Energy-Efficient Designs in Cyber-Physical Systems with a Control and Optimization Approach

by

Yingsong Huang

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
May 5, 2013

Keywords: CPS, smart grid, microgrids, wireless networks, video, energy efficiency, optimization, control, majorization, Lyapunov optimization

Approved by

Shiwen Mao, Chair, Associate Professor of Electrical and Computer Engineering
Prathima Agrawal, Professor of Electrical and Computer Engineering
Jitendra K. Tugnait, Professor of Electrical and Computer Engineering
Tin-Yau Tam, Professor of Mathematics

Abstract

In modern cyber-physical systems (CPS), new dimensions of freedom are enabled to energy efficient solutions. We first focus on the energy delivery side within the smart grid paradigm for smart, efficient and reliable energy delivery. We then explore the demand side with "green" wireless networks for multimedia streaming, in response to the drastic increasing demand in multimedia service in wireless networks.

In this dissertation, we first study energy management systems in smart grid. We design power scheduling policies for smoothing power profile in power distribution networks. The proposed power scheduling policies allow the operator to deploy generators, transformers and power transmission lines with smaller capacity in the grid, thus reducing the capital investment. In addition, the power consumption can be reduced during peak hours, and the average energy generation cost will also be minimized. We also propose a smart electric energy management system in microgrids (MGs). With the proposed algorithm, the MG achieves the fundamental requirements in smart grid with distributed renewable energy integration, energy storage systems management and residential power quality management, while keeping the compatibility to the legacy grid.

We then propose downlink power control frameworks for streaming multiple variable bit rate (VBR) videos in wireless cellular networks. We develop both centralized and low-complexity distributed algorithms, which optimally schedule the transmission power for the BS's, such that VBR videos can be delivered to mobile users without causing playout buffer underflow or overflow under wireless channel uncertainty. The proposed solutions achieve the quality of experience (QoE) requirements of users, as well as keeping the systems "green".

In this dissertation, we adopt a control and optimization approach for energy efficient design in CPS.The synergy of the advanced control and optimization methods in engineering systems provides new visions for practical solutions to bring a green world in the future.

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. Prathima Agrawal, Prof. Jitendra Tugnait, and Prof. Tin-Yau Tam for their valuable comments and time, who contribute their broad perspective in refining the ideas in this dissertation. I am also indebted to Prof. Alvin Lim for serving as the university reader, and reviewing my work.

I would like to thank Prof. Stuart Wentworth for his constant support and inspiration in my instructing the RF Systems Labs. I enjoyed the teaching experience as well as his sense of humor in the labs.

I want to take this opportunity to recognize all my fellow colleagues in the Electrical and Computer Engineering at Auburn University: Dr. In Keun Son, Dr. Donglin Hu, Bobby Black, Jing Ning, Yi Xu, Yu Wang, Zhifeng He and Zhefeng Jiang, 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.

Finally, I would like to extend my heart felt thanks to my parents, without their continuous support and tremendous care, the achievements of this dissertation could not be possible.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Enabling Energy Efficient Cyber-Physical Systems

Innovations in energy supply, delivery and consumption in physical systems are the most effective ways to fight against global warming and environmental degradation. The synergy of communication, networking, control, computation intelligence and physical components greatly extends the autonomy, efficiency, flexibility, reliability and adaptability of physical systems. Such an enhanced system as illustrated in Fig. 1.1 is termed a *cyber-physical system* (CPS). The emerging CPS will significantly enhance the capabilities of physical engineered systems and change how people interoperates with the physical world through real-time embedded systems by sensing, computation, optimization and control over communication networks, which makes the engineering systems reliable, secure, efficient and smart. Examples include smart electricity grid, smart building, smart manufacturing, and smart transportation [2]. The next generation of CPS brings new dimension of freedom to the energy efficient solution to the practical physical systems.

In this dissertation, we investigate energy efficient design of *smart grid* and *information and communications technology* (ICT) infrastructures. Electric power system is generally identified as the largest greenhouse gas emission source created by human-beings. Under traditional electricity grid structure, only $1/3$ of fuel energy, mainly from fossils, is converted to electricity and almost 8% is then further lost during the transmission on the power transmission lines. In addition, to manage the peak demand, about 20% of generation capacity is reserved. Those reserved capacity is only used in 5% of the time [3]. Thus, an efficient, reliable, flexible and economic electricity delivery system is needed for the new era. The next generation of electricity delivery network, called smart grid [4–6], is an evolution of the 20th centuary traditional grid, which is expected to solve the major inefficiency of the traditional electricity grid. A smart grid with two-way flow of both

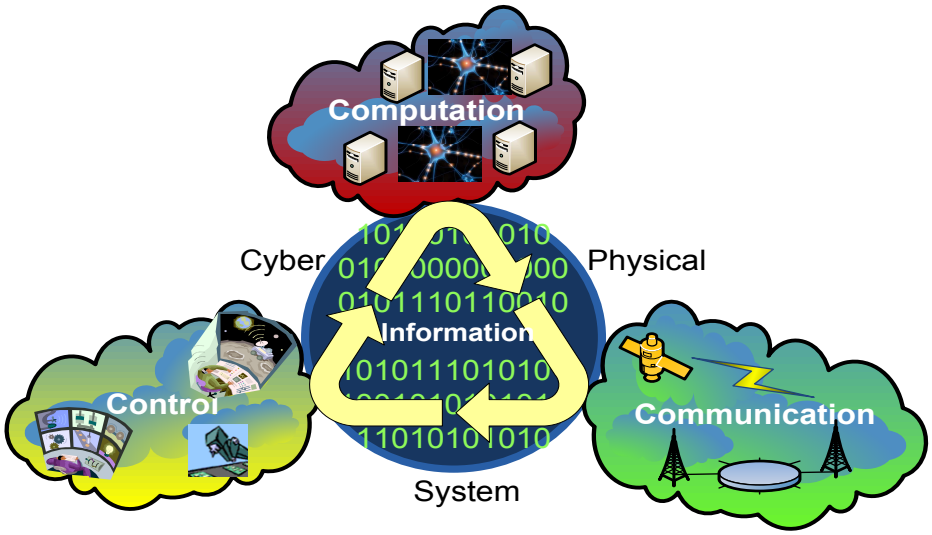


Figure 1.1: Cyber-physical system - Integration of communication, control and computation [1].

electricity and information provides full visibility and automatic control over the components and services in power networks. The ubiquitous sensing, monitoring and automatic control enable the ability for responding to a wide range of conditions and events, which allows the integration of the computational intelligence into the system to efficiently schedule power generation, transmission, distribution and usage. The smart grid can also faciliate the penertration of renewable energy, such as photovoltaics, wind, geothermal, and biofuels, which will greatly reduce resource depletion, increase sustainablity, lower greenhouse gas emissions and reduce air pollution.

Besides the enhancement of electricity delivery networks, it is equally important to implement energy efficient design at the demand side. Since the electricity supply continuously matches the demand under the current operation strategy of traditional grid, the emissions of power plants increase as the demand increases. However, it is not environment friendly to proportionally boost the electricity supply along with demand, especially when we consider that the nation wide electricity demand is estimated to increase by 41% by 2030 [7]. To alleviate the total energy consumption, there is a great need for energy efficient infrastructures, devices and consumption patterns. One of the fastest energy consumption growth comes from today's ICT infrastructure, which may be responsible for more than 10% of the total electrical power consumption [8], due to the tremendously wide spread of the Internet and mobile communication networks. According to a recent study by

Cisco, mobile data traffic will be expected to grow to 6.3 Exabytes per month by 2015, a $26-$fold increase over 2010 [9]. In addition, mobile video will generate much of the mobile traffic growth through 2015. Of the 6.3 Exabytes per month wireless data crossing the mobile network by 2015, 4.2 Exabytes will be related to video.

Furthermore, it is reported that the energy consumed by end-user equipment only contributes around 7% of the entire consumption, while the remaining 93% is consumed by mobile network components [10], of which more than 50% is used by the *base station* (BS) equipment [11]. Therefore, considerable savings on electrical bills could be achieved for wireless operators when the power of BS's is minimized for video streaming. The reduced electricity consumption will also bring about important improvement in the overall carbon footprint of the wireless industry and achieve the goal of "green" communications.

In this dissertation, we examine energy efficient design in CPS from two sides: energy delivery networks and energy demand using multimedia wireless networks as an example. We investigate the problems with a control and optimization theoretic approach, which involves Lyapunov optimization [12], majorization [13], nonlinear and convex optimization [14]. The synergy of these advanced mathematical tools brings about new visions for energy efficient solutions to practical engineering systems with performance bounds to bring a green world in the future.

1.2 Smart Grid

1.2.1 Traditional Electricity Grid

The electric power delivery system, i.e. electricity grid, was named as the greatest engineering achievement of the 20th century [15]. Generally, the electricity grid consists of three parts: power generation plants, power transmission networks, and power distribution networks, as shown in Fig. 1.2. The traditional grid is strictly a hierarchical system, in which the power plant is at the upstream to provide electric power to the user load at the downstream. The electric power is generated at central power plants normally driven by combustible engines that are mostly fueled by coal and gas. Due to economic and environmental considerations, the plants are usually located

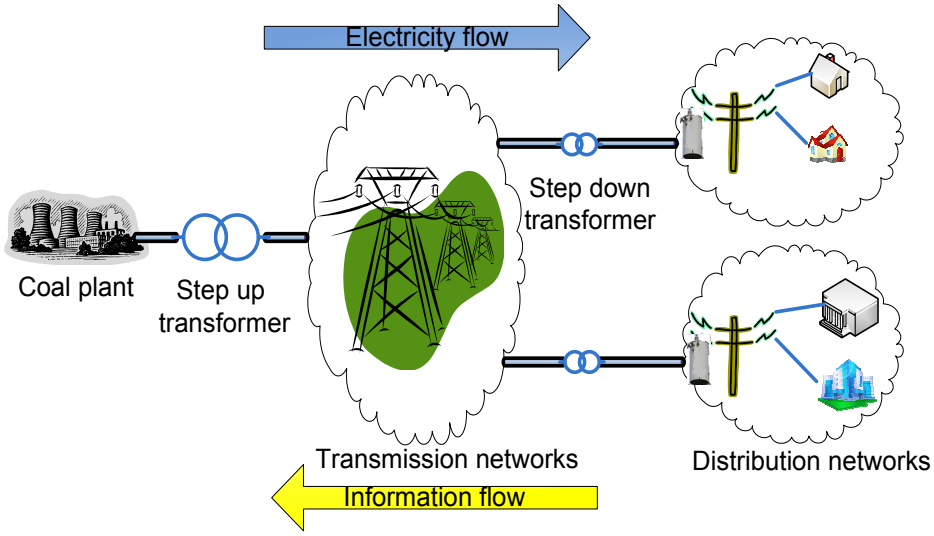


Figure 1.2: Traditional electricity grid with unidirectional power and information flow.

far away from users, and they are connected to users by transmission and distribution networks. The generated electric power is stepped up to a higher voltage (≥ 110 kV) by step up transformers to overcome the power transmission loss over a long distance. The transmission networks route the power to the substations, in which the voltage is stepped down from the transmission level to medium-voltage distribution level (< 33 kV). After the voltage regulation by the substations, the power flow is then forwarded into the distribution networks. Finally, when the power flow arrives at a service location, the voltage is further stepped down to the service level required by end users, which is generally 120 V, 240 V, or 480 V. It can be noticed that the electricity and information in this type of grid flows in a unidirectional fashion. The generated power strictly flows from the plant to the end users, while the information at the downstream is collected by the upstream components. For example, a distribution control center can acquire load information from end users, but the users generally have no idea about the status of the power generation and transmission.

The electricity grid must be operated to achieve real-time balance between generation and load. Otherwise, the grid frequency will drift up or down from the nominal value (typically 50Hz or 60 Hz). Today, the overall daily load profile in a given service area can be predicted well, and the day-ahead generating schedule can be developed based on the prediction. Thus, electricity

generation adopts a “load following” strategy. Due to the limitation of unidirectional flow structure, loads are not generally controlled directly, except for the case when there is insufficient generation available on peak time, and then the “load shifting” operation is executed to encourage users to shift load from on-peak to off-peak periods.

Increasing environmental concerns urge the high penetration of green and renewable energy resources, such as solar, wind, geothermal, tidal, and etc. The load-following strategy with the unidirectional electricity and information flows becomes awkward to meet the penetration of the new renewable resources into the grid. The electricity generated from renewable resources, such as wind and solar, is generally random, due to complex fluctuations of weather condition. Thus, it is hard to accurately predict the generation even in a short period. In addition, renewable energy sites are largely geographically distributed, due to the distribution of the renewable resources, the unidirectional electricity and information flow cannot provide the needed services and fast response to ensure that power generation matches load in real time, which may fail the load following strategy. To embrace the green electrification age, automated, distributed and advanced energy delivery networks should be promoted, which are enhanced by the two-way flow of electricity and information empowered by digital computation, communication and control technologies.

1.2.2 Smart Grid Evolution

Smart grid is a 21st century evolution of electricity delivery systems. Smart grid enhances the traditional power grid through communication, computation, and control technologies throughout the processes of electricity generation, transmission, and distribution. A key feature of smart grid is the two-way flow of electricity and realtime information through communication networks, which offers many benefits and flexibilities to both electricity consumers and providers. The US 2009 Recovery Act indicates that a smart grid will replace the traditional power grid system to improve energy efficiency and advance the liberalization of energy in North America [16].

The smart grid is illustrated in Fig. 1.3. In contrast to the traditional power grid, the power generation and power flow patterns in the smart grid exhibit more flexibility. The initial concept

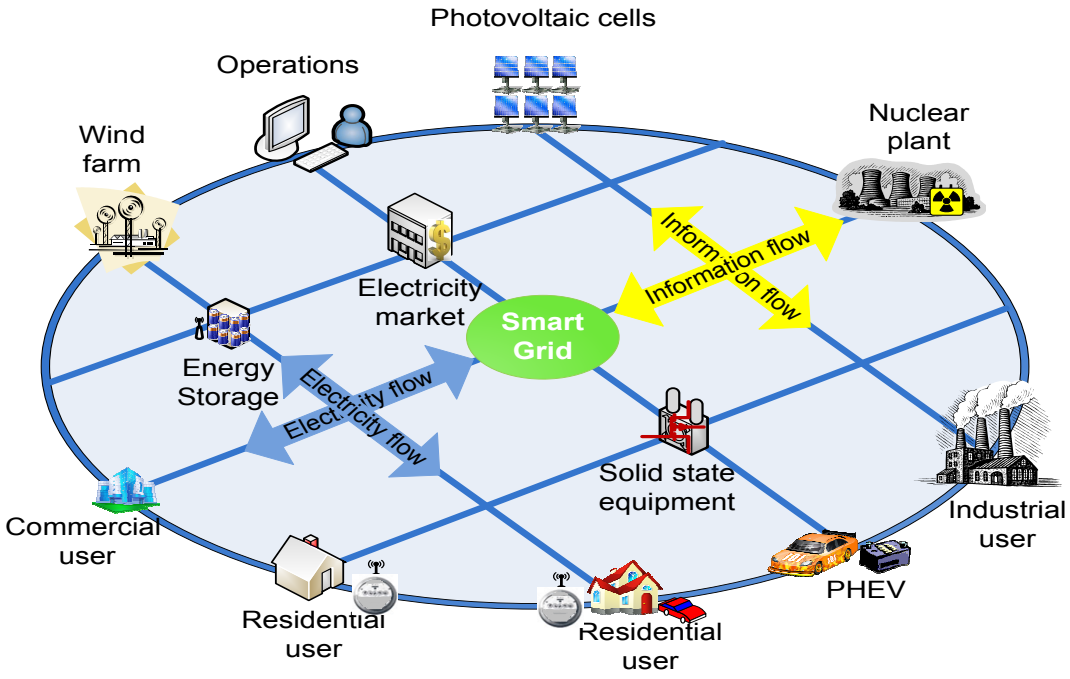


Figure 1.3: Smart grid with plug-and-play interfaces and two-way flow of power and information.

of smart grid begins with the introduction of *automated meter infrastructure* (AMI) systems in distribution networks. AMI enables the utilities to monitor the demand status of end users and impose certain control on the consumption and costs [3, 4, 17]. Future evolution with the integration of various new power electronics and information techniques provides real time sensing, monitoring and control for every corner of the power delivery system. For example, in power delivery networks, *phasor measurement unit* (PMU) are being deployed to synchronized measure the real-time phasor data at multiple points in the grid [18]. *Solid state transformer* (SST) can respond to signals from a facility or a household to change the voltage and other electric characteristics in the system [19]. On the user side, local renewable resources generation and storage systems, smart meters and smart facilities empower the pervasive sensing, monitoring and control of the power flow and power usage in response to the utility supply and market price fluctuations.

