^*(t). \qquad (4.19)$$

Theorem 4.1. *The solution to Prob-MA3 converges asymptotically to the solution to Prob-MA2.*

According to Theorem 1, we can solve Prob-MA3 in each time t using the information of the current time, while still achieving the optimal results in a certain amount of time slots.

4.4 Distributed Cooperative Power Scheduling for MGs

In Section 4.2.2, we formulate problem Prob-MG1, which can also be shown to be convex. We can solve Prob-MG1 with some convex optimization techniques [98], such as KKT conditions as in solving Prob-MA2 and Prob-MA3. In practical scenarios, a distributed algorithm is more appealing for reducing the computational complexity, reducing delay in realtime power scheduling, and enhancing scalability. Please see Chapter 3 for more discussions on the benefits of using distributed algorithms in the smart grid. In this section, we develop a distributed cooperative power scheduling algorithm for the MGs, by decomposing Prob-MG1 into multiple sub-problems to be solved by the MGCC in each MG.

4.4.1 Problem Reformulation

Recall the definition of Prob-MG1 for minimizing the cost in the MG network in (4.11). The constraints on the power storage levels $s_m(t)$, i.e., (4.8) and (4.9), can be merged as

$$\frac{s_{m,min}}{(1-\xi)} - s_{m,max} \leq s'_m(t) \leq s_{m,max} - s_{m,min}. \tag{4.20}$$

Substituting (4.20) into (4.5), we have a new constraint for $\sum_{k\in\mathbb{M}} p_{km}(t)$. Then Prob-MG1 (4.11) can be rewritten with the new constraint as

$$\begin{aligned} \text{min:} \quad \sum_{k\in\mathbb{M}} \Bigg(G_k \left(\sum_{m\in\mathbb{M}} \frac{p_{km}(t)}{\sigma_{km}} - p_k^*(t) I(\sigma_k) \right) + \\ \sum_{m\in\mathbb{M}} C_{km}(p_{km}(t)) \Bigg) \end{aligned} \tag{4.21}$$

$$\text{s.t.} \quad C_{m,min}(t) \leq \sum_{k\in\mathbb{M}} p_{km}(t) \leq C_{m,max}(t), \forall\, m \in \mathbb{M}, t \in \mathbb{T} \tag{4.22}$$

$$\sum_{m\in\mathbb{M}} \frac{p_{km}(t)}{\sigma_{km}} - p_k^*(t) I(\sigma_k) \leq g_{k,max}(t), \forall k \in \mathbb{M}, t \in \mathbb{T}, \tag{4.23}$$

where

$$\begin{cases} C_{m,min}(t) = \frac{s_{m,min}}{1-\xi} - s_{m,max} + d_m(t) - p_m^*(t) R(\sigma_m) \\ C_{m,max}(t) = s_{m,max} - s_{m,min} + d_m(t) - p_m^*(t) R(\sigma_m). \end{cases} \tag{4.24}$$

In Prob-MG1, the variables are $p_{km}(t)$ for each MG pair k and m. We next decompose Prob-MG1 into sub-problems using only local information with the *dual decomposition* technique [112].

4.4.2 Cooperative Distributed Power Scheduling for MGs

We first derive the Lagrangian of Prob-MG1 as follows.

$$\begin{aligned}
& L(\mathbf{P}(t), \vec{\lambda}_M(t), \vec{\beta}_M(t)) \\
=& \sum_{k\in\mathbb{M}} \left(G_k \left(\sum_{m\in\mathbb{M}} \frac{p_{km}(t)}{\sigma_{km}} - p_k^*(t) I(\sigma_k) \right) + \right. \\
& \qquad \left. \sum_{m\in\mathbb{M}} C_{km}(p_{km}(t)) \right) + \\
& \sum_{m\in\mathbb{M}} \lambda_m(t) \left(\sum_{k\in\mathbb{M}} p_{km}(t) - C_{m,max}(t) \right) + \\
& \sum_{m\in\mathbb{M}} \beta_m(t) \left(C_{m,min}(t) - \sum_{k\in\mathbb{M}} p_{km}(t) \right) \\
=& \sum_{k\in\mathbb{M}} \left(G_k \left(\sum_{m\in\mathbb{M}} \frac{p_{km}(t)}{\sigma_{km}} - p_k^*(t) I(\sigma_k) \right) + \right. \\
& \left. \sum_{m\in\mathbb{M}} \left(C_{km}(p_{km}(t)) + (\lambda_m(t) - \beta_m(t)) p_{km}(t) \right) \right) + \\
& \sum_{m\in\mathbb{M}} \left(\beta_m(t) C_{m,min}(t) - \lambda_m(t) C_{m,max}(t) \right),
\end{aligned} \tag{4.25}$$

where $\beta_m(t) \geq 0$ and $\lambda_m(t) \geq 0$ are the Lagrange multipliers associated with the two inequalities in constraint (4.22), respectively. We then decompose Prob-MG1 into M sub-problems $S(\vec{\lambda}_M(t), \vec{\beta}_M(t))$ for the MGs, termed as Prob-MG2.

$$\begin{aligned} \text{min:} \quad & S_k(\vec{\lambda}_M(t), \vec{\beta}_M(t)) \\ & = G_k \left(\sum_{m \in \mathbb{M}} \frac{p_{km}(t)}{\sigma_{km}} - p_k^*(t) I(\sigma_k) \right) + \\ & \quad \sum_{m \in \mathbb{M}} (C_{km}(p_{km}(t)) + (\lambda_m(t) - \beta_m(t)) p_{km}(t)) \qquad (4.26) \\ \text{s.t.} \quad & (4.23). \end{aligned}$$

The dual problem of Prob-MG1 is as follows [112].

$$\text{max:} \quad D(\vec{\lambda}_M(t), \vec{\beta}_M(t)) \qquad (4.27)$$

$$\text{s.t.} \quad \lambda_m(t) \geq 0, \beta_m(t) \geq 0, \forall\, m \in \mathbb{M}, \qquad (4.28)$$

where

$$\begin{aligned} & D(\vec{\lambda}_M(t), \vec{\beta}_M(t)) = \min \left\{ L(\mathbf{P}(t), \vec{\lambda}_M(t), \vec{\beta}_M(t)) \right\} \\ & = \sum_{k \in \mathbb{M}} S_k(\lambda_M(t), \beta_M(t)) + \\ & \quad \sum_{m \in \mathbb{M}} (\beta_m(t) C_{m,min}(t) - \lambda_m(t) C_{m,max}(t)) \,. \qquad (4.29) \end{aligned}$$

We thus decompose Prob-MG1 into M sub-problems each of which can be solved by the MGCC in each MG. Furthermore, because the primal problem (4.21) is convex and has feasible solutions for proper selections of $s_{m,min}$, $s_{m,max}$, and $g_{k,max}(t)$, *strong duality* holds [112], so that the optimal solution can be obtained from the dual problem (4.29).

For given $\vec{\lambda}_M(t)$ and $\vec{\beta}_M(t)$, the sub-problem $S_k(\vec{\lambda}_M(t), \vec{\beta}_M(t))$ for MG k is convex because the generation cost function $G_k(\cdot)$, the transmission cost function $C_{km}(\cdot)$, and constraint (4.23) are all convex as discussed in Section 4.2. The sub-problems can be solved by commonly used methods such as KKT conditions and the interior point method (IPM) [98]. After solving the sub-problems, the dual problem will be solved by the MGNC by gathering all the solutions to the sub-problems from the MGCCs. Furthermore, function $S_k(\cdot)$ is differentiable because $G(\cdot)$ and $C(\cdot)$ are both differentiable. We can use the following *gradient method* to obtain the dual variables $\lambda_m(t)$ and $\beta_m(t)$.

$$\begin{cases} \lambda_{m,t}(j+1) = \left[\lambda_{m,t}(j) - \delta\left(C_{m,max}(t) - \sum_{k\in\mathbb{M}} p^*_{km,t}(j)\right)\right]^+ \\ \beta_{m,t}(j+1) = \left[\beta_{m,t}(j) + \tau\left(C_{m,min}(t) - \sum_{k\in\mathbb{M}} p^*_{km,t}(j)\right)\right]^+, \end{cases} \tag{4.30}$$

where δ and τ are step-sizes; $p^*_{km,t}(j)$ is the solution to Prob-MG2 (4.26) for given $\lambda_{m,t}(j)$ and $\beta_{m,t}(j)$; and $[\cdot]^+$ is the projection onto the nonnegative orthant [112]. The dual variable $\vec{\lambda}_M(t)$ and $\vec{\beta}_M(t)$ will converge to the dual optimal $\vec{\lambda}^*_M(t)$ and $\vec{\beta}^*_M(t)$, respectively, since *strong duality* holds [112]. The optimal solution $\mathbf{P}^*(t)$ to Prob-MG1 can be acquired by solving each Prob-MG2 for $\vec{\lambda}^*_M(t)$ and $\vec{\beta}^*_M(t)$.

4.5 Optimal Hierarchical Power Scheduling for the Entire System

In this section, we summarize the analysis in Sections 4.3 and 4.4, and present the hierarchical power scheduling algorithms, termed HPS, for the entire power grid system. As discussed, HPS consists of two tiers: (i) online power distribution in the Macrogrid, and (ii) cooperative distributed power scheduling in the MGs. Furthermore, the lower tier algorithm consists of two parts: one for the MGNC and the other for each MGCC. The proposed algorithms are presented in Algorithms 4.1∼4.3, where $\epsilon_m > 0$ and $\varepsilon_m > 0$ are small tolerance values for termination conditions, for all $m \in \mathbb{M}$.

Note that the MC in the Macrogrid requires information on $p_{m,max}(t)$ to solve Prob-MA3. Each MG m can estimate $p_{m,max}(t)$ according to (4.31), where the first term refers

Algorithm 4.1: Online Power Distribution in the Macrogrid

Initialize $\hat{d}_i(0)$ and $\hat{p}_m(0)$, for all $i \in \mathbb{N}$ and $m \in \mathbb{M}$;
for *i=1:T* **do**
  Receive constraint $p_{m,max}(t)$ from the MGNC ;
  Solve Prob-MA3 (4.16) ;
  Send solution $p^*_m(t)$ to the MGNC ;
  Update $\hat{d}_i(t)$ and $\hat{p}_m(t)$ for all $i \in \mathbb{N}$ and $m \in \mathbb{M}$ as in (4.17) ;
  Exchange power $p^*_m(t)$ with MG m, for all m ;
  Distribute power $d^*_i(t)$ to Macrogrid user i, for all i ;
end

Algorithm 4.2: Distributed Cooperative Power Scheduling Algorithm for the MGNC

for *i=1:T* **do**
  Receive $p_{m,max}(t)$ from MG m and forward it to the MC, for all $m \in \mathbb{M}$;
  Receive $p^*_m(t)$ from the MC and forward it to MG m, for all $m \in \mathbb{M}$;
  Initialize $\vec{\lambda}_{M,t}(0) \geq 0$ and $\vec{\beta}_{M,t}(0) \geq 0$, and broadcast them to all the MGs ;
  repeat
    Receive $p^*_{km,t}(j)$ from the MGs ;
    Update $\vec{\lambda}_{M,t}(j)$ and $\vec{\beta}_{M,t}(j)$ using (4.30) ;
    Broadcast them to all the MGs, for all $k, m \in \mathbb{M}$;
  until ($|\lambda_{m,t}(j+1) - \lambda_{m,t}(j)| < \epsilon_m$ *and* $|\beta_{m,t}(j+1) - \beta_{m,t}(j)| < \varepsilon_m$);
  Broadcast $\vec{\lambda}^*_M(t)$ and $\vec{\beta}^*_M(t)$ to all the MGs ;
end

Algorithm 4.3: Distributed Cooperative Power Scheduling Algorithm for Each MGCC

for *i=1:T* **do**
  Estimate the maximum exchanged power with the Macrogrid $p_{m,max}(t)$ and report it to the MGNC ;
  Receive $p^*_m(t)$ and calculate constraints $C_{m,min}(t)$ and $C_{m,max}(t)$ using (4.24) ;
  repeat
    Receive $\vec{\lambda}_{M,t}(j)$ and $\vec{\beta}_{M,t}(j)$ and solve Prob-MG2 (4.26) ;
    Send solution $p^*_{km,t}(j)$ to the MGNC, for all $k, m \in \mathbb{M}$;
  until ($\vec{\lambda}^*_M(t)$ *and* $\vec{\beta}^*_M(t)$ *are received*);
  Calculate $s'_m(t)$ using (4.5) and charge or discharge the ESS accordingly, for all $m \in \mathbb{M}$;
  Transmit power $p^*_{km}(t)$ to MG m and exchange $p^*_m(t)$ with the Macrogrid, for all $k, m \in \mathbb{M}$;
end

$$\min\{|g_{m,max}(t) + \sum_{k\in\mathbb{M}} p^*_{km}(t-1) + (1-\xi)s_m(t-1) - \frac{s_{m,min}}{1-\xi} - d_m(t)|,$$
$$|-\sum_{k\in\mathbb{M}} p^*_{mk}(t-1) - (s_{m,max} - (1-\xi)s_m(t-1)) - d_m(t)|\}. \quad (4.31)$$

to the maximum possible amount of power transmitted to the Macrogrid from MG m, and the second term is the maximum possible amount of power that MG m can accept from the Macrogrid. The estimate of $p_{m,max}(t)$ is based on the power dispatching information of the last time slot. This works well for a short operation cycle, e.g., 15 minutes, because between two adjacent short time cycles, major grid related parameters such as generation and demand are usually closely correlated, while 15-min cycles are sufficient for power scheduling in a large Macrogrid with several MGs under current technology of information processing and communications.

The complexity of Algorithm 1 is related to the number of users and MGs, and the number of calculations solving the KKT equations of Prob-MA3 (4.16). According to [98], the complexity of Algorithm 1 is roughly $\mathcal{O}((N+M)^3)$. The complexity of Algorithm 2 and 3 is related to the product of the number of iterations of the dual variables and the number of calculations solving Prob-MG2 (4.26). And the complexity of Algorithm 2 and 3 is about $\mathcal{O}((2M)^3 \cdot M^3) = \mathcal{O}(M^6)$. The complexity analysis is quite conservative and thus, the complexity of the proposed algorithm is polynomial of the number of Macrogrid users and MGs, which can be processed easily within a 15-min cycle, based on the processing ability of current micro-computers.

The proposed cooperative distributed algorithm is also well suited for larger power systems due to the scalability. It is also worth noting that there is no information exchange directly between the MC and MGCCs. The MGNC connects the MC and MGCCs in the system so that there is only one information connection point between the Macrogrid and the MGNC, which increases the level of security and privacy protection.

4.6 Performance Evaluation

In this section, we present a trace-driven simulation study to evaluate the efficacy of the proposed HPS scheme. The simulation data and parameters for the Macrogrid are based on the power usage traces in the Southern California Edison (SCE) area recorded in 2011 [99]. The data for MGs are based on some statistical distributions, which are averaged over a large number of random runs.

We consider a power system as in Fig. 4.1 with a Macrogrid and four Microgrids. The Macrogrid supports 400 power users, while each MG supports 100 users. The user demand is based on the SCE trace and user utility function is defined as [74]

$$U(p_i(t), \omega_i(t)) = \begin{cases} \omega_i(t) \cdot p_i(t) - p_i(t)^2/8, \\ \qquad\qquad \text{if } 0 \leq p_i(t) \leq 4\omega_i(t) \\ 4 \cdot \omega_i(t), \qquad \text{if } p_i(t) \geq 4\omega_i(t). \end{cases}$$

As in [131], the power generated in each MG is independently chosen from a uniform distribution in [10 KW, 450 KW]. The generation cost function in the Macrogrid is assumed to be $G(x) = 4x^2$ and the MG generation cost function is $G_k(x) = 1.5x^2$. The transmission cost between the Macrogrid and MG m is assumed to be $C_m(x) = \theta_m x^2$; the transmission cost between MG k, and m is $C_{km}(x) = \frac{\theta_{km}}{2}x^2$, for $k \neq m$ and $C_{mm}(x) = 0$, for all m [125, 127]. θ_m and θ_{km} are transmission cost coefficients, which are defined as: $\theta_m = 1 - \sigma_m$, if $x > 0$, and $\theta_m = \frac{1-\sigma_m}{\sigma_m}$ otherwise; $\theta_{km} = \frac{1-\sigma_{km}}{\sigma_{km}}$.

In practice, the transmission loss ratios σ_m and σ_{km} usually differ from MG to MG because different factors such as distances. For practical considerations, we assume that $\{\sigma_{01}, \sigma_{02}, \sigma_{03}, \sigma_{04}\} = \{0.2, 0.6, 0.5, 0.3\}$, $\{\sigma_{10}, \sigma_{20}, \sigma_{30}, \sigma_{40}\} = \{0.5, 0.4, 0.67, 0.59\}$, $\sigma_{12} = 0.56$, $\sigma_{13} = 0.71$, $\sigma_{14} = 0.67$, $\sigma_{21} = 0.56$, $\sigma_{23} = 0.56$, $\sigma_{24} = 0.63$, $\sigma_{31} = 0.71$, $\sigma_{32} = 0.56$, $\sigma_{34} = 0.56$, $\sigma_{41} = 0.67$, $\sigma_{42} = 0.63$, $\sigma_{43} = 0.56$, and $\sigma_{mm} = 1$, for all m. The corresponding transmission cost coefficients are $\theta_{12} = 0.8$, $\theta_{13} = 0.4$, $\theta_{14} = 0.5$, $\theta_{21} = 0.8$, $\theta_{23} = 0.8$, $\theta_{24} = 0.6$, $\theta_{31} = 0.4$,

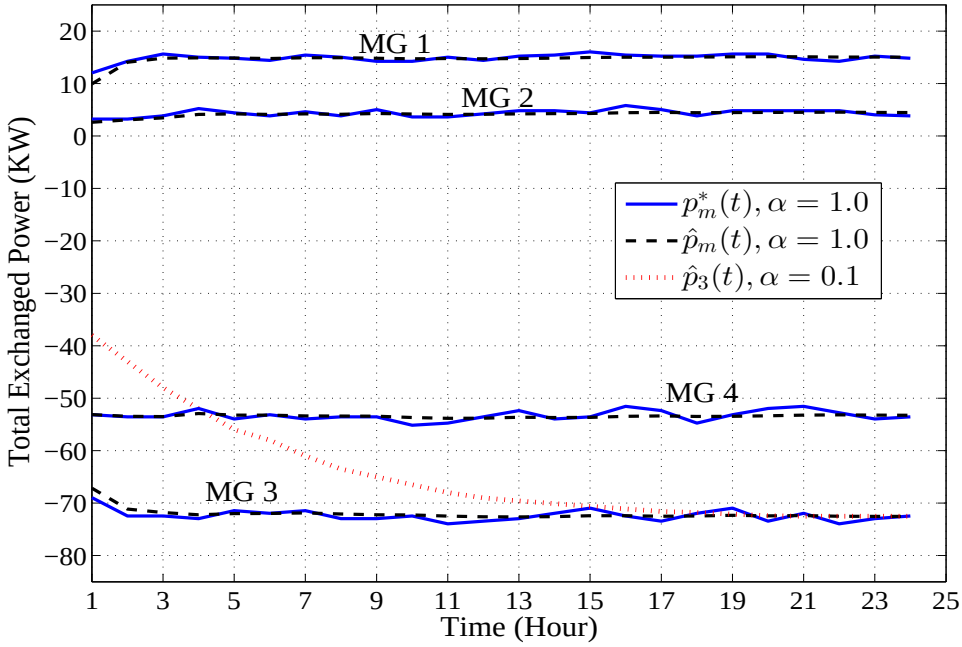

Figure 4.2: Convergence of $\hat{p}_m(t)$.

$\theta_{32} = 0.8$, $\theta_{34} = 0.8$, $\theta_{41} = 0.5$, $\theta_{42} = 0.6$, $\theta_{43} = 0.8$, and $\theta_{mm} = 1$, for all m. We also assume different storage capacities to the MGs as: $(s_{1,min}, s_{1,max}) = (30, 300)$ KW, $(s_{2,min}, s_{2,max}) = (50, 100)$ KW, $(s_{3,min}, s_{3,max}) = (60, 500)$ KW, and $(s_{4,min}, s_{4,max}) = (100, 200)$ KW. The algorithms are executed on 15-minute time slots.