One of the most important benefit from smart grid is that the two-way flow of electricity and information facilitates the deployment and management of the *distributed renewable energy resource* (DRER), such as wind farms and solar photovoltaic cells. Unlike the load following strategy, in which the supply continuously matches the demand, the real time information exchange

among DRERs, central power plants and end users provides a new way to match the demand to the available supply by regulating the power generation, as well as controlling the load service level of the users.

The deployment of the DRERs fundamentally alters the operation of power generation, which is conventionally controlled centrally. Furthermore, the increasing popularity of *plug-in hybrid electric vehicles* (PHEVs) serves as distributed energy storage system for residential users by *vehicle-to gird* (V2G) technology [20]. To cope with distributed generation, a concept of *virtual power plant* (VPP) is introduced [21], which clusters numerous DRERs with a total capacity comparable to a traditional power plant. The group of DRERs is managed by a central controller and appears like a virtual central power plant to the grid. VPP provides a promising paradigm to replace a conventional power plant by a cluster of local DRERs with more flexibility and efficiency.

It can be seen from Fig. 1.3 that the smart grid is organized like the Internet, which may be called the "Energy Internet" [22], in contrast to the strictly hierarchical structure of the traditional grid in Fig 1.2. All the components in the power delivery systems, including generation, transmission,distribution and consumption can be deployed and managed through plug-and-play interfaces. By the full duplex of electricity and information flows, configuration of the devices in the system may be customized to respond the grid status in real time. For example, energy storage systems may cooperate with DRERs to balance the supply and demand according to the power generation conditions. On the other hand, users may customize their demand for low cost energy consumption by responding to the realtime market price. The concept of Energy Internet envisions the highly flexible smart grid framework to facilitate a green and sustainable energy-based society, mitigating the growing energy crisis, and reducing the impact of greenhouse gas emissions.

1.2.3 Microgrid

The existing traditional electricity grid has often been cited as the most complex engineering system ever built. Thus to fundamentally overhaul the existing infrastructure is either unimplementable or economically inefficient. The transition to the smart grid would favor the strategy

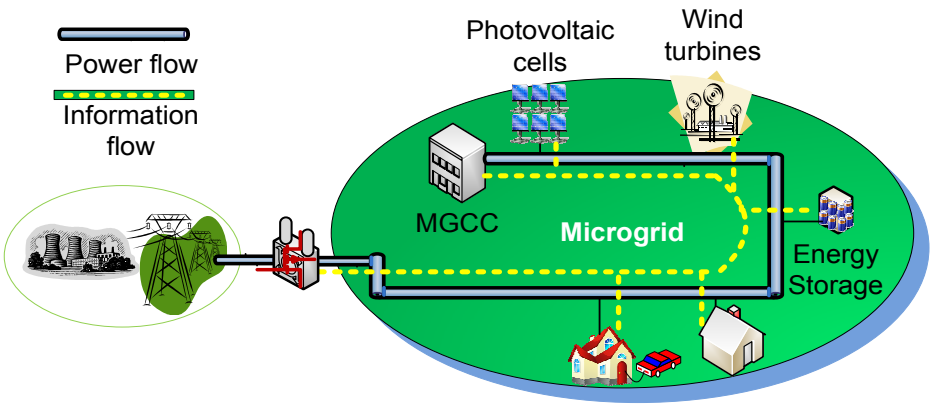


Figure 1.4: Microgrid: Localized cluster of DRERs, ESS's, information networks and residents.

based on the reuse and upgrade from the existing grid by adding capabilities and functionalities in a sustainable growth fashion. In a long period, the smart gird would coexist with the traditional grid, and also provide certain backward compatibility with the legacy systems.

With the concept of plug-and-play interface in the smart grid, a new grid paradigm called *microgrid* (MG) is regarded as a promising component for future smart grid deployment. Microgrids are interconnected networks that provide a localized cluster of renewable energy generation, storage, distribution for local demand, to achieve reliable and effective energy supply with small scale implementation of smart grid functionalities [4, 23]. A typical MG is shown in Fig. 1.4, which includes DRERs, *energy storage systems* (ESS's), wired/wireless networks for information delivery, an *MG central controller* (MGCC), and local users. An MG is centrally controlled and managed by the MGCC [23], which may exchange information with the local users via two-way information networks, such as a wireless network or a *power line communication* (PLC) system. There is a single common coupling point with the macrogrid. When disconnected, the MG works in an *islanded mode*, in which DRERs and ESS's continuously provide electricity to satisfy the local demands. When connected to the macrogrid, the MG may request extra electricity from the macrogrid or sell the excess energy back to the market [3].

The MGs are designed with the fundamental elements of smart grid, such as the integration of DRERs, intelligence core, two-way electricity and information flow, self-healing, and demand side management. This simplified design of the integration of DRERs and the ability to isolate

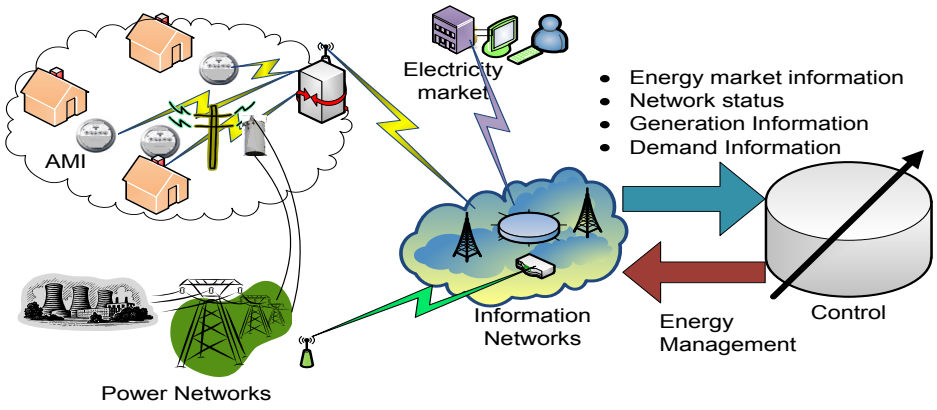


Figure 1.5: Energy management systems enhanced by smart grid paradigm.

the MG from the macrogrid in disturbance will yield highly reliable electricity supply. The island operation mode has the potential to provide a higher local reliability and efficiency than that provided by the general macrogrid. In addition, the MG works with the plug-and-play interface to the macrogrid, which minimizes the regulation effort of DRERs in the cooperative power scheduling and management with the macrogrid and enables sustainable evolution to the emerging smart grid.

1.2.4 Smart Energy Management Systems

The two-way flow of electricity and information infrastructure enables various innovative functions and management principles in practicing dynamic energy management systems, which significantly alter the nature of future power system operation and consumer behavior. It involves the incorporation of smart energy management based on advanced control and communications capabilities, computational intelligence, and smart devices that make the electricity gird “smart”, as shown in Fig. 1.5. The real time information flow from the grid, such as active/reactive power, voltage, phase, user demand, as well as energy price on the market, brings great flexibility in designing the new smart energy management system. With smart energy management systems, the grid achieves energy savings, operation cost reduction, demand and supply balance, emission control, peak load reduction and elimination of many inherent inefficiencies that may be caused by the conventional “load-following” strategy.

Among various designs of smart energy management, *demand response* (DR) serves an important function in smart grid. Demand response is the mechanisms to manage the demand from user side in response to the supply condition. Unlike the "load-following" strategy, which continuously matches the supply to the demand, demand response enables the "generation-following" strategy to match the demand to the available supply by controlling the service level, thus achieving better overall capacity utilization [24]. Currently, the research in demand response mainly focuses on two branches: direct load control and real time pricing.

Direct load control takes advantage of the scheduling flexibility of certain loads, which may be scheduled on and off remotely without degrading the satisfaction of end users. It is estimated that up to 33% residential loads, such as dishwasher, washer/dryer, and PHEV charging, could be rescheduled at some level without major impact on users [24]. Unlike the traditional energy management, the smart demand response mechanism has the capability to aggregate and precisely control the service level of individual load according to the grid status. The application of direct load control not only sheds load during peak demand hours, but also intends to actively promote new types of grid services that could reshape a demand profile to a nicely smoothed demand profile.

Real time pricing is another important method that encourages the users to reshape their electricity consumption pattern by various price strategies. The utilities change energy price based on the fluctuation in the cost of generation, the aggregated load demand, and other realtime states of the grid, thus providing immediate financial incentives to the users to regulate the demand side applications and perform load shifting. The pricing strategy jointly optimizes the cost of generation, user electricity payment and user electricity utilization, which increases economic and energy efficiency and delivers the fair prices to both the utilities and users.

In smart grid, both approaches rely on the information exchange between energy providers and consumers. The control center monitors the real time states of the power networks, as well as transmits control commands to the users by various communication options, including PLC, *wide area networks* (WAN) and *local area networks* (LAN). Smart meters provide the interface between LAN and *home area networks* (HAN), and serve as the gateway for security authentication and

command interpretation to ensure the command issued from the trusted controller. Smart meters may plan the service level of each smart load/facility, and send out the control commands with the service level of individual load through HAN. After the smart load obtains the command, it adjusts its service level according to the new requirement. The process resorts to the support of information transmission protocols, security authentication, energy plan intelligent core in smart meter, and remote control for smart facilities, which are still under development and standardization.

1.3 Energy Efficient Multimedia Networks

1.3.1 Multimedia Networks Architecture

A typical wireless video system is illustrated in Fig. 1.6, which generally consists of the video encoder/ decoder, the wireless/wireline networks, and mobile receiving/playout devices. To achieve QoE guaranteed and/or energy efficient wireless multimedia systems, various schemes have been developed, each of which focuses on one (or several) component(s) of the system .

Video contents provided by commercial video providers or individuals are encoded into compressed frames with different codecs. The picture coding basics are detailed reviewed in [25, 26]. The algorithms in the codecs play large role in the quality of the video, which is directly related to the QoE. Typical encoder and decoder block diagrams are illustrated in Fig. 1.7 and Fig. 1.8, respectively. These frameworks have been adopted in many video coding standards, such as H.264, and MPEG-4, which consist of motion estimation and compensation, *discrete cosine transform* (DCT), quantization, entropy coding, inverse quantization, and *inverse DCT* (IDCT).

After source coding, the encoded frames are then packetized by the network transmission protocols. The video packets are streamed toward the destination through wired and/or wireless networks. When frames arrive at the destination, the codec decodes the received frames. Since the video packets could be corrupted or lost during transmission, error control and concealment techniques at the codec may be applied to mitigate the impact of transmission errors. Then the reconstructed video frames are played out on screens at the receiving device. Multimedia-aware network protocols design has gained consideration for multimedia application support [27–33].

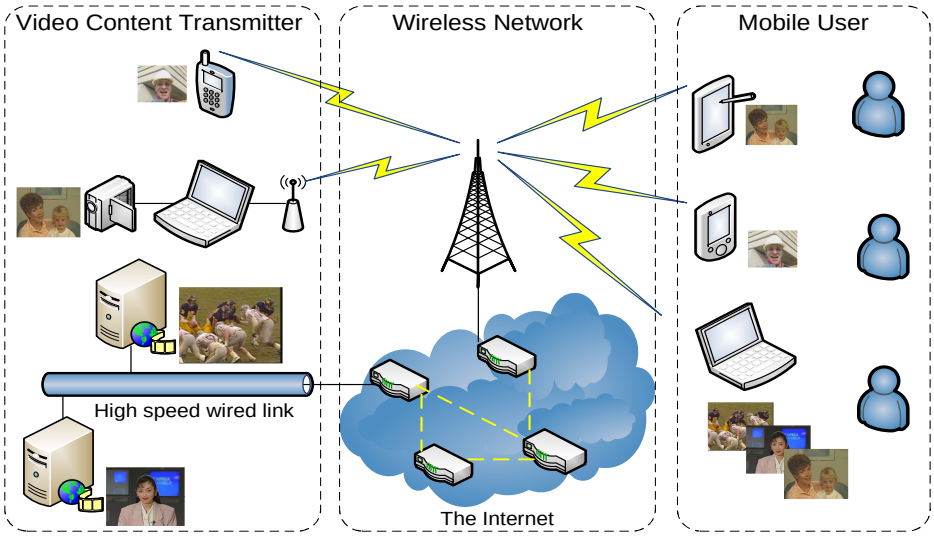


Figure 1.6: Overview of the wireless video networking system.

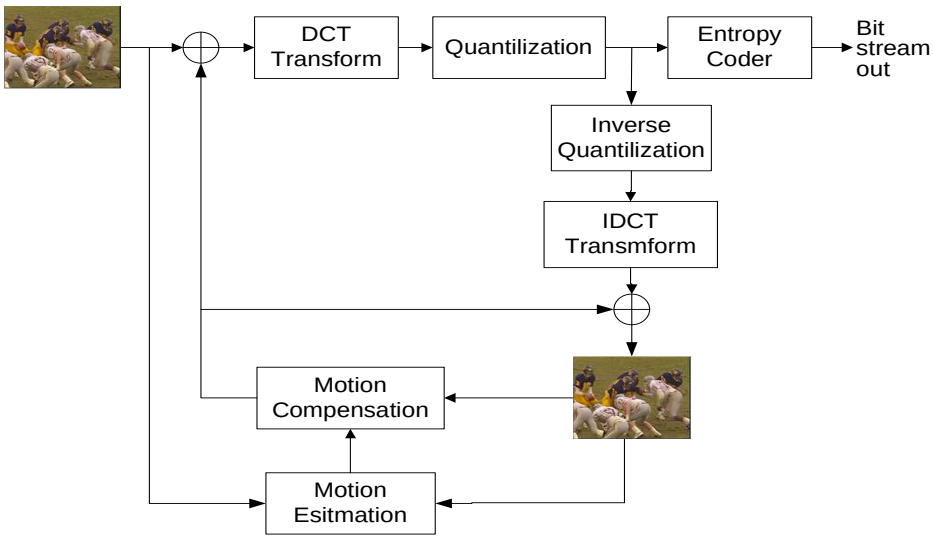


Figure 1.7: Block diagram of a typical video encoder.

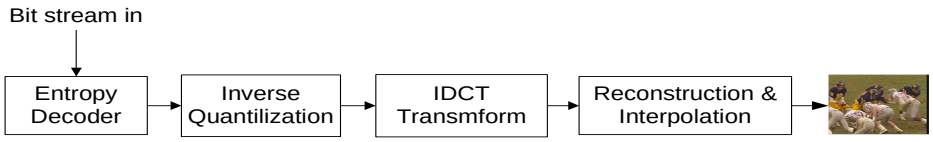


Figure 1.8: Block diagram of a typical video decoder.

During the video streaming process, energy is largely consumed by video codec encoding, network transmissions, receiver decoding, error mitigation, and playout. The energy consumption incurred at the encoder and decoder are mainly due to the processing of video data at the end

nodes, while encoding is usually more computation and energy intensive than decoding. The codec computation and network transmission use up the largest part of the overall energy, and are also critical components for the achievable QoE of video service [34]. It is important to address the "green" communication problem in video streaming by exploring the energy savings in both video codec and video transmission.

1.3.2 Energy Efficiency in Video Coding

In the past decades, the advances of wireless communications and networking technologies are much significant than that of the battery technology. Consequently, how to prolong the battery life of mobile devices becomes one of the major environmental and economical concerns. We focus on power efficient codec design in this section and will explore the energy efficient transmission in the next section.