4.6.1 HPS Performance

The HPS algorithms contain two iterative sequences: (i) $d_i(t)$ and $p_m(t)$ in the Macrogrid control; (ii) $p_{km,t}(j)$, $\vec{\lambda}_{M,t}(j)$ and $\vec{\beta}_{M,t}(j)$ in the MG control. In the first tier, the convergence of $d_i(t)$ and $p_m(t)$ is over multiple time slots; while in the second tier, the convergence is achieved within every time slot. Because the second tier control requires the solution $p^*_m(t)$ from the first tier control as a constraint, the convergence of $p_m(t)$ is critical for HPS performance. The parameter α in the updating equations (4.17) should be carefully selected. For the second tier, the MGNC needs to exchange $\vec{\lambda}_{M,t}(j)$ and $\vec{\beta}_{M,t}(j)$ for several iterations until they converge. For given termination conditions, the step-size parameters δ and τ will affect the speed of convergence of $\vec{\lambda}_M(t)$ and $\vec{\beta}_M(t)$. Intuitively, small δ and τ guarantee the

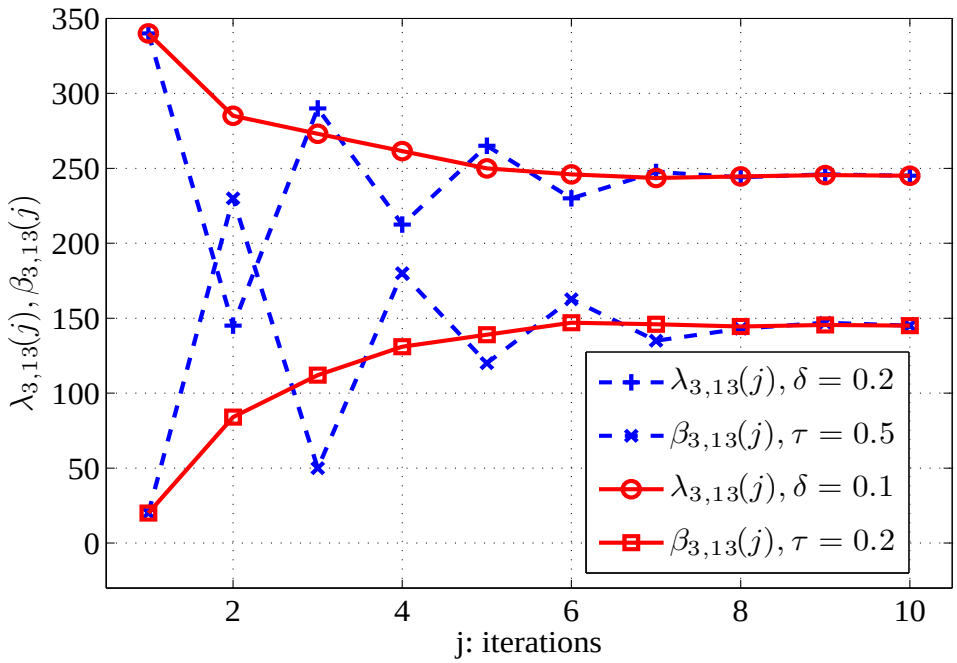

Figure 4.3: Convergence of $\lambda_{3,13}(j)$ and $\beta_{3,13}(j)$ with different δ and τ.

convergence, but may require more iterations. We illustrate the effect of the parameters on convergence in Figs. 4.2 and 4.3.

Fig. 4.2 shows that $\hat{p}_m(t)$ converges in every MG, and $p_m^*(t)$ fluctuates around $\hat{p}_m(t)$, as specified in Lemma 4.1. From the two sequences of $\hat{p}_3(t)$ with $\alpha = 1$ and $\alpha = 0.1$, it can be seen that a smaller α results in slower convergence. However, a large α may lead to larger variance in the transient phase. From a larger number of simulation runs, we set $\alpha = 1$ in our simulations. It is also worth noting that the MGs have different levels of $p_m^*(t)$. This is because different MGs have different generation levels, storage limits, and transmission cost coefficients. For example, $p_1^*(t)$ has a positive level of 13.5 KW, which means MG 1 requires a 13.5 KW load from the Macrogrid. This may due to low generation, small storage, or large transmission cost with the Macrogrid. On the other hand, $p_3^*(t)$ has a negative level of -70 KW, which means MG 3 transfers 70 KW to the Macrogrid. Furthermore, the sum of $p_m^*(t)$'s is negative, meaning that the Macrogrid acquires power from the MG network in this time frame.

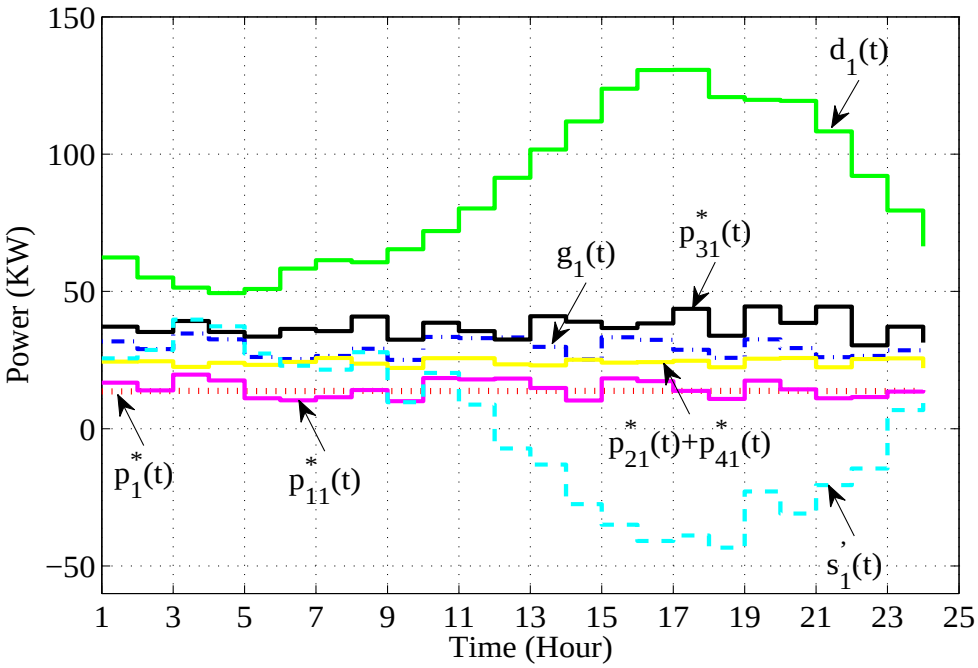

Figure 4.4: Power scheduling in MG 1.

In Fig. 4.3, the evolutions of $\lambda_{3,13}(j)$ and $\beta_{3,13}(j)$ in a time slot are plotted with different step sizes δ and τ. The curves with larger step sizes have larger variances and slightly slower convergence speed. Both $\lambda_{3,13}(j)$ and $\beta_{3,13}(j)$ have a very fast convergence in 6 to 8 iterations, which also indicates a fast convergence of $p^*_{3m}(t)$ as stated in Section 4.4.2. The fast convergence is due to the transformed constraints (4.24) of Prob-MG1, which further restricts the set of feasible $p^*_{km}(t)$'s. As a result, $\lambda_m(t)$ and $\beta_m(t)$ are forced to increase or decrease in a reverse direction, which reduces the number of iterations needed for convergence.

In addition to convergence performance, HPS can be further evaluated with respect to power scheduling. In Figs. 4.4 and 4.5, we present the power flows in MGs 1 and 3, respectively. We find in Fig. 4.4 that MG 1 has a very low power generation $g_1(t)$ from about 20 KW to 40 KW, such that it cannot support the power demand in the range of 50 KW to 130 KW with its own generation alone. However, it only requests less than 20 KW from the Macrogrid, but accepts more than 40KW from MG 3 and 20KW from MGs 2 and 4. Note that the transmission cost coefficient between the Macrogrid and MG 1 is $\theta_1 = 1 - \sigma_{01} = 0.8$, which is larger than the coefficients between MG 1 and the other MGs.

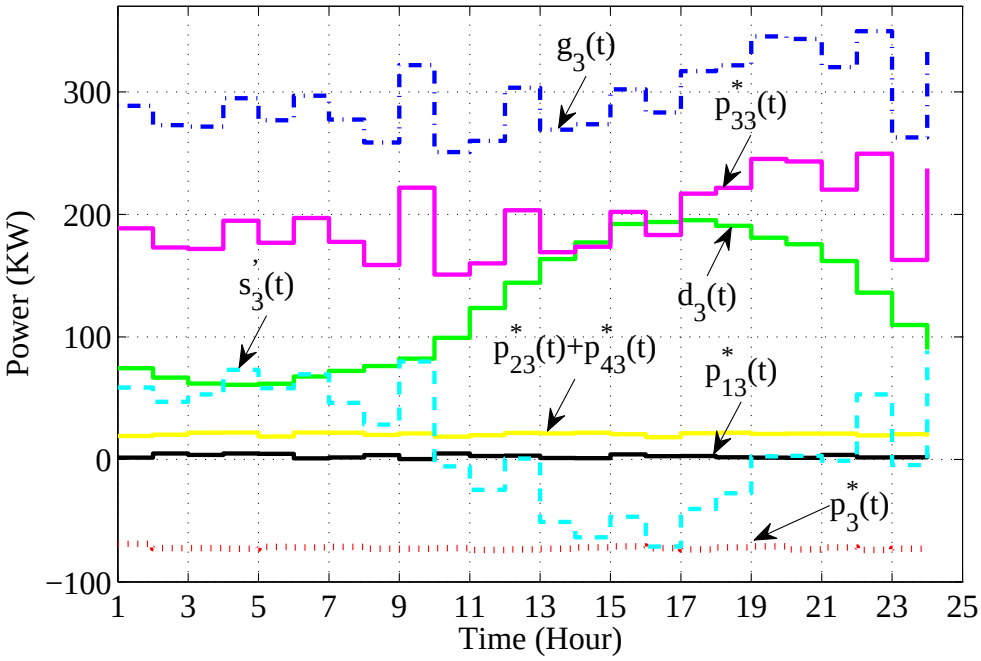

Figure 4.5: Power scheduling in MG 3.

To find a sufficient power flow to support its users, while keeping the cost low, MG 1 chooses to request more power from MG 3. MG 3 has a large generation and storage capacity, so that it can share much power with MG 1 and the Macrogrid. However, we observe a relatively low power flow around 20 KW between MG 3 and MG 2 and 4. This can also be explained by the objective to minimize the transmission cost. Actually, $\frac{\theta_{32}}{2}$ and $\frac{\theta_{34}}{2}$ are both 0.4, which is very close to $\theta_3 = \frac{1-\sigma_{30}}{\sigma_{30}} = 0.5$.

In real power systems, the transmission cost ratio σ and coefficient θ is usually affected simultaneously by many different factors, such as distance, the power gap, and the complexity of the system. Thus, in a real system, MG 1 may have shorter distance and smaller gap of power level with MG 3 compared to that with MG 2 and 4. The storage also plays an important role as a power buffer to enhance system stability and capacity. As a result, with all the key factors considered, HPS is able to achieve a balance in the power system, maximize the Macrogrid user utilities, smooth the load of the Macrogrid, and minimize the cost in the cooperative MGs.

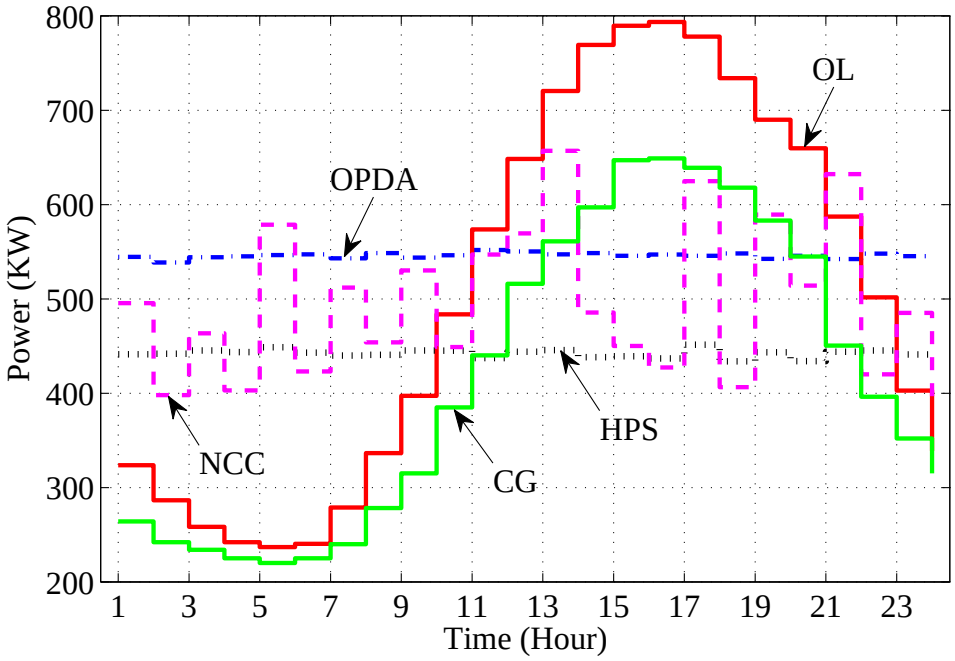

Figure 4.6: Macrogrid load under different power scheduling schemes.

4.6.2 Comparison with Existing Schemes

We next provide a comparison study of HPS versus several existing schemes [74, 125]. In Fig. 4.6, we show the load of the Macrogrid under five different power scheduling schemes. The original load (OL) is based on the SCE trace of a one day period in September, 2011. The online power distribution algorithm (OPDA) proposed in [74] has considered many factors including user utility and grid load variance in a Macrogrid, but no MG is involved in the model and algorithm. Thus, the OPDA curve in Fig. 4.6 is obtained by running OPDA in a Macrogrid with 800 users. A coalition game (CG) is used in [125] to minimize the power loss in an MG network, where power flows between MGs and Macrogrid are allowed. However, it does not consider smoothing the Macrogrid load. For comparison purpose, we also develop another scheme (termed as no cooperation control scheme (NCC)), which only allows power flow between each MG and the Macrogrid, while power trading among the MGs is not allowed.

In Fig. 4.6 , the OL curve has the largest peak. OPDA achieves an expected smooth load in the Macrogrid, because the variance term is explicitly minimized. OPDA also distributes power well such that the user demands can be satisfied. However, without the cooperative MGs, the Macrogrid generates 7.8% more power in total to achieve 31.7% of peak reduction. With NCC, the Macrogrid can exchange power with the MGs, which usually have random power generations. As a result, power generation in the Macrogrid under NCC is almost the same as that under OL. NCC achieves 17.2% peak reduction while causing 21.4% increase in the Macrogrid grid load variance. For CG, coalitions are formed among the MGs to minimize power loss. We find a 15.7% of generation reduction, 18.4% of peak reduction, and 40.9% of variance reduction. It achieves a fairly good result by exploiting the distributed generation from the cooperative MGs. However, it does not explicitly consider variance reduction; the resulting variance is actually still large.

With HPS, we jointly consider all the above factors. As a result, under HPS, the Macrogrid has a 97.1% variance reduction, 43.1% of peak reduction, and 13.1% of generation reduction. Compared to CG, HPS achieves considerably better results on peak reduction and variance reduction, and the Macrogrid generation reduction is only slightly lower (i.e., 13.1% versus 15.7% with CG).

4.7 Related Work

In a single MG, research works cover several main topics, including interface or coupling between an MG and the Macrogrid, DER dispatching and power support, and energy management [71, 72, 117–119]. In [71], the MG control strategies and energy management are examined from several aspects. In [117], a detailed report is presented to test the building and management of a hydrogen MG in Spain in a simple and reliable way. In [72], the authors present a control operation for a centralized controller for MGs, which maximizes its value by optimizing the production of local DGs and power exchanges with the main

distribution grid during interconnected operation. In [73], the authors introduce an economic power dispatching scheme for stable operation of an MG, while a multi-agent system is presented in [118] for DER energy management in an MG. The authors of [119] propose a multivariable digital control design methodology for the voltage regulation of an islanded single distributed generation (DG) unit MG and its dedicated load.

The problem of cooperative MGs has been considered in several recent papers [126, 127, 129, 130]. In [126], the authors present a decentralized control strategy modeling the MGs as a team of cooperative agents to minimize the costs of energy storage and the power exchanged among the MGs. The authors of [127] propose a game theoretic coalition to optimally reduce the total power losses in a MGs power system with power storage devices, and demonstrate the overhead of communications. In [129], the authors formulate the optimal decision making problem in cooperative MG networks as a linear quadratic Gaussian problem. There have been some recent works that consider the power flow between the MGs and the Macrogrid. For instance in [130], the authors present an optimal energy management framework for a cooperative network of heterogeneous MGs to achieve an efficient tradeoff between low operation cost and good energy service for customers.

4.8 Conclusion

In this chapter, we developed a hierarchical power scheduling scheme to optimally manage the power distribution in the smart grid with one Macrogrid and cooperative MGs. We first presented a formulation considering both the Macrogrid, which jointly considers user utility, generation cost, transmission cost, and grid load smoothing, and the MGs, which aims to minimize the cost of power generation and transmission within the MGs. We then decompose the problem into a two tier formulation and developed the corresponding online and distributed algorithms for solving both problems, which were proven to be asymptotically optimal. The proposed algorithms were validated with trace-driven simulations.

Chapter 5

Analysis of Solar Generation in Smart Grid with Simultaneous Inference of Nonlinear Time Series

5.1 Introduction

In recent years, as the development of the modern technologies in informatics, communication, control and computing, our living environment is becoming "smart." Smart Home and Smart City have gradually become part of our lives, and are no longer merely future concepts for the public. An important component of Smart City is the Smart Grid (SG), which is regarded as the next generation power grid to create a widely distributed energy generation and delivery network. The SG features the incorporation of power generation from renewable energy sources, especially solar and wind, which meanwhile requires a better energy management system in the SG [44].

As stated in previous chapters, energy management in SG has been studied in many previous works [44,76,132]. It is indicated in [132] that high efficient power management cannot be realized without a better forecast on the grid load and renewable power generation in SG. The problem of grid load forecasting has been studied by many researchers with different techniques such as state space models [133], Artificial neural networks and support vector machine [134], and nonparametric functional time series analysis [132]. And the prediction on the renewable energy generation in SG has also attracted some interests. Predictions on solar and wind generation can be found in [135] and [136] using support vector machines (SVM) regression and joint probability density function (JPDF) forecast respectively. Because of the weather dependence nature of the forecasting problem, statistical methods can be found in almost every related literature. On the other hand, the wide range of applications helps the improvement of the statistical theory on nonparametric analysis [137–139],

and non-stationary time-series analysis [140, 141]. Recently, the authors of [142] propose the method of constructing a simultaneous confidence bands for time-varying coefficients. These researches improve the understanding of the time-series, and provides better techniques in renewable energy generation prediction.

The previous work on predicting solar power generation [135] provides acceptable results using SVM regression. However, by simply trying different SVM kernels after some basic data processing statistically, it lacks a deep analysis of the solar power generation and weather data, and thus is limited in precise predictions of other data set. For example, the check of assumptions is missing on independence of variables and errors. Furthermore, the renewable energy generation is a function of weather variables, and is a stochastic process on nonlinear time series. Therefore, the associations between the weather variables and the power generation should be analyzed over a long time for a comprehensive understanding of their dynamic relations. For example, a coefficient varying by time overall may stay constant for short periods. These drawbacks will limit the applications for predictions in other cases. Therefore, a method that can show the dynamic property of the process is highly demanded for better predictions on solar power generation in different cases.

Motivated by this observation, we introduce simultaneous inference of nonlinear time series proposed in [142] for understanding comprehensively the deep and dynamic relationship between renewable power generation process and the weather variable processes. The simultaneous inference is based on the simultaneous confidence bands (SCB) [143] of time-varying coefficients in the local linear model, which we use for nonlinear time series analysis. It is based on the assumption of nonstationary processes for the error and weather variables, which matches the case of our problem where many weather variables are shown to have an obvious seasonal pattern, meaning the observations are not stationary assumed by many other forecast techniques. And the SCB shows the confidence bands of the coefficients over any length of time, which can be used to test if the coefficients are truly time-varying or not. This helps us to refine the model by omitting variables which that are not significant.

The main contribution of this chapter is the introduction of the local linear model and SCB for analyzing the solar power generation as a nonlinear time series. By checking the dynamic properties of the coefficients from its SCB, we are able to achieve a more comprehensive understanding of the model, and based on this, we can further refine the model and use it for predicting the renewable energy generation. As an example, we apply it in predictions on daily solar power generation. This method has a wide range of applications, which is not limited to analyze and predict the solar energy generation. It can also be used for predictions in different time scales, from minutes to months, depending on different purposes, and in other cases, such as wind power generation prediction.

The remainder of this chapter is organized as follows. We present the local linear model for Nonlinear time-series analysis in Section 5.2. The construction of SCB for time-varying coefficients with simulated results is introduced in Section 5.3.We use the simultaneous inference for analyzing a trace of solar intensity and weather data in Section 5.4, and review related works in Section 5.5. Section 5.6 concludes this chapter.

5.2 Local Linear Model for Nonlinear Time Series

We consider the power generation from a renewable source as a continuous-time stochastic process $Y(t)$ and a function of the meteorological variables $\vec{X}(t)$, which is a continuous-time covariate process. It follows that

$$Y(t) = f(\vec{X}(t)), t \in \mathbb{R}.$$

To identify the function $f(\cdot)$, it is straightforward to try a linear model first, as

$$Y(t) = \vec{X}^T(t)\vec{\beta}(t) + \epsilon(t), t \in \mathbb{R}, \tag{5.1}$$

where $\vec{X}(t) = (1, X_1(t), ..., X_{p-1}(t))^T$ and $\vec{\beta}(t) = (\beta_1(t), ..., \beta_p(t))^T$ are both $p \times 1$ vectors, and $\epsilon(t)$ is the error at time t. To use this model, we need to predict the regression coefficients

$\vec{\beta}(t)$ at each time t, so that we can calculate $Y(t)$ given the forecast on weather variables $\vec{X}(t)$. It is not difficult to obtain $\vec{\beta}(t)$ using parametric smoothing methods such as multiple linear regression. Although this model can indicate a certain level of interactions between the variables $\vec{X}(t)$ and the response $Y(t)$, it cannot be used to represent the real underlying process, especially when the method treat continuous-time series simply as discrete data points. As we will show in Section 5.4.4, the prediction with the linear model is not satisfactory, and cannot be used in practice.

From the linear model, we notice that the linearity is fairly strong between $\vec{X}(t)$ and $Y(t)$ within a short time period, i.e., several days or a week. If we take advantage of this property and consider the process as continuous-time, the model would be closer to the real process. For t_i close to t, we can have $\vec{\beta}(t_i) \approx \vec{\beta}(t) + (t_i - t)\vec{\beta}'(t)$, where $\vec{\beta}'(t)$ is the derivative of $\vec{\beta}(t)$, and thus for any time t_i close to t, we have the local linear model as [144]

$$Y(t_i) \approx \vec{X}^T(t_i)(\vec{\beta}(t) + (t_i - t)\vec{\beta}'(t)) + \epsilon(t_i), t_i \in t \pm h, \tag{5.2}$$

where the bandwidth h is the size of the local neighborhood. This model divides the time series into periods and creates linear models using local data. This way, we treat the data as a continuous-time series, and exploit the strong correlations between close time periods in weather dependent systems.

5.2.1 Local Linear Estimation

To identify the time-varying coefficients $\vec{\beta}(t)$, the least squares method for linear regression can be used. We also add some weights on the terms considering that contributions from different neighbors are different, which means a closer neighbor would have a stronger effect, while a further neighbor weaker effect. Usually a kernel function $K(\cdot)$ is assigned to each point, which is a symmetric density function defined on [-1,1] [144]. Here, we use a

popular Epanechnikov kernel.

$$K(a) = \begin{cases} 3(1-a^2)/4, & \text{if } |a| \leq 1 \\ 0, & \text{if } |a| > 1, \end{cases}$$

which decays fast for remote data point. We then have the following weighted least squares problem to solve,

$$\underset{\vec{\beta}(t),\vec{\beta}'(t)\in\mathbb{R}^p}{\arg\min} : \sum_{t_i \in t \pm h} (Y(t_i) - \vec{X}^T(t_i)(\vec{\beta}(t) - (t_i - t)\vec{\beta}'(t)))K\left(\frac{t_i - t}{h}\right). \tag{5.3}$$

At each time t, we solve for the coefficients $\hat{\vec{\beta}}_h(t)$ and $\hat{\vec{\beta}}'_h(t)$ under the bandwidth h. Suppose the total number of observations is n, we can pick t_i simply as $t_i = i/n, 1 \leq i \leq n$, and denote $Y(t_i)$ as y_i and $\vec{X}(t_i)$ as $\vec{x}_i$. From [138], we can solve (5.3) by calculating the following matrices $\mathbf{S}_k(t)$ and $\mathbf{R}_k(t)$:

$$\mathbf{S}_k(t) = \sum_{i=1}^{n} \vec{x}_i \vec{x}_i^T \left(\frac{t_i - t}{h}\right)^k K\left(\frac{t_i - t}{h}\right)/(nh) \tag{5.4}$$

$$\mathbf{R}_k(t) = \sum_{i=1}^{n} \vec{x}_i y_i \left(\frac{t_i - t}{h}\right)^k K\left(\frac{t_i - t}{h}\right)/(nh), \tag{5.5}$$

where $k = 0, 1, 2, ...$ We then have

$$\begin{pmatrix} \hat{\vec{\beta}}_h(t) \\ h\hat{\vec{\beta}}'_h(t) \end{pmatrix} = \begin{pmatrix} \mathbf{S}_0(t) & \mathbf{S}_1^T(t) \\ \mathbf{S}_1(t) & \mathbf{S}_2(t) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{R}_0(t) \\ \mathbf{R}_1(t) \end{pmatrix}. \tag{5.6}$$

5.2.2 Selection of Bandwidth

To solve problem (5.3) for the complete model using (5.4) to (5.6), we need to first fix bandwidth h. As discussed, h is the bandwidth determining the size of data used to estimate for a local linear model at time t. If h is too small, many useful points are not included for estimation, which may increases variance; if it is too large, more remote points are included,

which increases the computational complexity and cause large bias of the model. Therefore, it is important to choose a proper h.

Popular bandwidth selection techniques can be found in [144, 145] for different applications. The techniques are different for constant bandwidth and variable bandwidth. For constant bandwidth selection considered in our model, we adopt the generalized cross-validation (GCV) technique [145], which is suitable for a wide range of applications.

Similar to multiple linear regression, the coefficients $\vec{\beta}$ are estimated from the observed data $\vec{Y}$ and $\vec{X}$. Thus, a square hat matrix $\mathbf{H}(h)$ exists for $\hat{\vec{Y}} = \mathbf{H}(h)\vec{Y}$ [146], depending on the bandwidth h. Then we can choose the bandwidth h by

$$\hat{h} = \arg\min \left\{ \frac{|\hat{\vec{Y}} - \vec{Y}|^2}{n(1 - \mathrm{tr}\{\mathbf{H}(h)\}/n)^2} \right\}, \tag{5.7}$$

where, $\mathrm{tr}(\cdot)$ is the trace of the matrix, and n is the number of total observations.

5.3 Simultaneous Confidence Band for Time-varying Coefficients

In this section, we introduce the basic conditions and construction of SCB method proposed in [142], and then discuss its implications to further understand the modeling and predicting for the power generation process from the renewable energy sources based on the weather data.

5.3.1 Model Assumptions and Asymptotic Normality

Different from most current models for time series, the approach of SCB analysis assumes locally stationary processes for both $\vec{X}(t)$ and $\epsilon(t)$ [141]. The locally stationary process guarantees the stationary property for local time series, and is useful for local linear estimation. It actually belongs to a special class of non-stationary time series as

$$\vec{x}_i = \vec{G}(t_i, F_i), \quad \epsilon_i = H(t_i, F_i), \quad i = 1, 2, ..., n, \tag{5.8}$$

where $\vec{G}(t_i, F_i)$ and $H(t_i, F_i)$ are measurable functions well defined on $t_i \in [0,1]$, $F_i = (..., \xi_{i-1}, \xi_i)$ with $\{\xi_i\}_{i\in\mathbb{Z}}$ are independent and identically distributed (i.i.d.) random variables, and $\mathbb{E}(\epsilon_i|\vec{x}_i) = 0$. In our model for renewable power generation processes, we further assume that $\{\epsilon_i\}_{i\in\mathbb{Z}}$ are i.i.d. and dependent of $\{\xi_i\}_{i\in\mathbb{Z}}$.

Based on the above assumptions, the central limit theorem for $\hat{\vec{\beta}}(t)$ states that: supposing $nh \to \infty$ and $nh^7 \to 0$ [142], then

$$(nh)^{1/2}\{\hat{\vec{\beta}}(t) - \vec{\beta}(t) - h^2\vec{\beta}''(t)/10\} \to N\{0, \Sigma^2(t)\}, t \in (0,1) \tag{5.9}$$

where

$$\mu = \int_{\mathbb{R}} x^2 K(x)\, dx, \tag{5.10}$$

$$\Sigma(t) = (M^{-1}(t)\Lambda(t)M^{-1}(t))^{1/2}, \tag{5.11}$$

$$M(t) = \mathbb{E}(\vec{G}(t, F_0)\vec{G}(t, F_0)^T). \tag{5.12}$$

The covariance matrix $\Lambda(t)$ can be further approximated using techniques proposed in Section 5.3.2.

5.3.2 Simultaneous Confidence Band

Deriving from the central limit property and basic assumptions shown above, the $100(1-\alpha)\%$ asymptotic simultaneous confidence tube of $\vec{\beta}_C(t)$ can be constructed using the following formula:

$$\tilde{\beta}_{C,\tilde{h}}(t) + \hat{q}_{1-\alpha}\hat{\Sigma}_C(t)\mathcal{B}_s, \tag{5.13}$$

where $\tilde{\beta}_{C,\tilde{h}}(t)$ is the bias corrected estimator defined in (5.14), $\mathcal{B}_s = \{\vec{z} \in \mathbb{R}^s : |\vec{z}| \leq 1\}$ is the unit ball, and s is the rank of a matrix $C_{p\times s}$, which we use for choosing different linear combinations of $\beta(t)$, and $\vec{\beta}_C(t) = C^T\vec{\beta}(t)$. To obtain the SCB, we simply take $s = 1$ in (5.13), and the SCB is constructed similarly to the confidence interval of the coefficients

of the multiple linear regression: $\hat{\beta} \pm t_{\alpha/2,n-p} se(\hat{\beta})$, where $se(\hat{\beta})$ is the standard error of $\hat{\beta}$, and $t_{\alpha/2,n-p}$ is the upper $\alpha/2$ percentage point of the t_{n-2} distribution [146].

Similarly, the first term is the estimator of the time-varying coefficients corrected for bias by

$$\tilde{\beta}_{C,\tilde{h}}(t) = C^T \tilde{\beta}_{\tilde{h}}(t) = C^T \left(2\tilde{\vec{\beta}}_{\tilde{h}/\sqrt{2}}(t) - \hat{\vec{\beta}}_{\tilde{h}}(t) \right), \tag{5.14}$$

where the corrected estimator $\tilde{\beta}_{\tilde{h}}(t)$ can also be acquired by solving (5.3) using an corresponding kernel function $K^*(a) = 2\sqrt{2}K(\sqrt{2}a) - K(a)$ and an updated bandwidth $\tilde{h} = 2\hat{h}$ of the GCV selector $\hat{h}$.

The second term in (5.13) $\hat{q}_{1-\alpha}$ is actually the upper $\alpha/2$ percentage point of the normal distribution $N\{0, \Sigma^2(t)\}$ defined in (5.9), while the third term $\hat{\Sigma}_C(t)$ is the estimated stand error. The method of wild bootstrap is applied to obtain $\hat{q}_{1-\alpha}$. Firstly, generate a large number i.i.d. vectors $\vec{v}_1, \vec{v}_2, ...,\ N(0, I_s)$, where $\vec{v}_i \in \mathbb{R}^p$ and I_s denotes the $s \times s$ identity matrix, and then calculate $q = sup_{0 \le t \le 1} | \sum_{i=1}^n \vec{v}_i K^*((t_i - t)/\tilde{h})/(n\tilde{h})|$; repeat the previous step for a large number of times (say, 5000) to acquire the estimated $100(1-\alpha)\%$ quantile $\hat{q}_{1-\alpha}$ of q.

The estimate of the stand error in (5.13), $\hat{\Sigma}_C(t)$ is defined similarly as (5.11):

$$\hat{\Sigma}_C(t) = (C^T \hat{M}^{-1}(t) \hat{\Lambda}(t) \hat{M}^{-1}(t) C)^{1/2}, \tag{5.15}$$

and we shall estimate $\hat{M}(t)$ and $\hat{\Lambda}(t)$ respectively. From the definition of $M(t)$ in (5.12), it can be estimated by $\hat{M}(t) = S_0(t^*)$, where $S_0(\cdot)$ is defined in (5.4), and $t^* = max\{h, min(t, 1-h)\}$. To obtain $\hat{\Lambda}(t)$, we first define two $p \times 1$ vectors $\vec{Z}_i = \vec{x}_i \hat{\epsilon}_i$ and $\vec{W}_i = \sum_{j=-m}^m \vec{Z}_{i+j}$, a matrix $\Omega_i = \vec{W}_i \vec{W}_i^T/(2m+1)$, and a function $g(t,i) = K((t_i - t)/\tau)/\sum_{k=1}^n K((t_k - t)$, where m and τ can be simply chosen as $m = \lfloor n^{2/7} \rfloor$ and $\tau = n^{-1/7}$. Then $\hat{\Lambda}(t)$ can be calculated by

$$\hat{\Lambda}(t) = \sum_{i=1}^n g(t,i) \Omega_i. \tag{5.16}$$

Algorithm 5.1: Construction of SCB for Time-varying Coefficients

1 Find a proper bandwidth $\hat{h}$ from the GCV selector (5.7);
2 Let $\tilde{h} = 2\hat{h}$ and calculate $\tilde{\beta}_{C,\tilde{h}}(t)$ using (5.14) and (5.3);
3 Obtain the estimated $(1-\alpha)$th quantile $\hat{q}_{1-\alpha}$ via the bootstrap method;
4 Estimate $\hat{M}(t) = S_0(t^*)$ and $\hat{\Lambda}(t)$ by (5.16), and calculate $\hat{\Sigma}_C(t)$ according to (5.15);
5 Construct the $100(1-\alpha)\%$ SCB of $\vec{\beta}_C(t)$ using (5.13).

This way, we are able to calculate the SCB using all the estimates. The above steps for constructing the SCB are summarized in Algorithm 5.1.

5.3.3 Further Discussions

To make a better estimation and prediction using the local linear model, we need to understand the coefficients $\vec{\beta}(t)$ comprehensively. As our aim is to predict on a continuous-time process, we must learn more about the dynamics of the model, and especially, the time-varying coefficients $\vec{\beta}(t)$. This way, the predicting results are meaningful at any time point. Considering this, we apply the SCB analysis into predicting the power generation process based on the weather data.

Firstly, the SCB provides a dynamic and comprehensive view on $\vec{\beta}(t)$. In simple linear regression, the confidence interval provides a measure of the overall quality of the regression line [146]. Similarly, the SCB illustrates the overall pattern of $\vec{\beta}(t)$ and thus the accuracy of the model. Confidence bands with smaller width implies a better model with smaller variability, while too wide confidence bands are limited in use. Note that the SCB is constructed under an complete analysis on the continuous-time assumption, which is not merely the connections of the pointwise confidence intervals on different time points.

The SCB can also be used to test whether the coefficients $\vec{\beta}(t)$ are truly time-varying or not. If a horizontal line is covered by the SCB of a $\beta_k(t)$, we accept the hypothesis that $\beta_k(t)$ is constant and not time-varying. Furthermore, in different cases, we can construct the SCB for different linear combinations of $\beta_k(t)$s by setting different matrix $C_{p\times s}$. For example, if we set $C_{p\times 1} = [1, 1, 0, ...0]$, we get the SCB of $\vec{\beta}_C(t) = C^T\vec{\beta}(t) = \beta_1(t) + \beta_2(t)$; if we set

$C_{p\times 2} = [1, 0, 0, ...0; 0, 1, 0, ...0]$, the SCB of $\beta_1(t)$ and $\beta_2(t)$ turn to a tube as at any time t, because when $s = 2$, the unit ball $\mathcal{B}_2$ turn to a unit circle from a unit interval. This provides us a convenient way to further test the model.

5.3.4 Algorithm Performance for Simulated Processes

We now simulate a model with $\vec{X}(t)$ and $\epsilon(t)$ locally stationary processes discussed in Section 5.3.1, and construct the 90% and 95% SCB respectively for a given model, to test the correctness by comparing with the true results.

We use the following local linear model with time-varying coefficient:

$$y_i = \beta_1(i/n) + \beta_2(i/n)x_i + \epsilon_i, \tag{5.17}$$

where $\beta_1(t) = \cos(2\pi t)/4$, and $\beta_2(t) = \exp\{-(t - 1/2)^2\}/2$.

Define $H(t_i, F_i) = (1/2)\sum_{j=0}^{\infty} a(t_i)^j \xi_{i-j}$, $\vec{G}(t_i, F_i) = (1; \sum_{j=0}^{\infty} b(t)^j \varepsilon_{i-j})$, where ξ_k and ε_l are i.i.d. $N(0, 1)$. Then $\vec{x}_i$ and ϵ_i can be generated using (5.8), for $i = \{1, 2, ..., n\}$.

For the above setting, we generate 5000 samples of size 500, and for each sample SCB is constructed with bandwidths setting from 0.1 to 0.3 of step 0.025. We use 3000 and 5000 bootstrap samples to estimate $\hat{q}_{1-\alpha}$ for $\alpha = 0.1$ and $\alpha = 0.05$ to show the effect of the sample size on the results. The simulation results are shown in Table 5.1, where the coverage rate and width of SCB for $\beta_2(t)$ with different bandwidths h at 90% and 95% levels are listed. It shows that the coverage rate is close to the nominal level with most bandwidths.

And the bandwidth selected by GCV is 0.22, which yield fairly good results. It also shows that the 95% SCB is wider than the 90% and the width of SCB decreases as h increases, which indicates a better estimation for larger h. Besides, we notice that $\hat{q}_{1-\alpha}$ are affected by the bootstrap sample size, and its value affects the width of SCB directly. Therefore, for practical application, a large size of bootstrap samples is very important. According to our numerical studies, at least 5000 samples are suggested.

Table 5.1: The Coverage Probabilities of SCB for $\beta_2(t)$ and Quantiles of $\hat{q}$ at Nominal Level of 90% and 95%

	$\beta_2(t)$		3000 Samples		5000 Samples	
bandwidth	90%	95%	$\hat{q}_{0.90}$	$\hat{q}_{0.95}$	$\hat{q}_{0.90}$	$\hat{q}_{0.95}$
0.1	0.814	0.864	0.595	0.617	0.476	0.509
0.125	0.875	0.925	0.522	0.568	0.415	0.453
0.15	0.914	0.953	0.483	0.505	0.378	0.401
0.175	0.923	0.945	0.451	0.464	0.335	0.363
0.2	0.901	0.951	0.417	0.441	0.309	0.336
0.225	0.908	0.949	0.392	0.412	0.284	0.311
0.25	0.904	0.955	0.369	0.392	0.269	0.295
0.275	0.899	0.951	0.348	0.366	0.251	0.275
0.3	0.898	0.946	0.337	0.348	0.235	0.264

5.4 Application to Solar Energy Generation

In this section, we apply the SCB analysis to modeling the solar power generation process and predicting the generation based on the weather data, and compare results to other methods.