As the increasing demand of high quality video, high compression efficiency codecs are designed to enable higher resolution, which significantly increases the complexity of encoding algorithms. Moreover, the stringent delay requirements of video service usually keep the video device processor constantly busy for managing the high computation tasks. The processor may consume as much as $2/3$ of the total power of a mobile device [35]. Thus it is important to balance video quality and the computational complexity to achieve power-aware video coding.

Typical encoder and decoder block diagrams are illustrated in Fig. 1.7 and Fig. 1.8, respectively. Such approaches have been adopted in many video coding standards, such as H.264, and MPEG-4. The framework consists of motion estimation and compensation, DCT, quantization, entropy coding, inverse quantization, and IDCT. According to recent research [36], motion estimation/compensation constitutes more than 40% of the CPU workload, DCT/IDCT and quantization/inverse-quantization makes up over 16% of the CPU workload, and the entropy encoder, whose computational complexity largely depends on the coding bit rate, composes less than 10% of the CPU workload. Thus, it is important to explore the energy efficiency of these components that consume the most part of the processing power at a video codec.

There are many power-aware codec designs to balance power consumption and video quality. The main idea behind these techniques is that the diverse complexity of the video content may require different levels of compression. To achieve a certain video quality, slow motion and simple scenes require much less computation than high motion sports and movie streams. Thus it is theoretically feasible to obtain the optimal power efficiency by dynamically adjusting the computation complexity of the codec components for different videos, frames, *macro blocks* (MBs) and blocks, while keeping the video quality relatively constant at a certain level.

In [37], the authors present a configurable coding scheme, which adjusts the codec control parameters to achieve an optimal operation point on complexity-distortion curves based on exhaustive search and the Lagrangian multiplier method. In [38], a power-aware motion estimation algorithm is presented, which is adaptive to the battery status by a content-based subsample algorithm. When the battery is in the full capacity, all the processing elements in the motion estimation function are turned on to provide the best quality. On the other hand, when the battery capacity is decreased, some processing elements are disabled to extend the battery life with little quality degradation. In [39], the authors extend the functions of DCT/IDCT in a framework to decrease the power consumption by skipping the low energy MBs in DCT and all zero coefficients input data in IDCT. The combined method reduces, on average, 94% of power dissipation.

Another class of power-aware video codec design aims to dynamically adjust the voltage and frequency of the CPU for energy conservation. Various *dynamic voltage scaling* (DVS) algorithms are provided to determine the minimum energy consumption for processing video tasks under stringent delay requirements. With a DVS enabled processor, the voltage level and associated clock frequency are adapted to the time-varying video processing workload to save energy. The trade-off between reducing voltage level/clock frequency and increasing processing time is the core in the DVS-based design.

In [40], the authors derive the optimal voltage scheduling with linear programming. The algorithm calculates the optimal scheduling offline with knowledge of the precise complexity and arrival time of each decoding job, which may not be easy to acquire in real time. A heuristic

algorithm is then introduced by predicting the stochastic complexity of the workloads. In [41], a DVS algorithm is presented that adjusts both the clock frequency and the voltage level of the CPU to achieve energy efficiency for video content processing, while maintaining the QoS of the video. A comprehensive statistical analysis of the CPU workload is presented in [42] for multimedia applications. The statistical results show that there is large room for DVS to reduce energy consumption for multimedia streaming and the processor workload can also be accurately predicted with a moderate effort. The DVS system is based on the control theoretic framework. A PID-based DVS controller is developed to achieve a penalty controllable energy reduction, which can be incorporated into an online algorithm.

In summary, the power-aware codec design focuses on the tension between video quality and power consumption based on content diversity. The existing power-aware schemes extend the traditional video codec functions by jointly considering the video content and power constraint. The algorithms aim to adjust the codec parameters to minimize the power consumption while preserving good video quality. In addition, the hardware support for DVS technologies enables adaptive adjustment of the clock frequency and operating voltage level of the CPU, to accommodate varying codec workload. It should be noted that the power-aware codec design needs to jointly adjust large number of configurable parameters, which provides the context for applying effective globally optimal techniques and algorithms.

1.3.3 Energy Efficiency in Video Transmission

A typical video transmission path is shown in Fig. 1.9. The video frames generated by the codec are packetized and delivered through the network protocol stack (UDP/RTP/IP). The link layer schedules the packets with a MAC protocol (e.g., TDMA, FDMA, CDMA, or CSMA/CA) and passes the frames down to the physical layer, where channel coding, modulation and power allocation may be applied to overcome the time-varying and unreliable wireless channels. At the

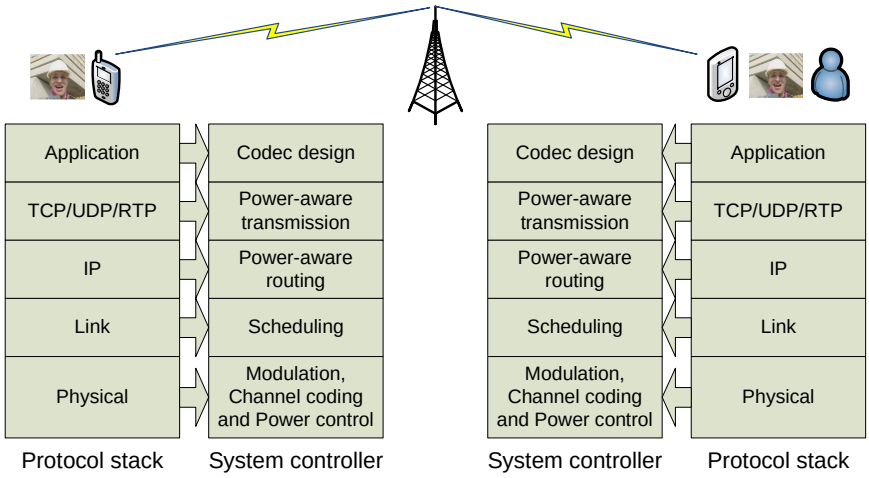


Figure 1.9: Protocol layers and system controls involved in wireless video transmission.

receiver, received video packets are decoded. The video bit stream is restored and then decompressed by the decoder. Error concealment techniques may be applied to mitigate the impact of delayed and corrupted video frames.

The key challenge of video over wireless networks is the time-varying wireless channel, which has a gain that varies over time due to channel fading, shadowing, and inter channel interference [43]. This causes random packet losses and delays. Thus, fixed resource allocation or scheduling schemes may not be sufficient to achieve the best video quality or energy efficiency. An adaptive, video content-aware resource allocation scheme is necessary for supporting energy efficient video streaming over wireless networks. While it is important to increase the bandwidth and throughput in wireless networks [44, 45], energy efficiency is also a critical factor for the success of multimedia applications over wireless networks.

It is reported that BS equipment consumes more than 50% of the total power in a typical cellular network. The energy efficiency and power control in cellular networks thus demand careful reexamination to achieve the goal of green communications. Power control in cellular networks has been widely studied for more than 15 years for voice or data applications. Many effective power control algorithms are proposed in the literature (see [46–50]), and the closely related admission control problems are also investigated [51–53]. The problem of video communications in wireless cellular networks brings about many new challenges to the BS power control problem,

since the power allocations need to achieve not only the target *Signal to Interference-plus-Noise Ratio* (SINR), but also the QoS or QoE of the streamed videos [54–58].

1.3.4 Joint Video Coding and Transmission Design

As shown in Fig. 1.9, the wireless video system is a complex system with many closely coupled control knobs and parameters. Clearly, a cross-layer design that jointly optimizes multiple parameters in different layers has the potential of achieving better energy efficiency and video performance, comparing to the traditional layered approach.

Consider power allocation as an example. Normally, a lower transmission rate requires a smaller transmit power, as well as high compression ratio at the video codec. However, as discussed in the previous section, a higher compression ratio incurs more intensive computations and consumes more power at the codec. Apparently, a cross-layer design framework would be useful in this case, where video coding in the application layer and the transmission schemes in the lower network layers are jointly considered and optimized [56, 59]. Specifically, the lower layer network protocols now have the opportunity to exploit the information from the video content and source coding parameters to optimize the transmission strategy. On the other hand, video coding in the application layer may also take advantage of the channel and network information, and thus can select the coding parameters to provide the best coding quality and be adaptive to the status of wireless networks and channels. This approach also provides future support for the "content-centric" multimedia network design [60–63].

A large number of designs that jointly consider video coding and transmission have been proposed in the literature. In [64], the authors adopt joint source coding and channel coding to minimize the total power consumption, while keeping the end-to-end video quality at a fixed level. A framework of joint source-channel coding and power adaptation is presented in [65], where error resilient source coding , channel coding and transmission power adaptation are jointly designed to optimize video quality, given constraints on the total transmission energy and delay. An algorithm

is introduced to find the minimal energy source coding and power allocation by adaptively allocating resources to different video segments based on their relative importance. In [66], the RF front-end circuit energy is controlled for wireless video transmission by adjusting parameters in physical layer and MAC layer. In [67], the authors investigate the transmission over bandwidth-limited multi-acess wireless uplink channel. Energy-efficient video communication is obtained by jointly adapting video summarization, coding schemes, modulation schemes, and packet transmission. In [23],the authors present a framework for joint network optimization, source adaption and deadline-driven scheduling for multi-user video streaming over wireless networks. Both the physical layer and application layer are jointly considered to maximize the total users' reception quality under the power consumption constraint. In [68], the authors investigate the joint optimization among source coding at application layers, ARQ scheme at data link layers and adaptive modulation and channel coding at physical layer. Within the delay-distortion framework, the parameters of above layers are jointly optimized to achieve the best quality of the received videos.

1.3.5 Video over Emerging Wireless Networks

The video streaming problem has also been considered for several emerging wireless networking paradigms. *Visual sensor networks* (VSN) also represent an important application of energy efficient multimedia networks. VSN consists of a large number of low-power camera nodes, which integrate the image sensor, embedded processor, and wireless transceiver. The development of VSN has brought about many potential applications, such as surveillance, environmental monitoring, smart homes/cities, and visual reality [69], to name a few. Due to the battery limitation, the life time of VSN camera nodes is limited by their energy consumption in wireless channel sensing, transmission, and video and image data processing. Energy efficiency is a critical issue in the design of VSN nodes, since they may not be recharged as often as smart phones, and are expected to operate over extended periods of time (e.g., on two AA batteries for one year [70]). Therefore power efficient designs are highly preferable at all the protocol layers in VSNs.

Comprehensive surveys of VSNs can be found in [69,70]. It has been shown that power-aware routing is highly effective in prolonging the lifetime of wireless sensor networks [71]. In [72, 73], the authors investigate the directional-control data fusion scheme to reduce the amount of sensory data transmission in sensor networks. When processing video data is allowed within the network, data fusion can be employed to reduce the redundancy among multiple video streams along the routing path, thus reducing the volume of transmitted video data and saving energy at the intermediate nodes. Power-aware transport layer designs are mainly based on de facto standard of TCP. In [74, 75], the authors incorporate a new error-recovery mechanism into TCP to avoid unnecessary retransmissions caused by AIMD, especially when the network is disconnected or there are losses due to high bit error. This scheme is shown to prolong the lifetime of wireless sensor networks.

In [76–79], the authors investigate the problem of video transmission over cognitive radio networks, where secondary users sense the licensed channels and aim to exploit the transmission opportunities in the spectrum holes. The uncertain channel availability condition brings about many unique challenges. These works investigate the challenging problem of video over cognitive radio networks with a cross-layer optimization approach, which leads to effective centralized or distributed algorithm design with performance guarantees.

1.4 Key Contributions

In this dissertation, we address the problem of energy efficient design in CPS, including smart grid and multimedia communication networks, with a control and optimization approach. The key contributions are summarized as follows.

First, We explore the problem of smooth electric power scheduling in power distribution networks [80]. The smooth power profile greatly simplifies the strategy for balancing the supply and demand in the grid. Moreover, since electricity generation and transmission systems are generally designed to accommodate peak electric power [4], the smooth demand profile has the advantage of

optimizing the assets and operation cost of the grid. We introduce a deterministic model to characterize the complex relationship between demand and supply. The deterministic model adopts cumulative electricity demand/supply curves, which characterize the time varying demand/supply relationship. A constrained nonlinear optimization problem is then formulated aiming to minimize the electric power variation, as well as satisfy users' power usage quality. We develop majorization-based algorithms for deriving smooth power schedules for the networks. We also design a distributed algorithm for supplying the power among the users. Although many existing work reveals the intrinsic connection between pricing policies and demand response, few of the existing work explicitly address the problem of smooth electric power scheduling. It is shown that the simple off-peak pricing scheme may not be effective in mitigating the demand peak problem, because simply shifting the off-peak period may generate a new *rebound* peak [81]. To the best of our knowledge, this is the first work that directly addresses the smooth optimal energy scheduling with majorization theory in power distribution networks. The solutions show the deterministic performance for smooth power scheduling, peak power and operating cost reduction through the enhancement of bidirectional communication flow, smart meters and smart facilities.

We next propose a comprehensive design of an energy management system in MG by taking advantage of the plug-and-play interfaces of smart grid [82–84]. We jointly consider renewable energy penetration, ESS management, residential demand management, and utility market participation in the MG and introduce the model of *Quality of Service in Electricity* (QoSE). The QoSE concept takes into account minimization of the MG operation cost, while maintaining the power usage quality of residents. We transform the QoSE control problem and ESS management problem into queue stability problems by introducing the QoSE virtual queues and battery virtual queues. The Lyapunov optimization method is applied to solve the problem and generate the online optimal ESS charge/discharge algorithm, adaptive residential load service and cost effective operation strategy on utility market, with hard performance bounds, which do not require any statistics and future knowledge of the electricity supply, demand and price processes. The proposed policy effectively reduces the MG operation cost and maintains the QoSE for the residents. With this new

energy management framework, the MG achieves the fundamental requirements of smart grid in DRERs integration, ESS's management, residential power quality management, and maintains the compatibility to the legacy grid.

Furthermore, we investigate the energy efficient design on the demand side as applications in wireless communication and networks. Among various green communication technologies, we focus on the energy efficiency of base stations for downlink video streaming. This is due to the expected surge in wireless video data, as well as the drastic increase in the deployment of BS's. Therefore, any small improvement in the energy efficiency of wireless video streaming system will be amplified by the huge volume of wireless video data and number of BS's deployed, and will result in considerable environmental impact.

Specifically, we design the energy efficient streaming for *variable bit rate* (VBR) video over wireless networks. VBR video offers stable and superior quality over *constant bit rate* (CBR) videos, however, the complexity statistics of the VBR video frames introduces great challenge in wireless network design. We first present analytical frameworks for streaming multiple VBR videos in a wireless cellular networks, where downlink capacities are limited by inter-cell/intracell interference [55, 56, 58]. By jointly considering the deterministic model for VBR video traffic, stringent playout delay constraint, BS peak power constraint, wireless channel uncertainty and finite playout buffer at the mobile users, we formulate the video streaming systems as nonlinear optimization problems with the objective to maximize the throughput under the QoE and power constraint. For the intracell interference situation, we analyze the convexity conditions of the problem and propose a two-step approach to maximize the streaming transmission throughput under power constraint, while maintaining the QoE. We also develop a distributed algorithm based on the dual decomposition technique. The more challenging problem involved in intercell interference is solved by a centralized branch-and-bound algorithm incorporating the Reformulation-Linearization Technique, which can produce optimal bounded solutions. We also propose a low-complexity distributed algorithm with fast convergence. The proposed solutions effectively make

use of the power in BS's to stream the VBR video in cellular networks, while preserving the QoE requirement.

We further study the energy efficient downlink multi-user VBR streaming in the wireless cellular networks with orthogonal channels by directly minimizing the power consumption to achieve green multimedia communications [57,85–87]. We present a cross-layer optimization and scheduling framework with the objective to minimize the BS's power consumption during steaming period while maintaining the QoE of video users. We develop a majorization-based solution approach to solve the formulated problem. We prove the proposed algorithm is unique and global optimum, and demonstrate that the proposed algorithm is also smoothness optimal. These research projects may bring about a new paradigm for the design of future green wireless multimedia networks.

1.5 Overview of the Dissertation

In this dissertation work, we focus on the energy efficient design in CPS from two sides: electricity delivery networks and wireless multimedia networks, which are integrated by the methodology of control and optimization theoretic design as indicated in Fig. 1.10. The rest of the dissertation is organized as follows.

We present the smooth electric power scheduling in power distribution networks in Chapter 2. We introduce an electricity supply/demand model that takes into account of time-varying demands and their deadlines. We formulate a constraint nonlinear optimization problem and incorporate the theory of majorization to develop algorithms that can compute smoothness optimal schedules. After the smooth power schedule is obtained, a distributed user benefit maximization load control scheme is used to allocate the scheduled power to individual users, while maximizing their level of satisfaction. We demonstrate the efficacy of the proposed algorithms by extensive simulations.

In Chapter 3, we propose a smart energy management framework in MG based on the concept of QoSE. The MGCC aims to minimize the MG operation cost and maintain the outage probability of quality usage, i.e., QoSE, below a target value, by scheduling electricity among renewable energy sources, energy storage systems, and macrogrid. We formulate the problem to a constrained

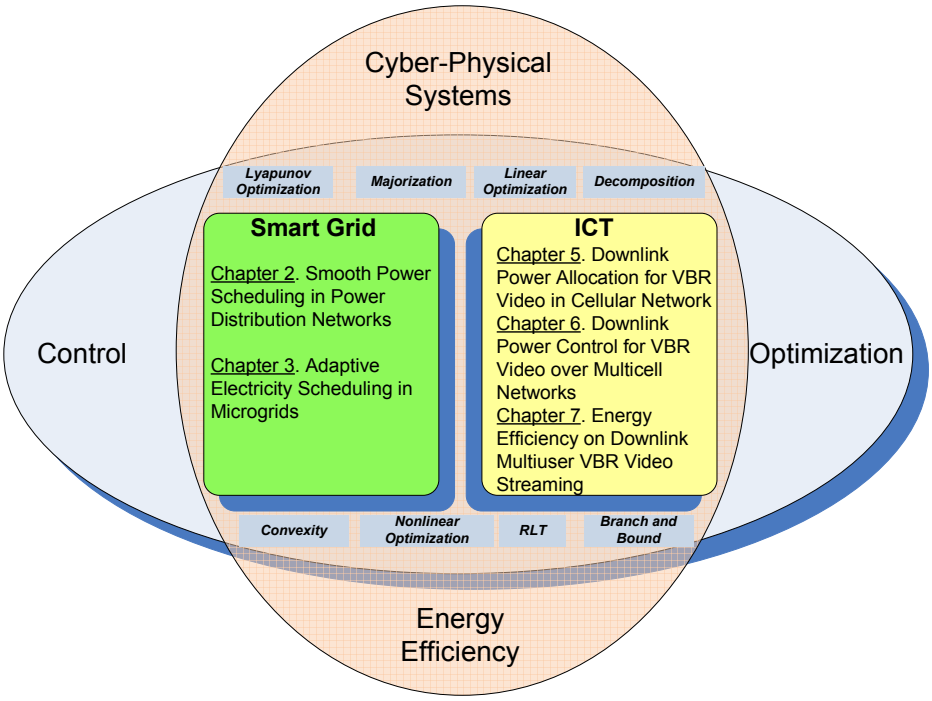


Figure 1.10: Overview of the dissertation.

stochastic programming problem and apply Lyapunov optimization technique to derive an adaptive electricity scheduling algorithm by introducing the QoSE virtual queues and energy storage virtual queues. We derive several hard performance bounds for the proposed algorithm and evaluate its performance with trace-driven simulations.

In Chapter 5, we shift the energy efficient design from electricity delivery networks to the demand side. We investigate the power control for the downlink VBR video streaming in the cellular networks with intracell interference. We consider a deterministic model for VBR video traffic and finite playout buffer at the mobile users. The objective is to derive the optimal downlink power allocation for the VBR video sessions, such that the video data can be delivered in a timely fashion without causing playout buffer overflow and underflow. The formulated problem is a non-linear nonconvex optimization problem. We analyze the convexity conditions for the formulated problem and propose a two-step approach to solve the problem. We also develop a distributed algorithm based on the dual decomposition technique. The performance of the proposed algorithms are validated with simulations using VBR video traces under realistic scenarios.

In Chapter 6, we study the more challenging problem of power control for streaming multiple VBR videos in multicell wireless networks under the intercell interference condition and derived both centralized and distributed algorithm for the solutions. The problem is formulated to find the optimal transmit powers for the base stations, such that VBR video data can be delivered to mobile users without causing playout buffer underflow or overflow. We formulate a nonlinear nonconvex optimization problem and prove the condition for the existence of feasible solutions. We then develop a centralized branch-and-bound algorithm incorporating the Reformulation-Linearization Technique, which can produce $(1-\epsilon)$-optimal solutions. We also propose a low-complexity distributed algorithm with fast convergence. Through simulations with VBR video traces under fading channels, we find the distributed algorithm can achieve a performance very close to that of the centralized algorithm.

In Chapter 7, we relax the channel constraint in the cellular wireless networks to the orthogonal channels and directly address power minimization strategy for multiuser VBR video streaming. We also adopt a deterministic model for VBR video traffic that incorporates video frame and playout buffer characteristics, and formulate a constrained stochastic optimization problem. We then develop a majorization-based solution approach. For the case of a single VBR video session with relaxed peak power constraint, we develop a power optimal algorithm with low complexity. We prove the power optimality of the proposed algorithm and the uniqueness of the global optimum, and demonstrate that the proposed algorithm is also smoothness optimal. For the case of multiuser VBR video streaming, we develop a heuristic algorithm that selectively suspends some video sessions when the peak power constraint is violated. In addition to the traditional VBR video streaming application, we also consider the case of interactive video streaming, and show that the proposed schemes can be easily adapted and applied.

We conclude the dissertation and present the future work in Chapter 8.

Chapter 2

Smooth Electric Power Scheduling in Power Distribution Networks

2.1 Introduction

The emergence of *Smart Grid* (SG) brings about many fundamental changes in electric power systems [4]. Various new power electronics and information techniques are greatly advancing the control and management of energy and resources in the power system. For example, *solid state transformers* (SST) can respond to signals from a facility or a household to change the voltage and other electric characteristics. On the user side, smart meters and smart facilities empower the pervasive monitoring and controlling at all levels of power usage in response to power supply and market price fluctuations [4]. The two-way flows of electricity and information in SG are instrumental to the control and optimization of energy and resource allocation in the grid to achieve efficient, green and robust energy systems.

Unlike the traditional grid, in which the electricity supply continuously matches user demands, the next generation power distribution system is based on a network structure [88] and is capable of allowing users to control their loads in response to the dynamics in the grid. *Demand response* (DR) is a technique to balance power generation and demand in the grid [89]. One of the important targets of DR is to reduce the peak demand by scheduling user requests. With the two-way information flow among provider, users and the market, various DR schemes based on real-time pricing and day-ahead load response concepts have been investigated recently [90–93]. Most of the existing DR schemes aim to maximize the social welfare or minimize the electricity payment under given demand requirements. Although revealing the intrinsic connection between pricing policies and demand response, the problem of smooth electric power scheduling is not explicitly addressed, although being the key issue in DR. It is shown that the simple off-peak pricing

scheme may not be effective in mitigating the demand peak problem, because simply shifting the off-peak period may generate a new *rebound* peak [81].

In this chapter, we address the challenging problem of *smooth electric power scheduling* in power distribution networks. The network model is shown in Fig. 2.1. We assume the end-users are equipped with smart meters and are capable of communicating with the distribution substation and the distribution control center (DCC) through a communication network, and receiving commands from the DCC to adjust the user's electric energy consumption level [3]. The DCC schedules electricity supply on daily basis, which is further divided into multiple time slots. The electricity usage requests at each user are classified into two categories: the *priority load* that must be satisfied in every time slot, and the *deferrable load* that should be satisfied before specific deadlines. Users may set the load for each type according to their preference (e.g., lighting, entertainment, laundry, or charging a plug-in hybrid electric vehicle (PHEV)) [89]. The DCC aggregates the demand profiles from the users through the aggregator [94] and smooths the aggregated electric power supply under the priority load and deferrable load deadline constraints.

In the smooth electric power scheduling problem, the objective is to minimize the power variation during a daily period, based on the concept of day-ahead load response. A *deterministic electricity supply/demand model* is introduced with cumulative electricity demand/supply curves, which characterize the demand/supply relationship during the day. We find the formulated problem suits well with the *majorization* theory, which concerns with the comparison and ordering of vectors with respect to the distribution of their elements [13]. Majorization has been used in solving optimization problems in the communications and networking area [57, 95, 96]. In this chapter, we present a majorization-based framework to develop two smooth electric power scheduling algorithms with low computational complexity. After the smooth electric power profile for the entire network is obtained, a user benefit maximization load control algorithm will be executed to allocate the total amount of supply to the individual users, while maximizing their satisfaction of electricity usage. The proposed algorithms can achieve the minimum peak power, thus requiring smaller capacity for the generators, transmission lines and transformers to support the same demand. Since

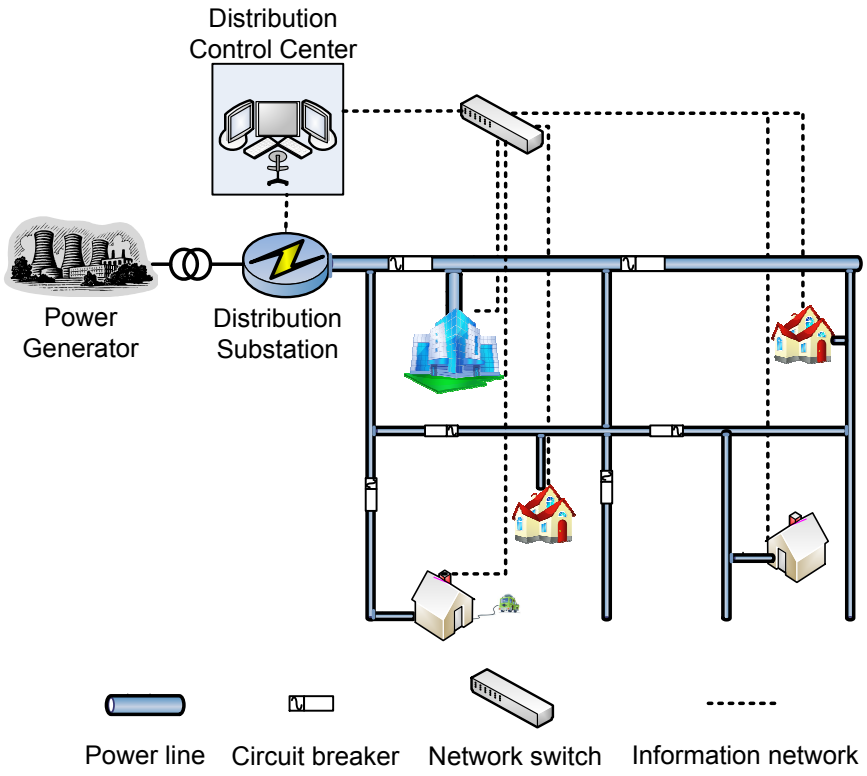


Figure 2.1: Illustration of the electricity distribution network.

electrical generation and transmission systems are generally designed to accommodate peak electric power [4], the smooth electric power schedule has the potential of optimizing the deployment and operation cost of the grid.

The remainder of this chapter is organized as follows. We first present the system model and problem statement in Section 2.2. The smooth electric power scheduling algorithms are described in Section 2.4 and their performance evaluated in Section 2.5. Related work is discussed in Section 2.6 and Section 2.7 concludes this chapter.

The notations used in this chapter are summarized in Table 2.1.

2.2 Problem Statement

2.2.1 Load Demand Profile

We consider a power distribution network with two-way flows of electricity and information. We assume N users in the power distribution network, which may generate residential, commercial

Table 2.1: Notation Table for Chapter 2

Symbol	*Description*
$\mathbb{R}$	set of the electricity consumption users in the power distribution networks
N	total number of local electricity consumption user
L	total number of slots
τ	slot period
$p_n(t)$	power consumption of user n in slot t
$e_{n,p}(t)$	priority load electric energy of user n in slot t
$e_{n,d}(t)$	deferrable load electric energy need to be fulfilled since last deadline of user n in slot t
$p_n^{max}(t)$	maximum power consumption of user n in slot t
$p_n^{min}(t)$	minimum power consumption of user n in slot t
$e_n(t)$	electricity usage of user n in slot t
$e_n^{max}(t)$	maximum electricity usage of user n in slot t
$e_n^{min}(t)$	minimum electricity usage of user n in slot t
E_n	total electricity usage during L slots for user n
$E_{max}(t)$	maximum electricity usage for the residential area in slot t
$E_{min}(t)$	minimum electricity usage for the residential area in slot t
$E(t)$	scheduled electricity usage for the residential area in slot t
$\vec{W}_{max}$	cumulative upper bound of electricity usage during L slots
$\vec{W}_{min}$	cumulative lower bound of electricity usage during L slots
$\vec{W}$	cumulative scheduled electricity usage during L slots
Φ	total electricity usage for all users during L slots
$\vec{P}$	feasible power supply scheduling during L slots
$\vec{P}^*$	smooth optimal power supply scheduling during L slots
$U_n(\cdot)$	utility function for user n
$h(t)$	electricity price at time slot t
ν	Lagrange multiplier
$\mathcal{L}$	Lagrange function
$\kappa(l)$	stepsize of step l in equation (2.7)
$\alpha(l)$	stepsize of step l in equation (2.8)

and industrial loads. Let $\mathbb{R} = \{1, 2, \cdots, N\}$ be the set of users. The electric demand of a user is daily based. Without loss of generality, we assume the one day period is divided into L time slots, each with length τ. Let $p_n(t)$ be the power consumption of user n in time slot t, which is time varying but remains constant within the time slot. Each user n knows its own total daily demand, i.e., $E_n = \sum_{t=1}^{L} p_n(t)\tau$, and wishes to schedule the demand over the one day period [91].

We assume the total demand E_n consists of two parts: the *priority load* and the *deferrable load*. The priority load should be strictly guaranteed in a time slot (e.g., for lighting), while the deferrable load can be served flexibly but with a specific deadline (e.g., charging a household battery or PHEVs). We define $e_{n,p}(t)$ and $e_{n,d}(t)$ as the electric energy for priority load in time slot t, and the deferrable load that must be satisfied by time slot t, respectively. The minimum demand of user n in time slot t, denoted by $e_n^{min}(t)$, is the sum of $e_{n,p}(t)$ and $e_{n,d}(t)$. Finally, let $e_n^{max}(t)$ be the maximum possible demand for user n, which is limited by the amount of deferrable loads that have not been satisfied yet, and the capacity of protective relays and switches of the users.

2.2.2 Cumulative Demand and Supply Curves

At the beginning of a day, the DCC will aggregate the individual demand profiles received by communicating with the smart meters and smart facilities via the communications network [3]. Let the total minimum electricity demand in time slot t be $E_{min}(t) = \sum_{n\in\mathbb{R}} e_n^{min}(t)$. We have $E_{min}(L) = \sum_{n\in\mathbb{R}} E_n = \Phi$, since the daily aggregated demand of all users, denoted by Φ, should finally be satisfied by the end of the day. We define the *cumulative minimum demand curve* $\vec{W}_{min}$ as

$$W_{min}(t) = \sum_{l=1}^{t} E_{min}(l), 1 \le t \le L. \tag{2.1}$$

We define the *cumulative maximum demand curve* $\vec{W}_{max}$ to represent the maximum amount of electricity demand that can be consumed up to t as

$$W_{max}(t) = \min\{W_{min}(t-1) + \sum_{n\in\mathbb{R}} [e_n^{max}(t) + \Delta e_n(t-1)], \Phi\}, 1 \le t \le L$$

where $\Delta e_n(t) = e_n^{max}(t) - e_n^{min}(t)$ is the deferrable load that can be served in slot t but with deadlines later than t. To incorporate the priority load in the model, $W_{max}(t)$ also satisfies $W_{max}(t) \ge W_{max}(t-1) + \sum_{n\in\mathcal{R}} e_{n,p}(t)$.