5.4.1 Data Description

As an application, we consider the data from the UMASS Trace Repository [147], which records the solar power generation by solar intensity in $watts/m^2$, and the data of several weather metrics from January, 2010 to February, 2013. It recorded the weather data every 5 minutes. Many weather parameters were observed in details. Here, we use five main variables of temperature, humidity, dew point, wind speed and precipitation. The data has been studied in [135], which studied the statistical connection between the weather variables and the solar power generation, and predicted the solar power generation using multiple linear regression and Support Vector Machines [148] regression. Our purpose is to investigate the

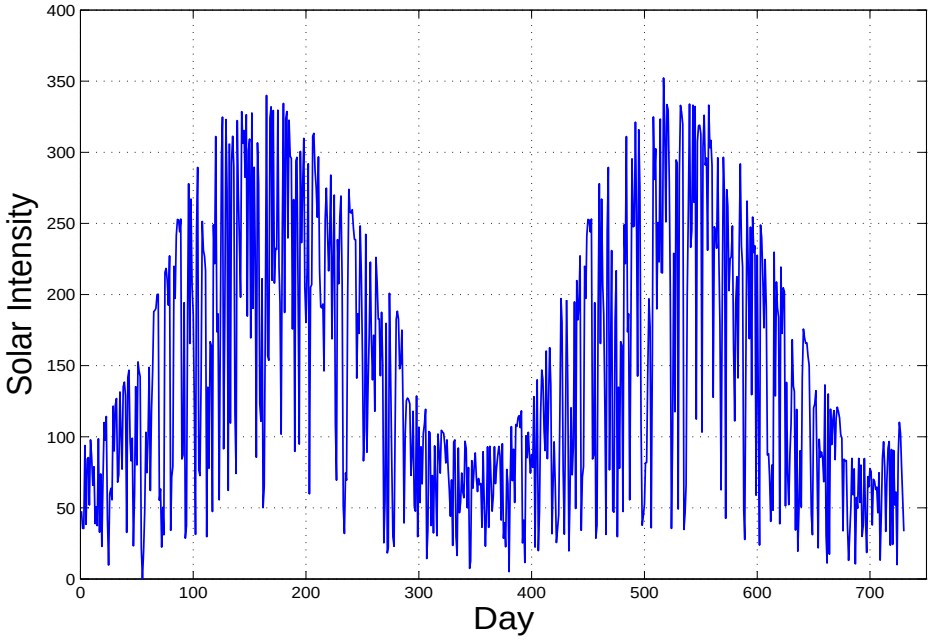

Figure 5.1: Daily Solar Intensity for 2011 and 2012.

dynamic association between the weather variables and the solar power generation, and to help better predict the solar power generation.

Here, we plot the daily solar intensity for 2011 and 2012 in Fig. 5.1. An apparent seasonal pattern is shown with the peak points in summer periods, and lowest points in winter time. It is helpful to consider and use the seasonal patterns for forecast. And it is interesting to see a similar pattern for daily observation. Similar patterns can also be seen for several weather variables, such as temperature, humidity, and dew point (See [135]). Fig. 5.1 also shows a strong relation between two days, which means the solar generation process is not i.i.d. As discussed in Section 5.3.1, we do not require i.i.d. of observations to construct the SCB.

5.4.2 Prediction Model

Based on (5.2), we use the following local linear model:

$$y_i = \beta_1(i/n) + \sum_{p=2}^{6} \beta_p(i/n) x_{p,i} + \epsilon_i, \quad \text{for } i = 1, ..., n, \tag{5.18}$$

where, y_i is the solar intensity, $x_{p,i}$, $p = 2, 3, 4, 5, 6$, represent the series of temperature in Fahrenheit, humidity in percentage, dew point in Fahrenheit, wind speed in miles per hour and precipitation in inches, respectively. We use $n = 730$ observations of 2011 and 2012 for local linear regression and the model is represented in a daily pattern. Note our model and analysis can be built on any time scale, we take daily pattern for an application example here. And $\beta_1(\cdot)$ is the intercept and $\beta_p(\cdot)$ are the associated coefficients for $x_{p,i}$.

5.4.3 Simultaneous Inference for Time-varying Coefficients

We now perform the SCB analysis. We center all the weather variables on their averages so that the intercept $\beta_1(\cdot)$ can be interpreted as the expected solar intensity. From GCV, we select the bandwidth $h = 0.25$. The 95% SCB of the coefficients $\beta_p(\cdot)$ are shown from Fig. 5.2 to 5.7. In each figure, the middle thick solid curve is the estimated series for the variable; the upper and lower solid curves are the envelops for the simultaneous confidence band for each variable. From the SCB, we are able to test whether a coefficients is significantly associated with the solar intensity, which equals to test:

$$H_0 : \beta_p(i/n) = 0, \forall i \in \{1, 2, ..., n\}; \text{v.s. } H_1 : \beta_p(i/n) \neq 0, \exists i \in \{1, 2, ..., n\}.$$

If the zero line is included in the SCB, we accept the hypothesis that the coefficient is not significant and could be omitted from the model; otherwise, we keep it in the model. We can also test whether the coefficients are constant, by attempting to include a constant horizontal line into the SCB. This is equal to testing:

$$H_0 : \beta_p(i/n) = c_p, \forall i \in \{1, 2, ..., n\}; \text{v.s. } H_1 : \beta_p(i/n) \neq c_0, \exists i \in \{1, 2, ..., n\},$$

where c_p is a constant of each p. If the line is covered, we accept that the coefficient is constant; otherwise, it is not. In Fig. 5.2, the curve indicates the expected solar intensity for two years, and illustrates an obvious seasonal pattern. The width of the 95% SCB of $\beta_1(t)$ is

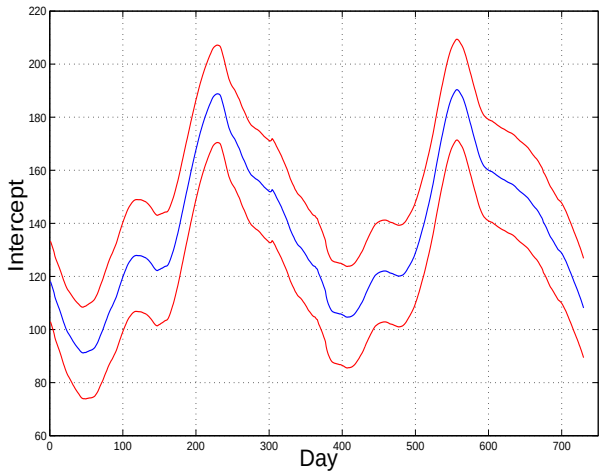

Figure 5.2: 95% SCB for Intercept.

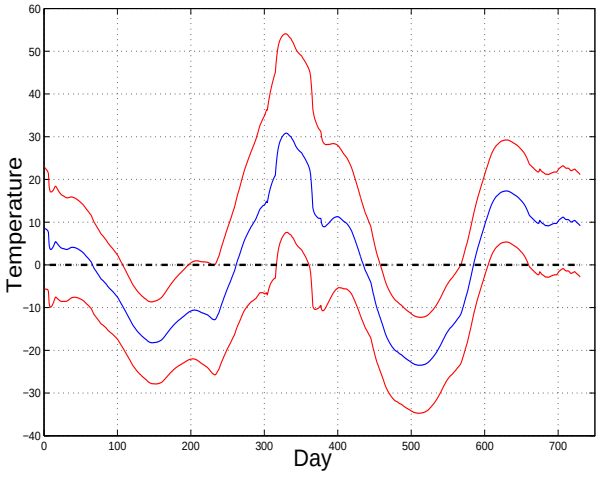

Figure 5.3: 95% SCB for Temperature.

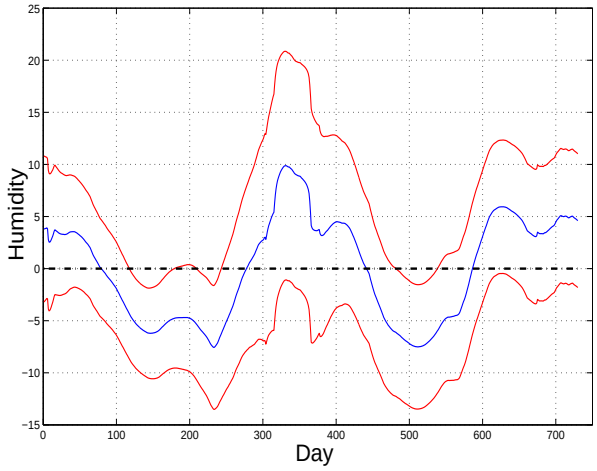

Figure 5.4: 95% SCB for Humidity.

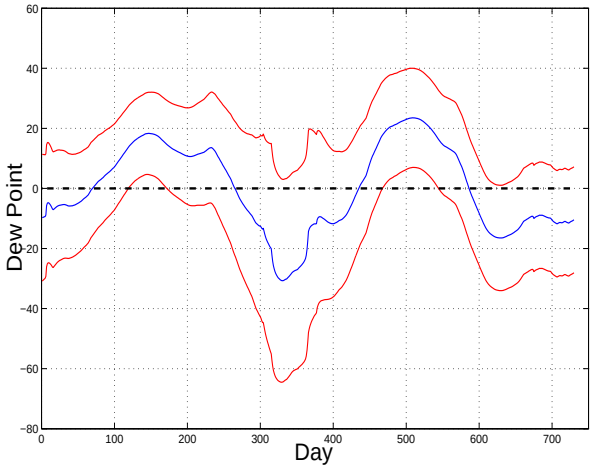

Figure 5.5: 95% SCB for Dew Point.

so narrow that no horizontal line can be covered, and even a higher level of 98% SCB cannot cover a horizontal line. We are confident that the solar power generation is time-varying, the same as the natural process.

As we center all the weather variables on their averages, the SCB of the $\beta_p(t)$ actually indicates the effect on the solar intensity. In each figure of Fig. 5.3 to 5.5, the zero line is not covered, while in Fig. 5.6 and 5.7, the zero line is covered by the 95% SCB. Therefore, we can conclude that for a level of 95%, temperature, humidity and dew point have a strong effect on solar generaion, but the effect from wind and precipitation are weak. Also, we accept $\beta_1(t)$ to $\beta_4(t)$ as time-varying coefficients, because a constant horizontal line cannot

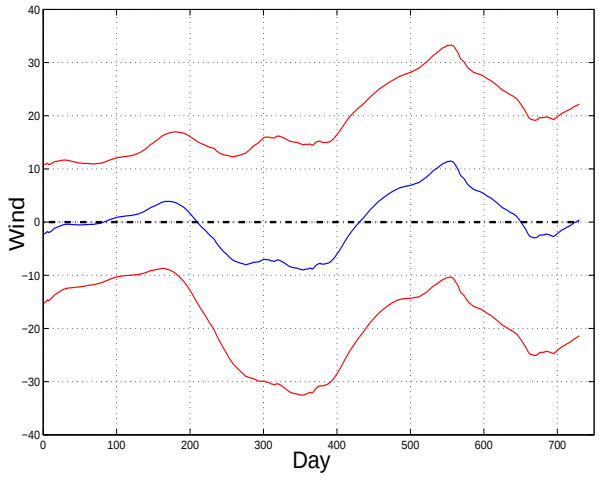

Figure 5.6: 95% SCB for Wind.

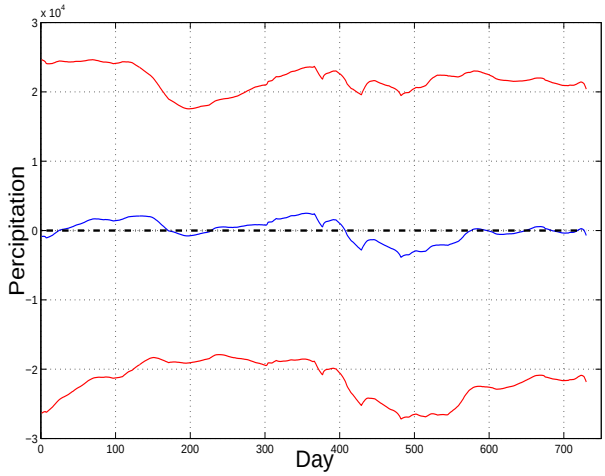

Figure 5.7: 95% SCB for Precipitation.

be covered entirely in those SCBs. Note the SCB associated with wind in Fig. 5.6 shows some variations. Although the zero line may not be covered by a narrower SCB, say 90% SCB, at the 95% significant level, we do not accept $\beta_5(t)$ as a non-zero function.The SCB associated with precipitation in Fig. 5.7 is also too wide for $\beta_6(t)$ to be accepted as a non-zero function.

It is interesting to point out that although the overall SCB does not cover the zero line entirely, for certain time periods, it contains the reference line. For example in Fig. 5.4, the SCB covers the zero line from $Day = 300$ to 400. It suggested that during Day 300 to Day 400, the humidity is not a significant covariate and can be removed from the original regression model.

5.4.4 Comparisons with Other Models on the Prediction Results

From the above discussions, we could exclude the variable of precipitation and simplify the model for better prediction. We use the model to predict the daily solar intensity for January and February in 2013. The weather information of the previous is used as the weather forecast, and the time for prediction is set from $day = 366$ to 423 which was estimated using data around January and February in 2012. In other words, the predictions

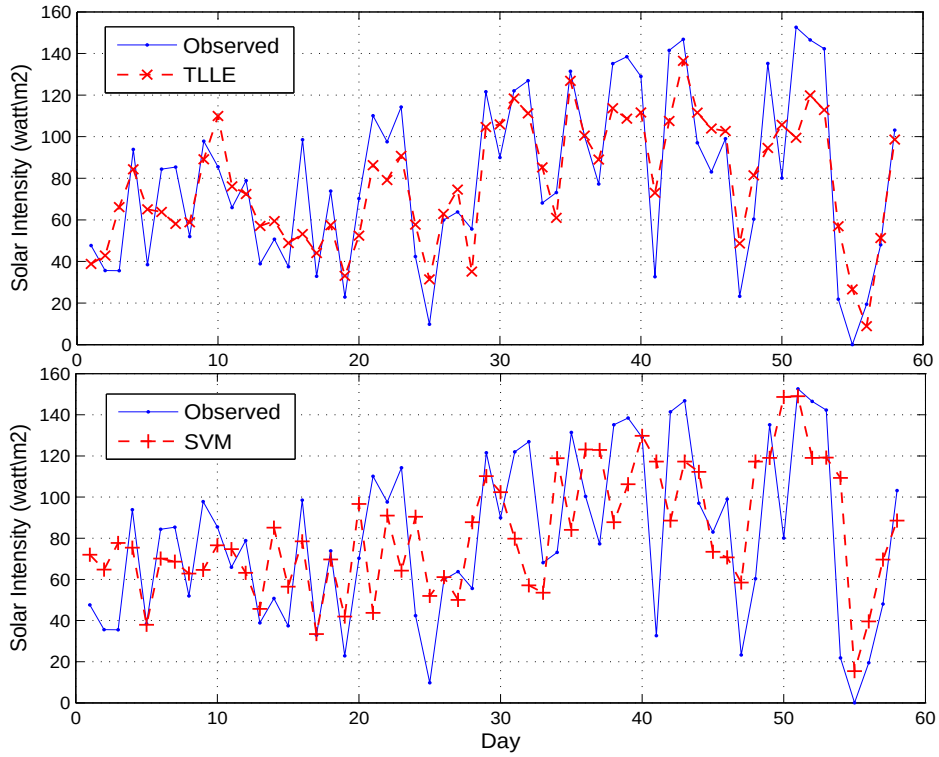

Figure 5.8: Comparisons of Predictions on Solar Intensity between TLLE and SVM

are made using the data around the same time in the previous year. The results are shown in Fig. 5.8.

As a comparison, we also present in Fig. 5.8 the predictions from the model proposed in [135]. The upper figure is the prediction curve made by the time-varying local linear estimation (TLLE) and the actual observations; the lower one shows the results from SVM regression used in [135]. We also perform the multiple linear regression (MLR). But the prediction is too poor to be shown as a comparison here. Actually, MLR uses the time merely as a common variable and the coefficients are not time-varying.

From Fig. 5.8, we can see that the TLLE predicted curve tracks the actual observations better than the SVM regression. And it is also shown in Table 5.2 that the root mean squares error between the predicted series and the observations for TLLE, SVM and MLR, which are 22.59 $watts/m^2$, 32.71 $watts/m^2$ and 53.35 $watts/m^2$ respectively. Note that the SVM regression depending highly on the selection of the parameters and the kernels, and thus, is

Table 5.2: Comparisons of RMS-Error in $watts/m^2$ between TLLE, SVM and MLR

	TLLE	SVM	MLR
RMS-Error	22.59	32.71	53.35

not practical in many cases lacking a comprehensive understanding of the real model. For example, the kernel function chosen for daily prediction is not guarantied to perform well in weekly prediction. However, the TLLE analyzes the model using simultaneous inference, which reflects the overall pattern of the dynamic pattern of the regression functions. Therefore, it can be used in many other applications, such as the hourly short-time solar power generation forecast where the time scale is set in hours.

5.5 Related Work

Load forecasting in traditional power grids has been widely studied for a long time [149]. Researchers apply different statistical methods for better prediction in different cases [133, 134]. Machine learning methods for short-term load forecasting can be found in [134]. In [133], hourly electricity load prediction is made based state space models.

As the development of the Smart Grid, forecasting on load and generation is still very important [44]. Different from the load forecast in the traditional power grid, researchers need to tackle the problems caused from the new type of power grid, such as more incorporation of renewable energy and electric vehicles [132, 135, 136]. The authors of [135] studies the statistical relationship between the weather variables and the solar generation, and predicts the solar power generation using Support Vector Machines regression. In [132], a clustering-Based nonparametric functional time series model is proposed to forecast the household-level electricity demand, for balancing the supply/demand in the low-voltage network. And a parametric approach for short-term multi-period JPDF forecast of wind generation is proposed in [136].

On the other hand, statistical theory on time-series analysis and time-varying coefficients are developing [140,142,150]. In [140], the authors propose a Gaussian approximation principle for nonstationary multiple time series with nearly optimal rates, while The simultaneous inference for time series and functional data are discussed in [142] and [150] respectively.

5.6 Conclusion

In this chapter, we propose the simultaneous inference for weather dependent power generation from renewable energy, such as solar energy and wind. We first introduce the local linear model for time series, and present the construction of the simultaneous confidence band for time-varying coefficients. And then for an application, we perform the SCB analysis to a trace of solar intensity and weather data. The presented model is also shown to outperform some existing methods for solar intensity prediction.

Chapter 6

Summary and Future Work

6.1 Summary

In previous chapters, we introduced the background of the Smart Grid (SG) and the coverage of the SG research. Out works focus on energy management in SG environment. We investigated the problems from energy distribution and generation forecasting, using convex optimization methods and the simultaneous inference. The synergy of these mathematical tools produces new solutions to the optimal energy management in the SG, which brings an advanced, efficient, green, clean, and sustainable power grid.

In Chapter 1, we presented a big picture of the new and emerging SG by introducting SG infrastructure and SG applications. The SG infrastructure is classified into smart power system, information technology, and communication system. And the SG applications include fundamental applications, emerging applications, and derived applications.

In Chapter 2, we presented a study of optimal real-time energy distribution in smart grid. With a formulation that captures the key design factors of the system, we first presented an offline algorithm that can solve the problem with optimal solutions. The proposed framework is quite general. It does not require any specific models for the electricity demand and supply processes, and only have some mild assumptions on the utility, cost, and price functions (e.g., convex and differentiable). We then developed an online algorithm that requires no future information about users and the grid, making it easy to be implemented in a real smart grid system. We also showed that the online solution converges to the offline optimal solution asymptotically and almost surely. The proposed online algorithm was evaluated with trace-driven simulations and was shown to outperform an existing benchmark scheme.

In Chapter 3, we presented a study of optimal distributed online energy distribution in the smart grid. We considered the problem from the aspect of the NIST standard. With a formulation that captures the key design factors of the system, we extended the work of a centralized online algorithm presented in Chapter 2, by decomposing the problem into many sub-problems that can be solved in a distributed manner, thus protecting users' privacy and achieving scalability. It inherits the advantages of online algorithms that requires no future information for a convergent solution, and the advantages of distributed algorithms, which solves the problem in a distributed manner with local information. Although user power usages are still exchanged with the PGO, the distributed online algorithm mitigates the privacy problem since it does not require disclosure of user's utility function and its parameters. The proposed algorithm is easy to be implemented in a real smart grid system. The distributed computation allows scalability for handling large systems. We then showed that the distributed online solution converges to the optimal offline solution asymptotically. The proposed distributed online algorithm was evaluated with trace-driven simulations and outperformed a benchmark scheme.

In Chapter 4, we developed a hierarchical power scheduling scheme to optimally manage the power distribution in the smart grid with one Macrogrid and cooperative MGs. Under some mild assumptions, we first formulated the cooperative MG problem as a convex optimization problem by capturing the key factors in a grid system, i.e., operation cost, power generation and transmission losses, user utility, distributed storage, and grid load smoothing. We then decomposed the original problem into a two-tier power control problem. The first-tier control is for the Macrogrid, aiming to maximize user utility, minimize power transmission cost from/to the Macrogrid, and smooth the grid load of the Macrogrid. The second-tier control is for each MG, aiming to minimize the cost of the MGs for power generation and transmission, while guaranteeing the power demand of MG users. It balances the power level with the Macrogrid and makes energy trading and storage decisions within the MG network. We then developed the corresponding online and distributed algorithms

for solving both problems, which were proven to be asymptotically optimal. The proposed algorithms were validated with trace-driven simulations.

In Chapter 5, we proposed the simultaneous inference for weather dependent power generation from renewable energy, such as solar energy and wind. We first introduced the local linear model for time series, and presented the construction of the SCB for time-varying coefficients. The SCBs depict the time dependent charactristics of the coefficients, which provided a new dynamic view on the model and the relationship between the weather variables and the power generation from solar and wind resources. We then performed the SCB analysis to a trace of solar intensity and weather data. The presented model was shown to outperform some existing methods for solar intensity prediction.

6.2 Future Work

The research on SG has been just a decade, and as a new power grid, there are still many problems open for research in SG. Here, we briefly extend our discussion on energy management in islanded microgrids and cooperative microgrids.

6.2.1 Energy Management for Islanded Microgrids

In Chapters 2 and 3, we focus on the energy distribution in SG environment, which can also be applied into the energy management in grid-connected MGs [151]. The mainstream researches on energy management in MGs also show preference on connected MGs. This is partly because an MG is connected to the macrogrid for most of time, and partly because the lack of infinite power bus incurs many new issues that are common in traditional power system. The intermittent power generation of renewable energy resources is usually hard to predict, which increases the difficulties of power management.

Without a previous prediction on the generation, energy management cannot be very efficient and effective in islanded MGs. The common solution to this problem of renewable energy generation is either an ESS or a backup generation. However, even with an ESS

of certain capacity, the charging and discharging activities cannot be optimized with no idea about the generation process. Therefore, the forecasting of renewable power generation is crucial for energy management in islanded MGs. On the other hand, many MGs are anticipated to meet the challenges caused by severe weather conditions. Energy management for islanded MGs in emergency scenarios should be considered in advance. In this subsection, we discuss the problems of *forecasting* and *energy management in emergency scenarios* for islanded MGs and present potential methods that may be used for solutions.

Forecasting in Islanded Microgrids

In the traditional power grid, power is generalized in centralized plants with large capacity. And power supply is following the demand all the time. This grid structure and operation mode require a previous knowledge of the possible demand, and that is why forecasting has been an important part for power control in the existing grids. The fundamental function of a power grid is to deliver power to end users. So demand load forecasting is important in any form of power grid, including both SG and MG. On the other hand, in an islanded MG, distributed generation with small capacity consists of the major power supply. To overcome the difficulties of the intermittence of generation from renewable energy resources, an ESS is necessary in MG. However, ESS works well under an optimal charging/discharging schedule for both energy efficiency improvement and cost reduction, which will be meaningless without knowledge of future generation. Therefore, both *generation forecasting* and *demand forecasting* are impportant problems in islanded MGs for better energy management.

Demand Forecasting Forecasting of power demand or grid load profile is used in current power system everyday. The research on the demand has lasted for many years, but the results are not satisfactory. The differences between predicted and real-time load profiles can be seen in traces for the New England area [152] and California SCE area [99]. By applying demand side management in the SG environment, the gap can be decreased. And demand forecasting in traditional power grid provides some techniques and experiences for

demand forecasting in MG. The popular methods and models already applied in load profile forecasting are summarized in [153]. Many of them can be helpful for MG forecasting, such as linear model, non-linear regression, machine learning, neutral networks, etc. For example, forecasting on power consumption of house-hold electric appliances can be applied directly.

The difference in MGs demand forecasting lies in the much smaller load, which can be easily affected by occasional activities from a small group. For example, in our current power grid, load demand is very large and thus not sensitive to fluctuations. It means that only large scale activities will affect the overall load profile, such as the Olympic Games. But in some islanded MGs, the load is aggregated from only a few demands, i.e., a hundred residents. The load profile is more sensitive to disturbances. Even some activity of a small group of people will have effects on the overall load. Therefore, more precise predictions along better DSM schemes on power demand are required in islanded MGs. Models on human activity power consumption predictions and time series models in statistics may be considered for possible solutions.

Generation Forecasting As stated before, generation forecasting is a new but challenging topic. It involves many uncertain processes that cannot be easily represented by a single mathematical model. For example, the solar energy generation is varying in different weather, seasons and locations. Also, the generation forecasting methods may not be evaluated and verified easily, because of uncertainties. It may take many years to test the forecasting accuracy, and modify accordingly.

For weather related renewable energy generation, such as wind and solar energy, the techniques applied in weather forecasting are helpful to find a model between the actual generation and weather variables, such as temperature, humidity and wind. But this requires data monitoring for a long time in different places. And a single model is not sufficient for precise forecasting. Thus, it is important to study the relationship between power generation and weather metrics. The methods used in weather forecast may be considered for possible solutions. The SCB method presented in Chapter 5 proposed a model based on

the dynamic relationship in statistics between solar energy generation and several weather variables. Although the trace used is observed from a single source and the predictions rely on the accuracy of the weather forecast, this work provides an novel idea for forecasting, based on which more precise models may be developed with more records of data.

Energy Management for Islanded MGs in Emergency

When MGs are disconnected and the power generation in MG is stopped by emergencies or weather disasters, such as hurricanes and tornadoes, the MGs may have to be islanded for some periods, from many hours to several days. It will be a challenge for the energy management in emergency. Different management schemes may be needed for different levels of emergency. Some parts of MG system may fail to work. Depending on the capability of generation and storage, the MGCC should be able to offer different power distribution and management plans. Also, the MGCC should be able to acquire and report additional information from the macrogrid. As an important topic in MG management, the energy management for islanded MGs in emergency is still under investigation and is an open problem.

6.2.2 Energy Management for Cooperative Microgrids without Macrogrid

In Chapter 4, we presented the power management for cooperative MGs and the Macrogrid, which brings a new perspective to the power grid composition. In future, the ultimate SG may be comprised of many MGs without the macrogrid of massive centralized power generation. It will be a highly flexible and sustainable system with optimal management at each level. Each MG will be a truly independent power system. Energy will be a common commodity that can be traded freely between MGs. However, this ultimate SG will be extremely hard to realize from the point of view of the current mainstream SG researchers and engineers. In such a system, there will be more flexible power flows than the system comprised of cooperative MGs and macrogrid. The possibility and stability of multiple power

flows within an MG from many outside MGs need further proofs and supports from experiments and testing. Also, the design of the control system, and energy management is also highly complicated. They all require more discussions, works, researches, and experiments before realization.

Appendices

Appendix A

Proofs in Chapter 2

A.1 Proof of Property 2.1

Proof. i) From (2.11), $g(\hat{\vec{p}}, c(t))$ is a continuous function of $\hat{\vec{p}}$. The continuity of $\vec{p}^*(\hat{\vec{p}}, c(t))$ could be guaranteed if all the four conditions of Theorem 2.2 from [154] are satisfied. The conditions are verified because Prob-ON is always feasible on a closed set and $\hat{\vec{p}}$ is bounded on a set $\mathbb{P}$ in our case. Therefore, $\vec{p}^*(\hat{\vec{p}}, c(t))$ is continuous with respect to $\hat{\vec{p}}$.

ii) Take $\hat{\vec{p}}_n$ as any sequence such that $\lim_{n\to\infty} \hat{\vec{p}}_n = \hat{\vec{p}}$. Then we have

$$\lim_{m\to\infty} E[p_i^*(\hat{\vec{p}}_n, c(t))] = E[\lim_{m\to\infty} p_i^*(\hat{\vec{p}}_n, c(t))] = E[p_i^*(\hat{\vec{p}}, c(t))],$$

which follows the Bounded Convergence Theorem since we already have the continuity of $\vec{p}^*(\hat{\vec{p}}, c(t))$ and the closed set $\mathbb{P}$ of p_i^* (see 2.2.1). Consequently, $E[g(\hat{\vec{p}}, c(t))]$ is also continuous.

□

A.2 Proof of Property 2.2

Proof. i) The differentiability of $g(\hat{\vec{p}}, c(t))$ follows directly from Theorem 4.1 in [155].

ii) Similar to the proof in Part ii) of Property 2.1, take any sequence $\hat{p}_{i,n}$ such that $\lim_{n\to\infty} \hat{p}_{i,n} = 0$. We have that

$$\left| \frac{g(\hat{\vec{p}} + \hat{p}_{i,n}\vec{e}) - g(\hat{\vec{p}}, c(t))}{\hat{p}_{i,n}} \right| = \alpha \left| p_i^*(\hat{\vec{p}} + \hat{p}_{0,n}\vec{e}, c(t)) - \hat{p}_i - p_{0,n} \right| \leq \alpha p_{max},$$

for $0 < p_{0,n} < \hat{p}_{i,n}$, which follows the Mean Value Theorem and part (ii) of Property 2.2. For each $i \in \mathbb{N}$,

$$\begin{aligned}
\frac{\partial}{\partial \hat{p}_i} E[g(\hat{\vec{p}}, c(t))] &= \lim_{n\to\infty} E\left[\frac{g(\hat{\vec{p}} + \hat{p}_{i,n}\vec{e}) - g(\hat{\vec{p}}, c(t))}{\hat{p}_{i,n}}\right] \\
&= E\left[\lim_{n\to\infty} \frac{g(\hat{\vec{p}} + \hat{p}_{i,n}\vec{e}) - g(\hat{\vec{p}}, c(t))}{\hat{p}_{i,n}}\right] \\
&= \alpha(E[p_i^*(\hat{\vec{p}}, c(t))] - \hat{p}_i).
\end{aligned}$$

□

A.3 Proof of Lemma 2.1

Proof. For two vectors $\vec{P}_i^1$, $\vec{P}_i^2$ and for any $i \in \mathbb{N}$, $0 < \theta < 1$, it follows from the variance definition and the strict convexity of quadratic function $f(x) = x^2$ that

$$\text{Var}(\theta\vec{P}_i^1 + (1-\theta)\vec{P}_i^2) \leq \theta\text{Var}(\vec{P}_i^1) + (1-\theta)\text{Var}(\vec{P}_i^2).$$

We conclude that $\text{Var}(\vec{P}_i)$ is strictly convex unless $\text{Var}(\vec{P}_i^1) = \text{Var}(\vec{P}_i^2)$. Since all the constraints of Prob-OFF are also convex, we conclude that Prob-OFF is a convex problem.

We next prove that Prob-OFF has a unique solution. Assume $\vec{P}_i^1$ and $\vec{P}_i^2$ are two optimal solutions to Prob-OFF. Because the objective function is concave, $\theta\vec{P}_i^1 + (1-\theta)\vec{P}_i^2$ is also optimal, for $0 < \theta < 1$. Note that we have three terms that are all concave (or convex) in (2.7). Thus $\theta\vec{P}_i^1 + (1-\theta)\vec{P}_i^2$ is optimal only if

$$U(\theta P_i^1(t) + (1-\theta)P_i^2(t)) = \theta U(P_i^1(t)) + (1-\theta)U(P_i^2(t)) \tag{A.1}$$

$$f\left(\theta\vec{P}_i^1 + (1-\theta)\vec{P}_i^2\right) \cdot \left(\theta \sum_{i\in\mathbb{N}} P_i^1(t) + (1-\theta)\sum_{i\in\mathbb{N}} P_i^2(t)\right)$$
$$= \theta f(\vec{P}_i^1)\sum_{i\in\mathbb{N}} P_i^1(t) + (1-\theta) f(\vec{P}_i^2)\sum_{i\in\mathbb{N}} P_i^2(t) \tag{A.2}$$

$$\text{Var}(\theta\vec{P}_i^1 + (1-\theta)\vec{P}_i^2) = \theta\text{Var}(\vec{P}_i^1) + (1-\theta)\text{Var}(\vec{P}_i^2), \forall\, i \in \mathbb{N}. \tag{A.3}$$

Since $U(\cdot)$ is assumed to be a strictly increasing function in Section 2.2.1, (A.1) holds true if and only if $P_i^1(t) = P_i^2(t)$, for all $i \in \mathbb{N}$, $t \in \{1, 2, \cdots, T\}$. Eqs. (A.2) and (A.3) are also sufficient for this result. Therefore, we conclude that Prob-OFF is a convex problem with a unique solution.

□

A.4 Proof of Lemma 2.2

Proof. We define several notations to be used in this proof. Define $\rho_i = \sqrt{\alpha}p_i$, function $\vec{\rho_i}^*(\hat{\vec{\rho}}, c(t))$, and $\rho_i^*(\hat{\vec{\rho}}, c(t)) = \sqrt{\alpha}p_i^*(\hat{\vec{p}}, c(t))$ for each $i \in \mathbb{N}$ and $p_i \in \mathbb{P}$. Also define

$$\text{dist}(\vec{p}^1, \vec{p}^2) = \sqrt{\sum_{i\in\mathbf{N}}(p_i^1 - p_i^2)^2} = \left|\sum_{i\in\mathbf{N}}(p_i^1 - p_i^2)\right|, \text{ for any} \vec{p}^1, \vec{p}^2 \in \mathbb{P}^N.$$

We next show the following two intermediate results that will be used to prove the lemma. The first result is that the solution of the next fixed point equation exits.

$$E[\vec{\rho_i}^*(\hat{\vec{\rho}}, c(t))] = \hat{\vec{\rho}}. \tag{A.4}$$

It follows Property 2.1 that $E[\vec{\rho_i}^*(\hat{\vec{\rho}}, c(t))]$ is a continuous function and it maps a convex compact subset of $\mathbb{P}^N$ to itself. Hence from Brouwer's Fixed Point Theorem in [156], the existence of the solution to (A.4) can be shown.

Secondly, we show that $E[\vec{\rho_i}^*(., c(t))]$ is a pseudo-contraction. Since $\mathbb{P}^N$ is a compact set, we need to show equivalently that for any two different $\hat{\vec{p}}^1$and $\hat{\vec{p}}^2 \in \mathbb{P}^N$,

$$\text{dist}(E[\vec{\rho_i}^*(\hat{\vec{\rho}}^1, c(t))], E[\vec{\rho_i}^*(\hat{\vec{\rho}}^2, c(t))]) < \text{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2).$$

Here, let $\hat{\vec{\rho}}^1$ be a solution to (A.4) and $\hat{\vec{\rho}}^2 \neq \hat{\vec{\rho}}^1$.

To prove this, we modify the Prob-ON to obtain a new problem New-Prob-ON as

$$\begin{aligned} \max: \quad & g_0(\vec{\rho}, \hat{\vec{\rho}}) \\ \text{subject to:} \quad & \frac{\rho_i}{\sqrt{\alpha}} \geq p_{i,min}, \forall\, i \in \mathbb{N} \\ & C\left(\sum_{i\in\mathbb{N}} \frac{\rho_i}{\sqrt{\alpha}}\right) \leq c(t), \forall\, t, \end{aligned} \tag{A.5}$$

where

$$g_0(\vec{\rho}, \hat{\vec{\rho}}) = \sum_{i\in\mathbb{N}} U(\frac{\rho_i}{\sqrt{\alpha}}, \omega_i) - f(\sum_{i\in\mathbb{N}} \frac{\rho_i}{\sqrt{\alpha}}) \sum_{i\in\mathbb{N}} \frac{\rho_i}{\sqrt{\alpha}} - \frac{\alpha}{2} \sum_{i\in\mathbb{N}} (\frac{\rho_i}{\sqrt{\alpha}} - \frac{\hat{\rho}_i}{\sqrt{\alpha}})^2.$$

For brevity, we drop the time index (t) in the remainder of this proof, when their meanings are clear in the context. Note that $\vec{\rho_i}^*(\hat{\vec{\rho}}, c(t))$ is the optimal solution for New-Prob-ON.

Now, we use Proposition 6.1 from [155] to achieve the Lipschitz continuity and acquire the Lipschitz constant of $\vec{\rho_i}^*(., c(t))$ in a neighborhood of $\hat{\vec{\rho}}^1$. Two conditions are necessary to hold the proposition: the Lipschitz continuity of the difference function in a neighborhood of $\hat{\vec{\rho}}^1$ and the second-order growth condition.

We define the difference function $\Delta g_0(\vec{\rho}, \hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2)$ as

$$\begin{aligned} \Delta g_0(\vec{\rho}, \hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2) &= g_0(\vec{\rho}, \hat{\vec{\rho}}^2) - g_0(\vec{\rho}, \hat{\vec{\rho}}^1) \\ &= \frac{1}{2} \sum_{i\in\mathbb{N}} (\hat{\rho}_i^1 - \hat{\rho}_i^2)(2\rho_i - \hat{\rho}_i^1 - \hat{\rho}_i^2). \end{aligned}$$

Then it follows that

$$\begin{aligned}&\mathrm{dist}(\Delta g_0(\vec{\rho}^1, \hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2), \Delta g_0(\vec{\rho}^2, \hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2))\\ &= \left|\sum_{i\in\mathbb{N}} (\hat{\rho}_i^1 - \hat{\rho}_i^2)(\rho_i^1 - \rho_i^2)\right| \leq \mathrm{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2)\mathrm{dist}(\vec{\rho}^1, \vec{\rho}^2),\end{aligned} \tag{A.6}$$

where the inequality holds from Cauchy-Schwarz inequality. Hence, the first condition of Proposition 6.1 in [155] holds.