For given demand curves $\vec{W}_{min}$ and $\vec{W}_{max}$, we aim to find a feasible electricity schedule $\vec{W}$, which is the *cumulative supply* of electricity to the users that satisfies constraints $W_{min}(t) \le$

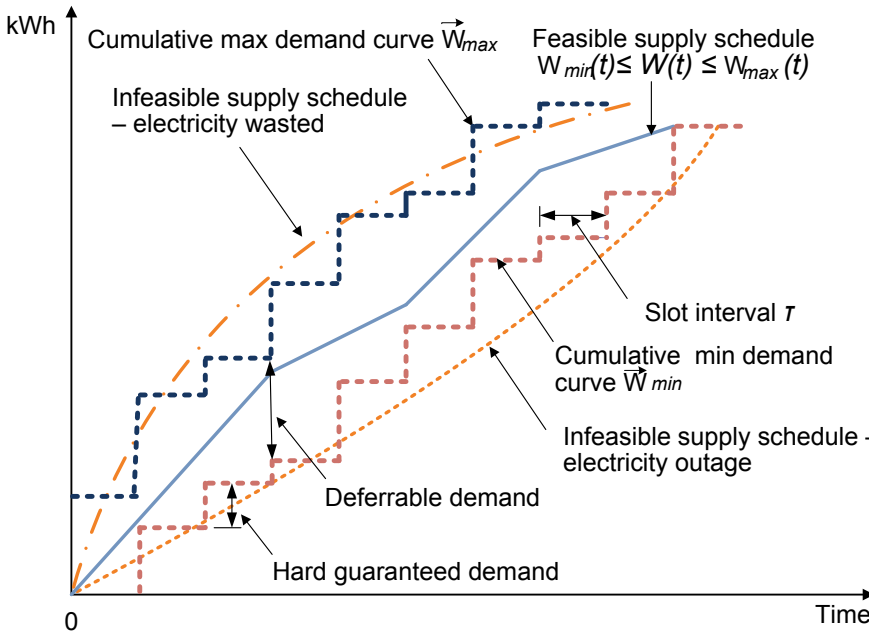


Figure 2.2: Cumulative demand and supply curves.

$W(t) \leq W_{max}(t)$, for all $1 \leq t \leq L$, and $W(L) = \Phi$ (i.e., the total demand should be satisfied by the end of the day). The three cumulative curves are illustrated in Fig. 2.2, which are all nondecreasing over time.

The proposed demand and supply model is quite general. It does not assume any mathematical model for either the supply or the demand. It is more practical than the complex statistical models for supply and demand used in the literature [4]. The cumulative curves represent the demand/supply status in the power distribution network. In each time slot t, $W_{min}(t)$ tracks the priority load and the deferrable load with deadline t, while $W_{max}(t)$ represents an upper bound of the possible consumption by time t. The gap between $W_{max}(t)$ and $W_{min}(t)$ may accommodate the future uncertainty of the electric power usage. The slope of $W(t)$, denoted by $P(t)$, corresponds to the scheduled electric power. The DCC aims to find an optimal schedule $W(t)$ for every time slot t to achieve a specific control target. A feasible power supply schedule $\vec{P} = [P(1), P(2), \cdots, P(L)]$ ensures that $\vec{W}$ lies between $\vec{W}_{min}$ and $\vec{W}_{max}$ for all the L time slots, thus preventing both outage events and energy waste.

It can be seen from Fig. 2.2 that the feasible electric power schedule may not be unique. Among various feasible schedules, we are interested in the one that distributes electricity most smoothly among the L time slots, i.e., the *smoothness optimal schedule*. Once the DCC obtains the smoothness optimal schedule, it can announce the schedule to the smart meters and smart utilities at the users' premises via the communication network, and the users can shape their demand to match the schedule (assuming cooperative users). Therefore, we can achieve smooth electricity generation, transmission and consumption, which is highly preferable for the grid design and operation [4].

2.2.3 Smooth Power Scheduling Problem

Based on the demand and supply model, we formulate the smooth power scheduling problem in this section. Let $\bar{P} = \Phi/(L\tau)$ be the average power consumption in the power distribution network through the daily period. The scheduled power for each time slot is $P(t) = W(t)/\tau$. The smoothness optimal schedule minimizes the variations of the supplied power over the entire period, i.e.,

$$\begin{aligned} \text{maximize:} \quad & \mathcal{S}(\vec{P}) \\ \text{subject to:} \quad & W_{min}(t) \leq W(t) \leq W_{max}(t), \text{ for all } t \\ & W(L) = W_{min}(L) = W_{max}(L) = \Phi \\ & P(t) \geq E_p(t)/\tau, \text{ for all } t, \end{aligned} \tag{2.2}$$

where $\mathcal{S}(\vec{P})$ is the *smoothness* of a schedule $\vec{P}$, $E_p(t) = \sum_{n\in\mathbb{R}} e_{n,p}(t)$ is the total priority load in time slot t.

Generally, smoothness can be measured by different metrics, such as variance, cumulative absolute difference, etc. Each smoothness measure leads to a different objective function in problem (2.2), while the solution to the problem will then depend on the specific form of the objective

function. In addition, the smoothness measures are generally nonlinear, making the problem non-trivial to solve. In this chapter, we resort to a mathematic theory of majorization [13], which explicitly addresses the unique mathematical notion for smoothness. Applying majorization theory, we will see that for an arbitrary smoothness objective function in problem (2.2) that satisfies the Schur-convex properties [13], the problem can be solved by a universal algorithm in polynomial time. For brevity in the deduction, we minimize variance in the rest of the chapter, while the solution algorithms developed in Section 2.4 apply to any objective function that is Schur-convex.

We first consider the case where deferrable load is the dominant component [97], i.e. $E_p(t) \approx 0$. Problem (2.2) is then reduced to problem (2.3).

$$\begin{aligned} \text{minimize:} \quad & \textstyle\sum_{t=1}^{L} \left[P(t) - \bar{P}\right]^2 / L \\ \text{subject to:} \quad & W_{min}(t) \leq W(t) \leq W_{max}(t), \text{ for all } t \\ & W(L) = \Phi \\ & P(t) \geq 0, \text{ for all } t. \end{aligned} \tag{2.3}$$

This problem fits well with the *majorization* theory, since the objective function is Schur-convex [13]. We briefly review Majorization preliminary in Section 2.3. Applying majorization, we will design a smooth electric power scheduling algorithm for solving problem (2.3) in Section 2.4.1. We will then extend the algorithm for solving problem (2.2) in Section 2.4.3.

2.3 Majorization Preliminaries

Majorization theory [13] describes the "less spread out" or "more nearly equal" properties of the elements of a vector comparing to the elements of another vector. It concerns with the problem of ordering vectors with nonnegative, real elements, as well as order-preserving functions. For simplicity, all the vectors in this section are row vectors.

Definition 2.1. *For two n-dimensional vectors $\vec{X} = (x_1, x_2, \cdots, x_n)$ and $\vec{Y} = (y_1, y_2, \cdots, y_n)$, with elements sorted in the non-increasing order as $x_1 \geq x_2 \geq \cdots \geq x_n \geq 0$ and $y_1 \geq y_2 \geq$*

$\cdots \geq y_n \geq 0$. *$\vec{X}$ is said to be* majorized *by $\vec{Y}$, denoted as $\vec{X} \prec \vec{Y}$, if (i) $\sum_{i=1}^{t} x_i \leq \sum_{i=1}^{t} y_i$, $t = 1, 2, \cdots, n-1$, and (ii) $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ [13].*

Definition 2.2. *A real-valued function ϕ defined on a set $\mathcal{A} \subset \mathcal{R}^n$ is said to be* Schur-convex *on $\mathcal{A}$ if $\vec{X} \prec \vec{Y}$ on $\mathcal{A} \Rightarrow \phi(\vec{X}) \leq \phi(\vec{Y})$ [13].*

Schur-convex functions have the "order-preserving" property, which bridges majorization to optimization. Schur-convex functions can be validated with the following fact.

Fact 2.1. *If ϕ is symmetric and convex, then ϕ is Schur-convex. Consequently, $\vec{X} \prec \vec{Y}$ implies $\phi(\vec{X}) \leq \phi(\vec{Y})$ [13].*

This fact provides connection between ordering and its order-preserving functions. By this fact, we may solve the minimization optimization problem by generating the most possible "spread out" vector as the solution. If $\phi = \sum g$ and g is continuous convex, then we have the following strong fact:

Fact 2.2. *$\sum g(x_i) \leq \sum g(y_i) \Leftrightarrow \vec{X} \prec \vec{Y}$ holds for all continuous convex function $g : \mathcal{R} \to \mathcal{R}$ [13].*

Lemma 2.1. *Let $\vec{X} = (\vec{X}_1, \cdots, \vec{X}_K)$, and $\vec{Y} = (\vec{Y}_1, \cdots, \vec{Y}_K)$, where each element has dimension J_i and satisfying $\vec{X}_i \prec \vec{Y}_i$. Then $\vec{X} \prec \vec{Y}$.*

Proof. Let g be the continuous convex function $\mathcal{R} \to \mathcal{R}$. By Fact 2.2, $\vec{X}_i \prec \vec{Y}_i \Leftrightarrow \sum_{j=1}^{J_i} g(x_i^j) \leq \sum_{j=1}^{J_i} g(y_i^j) \Rightarrow \sum_{i=1}^{K} \sum_{j=1}^{J_i} g(x_i^j) \leq \sum_{i=1}^{K} \sum_{j=1}^{J_i} g(y_i^j) \Leftrightarrow \vec{X} \prec \vec{Y}$. □

Observation 2.0.1. *Let $\vec{X} = (\bar{x}, \cdots, \bar{x}), \vec{Y} = (y_1, \cdots, y_n), \vec{Z} = (z_1, \cdots, z_n)$, where $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} z_i = n\bar{x}$. If the elements in each vector is non decreasing, we may plot the normalized points of $\vec{X}/(n\bar{x}), \vec{Y}/(n\bar{x})$ and $\vec{Z}/(n\bar{x})$ on Fig. 2.3. If we explain the element of vectors as the income of individual, this is Lorenz Curves, which evaluate the social income inequality [98]. The curves show the normalized cumulative proportion of the income versus the cumulative percentage of population. The normalized $\vec{X}$ forms the straight curve A, which corresponds to the equal*

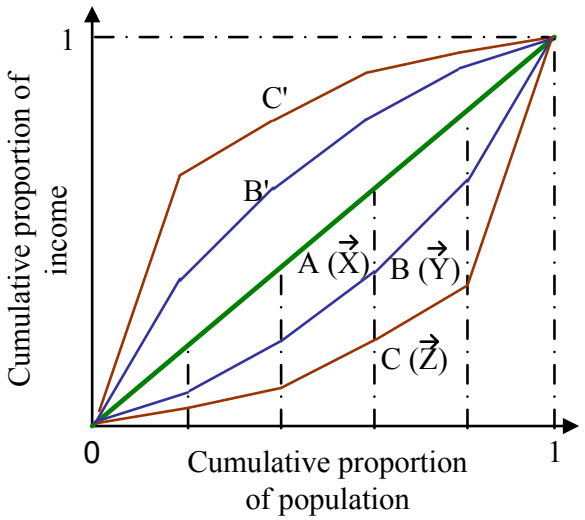


Figure 2.3: Lorenz Curves.

distribution. Normalized $\vec{Y}$ and normalized $\vec{Z}$ represent the unequal distribution and bent in the middle, and are denoted curve B and C, respectively. We call these bow curves in the convex shape. By the fact that the Lorenz curves are bent more, concentration increases, the bow curve inside represents more even distribution [13], which leads to $\vec{X} \prec \vec{Y} \prec \vec{Z}$. Similarly, we may mutate the B and C to B' and C' by only changing the order of the elements. We call these bow curves in the concave shape. Since the order of the vectors plays no role in majorization, $\vec{X} \prec \vec{Y} \prec \vec{Z}$ still holds for concave shape.

Theorem 2.1. *The objective function of problems (2.2) and (2.3) is Schur-convex.*

Proof. The proof follows the Fact 2.1 straightforward, due to the symmetric, increasing and convex of the objective function in problem (2.2) and (2.3). □

2.4 Smooth Electric Power Scheduling

2.4.1 SEPS-DL Algorithm

We first develop a smooth electric power scheduling for deferrable load algorithm (SEPS-DL) based on majorization. With Theorem 2.1, we convert the optimization problem (2.3) into an ordering problem of vectors, each representing a feasible schedule. Thus, we solve problem (2.3)

by finding the most evenly distributed electric power schedule that is feasible for the entire period. Obviously, the most evenly distributed schedule is $\vec{P}^{opt} = [\Phi/(L\tau), \cdots, \Phi/(L\tau)]$, corresponding to having the average power consumption $\bar{P}$ in each time slot. However, due to time varying user demands, $\vec{P}^{opt}$ may not be feasible. In general, each feasible schedule is piece-wise linear with a set of *power changing points*, where the scheduled power increases or decreases to prevent outage events or electric energy waste.

The proposed SEPS-DL algorithm can generate a feasible piece-wise linear schedule, which keeps each piece as long as possible into the future and keeps the power variation as small as possible. The algorithm is illustrated in Fig. 2.4. Starting from t_{start}, SEPS-DL first computes two *probe lines*:

- One probe line from t_{start} to the next corner point of $W_{max}(t)$, which can go the furthest into the future without causing outage events or energy waste (e.g., lines P_1P_2 in Case 1 and P_5P_6 in Case 2 of Fig. 2.4). The power of this probe line is $P_{max}(t) = \frac{W_{max}(t)-W(t_{start})}{t-t_{start}}$.
- The other probe line from t_{start} to the next corner point of $W_{min}(t)$, which can go the furthest into the future without causing outage events or energy waste (e.g., lines P_1P_3 in Case 1 and P_5P_7 in Case 2). The power of this probe line is $P_{min}(t) = \frac{W_{min}(t)-W(t_{start})}{t-t_{start}}$.

All feasible schedules should reside between the two probe lines in order to go farther into the future (i.e., to be smooth). Moreover, when the two probe lines are ended, they must hit *both* on either $W_{max}(t)$ or both on $W_{min}(t)$. Otherwise, we can always adjust one of the probe lines to make it go even further into the future. For example, see lines P_1P_3 and P_1P_4' in Case 1 of Fig. 2.4. We can use line P_1P_2 (which goes farther into the future) to replace line P_1P_4', and both probe lines hit $W_{min}(t)$ eventually. In Case 2 in the figure, both probe lines P_5P_6 and P_5P_7 hit $W_{max}(t)$.

If both probe lines hit $W_{min}(t)$ (i.e., Case 1 in Fig. 2.4), any feasible schedule for this interval will also hit $W_{min}(t)$, since it must lie between the two probe lines. We then trace back the *upper* probe line (i.e., line P_1P_2) to find the latest time when the schedule just satisfies the maximum

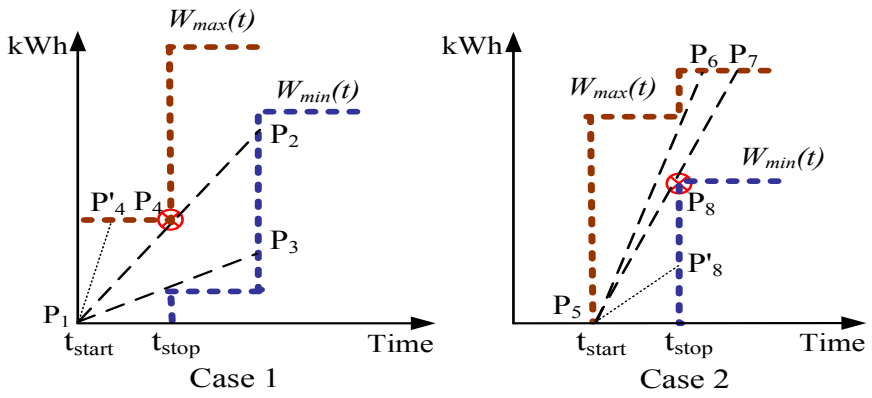


Figure 2.4: Illustrate the operation of the SEPS-DL algorithm.

demand (i.e., point P_4 at time t_{stop}). Then segment P_1P_4 will be chosen as the schedule for the interval [t_{start}, t_{stop}), with power $\frac{W_{max}(t_{stop})-W(t_{start})}{t_{stop}-t_{start}}$.