Next, we show that the second condition also holds. In our case, the second-order growth condition requires that there exists a positive constant a such that

$$g_0(\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t)), \hat{\vec{\rho}}^1) - g_0(\vec{\rho}, \hat{\vec{\rho}}^1) \geq a(\mathrm{dist}(\vec{\rho}, \vec{\rho}^*(\hat{\vec{\rho}}^1, c(t))))^2.$$

We find a sufficient condition for this second-order growth condition in [157], in which Theorem 6.1 states that if the Slater qualification hypothesis holds, the second-order growth condition (Theorem 6.1 (v)) is equivalent with three other conditions (Theorem 6.1 (vi)-(viii)). Because the Slater qualification hypothesis could be satisfied if we carefully choose $p_{i,min}$ (see Section 2.3). We thus verify that an equivalent condition Theorem 6.1 (vii) is satisfied. For this, define:

$$\begin{aligned}&L(\vec{\rho}, \lambda, (\nu_i : i \in \mathbb{N})) = g_0(\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t)), \hat{\vec{\rho}}^1) - g_0(\vec{\rho}, \hat{\vec{\rho}}^1)\\ &+\lambda\left(\frac{1}{c(t)}C\left(\sum_{i\in\mathbb{N}}\frac{\rho_i}{\sqrt{\alpha}}\right) - 1\right) - \sum_{i\in\mathbb{N}}\nu_i\left(\frac{\rho_i}{\sqrt{\alpha}} - p_{i,min}\right).\end{aligned} \tag{A.7}$$

Then, we can write the function φ in Theorem 6.1 (vii), for any $\vec{d} \in \mathbb{R}^N$, as

$$\varphi_{\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t))}(\vec{d}) = \vec{d}'\frac{\partial^2}{\partial\vec{\rho}^2}L(\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t)), \lambda^*, (\nu_i^* : i \in \mathbb{N}))\vec{d},$$

where λ^* and $(\nu_i^* : i \in \mathbb{N})$ are the optimal Lagrange multipliers and variables. Substituting (A.7), we have that

$$\varphi_{\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t))}(\vec{d}) = \sum_{i\in\mathbb{N}} d_i^2 \left(1 - \frac{1}{\alpha} U''\left(\frac{\rho_i^*(\hat{\vec{\rho}}^1, c(t))}{\sqrt{\alpha}}, \omega_i\right)\right.$$
$$+\frac{1}{\alpha}\left(2f'\left(\sum_{i\in\mathbb{N}} \frac{\rho_i^*(\hat{\vec{\rho}}^1, c(t))}{\sqrt{\alpha}}\right)\right.$$
$$\left.\left.+f''\left(\sum_{i\in\mathbb{N}} \frac{\rho_i^*(\hat{\vec{\rho}}^1, c(t))}{\sqrt{\alpha}}\right)\left(\sum_{i\in\mathbb{N}} \frac{\rho_i^*(\hat{\vec{\rho}}^1, c(t))}{\sqrt{\alpha}}\right)\right)\right.$$
$$\left.+\frac{\lambda^*}{\alpha} C''\left(\sum_{i\in\mathbb{N}} \frac{\rho_i^*(\hat{\vec{\rho}}^1, c(t))}{\sqrt{\alpha}}\right)/c(t)\right).$$

Since λ^* is the optimal Lagrange multiplier, $\lambda^* \geq 0$. Also U is a strictly increasing, concave function, and C and f are strictly increasing, convex functions. Moreover, $p_i \in \mathbb{P}$ so that ρ_i lies in a closed set $\mathbb{P}_0$, for $i \in \mathbb{N}$. Therefore, there exist positive constants $\xi_{U''}$, $\xi_{f'}$, and $\xi_{f''}$ such that $U''(\rho_i) \leq -\xi_{U''}$, $f'(\rho_i) \geq \xi_{f'}$, and $f''(\rho_i) \geq \xi_{f''}$, for all $\rho_i \in \mathbb{P}_0$. So we have that for any $\vec{d} \in \mathbb{R}^N$,

$$\varphi_{\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t))}(\vec{d}) \geq \left(1 + \frac{\xi}{\alpha}\right) \sum_{i\in\mathbb{N}} d_i^2 > \sum_{i\in\mathbb{N}} d_i^2, \tag{A.8}$$

where $\xi = \xi_{U''} + 2\xi_{f'} + \xi_{f''}$ is a positive constant. Now, we have verified the condition of Theorem 6.1 (vii) and hence from Theorem 6.1 of [157], Theorem 6.1 (v) is satisfied, which equals to the second-order growth condition. Thus, for proposition 6.1 of [155], both conditions are satisfied. We could use it safely and conclude that:

$$\mathrm{dist}(\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t)), \vec{\rho}^*(\hat{\vec{\rho}}^2, c(t))) \;\leq\; \left(1 + \frac{\xi}{\alpha}\right)^{-1} \mathrm{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2) < \mathrm{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2).$$

Thus, we can conclude that

$$E\left[\left(\mathrm{dist}(\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t)), \vec{\rho}^*(\hat{\vec{\rho}}^2, c(t)))\right)^2\right] < \left(\mathrm{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2)\right)^2.$$

Further, we have that

$$\begin{aligned}
&\text{dist}(E[\vec{\rho_i}^*(\hat{\vec{\rho}}^1, c(t))], E[\vec{\rho_i}^*(\hat{\vec{\rho}}^2, c(t))]) \\
&= \sqrt{\sum_{i\in\mathbb{N}} \left(E\left[\rho_i^*(\hat{\vec{\rho}}^1, c(t)) - \rho_i^*(\hat{\vec{\rho}}^2, c(t))\right]\right)^2} \\
&\leq \sqrt{\sum_{i\in\mathbb{N}} E\left[\left(\rho_i^*(\hat{\vec{\rho}}^1, c(t)) - \rho_i^*(\hat{\vec{\rho}}^2, c(t))\right)^2\right]} \\
&= \sqrt{E\left[\left(\text{dist}(\vec{\rho}^*(\hat{\vec{\rho}}^1, c(t)), \vec{\rho}^*(\hat{\vec{\rho}}^2, c(t)))\right)^2\right]} \\
&< \text{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2).
\end{aligned}$$

The first inequality is due to Jensen's inequality. This proves our second intermediate result.

Then, suppose that the fixed point equation (A.4) has two distinct solutions $\hat{\vec{\rho}}^1$ and $\hat{\vec{\rho}}^2$. We have that

$$\text{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2) \quad = \quad \text{dist}(E[\vec{\rho_i}^*(\hat{\vec{\rho}}^1, c(t))], E[\vec{\rho_i}^*(\hat{\vec{\rho}}^2, c(t))]) < \text{dist}(\hat{\vec{\rho}}^1, \hat{\vec{\rho}}^2),$$

which is an contradiction. This implies that (A.4) has at most one solution. We conclude that (2.17) has a unique solution.

□

A.5 Proof of Lemma 2.3

Proof. Given that $c(t)$ is an ergodic process, the updating function (2.14) can be considered as a stochastic approximation update equation. We can apply Theorem 1.1 of Chapter 6 in [158] for the convergence proof. We verify the assumptions in Theorem 1.1 of Chapter 6 in [158] in the following.

We first list the variables used in Theorem 1.1 and correspond them to our problem and our notation style: $\vec{\theta}_t = \hat{\vec{p}}^*(t)$, $\xi_t = c(t+1)$, $(Y_t)_i = \alpha(p_i^*(\hat{\vec{p}}^*(t), c(t+1)) - \hat{p}_i^*(t)), \forall\ i \in \mathbb{N}$, $\epsilon_t = \frac{1}{t+\alpha}$, $g(\hat{\vec{p}}, c(t)) = \alpha(p_i^*(\hat{\vec{p}}, c(t)) - \hat{p}_i^*), \forall\ i \in \mathbb{N}$, $\delta\vec{M} = \vec{0}$, $\vec{\beta}_t = \vec{0}$ and $\vec{Z}_t = \vec{0}$ for each t.

Now we need to verify that all the assumptions in Chapter 6 of [158] from (A.1.1) to (A.1.8) are satisfied. According to Property 2.1, $E[Y_t]$ is a continuous function of $\hat{p}_i^*(t)$ and $\vec{p}_i^*(\hat{\vec{p}}^*(t-1), c(t)) \in \mathbb{P}^N$ for any t. Thus, (A.1.1) is satisfied. (A.1.2) also follows from Property 2.1 that $g(\hat{\vec{p}}, c(t))$ is a continuous function of $\hat{\vec{p}}$, which guarantees (A.1.7) as well. For (A.1.3), we can take the following form of the function

$$(\bar{g}(\hat{\vec{p}}^*(t)))_i = \alpha(E[p_i^*(\hat{\vec{p}}^*(t), c(t+1))] - \hat{p}_i^*(t)).$$

According to [158], (A.1.3) holds due to the strong law of large numbers, because $c(t)$ is an ergodic process. Since $\vec{\beta}_t = \vec{Z}_t = \vec{0}$ for each t, we have both (A.1.4) and (A.1.5) hold true. For (A.1.6), it holds because $g(\hat{\vec{p}}, c(t))$ is bounded. Hence, all the assumptions are satisfied. It follows Theorem 1.1 in [158] and Property 2.2 that $\hat{p}_i(t)$ converges almost surely to the unique solution of $E[\vec{p}^*(\hat{\vec{p}}, c(t))] = \hat{\vec{p}}$.

□

A.6 Proof of Lemma 2.4

Proof. Rewrite (2.14) and sum from $t = 1$ to T. We have

$$\sum_{t=1}^{T} \left(\frac{t+\alpha}{\alpha} \right) (\hat{p}_i(t) - \hat{p}_i(t-1)) = \sum_{t=1}^{T} (p_i^*(t) - \hat{p}_i(t-1)).$$

Expanding the sum on the LHS, it follows that

$$\begin{aligned} & \frac{1}{\alpha} \left(T \cdot \hat{p}_i(T) - \sum_{t=1}^{T} \hat{p}_i(t-1) \right) - (\hat{p}_i(T) - \hat{p}_i(1)) \\ = & \sum_{t=1}^{T} (p_i^*(t) - \hat{p}_i(T) + \hat{p}_i(T) - \hat{p}_i(t-1)). \end{aligned}$$

Take limit over T on both sides and it follows that

$$\begin{aligned}
&\lim_{T\to\infty} \frac{T\cdot\hat{p}_i(T) - \sum_{t=1}^{T}\hat{p}_i(t-1)}{\alpha\cdot T} - \lim_{T\to\infty}\frac{\hat{p}_i(T)-\hat{p}_i(1)}{T} \\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(p_i^*(t)-\hat{p}_i(T)+\hat{p}_i(T)-p_i^*(t-1)).
\end{aligned}$$

The second term of the LHS is zero as $T\to\infty$. Rearranging the terms, we have

$$\begin{aligned}
&\lim_{T\to\infty}\left(\frac{1-\alpha}{\alpha}\right)\left(\hat{p}_i(T)-\frac{1}{T}\sum_{t=1}^{T}\hat{p}_i(t-1)\right) \\
= &\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(p_i^*(t)-\hat{p}_i(T)).
\end{aligned}$$

Since the sequence $\hat{p}_i(t)$ converges as shown in Lemma 2.3,

$$\lim_{T\to\infty}\left(\hat{p}_i(T)-\frac{1}{T}\sum_{t=1}^{T}\hat{p}_i(t-1)\right)=0$$

, and the LHS will be zero. Thus the limit on the RHS will also be zero. □

A.7 Proof of Theorem 2.1

Proof. The proof is equivalent to showing that $\lim_{T\to\infty}\frac{1}{T}(\Psi(\mathbf{p}^*)-\Psi(\mathbf{P}^*))=0$ holds true almost surely, where $\mathbf{p}^*$ is online optimal solution and $\mathbf{P}^*$ is the offline optimal solution. Recall that $\lambda^*(t)$ and $\nu_i^*(t)$ are the non-negative multipliers that satisfy the KKT conditions of the online problem (see (2.15)). We define a new differentiable concave function $\Phi(\cdot)$ as

follows.

$$\begin{aligned}\Phi(\mathbf{P}^*) &= \sum_{t=1}^{T}\sum_{i\in\mathbb{N}} U(P_i^*(t),\omega_i(t)) - \sum_{t=1}^{T} f\left(\sum_{i\in\mathbb{N}} P_i^*(t)\right)\sum_{i\in\mathbb{N}} P_i^*(t) - \\ &\quad \frac{\alpha T}{2}\sum_{i\in\mathbb{N}}\mathrm{Var}(\vec{P}_i^*) - \sum_{t=1}^{T}\lambda^*(t)\left(\frac{C(\sum_{i\in\mathbb{N}} P_i^*(t))}{c(t)} - 1\right) + \\ &\quad \sum_{t=1}^{T}\sum_{i\in\mathbb{N}}\nu_i^*(t)(P_i^*(t) - P_{i,min}(t)). \end{aligned} \tag{A.9}$$

Note that the sum of the first three terms on the RHS of (A.9) is equal to $\Psi(\mathbf{P}^*)$, while the last two terms on the RHS of (A.9) are both non-negative. It follows that

$$\Psi(\mathbf{P}^*) \leq \Phi(\mathbf{P}^*). \tag{A.10}$$

Furthermore, with the concave and differentiable properties of function $\Phi(\cdot)$, we have [98]

$$\Phi(\mathbf{P}^*) \leq \Phi(\mathbf{p}^*) + \bigtriangledown\Phi(\mathbf{p}^*) \bullet (\mathbf{P}^* - \mathbf{p}^*), \tag{A.11}$$

where $\bullet$ denotes the inner product operation. Combining (A.10) and (A.11), we have

$$\begin{aligned}
\Psi(\mathbf{P}^*) \;&\leq\; \Phi(\mathbf{P}^*) \leq \Phi(\mathbf{p}^*) + \bigtriangledown\Phi(p^*) \bullet (P^* - p^*) \qquad (A.12)\\
&= \sum_{t=1}^{T}\sum_{i\in\mathbb{N}} U(p_i^*(t),\omega_i(t)) - \\
&\sum_{t=1}^{T} f\left(\sum_{i\in\mathbb{N}} p_i^*(t)\right)\sum_{i\in\mathbb{N}} p_i^*(t) - \frac{\alpha T}{2}\sum_{i\in\mathbb{N}}\mathrm{Var}(\vec{p_i^*}) - \\
&\sum_{t=1}^{T}\lambda^*(t)\left(\frac{C(\sum_{i\in\mathbb{N}} p_i^*(t))}{c(t)} - 1\right) + \\
&\sum_{t=1}^{T}\sum_{i\in\mathbb{N}}\nu_i^*(t)(p_i^*(t) - p_{i,min}(t)) + \\
&\sum_{t=1}^{T}\sum_{i\in\mathbb{N}}(P_i^*(t) - p_i^*(t))\left(U'(p_i^*(t),\omega_i(t)) - g\left(\sum_{i\in\mathbb{N}} p_i^*(t)\right) + \right.\\
&\left.\frac{\alpha}{T}\sum_{k=1}^{T} p_i^*(k) - \alpha p_i^*(t) - \lambda^*(t)\frac{C(\sum_{i\in\mathbb{N}} p_i^*(t))}{c(t)} + \nu_i^*(t)\right).
\end{aligned}$$

As $\lambda^*(t)$ and $\nu_i^*(t)$ are the Lagrange multipliers and variables of Prob-ON, we can substitute (2.15) into the above inequality (A.12) to have

$$\begin{aligned}
\Psi(\mathbf{P}^*) \;&\leq\; \sum_{t=1}^{T}\sum_{i\in\mathbb{N}} U(p_i^*(t),\omega_i(t)) - \\
&\sum_{t=1}^{T} f\left(\sum_{i\in\mathbb{N}} p_i^*(t)\right)\sum_{i\in\mathbb{N}} p_i^*(t) - \frac{\alpha T}{2}\sum_{i\in\mathbb{N}}\mathrm{Var}(\vec{p_i^*}) + \\
&\sum_{t=1}^{T}\sum_{i\in\mathbb{N}}\alpha(P_i^*(t) - p_i^*(t))\left(\frac{1}{T}\sum_{k=1}^{T} p_i^*(k) - \hat{p}_i(t-1)\right).
\end{aligned}$$

Adding $-\hat{p}_i(T) + \hat{p}_i(T)$ to the last component of the RHS of the above inequality, we have

$$\frac{1}{T}\sum_{k=1}^{T} p_i^*(k) - \hat{p}_i(t-1) = \frac{1}{T}\sum_{k=1}^{T} p_i^*(k) - \hat{p}_i(T) + \hat{p}_i(T) - \hat{p}_i(t-1). \qquad (A.13)$$

From Lemma 2.4, the limit of the above equation is zero for all users. We can take limit of (A.13) and it follows that

$$\lim_{T\to\infty}\frac{\Psi(\mathbf{P}^*)}{T} \leq \lim_{T\to\infty}\frac{1}{T}\left(\sum_{t=1}^{T}\sum_{i\in\mathbb{N}}U(p_i^*(t),\omega_i(t)) - \right.$$
$$\left.\sum_{t=1}^{T}f\left(\sum_{i\in\mathbb{N}}p_i^*(t)\right)\sum_{i\in\mathbb{N}}p_i^*(t) - \frac{\alpha T}{2}\sum_{i\in\mathbb{N}}\mathrm{Var}(\vec{p_i^*})\right) = \lim_{T\to\infty}\frac{\Psi(\mathbf{p}^*)}{T}.$$

Thus $\lim_{T\to\infty}\frac{1}{T}(\Psi(\mathbf{P}^*)-\Psi(\mathbf{p}^*))\leq 0$ holds for all users. Because $\mathbf{P}^*$ is optimal to the offline problem and $\Psi(\mathbf{P}^*)$ is the offline objective value, we also have $\Psi(\mathbf{P}^*)\geq\Psi(\mathbf{p}^*)$. We conclude that Theorem 2.1 holds true. □

Appendix B

Proofs in Chapter 3

B.1 Proof of Theorem 3.2

Proof. Substituting (3.16) and (3.17) into (3.19), we have

$$\begin{aligned}
D(\lambda(t)) &= \sum_{i\in\mathbb{N}} S_i(\lambda(t)) + R(\lambda(t)) \\
&= \max\Bigg\{\sum_{i\in\mathbb{N}} (U(p_i(t),\omega_i(t)) - \\
&\quad \frac{\alpha}{2}(p_i(t)-\hat{p}_i(t-1))^2 - \lambda(t)p_i(t)\Big) + \\
&\quad \lambda(t)g(t) - C(g(t))\Bigg\} \\
&= \max\Bigg\{\sum_{i\in\mathbb{N}} (U(p_i(t),\omega_i(t)) - \\
&\quad \frac{\alpha}{2}(p_i(t)-\hat{p}_i(t-1))^2\Big) - \\
&\quad C(g(t)) - \lambda(t)\left(\sum_{i\in\mathbb{N}} p_i(t) - g(t)\right)\Bigg\} \\
&= \max\{L(\vec{p}(t), g(t), \lambda(t))\}.
\end{aligned}$$

Comparing function $L(\vec{p}(t), g(t), \lambda(t))$ with the centralized online problem Prob-ON (3.6), and its constraint (3.7), it can be seen that function $L(\vec{p}(t), g(t), \lambda(t))$ is actually the Lagrangian of Prob-ON and $\lambda(t)$ is the Lagrange multiplier. And function $D(\lambda(t))$ is then the dual function for Prob-ON. Similar to the Prob-OFF case, the Slater's condition holds true here again by careful choices of $p_i(t)$ and $g(t)$. Therefore, the distributed online subproblems (3.16) and (3.17) have the same solution as the centralized online problem Prob-ON.

On the other hand, Theorem 3.1 has proved that the solution of Prob-ON is optimal and asymptotically convergent to the offline optimal solution. We then conclude that the solution of the distributed online subproblems is also optimal and converges asymptotically to the optimal offline solution. □

Appendix C

Proofs in Chapter 4

C.1 Proof of Lemma 4.1

Proof. We prove (4.18) in this section, while (4.19) can be proved in the same way. Rewrite the first update equation in (4.17) and sum up from $t = 1$ to T. We have

$$\sum_{t=1}^{T} \left(\frac{t+\alpha}{\alpha} \right) (\hat{d}_i(t) - \hat{d}_i(t-1)) = \sum_{t=1}^{T} (d_i^*(t) - \hat{d}_i(t-1)).$$

Expanding the sum on the left-hand-side (LHS), some terms can be canceled. Also, adding term $-\hat{d}_i(T) + \hat{d}_i(T)$ to the right-hand-side (RHS), it follows that

$$\begin{aligned} & \frac{1}{\alpha} \left(T \cdot \hat{d}_i(T) - \sum_{t=1}^{T} \hat{d}_i(t-1) \right) - (\hat{d}_i(T) - \hat{d}_i(1)) \\ = & \sum_{t=1}^{T} (d_i^*(t) - \hat{d}_i(T) + \hat{d}_i(T) - \hat{d}_i(t-1)). \end{aligned}$$

Taking limit over T on both sides, we have

$$\begin{aligned} & \lim_{T\to\infty} \frac{T \cdot \hat{d}_i(T) - \sum_{t=1}^{T} \hat{d}_i(t-1)}{\alpha \cdot T} - \lim_{T\to\infty} \frac{\hat{d}_i(T) - \hat{d}_i(1)}{T} \\ = & \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} (d_i^*(t) - \hat{d}_i(T) + \hat{d}_i(T) - d_i^*(t-1)). \end{aligned}$$

The second term of the LHS goes to zero. Rearranging the remaining terms, we have

$$\begin{aligned}&\lim_{T\to\infty}\left(\frac{1-\alpha}{\alpha}\right)\left(\hat{d}_i(T)-\frac{1}{T}\sum_{t=1}^{T}\hat{d}_i(t-1)\right)\\ =&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(d_i^*(t)-\hat{d}_i(T)).\end{aligned}\tag{C.1}$$

Due to the first updating function in (4.17), $\hat{d}_i(t)$ is convergent as $t\to\infty$. Thus, the LHS of (C.1) is zero. It follows (C.1) that (4.18) holds true. □

C.2 Proof of Theorem 4.1

Proof. The convergence of the online solution is equivalent to the convergence of the online objective value to that of the offline problem Prob-MA2. Thus, we next prove that $\lim_{T\to\infty}\frac{1}{T}(\mathbf{F}(\mathbf{d}^*,\mathbf{p}^*)-\mathbf{F}(\tilde{\mathbf{d}},\tilde{\mathbf{p}}))=0$, where $(\mathbf{d}^*,\mathbf{p}^*)$ is the solution to Prob-MA3 and $(\tilde{\mathbf{d}},\tilde{\mathbf{p}})$ is the solution to Prob-MA2.

It can be shown that Prob-MA3 is also a convex optimization problem, and the Slater's condition is also satisfied. We derive the KKT conditions of Prob-MA3 as follows.

$$\begin{cases}I_{\mathbb{N}}\left(U'(d_i^*(t),\omega_i(t))-\alpha(d_i^*(t)-\hat{d}_i(t-1))+\nu_i^*(t)\right)+\\ I_{\mathbb{M}}\left(-C'(p_m^*(t))-\alpha(p_m^*(t)-\hat{p}_m(t-1))+\gamma_m^*(t)-\right.\\ \left.\rho_m^*(t)\right)-\mu^*(t)G'(l^*(t))/b_{max}(t)=0\\ \mu^*(t)\left(G(l^*(t))/b_{max}(t)-1\right)=0\\ \nu_i^*(t)\left(d_i^*(t)-d_{i,min}(t)\right)=0\\ \gamma_m^*(t)\left(p_m^*(t)+p_{m,max}(t)\right)=0\\ \rho_m^*(t)\left(p_m^*(t)-p_{m,max}(t)\right)=0\\ \mu^*(t),\nu_i^*(t),\gamma_m^*(t),\rho_m^*(t)\geq 0,\forall\ i\in\mathbb{N},\ m\in\mathbb{M},t\ \in\mathbb{T},\end{cases}\tag{C.2}$$

where the non-negative Lagrangian multipliers $\mu^*(t)$, $\nu_i^*(t)$, $\gamma_m^*(t)$, and $\rho_m^*(t)$ are the dual points where the KKT conditions are satisfied and the optimal value is achieved.

To prove the theorem, we need another differentiable concave function $\mathbf{H}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}})$ as defined in (C.3).

$$
\begin{aligned}
&\mathbf{H}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}}) \\
&= \sum_{t=1}^{T} \left(\sum_{i \in \mathbb{N}} U(\tilde{d}_i(t), \omega_i(t)) - \sum_{m \in \mathbb{M}} C_m(\tilde{p}_m(t)) \right) \\
&- \frac{\alpha T}{2} \left(\sum_{i \in \mathbb{N}} Var(\vec{\tilde{d}}_{i,T}) + \sum_{m \in \mathbb{M}} Var(\vec{\tilde{p}}_{m,T}) \right) \\
&- \sum_{t=1}^{T} \mu^*(t) \left(\frac{G(\tilde{l}(t))}{b_{max}(t)} - 1 \right) + \sum_{t=1}^{T} \sum_{i \in \mathbb{N}} \nu_i^*(t)(\tilde{d}_i(t) - d_{i,min}(t)) \\
&+ \sum_{t=1}^{T} \sum_{m \in \mathbb{M}} (\gamma_m^*(t)(\tilde{p}_m(t) + p_{m,max}(t)) - \rho_m^*(t)(\tilde{p}_m(t) - p_{m,max}(t))) \\
&= \mathbf{F}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}}) \\
&+ \sum_{t=1}^{T} \mu^*(t) \left(1 - \frac{G(\tilde{l}(t))}{b_{max}(t)} \right) + \sum_{t=1}^{T} \sum_{i \in \mathbb{N}} \nu_i^*(t)(\tilde{d}_i(t) - d_{i,min}(t)) \\
&+ \sum_{t=1}^{T} \sum_{m \in \mathbb{M}} ((\gamma_m^*(t) - \rho_m^*(t))\tilde{p}_m(t) + (\gamma_m^*(t) + \rho_m^*(t))p_{m,max}(t)) \,. \qquad \text{(C.3)}
\end{aligned}
$$

Recall that $G(\tilde{l}(t))$ is the generation cost and $b_{max}(t)$ is the maximum generation cost in the Macrogrid. Therefore the second term on the RHS of (C.3) is non-negative. Following (C.2), the last two terms on the RHS of (C.3) are both non-negative either. It follows that

$$\mathbf{F}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}}) \leq \mathbf{H}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}}). \qquad \text{(C.4)}$$

Due to the concavity and differentiability of $\mathbf{H}(\cdot)$, we have

$$\mathbf{H}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}}) \leq \mathbf{H}(\mathbf{d}^*, \mathbf{p}^*) + \bigtriangledown \mathbf{H}(\mathbf{d}^*, \mathbf{p}^*) \bullet ((\tilde{\mathbf{d}}, \tilde{\mathbf{p}}) - (\mathbf{d}^*, \mathbf{p}^*)), \qquad \text{(C.5)}$$

where $\bullet$ denotes the inner product operation. According to (C.4) and (C.5), we can derive inequality (C.6).

$$
\begin{aligned}
\mathbf{F}(\tilde{\mathbf{d}},\tilde{\mathbf{p}}) \;&\leq\; \mathbf{H}(\mathbf{d}^*,\mathbf{p}^*) + \bigtriangledown\mathbf{H}(\mathbf{d}^*,\mathbf{p}^*) \bullet ((\tilde{\mathbf{d}},\tilde{\mathbf{p}}) - (\mathbf{d}^*,\mathbf{p}^*)) \\
&= \mathbf{F}(\mathbf{d}^*,\mathbf{p}^*) \\
&+ \sum_{t=1}^{T} \mu^*(t) \left(1 - \frac{G(l^*(t))}{b_{max}(t)}\right) + \sum_{t=1}^{T}\sum_{i\in\mathbb{N}} \nu_i^*(t)(d_i^*(t) - d_{i,min}(t)) \\
&+ \sum_{t=1}^{T}\sum_{m\in\mathbb{M}} ((\gamma_m^*(t) - \rho_m^*(t))p_m^*(t) + (\gamma_m^*(t) + \rho_m^*(t))p_{m,max}(t)) \\
&+ \sum_{t=1}^{T}\sum_{i\in\mathbb{N}} (\tilde{d}_i(t) - d_i^*(t)) \cdot \\
&\left(U'(d_i^*(t),\omega_i(t)) + \nu_i^*(t) - \mu^*(t)\frac{G'(l^*(t))}{b_{max}(t)} + \frac{\alpha}{T}\sum_{k=1}^{T} d_i^*(k) - \alpha d_i^*(t)\right) \\
&+ \sum_{t=1}^{T}\sum_{m\in\mathbb{M}} (\tilde{p}_m(t) - p_m^*(t)) \cdot \\
&\left(-C_m'(p_m^*(t)) + \gamma_m^*(t) - \rho_m^*(t) - \mu^*(t)\frac{G'(l^*(t))}{b_{max}(t)} + \frac{\alpha}{T}\sum_{k=1}^{T} p_m^*(k) - \alpha p_m^*(t)\right). \qquad \text{(C.6)}
\end{aligned}
$$

Substituting (C.2) into inequality (C.6) , we have

$$
\begin{aligned}
&\mathbf{F}(\tilde{\mathbf{d}},\tilde{\mathbf{p}}) \leq \mathbf{F}(\mathbf{d}^*,\mathbf{p}^*) + \\
&\sum_{t=1}^{T}\sum_{i\in\mathbb{N}} \alpha(\tilde{d}_i(t) - d_i^*(t)) \left(\frac{1}{T}\sum_{k=1}^{T} d_i^*(k) - \hat{d}_i(t-1)\right) + \\
&\sum_{t=1}^{T}\sum_{m\in\mathbb{M}} \alpha(\tilde{p}_m(t) - p_m^*(t)) \left(\frac{1}{T}\sum_{k=1}^{T} p_m^*(k) - \hat{p}_m(t-1)\right).
\end{aligned}
$$

Adding $-\hat{d}_i(T) + \hat{d}_i(T)$ and $-\hat{p}_m(T) + \hat{p}_m(T)$ to the last two terms on the RHS of the above inequality, respectively, taking limit over T on both sides, and applying Lemma 4.1, we have

$$
\lim_{T\to\infty} \frac{\mathbf{F}(\tilde{\mathbf{d}},\tilde{\mathbf{p}})}{T} \leq \lim_{T\to\infty} \frac{\mathbf{F}(\mathbf{d}^*,\mathbf{p}^*)}{T}. \qquad \text{(C.7)}
$$

On the other hand, $(\tilde{\mathbf{d}}, \tilde{\mathbf{p}})$ is the optimal solution to Prob-MA2 and thus $\mathbf{F}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}})$ is the optimal objective value of Prob-MA2. Since it is a maximization problem, we have

$$\mathbf{F}(\tilde{\mathbf{d}}, \tilde{\mathbf{p}}) \geq \mathbf{F}(\mathbf{d}^*, \mathbf{p}^*). \tag{C.8}$$

Considering both (C.7) and (C.8), we conclude that Theorem 4.1 holds true. □

Appendix D

Acronyms

AGL Actual Grid Load

AMI Advanced Metering Infrastructure

BESS Battery Energy Storage System

CAES Compressed Air Energy Storage

CG Coalition Game

CHP Combined Heat and Power

COA Centralized Online Algorithm

CPP Critical-peak Pricing

CSP Concentrating Solar Power

DC Direct Current

DER Distributed Energy Resource

DG Distributed Generation

DOA Distributed Online Algorithm

DOE U.S. Department of Energy

DPA Dynamic Pricing Algorithm

DR Demand Response

DSM Demand Side Management

EC Electrochemical Capacitors

ED Energy Distributor

EP Energy Provider

EPRI Electric Power Research Institute

ESS Energy Storage System

EV Electric Vehicle

FACTS Flexible AC Transmission Systems

G2V Grid-to-Vehicle

GCV Generalized Cross-Validation

GHG Greenhouse Gas

GPS Global Positioning System

GRS Grid Energy Storage

HVDC High Voltage Direct Current

HPS Hierarchical Power Scheduling

IEA International Energy Agency

IEEE Institute of Electrical and Electronics Engineers

IPM Interior Point Method

JPDF Joint Probability Density Function

KKT KarushKuhnTucker

MG Microgrid

MGCC Microgrid Control Center

MGNC Microgrid Network Controller

MLR Multiple Linear Regression

NCC No Cooperation Control

NIST National Institute of Standard and Technology

OL Original Load

OPDA Online Power Distribution Algorithm

ORPA Optimal Real-time Pricing Algorithm

PCC Point of Common Coupling

PDR Peak Day Rebates

PEV Plug-In Electric Vehicles

PGO Power Grid Operator

PHEV Plug-In Hybrid Electric Vehicles

PLC Power Line Communication

PLP Peak Load Pricing

PMU Phasor Measurement Unit

PV Photovoltaic

RC Real Consumption

RTP Real-time Pricing

SC Smart City

SCB Simultaneous Confidence Bands

SCE Southern California Edison

SG Smart Grid

SH Smart Home

SMES Superconducting Magnetic Energy Storage

SVM Support Vector Machines

TES Thermal Energy Storage

TOU Time of Use

V2G Vehicle-to-Grid

VPP Virtual Power Plant

WMN Wireless Mesh Network

WSN Wireless Sensor Network

Appendix E

Publications

Book

1. **Yu Wang**, Shiwen Mao, and R.M. Nelms, *Online Algorithms for Optimal Energy Distribution in Microgrids.* Springer Briefs Series, New York, NY: Springer, June, 2015.

Journal & Magazine Publications

1. **Yu Wang**, Shiwen Mao, and R.M. Nelms, "On hierarchical power scheduling for the macrogrid and cooperative microgrids, *IEEE Transactions on Industrial Informatics,* Special Issue on New Trends of Demand Response in Smart Grid, to appear. DOI: 10.1109/TII.2015.2417496.

2. **Yu Wang**, Shiwen Mao, and R.M. Nelms, Asymptotic optimal online energy distribution in the smart grid, invited paper, *E-Letter of the IEEE Communications Society Multimedia Communications Technical Committee* (MMTC), Special Issue on Smart Grid, vol. 9, no. 4, pp.33-36, July 2014.

3. **Yu Wang**, Shiwen Mao, and R.M. Nelms, Distributed online algorithm for optimal real-time energy distribution in the smart grid, *IEEE Internet of Things Journal*, vol.1, no.1, pp.70-80, Feb. 2014.

4. **Yu Wang**, Shiwen Mao, and R. M. Nelms, "An online algorithm for optimal real-time energy distribution in smart grid," *IEEE Transactions on Emerging Topics in Computing*, Special Issue on Cyber-Physical Systems, vol.1, no.1, pp.10-21, July 2013.

Conference Publications

1. **Yu Wang**, Guanqun Cao, Shiwen Mao, and R.M. Nelms, Analysis of solar generation and weather data in smart grid with simultaneous inference of nonlinear time series, in *Proc. 2015 International Workshop on Smart Cities and Urban Informatics* (SmartCity 2015), Hong Kong, P.R. China, Apr. 2015, pp.672-677.

2. **Yu Wang**, Shiwen Mao, and R.M. Nelms, Optimal hierarchical power scheduling for cooperative microgrids, poster paper, in *Proc. IEEE MASS 2014, Philadelphia*, PA, Oct. 2014,

3. **Yu Wang**, Shiwen Mao, and R. M. Nelms, "A distributed online algorithm for optimal real-time energy distribution in smart grid," in *Proc. IEEE GLOBECOM 2013*, pp.1644-1649, Atlanta, GA, December 2013.

Bibliography

[1] Wald ML (2013) The blackout that exposed the flaws in the grid. In: The New York Times. [online] Available: http://www.nytimes.com/2013/11/11/booming/the-blackout-that-exposed-the-flaws-in-the-grid.html?_r=0. Accessed March 2015

[2] Hurricane Sandy (2015). [online] Available: http://en.wikipedia.org/wiki/Hurricane_Sa -ndy. Accessed Mar 2015

[3] National Institute of Standards and Technology (2010) NIST framework and roadmap for smart grid interoperability standards, release 1.0. [online] Available: http://www.nist.gov/public_affairs/releases/upload/smartgrid_interoperability_final.pd -f. Accessed March 2015

[4] European Committee for Electrotechnical Standardization (CENELEC) (2009) Smart meters coordination group: report of the second meeting held on 2009-09-28 and approval of SM-CG work program for EC submission

[5] Federation of German Industries (BDI e.V.) (2010) Internet of energy–ICT for energy markets of the future. [online] Available: http://www.bdi.eu/BDI_english/103.htm. Accessed March 2015

[6] State Grid Corporation of China (2010) SGCC framework and roadmap for strong and smart grid standards

[7] Japan (2010) Japans roadmap to international standardization for smart grid and collaborations with other countries

[8] Farhangi H (2010) The path of the smart grid. IEEE Power Energy Mag 8(1):18–28

[9] The shift project data portal (2012) World electricity production from all energy sources in 2012 (TWh). [online] Available: http://www.tsp-data-portal.org/Breakdown-of-Electricity-Generation-by-Energ
y-Source#tspQvChart. Accessed Mar 2015

[10] Energy Information Administration (EIA) (2014) Monthly energy review. [online] Available:http://www.eia.gov
/totalenergy/data/monthly/#electricity. Accessed Mar 2015

[11] Zareipour H, Bhattacharya H, Canizares C (2004) Distributed generation: current status and challenges. NAPS 04:1–8

[12] Pepermans G, Driesen J, Haeseldonckx et al (2005) Distributed generation: definition, benefits and issues. Energy Policy 33:787–798

[13] Molderink A, Bakker V, Bosman M et al(2010) Management and control of domestic smart grid technology. IEEE Trans. Smart Grid 1(2):109–119

[14] Carrasco J, Franquelo L, Bialasiewicz J et al (2006) Power electronic systems for the grid integration of renewable energy sources: a survey. IEEE Trans. Industrial Electronics 53(4):1002–1016

[15] Lopes J, Hatziargyriou N, Mutale J et al (2007) Integrating distributed generation into electric power systems: a review of drivers, challenges and opportunities. Electr Power Syst Res 77(9):1189–1203

[16] International Energy Agency (2002) Distributed generation in liberalised electricity ma -rkets. [online] Available: http://gasunie.eldoc.ub.rug.nl/FILES/root/2002/3125958/31 -25958.pdf. Accessed Mar 2015

[17] Coster EJ, Myrzik M, Kruimer B et al (2011) Integration issues of distributed generation in distribution grids. Proceedings of the IEEE 99(1):28–39

[18] Pudjianto D, Ramsay C, Strbac G (2007) Virtual power plant and system integration of distributed energy resources. IET Renew Power Gener 1(1):10–16

[19] Ruiz N, Cobelo I, Oyarzabal J (2009) A Direct load control model for virtual power plant management. IEEE Trans. Power Systems 24(2):959–966

[20] Lombardi P, Powalko M, Rudion K (2009) Optimal operation of a virtual power plant. In: Proc. Power & Energy Society General Meeting. IEEE, New York, pp 1–6

[21] Raab A, Ferdowsi M, Karfopoulos E et al (2011) Virtual power plant control concepts with electric vehicles. In Proc. 16th International Conference on Intelligent System Applications to Power Systems. IEEE, New York, pp 1–6

[22] Jansen B, Binding C, Sundstrom O, Gantenbein D (2010) Architecture and communication of an electric vehicle virtual power plant. In: Proc. SmartGridComm'10. IEEE, New York, pp 149–154

[23] Xiangjiaba–Shanghai HVDC system. [online] Available: http://en.wikipedia.org/wiki/ -Xiangjiaba%E2%80%93Shanghai_HVDC_system. Accessed March 2015

[24] Gemmell B, Dorn J, Retzmann D et al (2008) Prospects of multilevel VSC technologies for power transmission. In: Proc. Transmission and Distribution Conference and Exposition'08. IEEE/PES, USA, pp 1–16

[25] Li F, Qiao W, Sun H et al (2010) Smart transmission grid: Vision and framework. IEEE Trans. Smart Grid 1(2):168–177

[26] Takuno T, Koyama M, Hikihara T (2010) In-home power distribution systems by circuit switching and power packet dispatching. In: Proc. SmartGridComm10. IEEE, New York, pp 427–430

[27] Tashiro K, Takahashi R, Hikihara T (2012) Feasibility of power packet dispatching at in-home DC distribution network. In: Proc. SmartGridComm'12. IEEE, New York, pp 401–405

[28] Gellings CW (2009) The smart grid: enabling energy efficiency and demand response. The Fairmont Press, USA

[29] Denholm P, Mehos M (2011) Enabling greater penetration of solar power via the use of CSP with thermal energy storage. NREL Report No. TP-6A20-52978. Golden, CO: NREL.

[30] Register C (2015) The battery revolution: a technology disruption, economics and grid level application discussion with eos energy storage. [online] Available: http://www.forbes.com/sites/chipregister1/2015/01/13/the-battery-revolution-a-techn-ology-disruption-economics-and-grid-level-application-discussion-with-eos-energy-stor-age/. Accessed Mar 2015

[31] U.S. Department of Energy (2013) Grid energy storage. [online] Available: http://energy.gov/sites/prod/files/2013/12/f5/Grid%20Energy%20Storage%20Decemb-er%202013.pdf. Accessed March 2015

[32] He Y, Venkatesh B, Guan L (2012) Optimal scheduling for charging and discharging of electric vehicles. IEEE Trans. Smart Grid 3(3):1095–1105

[33] Su W, Eichi H, Zeng W et al (2012) A survey on the electrification of transportation in a smart grid environment. IEEE Trans Ind Inf 8(1):1–10

[34] Conway E (2003) World's biggest battery switched on in Alaska. [online] Available: http://www.telegraph.co.uk/technology/3312118/Worlds-biggest-battery-switched-on-in-Alaska.html. Accessed March 2015

[35] Hill C, Such M, Chen D et al (2012) Battery energy storage for enabling integration of distributed solar power generation. IEEE Trans. Smart Grid 3(2):850–857

[36] Hart DG (2008) Using AMI to realize the smart grid. In: Proc. Power and Energy Society General Meeting 2008 – Conversion and Delivery of Electrical Energy in the 21st Century. IEEE, New York, pp 1-2

[37] Gungor V, Lu B, Hancke G (2010) Opportunities and challenges of wireless sensor networks in smart srid. IEEE Trans Ind Electron 57(10)3557–3564

[38] Armenia A, Chow J (2010) A flexible phasor data concentrator design leveraging existing software technologies. IEEE Trans Smart Grid 1(1):73–81

[39] IEEE (2011) P2030/D7.0 draft guide for Smart Grid interoperability of energy technology and information technology operation with the electric power system (EPS), and end-use applications and loads. IEEE Std. 2030-2011, Sep. 2011.