If both probe lines hit $W_{max}(t)$ (i.e., Case 2 in Fig. 2.4), any feasible schedule for this interval will also hit $W_{max}(t)$. We then trace back the *lower* probe line (i.e., line P_5P_7) to find the latest time when the schedule just satisfies the minimum demand (i.e., point P_8 at time t_{stop}). Then segment P_5P_8 will be chosen as the schedule for the interval [t_{start}, t_{stop}), with power $\frac{W_{min}(t_{stop})-W(t_{start})}{t_{stop}-t_{start}}$.

After the schedule for $[t_{start}, t_{stop})$ is determined, we set $t_{start} = t_{stop}$ and repeat the above procedure to find the schedule for the next time interval.

As shown in Algorithm 1, SEPS-DL probes for the longest feasible power starting from t_{start} in Steps 4–10. In Steps 11–14, the power for the interval $[t_{start}, t_{stop})$ is determined depending on which of the two cases it is as illustrated in Fig. 2.4. Steps 16–17 are for the case that the power does not change in the time slot. Step 19 resets the variables to start the computation for the next segment of $P(t)$.

2.4.2 Performance of SEPS-DL

The proposed SEPS-DL algorithm is very easy to implement. It can be shown that SEPS-DL has the following properties.

Theorem 2.2. *SEPS-DL is smoothness optimal.*

Proof. See Appendix A.1. □

Algorithm 1: Smooth Electric Power Scheduling for Deferrable Load Demand

1 DCC aggregates the demand from all the users by information networks and calculates $\vec{W}_{min}$, $\vec{W}_{max}$ for the whole scheduling period ;
2 $t = 1, t_{start} = 0, t_{stop} = t_{c_1} = t_{c_2} = 1, P_{min} = 0, P_{max} = \infty$;
3 **while** *some time slots are not scheduled* **do**
4 Calculate $P_{max}(t)$ and $P_{min}(t)$ over interval $[t_{start}, t]$;
5 **if** $P_{min} \leq P_{min}(t)$ & $P_{min}(t) \leq \min\{P_{max}, P_{max}(t)\}$ **then**
6 $P_{min} = P_{min}(t)$ and $t_{c_1} = t$;
7 **end**
8 **if** $P_{max} \geq P_{max}(t)$ & $P_{max}(t) \geq \max\{P_{min}, P_{min}(t)\}$ **then**
9 $P_{max} = P_{max}(t)$ and $t_{c_2} = t$;
10 **end**
11 **if** $P_{min} > \min\{P_{max}, P_{max}(t)\}$ **then**
12 Select P_{min} from t_{start} to $t_{stop} = t_{c_1}$;
13 **else if** $P_{max} < \max\{P_{min}, P_{min}(t)\}$ **then**
14 Select P_{max} from t_{start} to $t_{stop} = t_{c_2}$;
15 **else**
16 $t++$;
17 CONTINUE ;
18 **end**
19 $t_{start} = t_{stop}, t_{stop} = t_{c_1} = t_{c_2} = t_{start} + 1, t = t_{start} + 1, P_{min} = 0, P_{max} = \infty$;
20 **end**

Corollary 2.2.1. *The optimal power schedule is unique.*

Proof. Suppose $\vec{P}^*$ is not unique. Then there exists $\vec{P}' \prec \vec{P}_k$, for all k, and $\vec{P}' \neq \vec{P}^*$. $\vec{P}'$ must have a different set of power changing points from that of $\vec{P}^*$. According to the proof of Theorem 2.2, we can construct an auxiliary schedule $\vec{P}_1$, such that $\vec{P}^* \prec \vec{P}_1 \prec \vec{P}'$, which contradicts the assumption that $\vec{P}'$ is optimal. □

Theorem 2.3. *The complexity of SEPS-DL is* $\mathcal{O}(L^2)$.

Proof. In the worst case, the SEPS-DL algorithm computes the optimal schedule for each time slot by probing the full length of the remaining power sequence (as in Steps 4–10 in Algorithm 1). The worst case execution time is $\sum_{i=L}^{1} i = \frac{L(L+1)}{2} \Rightarrow \mathcal{O}(L^2)$. □

Theorem 2.4. *The smooth electric power schedule computed by SEPS-DL has the smallest peak power.*

Proof. From Theorem 2.2, we have $\vec{P}^* \prec \vec{P}_k$, for all k. We may reorder the elements in both $\vec{P}^*$ and $\vec{P}_k$ to a non-increasing order. According to the definition of Majorization in Definition 2.3, the first element in the re-ordered $\vec{P}^*$ is not greater than that of $\vec{P}_k$, which means the largest element in $\vec{P}^*$ is not greater than that of $\vec{P}_k$. Thus, the schedule generated by SEPS-DL has the lowest peak power. □

Corollary 2.4.1. *The SEPS-DL achieves highest load factor.*

Proof. The load factor in electric power grid is defined as in [99]

$$\text{Load factor(\%)} = \frac{AveragePower}{PeakPower} \times 100\%$$

By Theorem 2.4, the SEPS-DL generates the lowest peak power, which achieves the highest load factor during the scheduled period. □

Theorem 2.4 and Corollary 2.4.1 is highly preferable for grid design and operation. A lower peak power allows the operator to deploy generators, transformers and power transmission lines with smaller capacity in the grid, thus reducing the capital investment. In addition, the grid may be alleviated of the power usage burden during peak hours, and the electric energy usage quality of users can be improved.

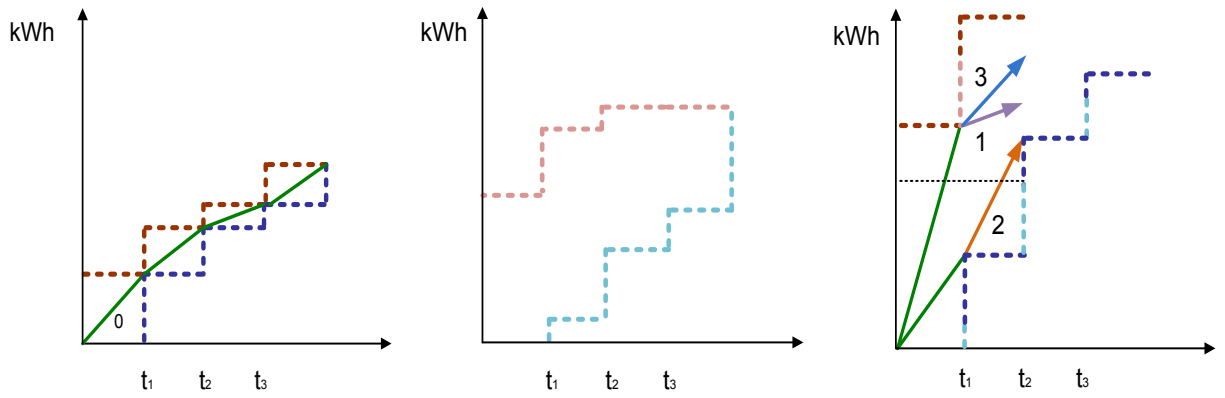
Theorem 2.5. *SEPS-DL is generation operating cost optimal.*

Proof. See Appendix A.2.

□

2.4.3 Extension to the General Case

We next extend SEPS-DL to solve the general case problem (2.2). With the priority load, a feasible power schedule should satisfy $P(t) \geq E_p(t)/\tau$ in every time slot. The presence of priority load enforces new constraints on the feasibility of the schedules. For example, consider the aggregated cumulative priority load curves and deferrable load curves shown in Fig. 2.5(a)



(a) Cumulative priority load curves (b) Cumulative deferrable load curves (c) Aggregated cumulative load curves

Figure 2.5: Smooth power scheduling with priority load. (Note that although not deferrable, the cumulative priority load can also be represented with two curves as shown in (a), where the maximum and minimum curves meet at the corners.)

and Fig. 2.5(b). According to the definition of feasible power supply schedule in SEPS-DL, both segments 1 and 2 in Fig. 2.5(c) are feasible. However, it can be seen that segment 1 actually cannot provide enough power to satisfy the priority load in time slot t_2, since its slope is smaller than the required slope (i.e., that of segment 0 in Fig. 2.5(a)).

To solve this problem, we develop the general smooth electric power scheduling algorithm (GSEPS), which is based on SEPS-DL. Specifically, SEPS-DL assumes no priority load. During the execution, the generated power segment is compared with the priority load. If the SEPS-DL generated power segment $P^*(t)$ is less than the priority load in time slot t, GSEPS will increase $P^*(t)$ to the priority load (e.g., see segment 3 in Fig. 2.5). Then SEPS-DL will continue to compute further segments of the schedule by setting the new starting point to t, until the entire schedule is computed.

2.4.4 Electric Power Allocation Among Individual Users

After the smooth electric power schedule is obtained, the DCC announces the schedule to all the users and requests them to control their loads to match the supplied electric energy $W^*(t) = P^*(t)\tau$ in each time slot t. To divide the total supply $W^*(t)$ among the N users, we assume a *benefit function* $U_n(p_n(t))$ for each user n, which is a nondecreasing concave function [92] and

represents the level of satisfaction of the user when receiving $p_n(t)$ in time slot t. We then develop an algorithm that maximizes the sum of the benefit functions of all users in the power distribution network. The maximization of the total benefit under the smooth schedule constraint can be formulated in each time slot t as follows:

$$\begin{aligned} \text{maximize:} \quad & \textstyle\sum_{n\in\mathbb{R}} U_n(p_n(t)) \\ \text{subject to:} \quad & p_n^{min}(t) \le p_n(t) \le p_n^{max}(t), \text{ for all } n \\ & \textstyle\sum_{n\in\mathbb{R}} p_n(t) = W^*(t)/\tau, \end{aligned} \tag{2.4}$$

where $p_n^{max}(t)$ and $p_n^{min}(t)$ are the maximum and minimum power consumptions of user n in slot t, respectively.

Problem (2.4) is a convex optimization problem, which can be solved effectively with a convex optimization solver. In case that DCC may not know the exact parameters of individual utility functions in practice, we develop a distributed user benefit maximization load control algorithm (DUBMLC) based on *dual decomposition* [100] to solve problem (2.4). For brevity, we omit the time slot notation t in following equations.

First, we introduce the non-negative Lagrange multiplier ν, and derive the Lagrange function:

$$\begin{aligned} \mathcal{L}(p_n, \nu) &= \sum_{n\in\mathbb{R}} U_n(p_n) + \nu\left(W - \sum_{n\in\mathbb{R}} p_n\right) \\ &= \sum_{n\in\mathbb{R}} \mathcal{L}_n(p_n, \nu) + \nu W, \end{aligned} \tag{2.5}$$

where $\mathcal{L}_n(p_n, \nu) = U_n(p_n) - \nu p_n$. Observing that $\mathcal{L}_n(\cdot)$ only depends on user n's local information, we have the dual decomposition for each user n. Each user solves subproblem (2.6) for given Lagrange multipliers $\tilde{\nu}$:

$$p_n(\tilde{\nu}) = \arg\max_{p_n^{min}\le p_n\le p_n^{max}} \mathcal{L}_n(p_n, \tilde{\nu}), \text{ for all } n. \tag{2.6}$$

Subproblem (2.6) can be solved with the subgradient method [14], since $\mathcal{L}_n$ is strictly concave. User n iteratively updates its power p_n until p_n converges, as:

$$p_n(l+1) = \left[p_n(l) + \kappa(l) \cdot \frac{\partial \mathcal{L}_n(p_n)}{\partial p_n} \right]^+ \tag{2.7}$$

where $[\cdot]^+$ denotes the projection of p_n onto the range $[p_n^{min}, p_n^{max}]$, and $\kappa(l)$ is the step size varies in each step l according to the Armijo Rule [14]. The solution p_n can be solved locally by the users and converges to the optimal solution of $\tilde{p}_n$ for all n as $l \to \infty$.

For a given optimal solution $\tilde{p}_n$, the master dual problem is to solved by DCC:

$$\begin{aligned} \text{minimize:} \quad & \mathcal{L}(\tilde{p}_n, \nu), \\ \text{subject to:} \quad & \nu \geq 0, \text{ for all } n. \end{aligned}$$

We can also apply the subgradient method to iteratively update the multipliers as:

$$\nu(l+1) = \max \left\{ \nu(l) - \alpha(l) \cdot \frac{\partial \mathcal{L}(\nu)}{\partial \nu}, 0 \right\}, \tag{2.8}$$

where $\alpha(l)$ is the step size. The Lagrange multipliers converges to the optimal as $l \to \infty$, since problem (2.4) is a convex problem, the duality gap is zero and the solution of (2.6) is unique. The primal variable p_n will also converge to the optimal solutions [100].

The distributed user benefit maximization load control (DUBMLC) algorithm is presented in Algorithm 2. With DUBMLC, each user greedily maximizes its own benefit by solving (2.6) with current "price" ν, which is controlled by DCC through the master due problem (2.8). Due to convexity of the problem, the optimization gap is zero and the optimal total maximum benefit is reached when the algorithm converges.

Combining of GSEPS and DUBMLC, we now present the General Smooth Electric Power Scheduling Policy. Specifically, at the beginning of each period, which can be daily based or be an arbitrary period of time, the users send their slotted demand profiles to the DCC through

Algorithm 2: Distributed User Benefit Maximization Load Control Algorithm

1 $l = 0$ and the DCC initializes nonnegative parameter $\nu(l)$;
2 The DCC announces the parameters to the users via the communications network ;
3 Each user locally solves problem (2.6) as in (2.7) to obtain its requested power ;
4 Each user sends its requested power to the DCC via information networks ;
5 The DCC updates the parameters $\nu(l)$ as in (2.8) and announces the new value $\nu(l+1)$ to all users ;
6 $l = l + 1$ and go to Step 3, until the solution converges ;

the communications network. After aggregating all the demand profiles, the DCC calculates the deterministic cumulative supply/demand curves for the power distribution networks and executes GSEPS to compute the smooth power profile. After that, DCC let the users to control their electricity usage to match the smooth schedule with DUBMLC. The load control does not necessarily need to be fully executed at the exact beginning of the period. It may operate at some time ahead of the scheduled time slot. If a user requests a usage exceeding the planned level, the DCC may allow the distribution substation to temporally fulfill the excess demand but charging a penalty price based on the electric energy availability of the power distribution network.

2.5 Simulation Evaluation

In this section, we evaluate the proposed algorithms by simulating an electric power distribution network with 250 independent users. We assume a daily period slotted into $L = 144$ time slots (i.e., $\tau = 10$ min). The demand for each user during the period is randomly distributed from 35 kWh to 50 kWh. The DCC aggregates the load profiles and generates the cumulative supply/demand curves at the beginning of the period. We adopt a benefit function $U_n(t) = k_1 q_n(t) - \frac{1}{2} k_2 q_n(t)^2$ [101], where $q_n(t) \in [0, 1]$ is the normalized value of power supply $p_n(t)$. With this $U_n(\cdot)$, problem (2.4) becomes a quadratic programming problem, which can be effectively solved with the proposed distributed algorithm. Without loss of generality, we set $k_1 = k_2 = 1$ in the simulations.