[40] Chen M, Mao S, Zhang Y et al (2014) Big data: related technologies, challenges and future prospects. Springer, New York

[41] Akyildiz I, Wang X (2005) A survey on wireless mesh networks. IEEE Communications Magazine 43(9):23–30

[42] Gungor V, Lambert F (2006) A survey on communication networks for electric system automation. Computer Networks 50(7):877897

[43] Akyol B, Kirkham H, Clements S et al (2010) A survey of wireless communications for the electric power system. [online] Available: https://www.pnnl.gov/nationalsecurity/technical/secure_cyber_systems/pdf/power_gri-d_wireless.pdf. Accessed Mar 2015

[44] Fang X, Misra S, Xue G, et al (2012) Smart grid – the new and improved power grid: a survey. IEEE Commun Surveys & Tutorials 14(99):944–980

[45] Erol-Kantarci M, Mouftah, H (2011) Wireless Sensor Networks for cost-efficient residential energy management in the smart grid. IEEE Trans. Smart Grid 2(2):314–325

[46] Lu B, Habetler T, Harley R et al (2007) Energy evaluation goes wireless. IEEE Industry Applications Magazine 13(2):17–23

[47] Ferreira H, Lampe L, Newbury J et al (2010) Power line communications: theory and applications for narrowband and broadband communications over power lines. Wiley, New York

[48] Galli S, Scaglione A, Wang Z (2010) Power line communications and the smart grid. In: Proc. 2010 First IEEE International Conference on Smart Grid Communications. IEEE, New York, pp 303–308

[49] Vardakas J, Zorba N, Verikoukis C (2015) A survey on demand response programs in smart grids: pricing methods and optimization algorithms. IEEE Commun Surv Tutorials 17(1):152–178

[50] U.S. Dept. Energy (2006) Benefits of demand response in electricity markets and recommendations for achieving them. Report to the United States Congress, Washington, USA

[51] Atwa Y, El-Saadany E, Salama M et al (2010) Optimal renewable resources mix for distribution system energy loss minimization. IEEE Trans Power Syst 25(1):360–370

[52] Bakker V, Bosman M, Molderink A et al (2010) Demand side load management using a three step optimization methodology. In: Proc. SmartGridComm'10. IEEE, New York, pp 431–436

[53] Gormus S, Kulkarni P, Fan Z (2010) The power of networking: how networking can help power management. In Proc. SmartGridComm'10. IEEE, New York, pp 561-565

[54] Brown R (2002) Electric power distribution reliability. Marcel Dekker, New York

[55] Moslehi K, Kumar R (2010) A reliability perspective of the smart grid. IEEE Trans Smart Grid 1(1):57–64

[56] Metke A, Ekl R (2010) Security technology for smart grid networks. IEEE Trans Smart Grid 1(1):99–107

[57] McDaniel P, McLaughlin S (2009) Security and privacy challenges in the smart grid. IEEE Secur Priv 7(3):75–77

[58] Cho H, Yamazaki T, Hahn M (2010) Aero: Extraction of users activities from electric power consumption data. IEEE Trans Consum Electron 56(3):2011–2018

[59] Li H, Mao R, Lai L et al (2010) Compressed meter reading for delay-sensitive and secure load report in smart grid. In: Proc. SmartGridComm10. IEEE, New York, pp 114–119

[60] Efthymiou C, Kalogridis G (2010) Smart grid privacy via anonymization of smart metering data. In: Proc. smartGridComm10. IEEE, New York, pp 238–243

[61] Liu Y, Ning P, Reiter M (2009) False data injection attacks against state estimation in electric power grids. ACM CCS:21–32

[62] Xie L, Mo Y, Sinopoli B (2010) False data injection attacks in electricity markets. In: Proc. SmartGridComm'10. IEEE, New York, pp 226–231

[63] Dan G and Sandberg H(2010) Stealth attacks and protection schemes for state estimators in power systems. In: Proc. SmartGridComm10. IEEE, New York, pp 214–219

[64] Clement-Nyns K, Haesen E, Driesen J (2010) The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. IEEE Trans Power Syst 25(1):371–380

[65] Clement K, Haesen E, Driesen J (2009) Coordinated charging of multiple plug-in hybrid electric vehicles in residential distribution grids. In: Proc. PSCE'09. IEEE/PES, New York, pp 1–7

[66] Papadopoulos P, Jenkins N, Cipcigan L et al (2013) Coordination of the charging of electric vehicles using a multi-agent system. IEEE Trans Smart Grid 4(4):18021809

[67] Wang M, Liang H, Zhang R et al (2014) Mobility-aware coordinated charging for electric vehicles in VANET-enhanced smart grid," IEEE J Sel Areas in Commun 32(7):1344–1360

[68] Lund H, Kempton W (2008) Integration of renewable energy into the transport and electricity sectors through V2G. Energy Policy 36(9):35783587

[69] Lasseter RH (2002) MicroGrids. In: Proc. 2002 Power Engineering Society Winter Meeting. IEEE, New York, 1:305-308

[70] Hatziargyriou N, Asano H, Iravani R (2007) Microgrids. IEEE Power Energy Mag 5(4):78–94

[71] Katiraei F, Iravani R, Hatziargyriou N et al (2008) Microgrids management. IEEE Power Energy Mag 6(3):54-65

[72] Tsikalakis A, Hatziargyriou N (2008) Centralized control for optimizing microgrids operation. IEEE Trans Energy Conver 23(1):241–248

[73] Ahn S, Nam S, Choi J et al (2013) Power scheduling of distributed generators for economic and stable operation of a microgrid. IEEE Trans Smart Grid 4(1):398–405

[74] Wang Y, Mao S, Nelms RM (2013) Online algorithm for optimal realtime energy distribution in the smart grid. IEEE Trans Emerg Top Comput 1(1):10–21

[75] Wang Y, Mao S, Nelms RM (2014) Asymptotic optimal online energy distribution in the smart grid. E-Letter IEEE Commun Soc Multimedia Commun Tech Committee (MMTC) 9(4):33–36

[76] Wang Y, Mao S, Nelms RM (2014) Distributed online algorithm for optimal real-time energy distribution in the smart grid. IEEE Int Things J 1(1):70–80

[77] Wang Y, Mao S, Nelms RM (2015) On hierarchical power scheduling for the macrogrid and cooperative microgrids. IEEE Trans on Ind Inf Appear. DOI: 10.1109/TII.2015.2417496

[78] Wang Y, Mao S, Nelms RM (2013) A distributed online algorithm for optimal real-time energy distribution in smart grid. In: Proc. IEEE GLOBECOM 2013, Atlanta, GA, pp 1644–1649

[79] Wang Y, Mao S, Nelms RM (2014) Optimal hierarchical power scheduling for cooperative microgrids. In: Proc. IEEE MASS 2014, Philadelphia, PA, pp 497–498.

[80] Huang Y, Mao S, Nelms RM (2015) Smooth scheduling for electricity distribution in the smart grid. IEEE Sys J Appear. DOI: 10.1109/JSYST.2014.2340231

[81] Huang Y, Mao S, Nelms RM (2014) Adaptive electricity scheduling in microgrids. IEEE Trans on Smart Grid 5(1):270–281

[82] Huang Y, Mao S (2014) On Quality of Usage provisioning for electricity scheduling in microgrids. IEEE Syst J 8(2):619-628

[83] Huang Y, Mao S, Nelms RM (2013) Adaptive electricity scheduling in microgrids. In: Proc. IEEE INFOCOM 2013, Turin, Italy, pp 1142-1150

[84] Huang Y, Mao S, Nelms RM (2012) Smooth electric power scheduling in power distribution networks. In: Proc. IEEE GLOBECOM 2012 - Workshop on Smart Grid Communications: Design for Performance, Anaheim, CA, pp 1469-1473

[85] Huang Y, Mao S (2012) Adaptive electricity scheduling with quality of usage guarantees in microgrids. In: Proc. IEEE GLOBECOM 2012, Anaheim, CA, pp 5160-5165

[86] Pecas Lopes J, Moreira C, Madureira A (2006) Defining control strategies for microgrids islanded operation. IEEE Trans Power Syst 21(2):916–924

[87] ZTE (2015) [online] Available: http://enterprise.zte.com.cn/us/. Accessed March 2015

[88] Logenthiran T, Srinivasan D, Shun T (2012) Demand side management in smart grid using heuristic optimization. IEEE Trans. Smart Grid 3(3):1244–1252

[89] Roozbehani M, Dahleh M, Mitter S (2012) Volatility of power grids under real-time pricing. IEEE Trans Power Syst (27(4):1926–1940

[90] O'Neill D, Levorato M, Goldsmith A et al (2010) Residential demand response using reinforcement learning. In: Proc. SmartGridComm'10. IEEE, New York, p409–414

[91] Samadi P, Mohsenian-Rad AH, Schober R et al (2012) Advanced demand side management for the future smart grid using mechanism design. IEEE Trans Smart Grid 3(3):1170–1180

[92] H. Mohsenian-Rad A, Leon-Garcia A (2010) Optimal residential load control with price prediction in real-time electricity pricing environments. IEEE Trans Smart Grid 1(2):120–133

[93] Samadi P, Mohsenian-Rad AH, Schober R et al (2010) Optimal real-time pricing algorithm based on utility maximization for smart grid. In: Proc. SmartGridComm'10. IEEE, New York, pp 415–420

[94] Albers S (2003) Online algorithms: a survey. Mathematical Programming 97:3–26

[95] Joseph V, Veciana G (2012) Jointly optimizing multi-user rate adaptation for video transport over wireless systems: Mean-fairness-variability tradeoffs. In: Proc. INFOCOM'12. IEEE, New York, pp 567–575

[96] Borenstein S, Jaske M, Rosenfeld A (2002) Dynamic pricing, advanced metering and demand response in electricity markets. Center for the Study of Energy Markets, Berkeley, CA, 2002

[97] Boisvert R, Cappers P, Neenan B (2002) The benefits of customer participation in wholesale electricity markets. The Electricity Journal 15(3):41–51

[98] Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge Univ. Press, Cambridge, UK

[99] Southern California Edison (2011) 2011 Static load profiles. [online] Available: http://www.sce.com/005_regul_info/eca/DOMSM11.DLP. Accessed March 2015

[100] CAISO (2011) CAISO daily report archives. [online] Available: http://www.ferc.gov/market-oversight/mkt-electric/california/caiso-archives.asp. Accessed March 2015

[101] Ipakchi A, Albuyeh F (2009) Grid of the future. IEEE Power Energy Mag 7(2):52–62

[102] Maharjan S, Zhu Q, Zhang Y, et al (2013) Dependable demand response management in the smart grid: A Stackelberg game approach. IEEE Trans Smart Grid 4:120–132

[103] Zhang Y, Yu R, Yao W (2011) Home M2M networks: architectures, standards, and QoS improvement. IEEE Communication Magzine 49:44–52

[104] Zhang Z, Li F (2010) Scheduling unit-length packets with soft deadlines. In: Proc. INFOCOM'10. IEEE, New York, pp 1–5

[105] Salodkar N, Karandikar A, Borkar V (2010) A stable online algorithm for energy-efficient multiuser scheduling. IEEE Trans Mobile Comput 9(10):1391–1406

[106] Buchbinder N, Lewin-Eytan L, Menache I et al (2012) Dynamic power allocation under arbitrary varying channel–An online approach. IEEE/ACM Trans Networking 20(2):477–487

[107] NIST (2012) NIST framework and roadmap for smart grid interoperability standards, release 2.0. National Institute of Standards and Technology, Tech. Rep., Feb. 2012

[108] NIST (2009) Guidelines for the use of wireless communications: Smart grid priority action plan. Dec. 2009

[109] NIST (2010) Guidelines for smart grid cyber security: Vol. 2, privacy and the smart grid. Tech. Rep. 7628, Aug. 2010

[110] Quinn E (2008) Privacy and the new energy infrastructure. Social Science Research Network, 2008

[111] Fahrioglu M, Alvarado F (2001) Using utility information to calibrate customer demand management behavior models. IEEE Trans Power System 16(2):317–322

[112] Palomar D, Chiang M(2006) A tutorial on decomposition methods for network utility maximization. IEEE J Sel Areas Commun 24(8):1439–1451

[113] Salinas S, Li M, Li P (2013) Multi-objective optimal energy consumption scheduling in smart grids. IEEE Trans Smart Grid 4(1):341–348

[114] Salinas S, Li M, Li P et al (2013) Dynamic energy management for the smart grid with distributed energy resources. IEEE Trans Smart Grid 4(4):2139–2151

[115] Gungor V, Sahin D, Kocak T et al (2011) Smart grid technologies: communication technologies and standards. IEEE Trans Ind Inf 7(4):529–539

[116] Gungor V, Sahin D, Kocak T et al (2013) A survey on smart grid potential applications and communication requirements. IEEE Trans Ind Inf 9(1):28–42

[117] Valverde V, Rosa F, Bordons C (2013) Design, planning and management of a hydrogen-based microgrid. IEEE Trans Ind Inf 9(3):1398–1404

[118] Logenthiran T, Srinivasan D, Wong D (2008) Multi-agent coordination for DER in MicroGrid. In Proc. IEEE ICSET'08, Singapore, pp. 77–82.

[119] Bahrani B, Saeedifard M, Karimi A et al (2013) A multivariable design methodology for voltage control of a single-dg-unit microgrid. IEEE Trans Ind Inf 9(2):589–599

[120] Kim JY, Jeon HJ, Kim KS et al (2010) Cooperative control strategy of energy storage system and microsources for stabilizing the microgrid during islanded operation. IEEE Trans Power Electron 25(12):3037–3048

[121] Balaguer I, Lei Q, Yang S et al (2011) Control for grid-connected and intentional islanding operations of distributed power generation. IEEE Trans Power Electron 58(1):147–157

[122] Di Silvestre ML, Graditi G, Sanseverino ER (2014) A generalized framework for optimal sizing of distributed energy resources in micro-grids using an indicator-based swarm approach. IEEE Trans Power Electron 10(1):152–162

[123] Vandoorn T, Renders B, Degroote L et al (2011) Active load control in islanded microgrids based on the grid voltage. IEEE Trans. Smart Grid 2(1):139–151

[124] Matamoros J, Gregoratti D, Dohler M (2012) Microgrids energy trading in islanding mode. In Proc. IEEE SmartGridComm'12, Tainan City, Taiwan, pp 49–54

[125] Saad W, Han Z, Poor H (2011) Coalitional game theory for cooperative micro-grid distribution networks. In Proc. IEEE ICC'11, Kyoto, Japan, pp 1–5

[126] Dagdougui H, Sacile R (2014)Decentralized control of the power flows in a network of smart microgrids modeled as a team of cooperative agents. IEEE Trans Contr Syst Technol 22(2):510–519

[127] Wei C, Fadlullah Z, Kato N et al (2014) On optimally reducing power loss in microgrids with power storage devices. IEEE J Sel Areas Commun 32(7):1361–1370

[128] Zhang S, Yang J, Wu X et al (2014) Dynamic power provisioning for cost minimization in islanding micro-grid with renewable energy. In Proc. IEEE ISGT'14, Washington, DC, pp 1–5

[129] Minciardi R, Sacile R (2012) Optimal control in a cooperative network of smart power grids. IEEE Sys J 6(1):126–133

[130] Nguyen D, Le L (2013) Optimal energy management for cooperative microgrids with renewable energy resources. In Proc. IEEE SmartGridComm'13, Vancouver, BC, Canada, pp 678–683

[131] Li H, Zhang W (2010) QoS routing in smart grid. In Proc. IEEE GLOBECOM'10, Miami, FL, pp 1–6

[132] Chaouch M (2014) Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves. IEEE Trans Smart Grid 5(1):411–419 2014.

[133] Dordonnat V, Koopman S, Ooms M (2012) Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. Comput Stat Data Anal 56(11):3134–3152

[134] Hippert H, Pedreira C, Souza R (2001) Neural networks for short-term load forecasting: A review and evaluation. IEEE Trans Power Syst 16(1):44–55

[135] Sharma N, Sharma P, Irwin D, et al (2011) Predicting solar generation from weather forecasts using machine learning. In Proc. IEEE SmartGridComm'11, pp 528–533

[136] Zhu S, Yang M, Liu M et al (2013) One parametric approach for short-term jpdf forecast of wind generation. In Procc. IEEE Industry Applications Society Annual Meeting 2013, pp 1–7

[137] Ferraty F, Vieu P (2006), Nonparametric Functional Data Analysis: Theory and Practice. Springer-Verlag, New York

[138] Hardle W (1992) Applied nonparametric regression. Econometric Theory 8(3):413–419

[139] Cao G, ang L, Todem D (2012) Simultaneous inference for the mean function based on dense functional data. Journal of Nonparametric Statistics 24(2):359–377

[140] Zhou Z, Wu W (2009) Local linear quantile estimation of nonstationary time series. Ann Statist 37:2696–2729

[141] Draghicescu D, Guillas S, Wu W (2009) Quantile curve estimation and visualization for nonstationary time series. J Comput Graph Statist 18:1–20

[142] Zhou Z, Wu W (2010) Simultaneous inference of linear models with time varying coefficients. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72:513–531

[143] Wang Y, Cao G, Mao S et al (2015) Analysis of solar generation and weather data in smart grid with simultaneous inference of nonlinear time series. In Proc. 2015 International Workshop on Smart Cities and Urban Informatics, Hong Kong, P.R. China, Apr. 2015

[144] Fan J, Gijbels I (1996) Local Polynomial Modelling and Its Applications. Chapman and Hall, New York

[145] Craven P, Wahba G (1979) Smoothing noisy data with spline functions. Numer Math 31:377–403

[146] Montgomery D, Peck E, Vining G (1991) Introduction to Linear Regression Analysis, 4th ed. Wiley, New York

[147] [online] Available: http://traces.cs.umass.edu/. Accessed March 2015

[148] Cristianini N, Shawe-Taylor J (2000) An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, UK

[149] Hong T (2010) Short term electric load forecasting. Ph.D. dissertation, Graduate Program of Operation Research and Dept. Electrical and Computer Engineering, North Carolina State Univ., Raleigh, NC, USA

[150] Cao G (2014) Simultaneous confidence bands for derivatives of dependent functional data. Electronic Journal of Statistics 8(2):2639–2663, 2014

[151] Wang Y, Mao S, Nelms RM (2015) Online Algorithms for Optimal Energy Distribution in Microgrids. Springer Briefs Series, Springer, New York

[152] Federal Energy Regulatory Commission (2015) Electric power markets: New England (ISO-NE). [online] Available: http://www.ferc.gov/market-oversight/mkt-electric/new-england.asp. Accessed 2015

[153] Hernandez L, Baladron C, Aguiar J (2014) A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. IEEE Communications Surveys & Tutorials 16(3):1460–1495

[154] Fiacco A, Ishizuka Y (1990) Sensitivity and stability analysis for nonlinear programming. Annals of Operation Research, p215–236

[155] Bonnans J, Shapiro A (1998) Optimization problems with perturbations: a guided tour. SIAM REVIEW 40(2):228–264

[156] Rudin W (1991) Function analysis, 2nd ed. McGraw-Hill Inc., New York

[157] Bonnans J, Ioffe A (1995) Quadratic growth and stability in convex programming problems with multiple solutions. Journal of Convex Analysis 2(1/2):41–57

[158] Kushner H, Yin G (2003) Stochastic approximation and recursive algorithms and applications, 2nd ed. Springer, Berlin