We first examine the performance of SEPS-DL and GSEPS. For SEPS-DL, all the electric energy demand is deferrable. For GSEPS, we assume 50% of the demand is deferrable and the

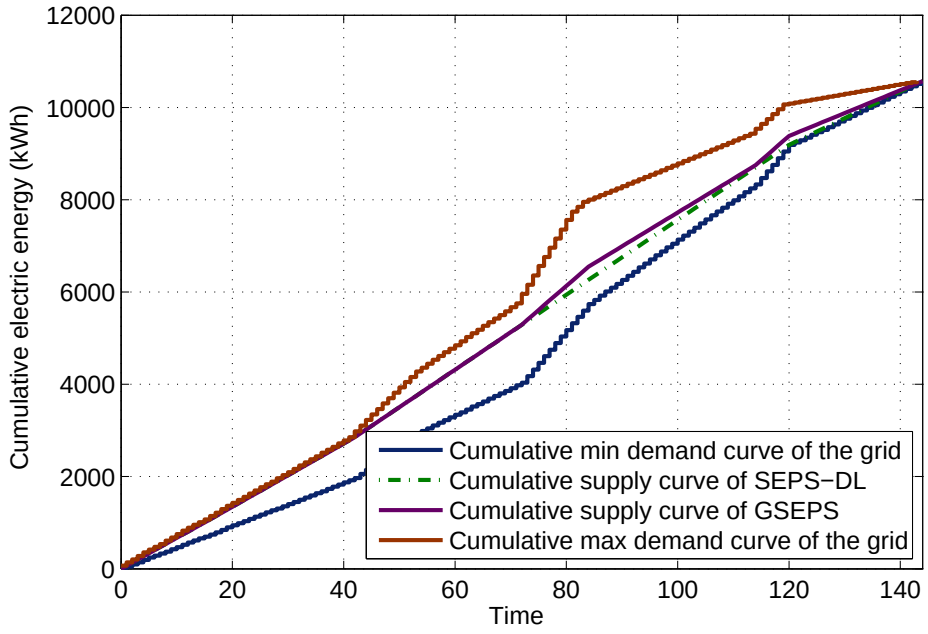


Figure 2.6: Comparison of the power schedules achieved by SEPS-DL an GSEPS.

deadlines are randomly distributed during the daily period. The cumulative demand curves and the computed schedules are plotted in Fig. 2.6. We find that both electric power schedules lie between $\vec{W}_{min}$ and $\vec{W}_{max}$, meaning they are feasible and satisfying the user demands in the entire period. In some time slots, e.g., slots $[70,\ 80]$ and $[110,\ 120]$, GSEPS requests a larger electric power than SEPS-DL. This is due to the hard requirement for the priority load, which temporally forces GSEPS to increase the electric power supply.

After the smooth power schedule is obtained, DUBMLC is executed to divide the power to individual users in each time slot. For better illustration, we only plot the power convergence curves for six users in Fig. 2.7. The curves for other users are similar. We find that all the curves converge to the optimal values very quickly; after one step there is no significant variation in the electric powers of individual users.

We next compare the proposed algorithms with two alternatives:

- A "supply until deadline" scheme (SUDP), in which the deferrable load demand is served until the last minute.

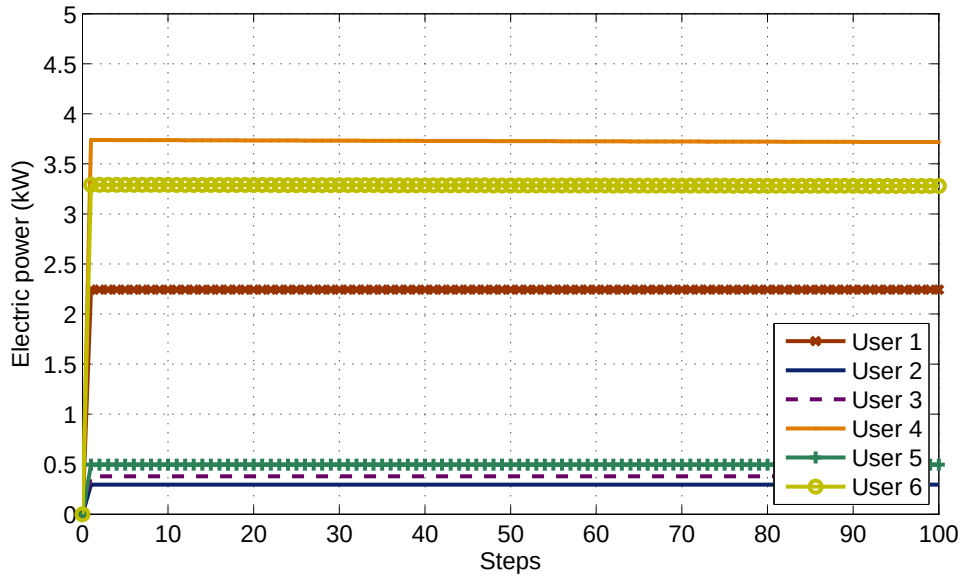


Figure 2.7: Convergence of the individual power allocation achieved by DUMLC.

- The "utility maximization real-time pricing" scheme (UMRP) that is introduced in [102], which solves the demand side management problem with a real-time pricing strategy and has been widely cited in the smart grid research community.

In particular, the UMRP scheme maximizes the social welfare in a smart grid at each time slot independently, i.e.,

$$\max \sum_{n \in \mathbb{R}} U_n(p_n(t)) - g(P(t), \theta(t)), \text{ for all } t,$$

subject to the total power generation constraint and user power consumption constraints. This algorithm can be extended to our simulation scenario. In this work, the operating cost of power generation is evaluated by $g(P(t), \theta(t)) = (\beta_1 + \beta_2 P(t) + \beta_3 P(t)^2)\theta(t)$, which is generalized from [103].We let $\beta_1 = 120.0, \beta_2 = 6, \beta_3 = 0.04$ for a generator [103] and $\theta(t)$ be uniformly distributed in $[0.5, 2.5]$. This yields a quadratic programming problem, which can be solved in either a centralized or distributed manner [102]. In the simulations, we use a centralized interior-point method to solve the problem.

The peak power, load factor and power variation achieved by the above algorithms for networks with different numbers of users are presented in Fig. 2.8, Fig. 2.9, and Fig. 2.10, respectively. Each number is the average of 100 simulation runs with different random seeds, with the 95% confidence interval plotted at the top of each bar. We observe that SEPS-DL and GSEPS can significantly reduce both peak power and power variation in the power distribution network. For example, for the network with 1000 users, SEPS-DL and GSEPS achieves peak powers 1947 kW and 2560 kW, respectively, which are only 55% and 73% of the corresponding SUDP and UMRP peak powers. We also notice that the load factors achieved by SEPS-DL and GSEPS are more than 100% and 50% larger than those of SUDP and UMRP under all cases. Similiarly, we find in Figure 2.10 that the SEPS-DL and GSEPS schedules are much more smoother than both SUDP and UMRP schedules. Therefore, to design the power generation, transmission and distribution infrastructure for this 1000-user site, we may select components, such as transformers and transmission lines, based on the capacity specifications of 1947 kW and 2560 kW, respectively (with SEPS-DL and GSEPS), instead of 3513 kW with SUDP and UMRP. As the network size increases, the performance gap increases as well. For the smallest gap case when network size is 100, the power reduction is 176 kW for SEPS-DL and 110 kW for GSEPS, which meets the requirement of 100 kW minimum energy reductions for the DR products at *ISONE* [104].

It is interesting to observe that UMRP and SUDP have almost identical performance in the simulations. This is largely due to the choice of the object function $\sum_{n\in\mathbb{R}} U_n(p_n(t)) - g(P(t), \theta(t))$. In the simulation with the above utility functions and cost functions, the effect of the total power decrement on the cost functions dominates the effect of the individual user power increments in their utility functions. Thus UMRP attempts to reduce the total power generation, while only maintaining the minimum user satisfaction level. This strategy indeed degenerates UMRP to SUDP, both with similar performance. Although UMRP maximizes the welfare of the distribution network, it does not aim to smooth the power schedules. It would be helpful to carefully introduce some coefficients to balance the contributions of utility and cost to the welfare. This phenomenon also verifies our motivation of the work that the real-time pricing with utility maximization may

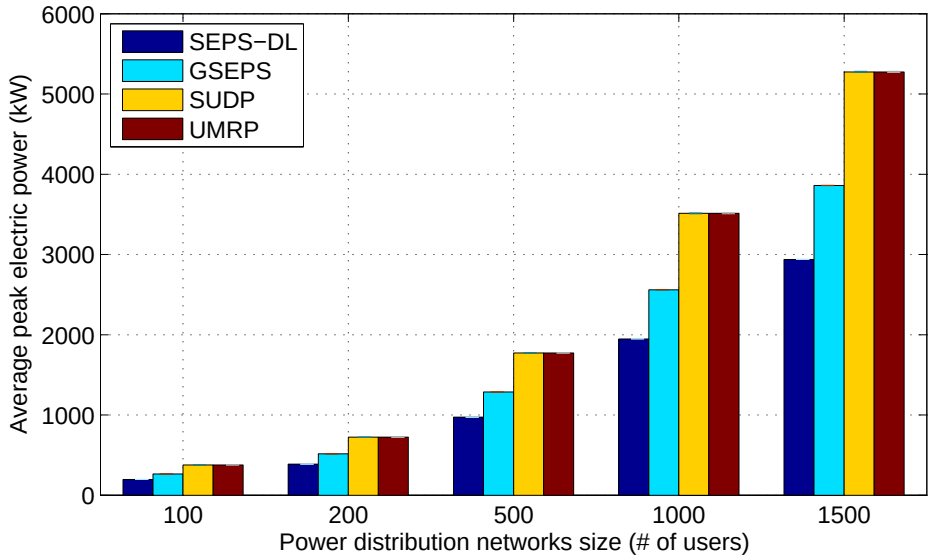


Figure 2.8: Peak electric powers achieved by SEPS-DL, GSEPS, SUDP and UMRP.

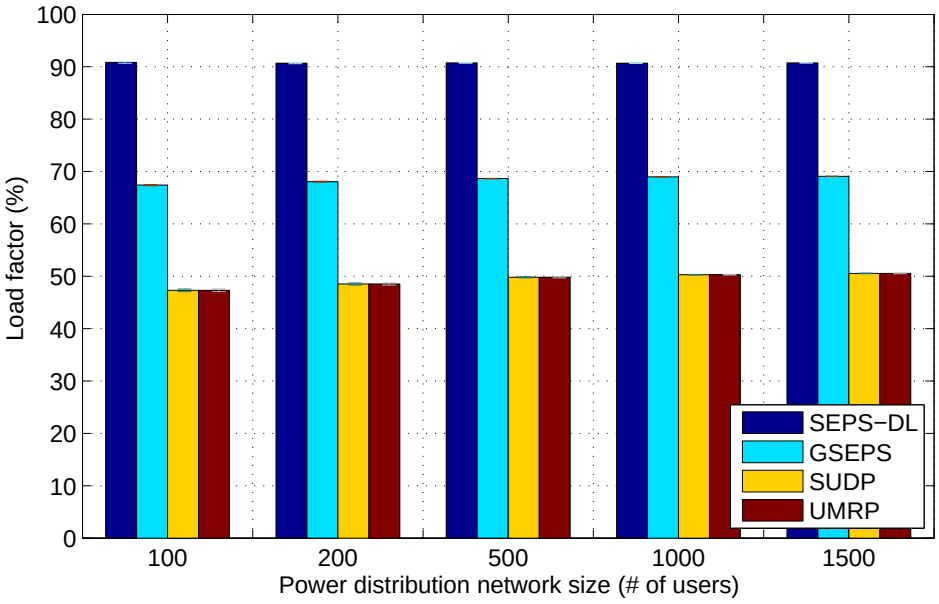


Figure 2.9: Load factor achieved by SEPS-DL, GSEPS, SUDP and UMRP.

not automatically solve the smooth electric power scheduling and peak power reduction problems. Compared to UMRP, the proposed algorithms directly target at the smoothness optimization problem and are robust to various configurations of the distribution network.

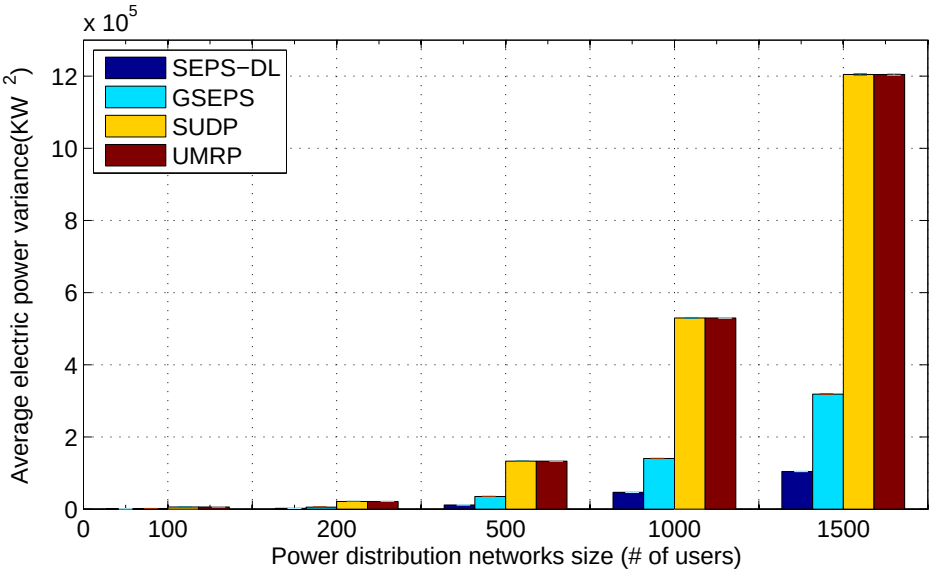


Figure 2.10: Variances of the power schedules achieved by SEPS-DL, GSEPS, SUDP and UMRP.

We finally compare the average generator operating cost of SEPS-DL, GSEPS, SUDP and UMRP. Similar to the previous scenario, we assume a load site with average aggregated demand 100 MW, which is scheduled by SEPS-DL, GSEPS, SUDP and UMRP based on a daily period. The average generator operating costs are shown in Table 2.2. It can be seen that both SEPS-DL and GSEPS obtain smaller operating costs than SUDP and UMRP for serving the same load site. It is not surprising to see that SUDP and UMRP obtain similar results, due to the effect of the objective function. Also note that UMRP is not operating cost optimal, although it seeks to maximize the social welfare. This is due to the fact that the operating cost is not independent from time slot to time slot; greedily optimizing social welfare in each time slot does not guarantee optimality over the entire period. Suppressing user satisfaction level some time slots, may result in high power allocations at a later time slot to meet all the delayed demands right before their deadline, leading to larger peak power and electricity generation cost. On the contrary, SEPS-DL and GSEPS optimize power scheduling over the entire period, and can flexibly serve the demand to achieve smooth electric power scheduling and low operating cost.

Table 2.2: Average operating cost of power generation

	SEPS-DL	GSEPS	SUDP	UMRP
$\$/hour$	1653	1669	1746	1746

2.6 Related Work

SG is regarded as the next generation power grid that exploits a coexisting communications network for better control and optimization of power generation and distribution. In SG, information technologies and computational intelligence are integrated across electricity generation, transmission, distribution and consumption to achieve green, reliable, efficient and sustainable energy goals. Comprehensive surveys of SG technologies can be found in [3–5, 105].

The emergence of SG attracts new interest in evolving the next generation of power distribution systems [88]. Demand response is an important power distribution paradigm to reduce the peak demand and smooth demand profiles in the grid by shaping the user demands. Various implementation issues of demand response in SG are examined in [104]. Real-time pricing and direct load control are two important ways to shape user demand profile.

Due to the real-time communications and control through two-way information flows in SG, new design approaches in demand response are being developed recently [90–93], which are based on optimization and game theory approaches. In [90], the authors proposed an optimal and automatic residential energy consumption scheduling framework to achieve the trade-off between minimization of electricity payment and appliance operation waiting time. In [91], a game theory approach is used to control the power demand at peak hours by dynamic pricing strategies. In [92], the authors studied demand response for households based on utility maximization, and showed that there exist time-varying prices that may achieve social optimality. A recent work [93] introduced the framework for optimal resource allocation under the uncertainty in the two-way information network and provided a decentralized algorithm that can be implemented in practice. Although providing some interesting methods to achieve cost efficient electricity usage, these work do not explicitly address the problem of smooth electric power scheduling.

Two recent works [106, 107] studied the problem of reducing the peak-to-average ratio (PAR) of the electric energy consumption in SG. In [106], the authors introduced a game theory framework for a distributed algorithm to minimize the total energy payment and reduce PAR. However, users need to broadcast control messages to announce their new schedules to the entire network. The control overhead could be considerable. In [107], energy storage devices were incorporated in the SG and users' cost and PAR are minimized with a distributed algorithm. Although the distributed algorithm only needs to exchange information with the energy provider, the achieved Nash equilibrium cannot be guaranteed to be socially optimal. In addition, this work does not consider the operating cost of the energy provider.

Majorization is a useful tool for problems involving vectors [13]. It has been used in solving optimization problems in the communication and networking area [57,95,96]. In [95], majorization is applied to variable-bit-rate (VBR) video smoothing over a wired CBR link. Ref. [96] presented an optimal transmission algorithm over a wireless multiple-input single-output (MISO) link based on majorization. In our recent work [57], we adopted majorization for power efficient VBR video streaming over a cellular network.

Our work differs from these existing efforts by introducing the mathematic theory of majorization to solve the smoothness scheduling problem in electric distribution networks, while explicitly targeting at the unique mathematical notion of smoothness. To the best of our knowledge, this is the first work that introduces maojorization into electric energy management in power grid, which jointly considers smooth power scheduling, electric usage quality provisioning on the user side, and grid operating cost on the electric energy provider side. The effective electric power smoothing solution provides a highly competitive solution for future SG design and operations.

2.7 Conclusions

In this chapter, we addressed the problem of smooth electric power scheduling in a power distribution network. We introduced a deterministic model to characterize the complex relationship

between demand and supply. A constrained nonlinear optimization problem is formulated aiming to minimize the electric power variation and satisfy user power usage quality. We developed majorization-based algorithms for deriving smoothness optimal schedules for the network, and a distributed algorithm for dividing the power supply among the users. Our simulation study shows that the proposed algorithms can effectively reduce the peak power, minimize the power variation, and reduce the operating cost of the grid, while satisfying user power usage quality.

Chapter 3

Adaptive Electricity Scheduling in Microgrids

3.1 Introduction

In this chapter, we designed a smart energy management systems in *Microgrid* (MG) by taking advantage of the plug-and-play interfaces of smart grid. MG is a promising component for future SG deployment. Due to the increasing deployment of *distributed renewable energy resources* (DRERs), MG provides a localized cluster of renewable energy generation, storage, distribution and local demand, to achieve reliable and effective energy supply with simplified implementation of SG functionalities [4, 108]. We review the typical MG architecture in Fig. 3.1, consisting of DRERs (such as wind turbines and solar photovoltaic cells), *energy storage systems* (ESS), a communication network (e.g., wireless or powerline communications) for information delivery, an *MG central controller* (MGCC), and local residents. The MG has centralized control with the MGCC [108], which exchanges information with local residents, ESS's, and DRERs via the information network. There is a single common coupling point with the macrogrid. When disconnected, the MG works in the *islanded mode* and DRERs and ESS's provide electricity to local residents. When connected, the MG may purchase extra electricity from the macrogrid or sell excess energy back to the market [3].

The balance of electricity demand and supply is one of the most important requirements in MG management. Instead of matching supply to demand, smart energy management matches the demand to the available supply using control technology or off-peak pricing to achieve more efficient capacity utilization [4]. In this chapter, we develop a novel access control framework for MG energy management, exploiting the two-way flows of electricity and information. In particular, we consider two types of electricity usage: (i) a pre-agreed *basic usage* that is "hard"-guaranteed, such as basic living usage, and (ii) extra elastic *quality usage* exceeding the pre-agreed level for

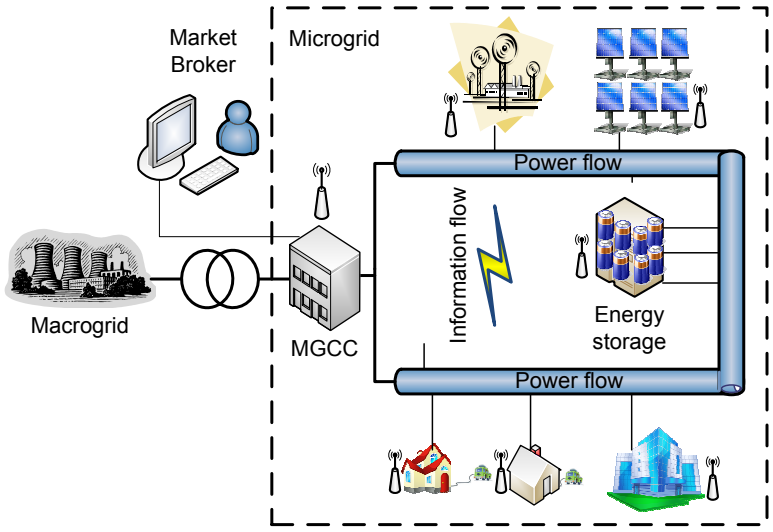


Figure 3.1: Illustrate the microgrid architecture.

more comfortable life, such as excessive use of air conditioners or entertainment devices. In practice, residents may set their load priority and preference to obtain the two types of usage [89]. The basic usage should be always satisfied, while the quality usage is controlled by the MGCC according to the grid status, such as DRER generation, ESS storage levels and utility prices. The MGCC may *block* some quality usage demand if necessary. This can be implemented by incorporating smart meters, smart loads and appliances that can adjust and control their service level through communication flows [3]. To quantify residents' satisfaction level, we define the outage percentage of the quality usage as *Quality of Service in Electricity* (QoSE), which is specified in the service contracts. For example, the residents may customize their outage risk of quality usage in return for paying an insurance premium, which is differentiated according to local residents preferences [109] [109]. The MGCC adaptively schedules electricity to keep the QoSE below a target level, and accordingly dynamically balance the load demand to match the available supply.

In this chapter, we investigate the problem of smart energy scheduling by jointly considering renewable energy distribution, ESS management, residential demand management, and utility market participation, aiming to minimize the MG operation cost and guarantee the residents' QoSE. The MGCC may serve some quality usage with supplies from the DRERs, ESS's and macrogrid. On the other hand, the MG can also sell excessive electricity back to the macrogrid to compensate

for the energy generation cost. The electricity generated from renewable sources is generally random, due to complex weather conditions, while the electricity demand is also random due to the random consumer behavior, and so do the purchasing and selling prices on the utility market. It is challenging to model the random supply, demand, and price processes for MG management, and it may also be costly to have precise, real-time monitoring of the random processes. Therefore, a simple, low cost, and optimal electricity scheduling scheme that does not rely on any statistical information of the supply, demand, and price processes would be highly desirable.

We tackle the MG electricity scheduling problem with a *Lyapunov optimization* approach, which is a useful technique to solve stochastic optimization and stability problems [12]. We first introduce two virtual queues: QoSE virtual queues and battery virtual queues to transform the QoSE control problem and battery management problem to queue stability problems. Second, we design an adaptive MG electricity scheduling policy based on the Lyapunov optimization method and prove several deterministic (or, "hard") performance bounds for the proposed algorithm. The algorithm can be implemented *online* because it only relies on the current system status, without needing any future knowledge of the energy demand, supply and price processes and any future information. The proposed algorithm also converges exponentially due to the nice property of Lyapunov stability design [110]. The algorithm is evaluated with trace-driven simulations and is shown to achieve significant efficiency on MG operation cost while guaranteeing the residents' QoSE.

The remainder of this chapter is organized as follows. We present the system model and problem formulation in Section 3.2. An adaptive MG electricity scheduling algorithm is designed and analyzed in Section 3.3. Simulation results are presented and discussed in Section 3.4. We discuss related work in Section 3.5. Section 3.6 concludes the chapter.

The notations used in this chapter are summarized in Table 3.1.

3.2 System Model and Problem Formulation

3.2.1 System Model

Overview

We consider the electricity supply and consumption in an MG as shown in Fig. 3.1. We assume that the MG is properly designed such that a portion of the electricity demand related to basic living usage (e.g., lighting) from the residents, termed *basic usage*, can be guaranteed by the minimum capacity of the MG. There are randomness in both electricity supply (e.g., weather change) and demand (e.g., entertainment usage in weekends). To cope with the randomness, the MG works in the *grid-connected* mode and is equipped with ESS's, such as electrochemical battery, superconducting magnetic energy storage, flywheel energy storage, etc. The ESS's store excess electricity for future use.

The MGCC collects information about the resident demands, DRER supplies, and ESS levels through the information network. When a resident demand exceeds the pre-agreed level, a *quality usage* request will be triggered and transmitted to the MGCC. The MGCC will then decide the amount of quality usage to be satisfied with energy from the DRERS, the ESS's, or by purchasing electricity from the macrogrid. The MGCC may also decline some quality usage requests. The excess energy can be stored at the ESS's or sold back to the macrogrid for compensating the cost of MG operation.

Without loss of generality, we consider a time-slotted system. The time slot duration is determined by the timescale of the demand and supply processes, as well as how frequent the MG can switch on and off to the macrogrid.

Energy Storage System Model

The system model is shown in Fig. 3.2. Consider a battery farm with K independent battery cells, which can be recharged and discharged. We assume that the batteries are not leaky and do not consider the power loss in recharging and discharging, since the amount is usually small. It is

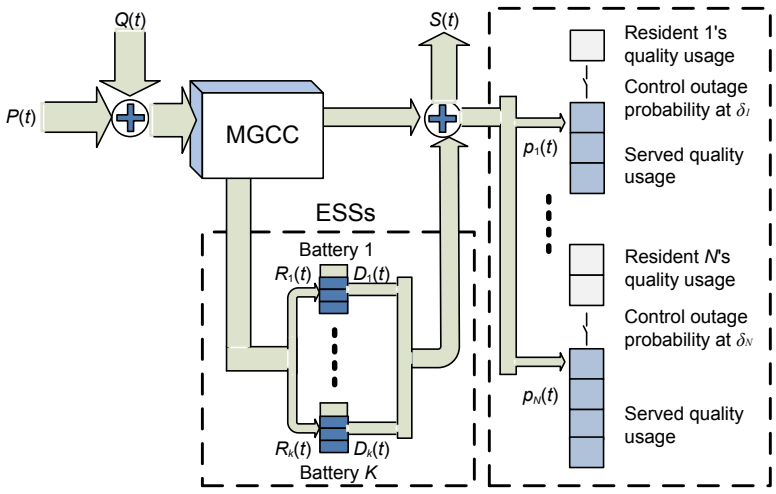


Figure 3.2: The system model considered in this chapter.

easy to relax this assumption by applying a constant percentage on the recharging and discharging processes. For brevity, we also ignore the aging effect of the battery and the maintenance cost, since the cost on the utility market dominates the operation cost of MGs.

Let $E_k(t)$ denote the energy level of the kth battery in time slot t. The capacity of the battery is bounded as

$$E_k^{min} \leq E_k(t) \leq E_k^{max}, \forall\ k, t, \tag{3.1}$$

where $E_k^{max} \geq 0$ is the maximum capacity. $E_k^{min} \geq 0$ is the minimum energy level required for battery k, which may be set by the battery deep discharge protection settings. The dynamics over time of $E_k(t)$ can be described as

$$E_k(t+1) = E_k(t) - D_k(t) + R_k(t), \forall\ k, t, \tag{3.2}$$

where $R_k(t)$ and $D_k(t)$ are the recharging and discharging energy for battery k in time slot t, respectively. The charging and discharging energy in each time slot are bounded as

$$\begin{cases} 0 \leq R_k(t) \leq R_k^{max}, & \forall\ k, t \\ 0 \leq D_k(t) \leq D_k^{max}, & \forall\ k, t. \end{cases} \tag{3.3}$$

In each time slot t, $R_k(t)$ and $D_k(t)$ are determined such that (3.1) is satisfied in the next time slot.

Usually the recharging and discharging operations cannot be performed simultaneously, which leads to

$$\begin{cases} R_k(t) > 0 \Rightarrow D_k(t) = 0, & \forall\ k, t \\ D_k(t) > 0 \Rightarrow R_k(t) = 0, & \forall\ k, t. \end{cases} \tag{3.4}$$

Energy Supply and Demand Model

Consider N residents in the MG; each generates basic and quality electricity usage requests, and each can tolerate a prescribed *outage probability* δ_n for the requested quality usage part. The MGCC adaptively serves quality usage requests at different levels to maintain the QoSE as well as the stability of the grid. The service of quality usage can be different for different residents, depending on individual service agreements.

Let λ_n be the *average quality usage arrival rate*, and δ_n a prescribed outage tolerance (i.e., a percentage) for user n. The average *outage rate* for the quality usage, ρ_n, should satisfy

$$\rho_n \leq \delta_n \cdot \lambda_n. \tag{3.5}$$

At each time t, the quality usage request from resident n is $\alpha_n(t) \in [0, \alpha_n^{max}]$ units, which is an i.i.d random variable with a general distribution. The average rate is $\lambda_n = \lim_{t\to\infty}(1/t)\sum_{\tau=0}^{t-1}\alpha_n(\tau)$ by the law of large number.

The DRERs in the MG generate $U(t)$ units of electricity in time slot t. $U(t)$ can offer enough capacity to support the pre-agreed *basic usage* in the MG, which is guaranteed by islanded mode MG planning. due to complex weather conditions. The electric is transmitted over power transmission lines. Without loss of generality, we assume the power transmission line is not subject to outages and the transmission loss is negligible. Let $\alpha_n^b(t)$ be the pre-agreed *basic usage* for resident n in time slot t, which can be fully satisfied by $U(t)$, i.e., $\sum_{n=1}^{N}\alpha_n^b(t) \leq U(t)$, for all t. In addition, some *quality usage* request $\alpha_n(t)$ may be satisfied if $P(t) = U(t) - \sum_{n=1}^{N}\alpha_n^b(t) \geq 0$.

Let $p_n(t)$ be the energy allocated for the quality usage of resident n. We have

$$0 \leq p_n(t) \leq \alpha_n(t). \tag{3.6}$$

We define a function $I_n(t) \geq 0$ to indicate the amount of quality usage outage for resident n, as $I_n(t) = \alpha_n(t) - p_n(t)$. Then the average outage rate can be evaluated as $\rho_n = \lim_{t\to\infty}(1/t)\sum_{\tau=0}^{t-1} I_n(t)$.

The MGCC may purchase additional energy from the macrogrid or sell some excess energy back to the macrogrid. Let $Q(t) \in [0, Q_{max}]$ denote the energy purchased from the macrogrid and $S(t) \in [0, S_{max}]$ the energy sold on the market in time slot t, where Q_{max} and S_{max} are determined by the capacity of the transformers and power lines. Since it is not reasonable to purchase and sell energy on the market at the same time, we have the following constraints

$$\begin{cases} Q(t) > 0 \Rightarrow S(t) = 0, & \forall\, t \\ S(t) > 0 \Rightarrow Q(t) = 0, & \forall\, t. \end{cases} \tag{3.7}$$

To balance the supply and demand in the MG, we have

$$P(t) + Q(t) + \sum_{k=1}^{K} D_k(t) - S(t) - \sum_{k=1}^{K} R_k(t) = \sum_{n=1}^{N} p_n(t), \forall\, t. \tag{3.8}$$

Utility Market Price Model

The price for purchasing electricity from the macrogrid in time slot t is $C(t)$ per unit. The purchasing price depends on the utility market state, such as peak/off time of the day. We assume finite $C(t) \in [C_{min}, C_{max}]$, which is announced by the utility market at the beginning of each time slot and remains constant during the slot period [111]. Unlike prior work [111], we do not require any statistic information of the $C(t)$ process, except that it is independent to the amount of energy to be purchased in that time slot.

If the MGCC determines to sell renewable energy on the utility market, the selling price from the market broker is denoted by $W(t) \in [W_{min}, W_{max}]$ in time slot t, which is also a stochastic process with a general distribution and mean λ_n. We also assume $W(t)$ is known at the beginning of each time slot and independent to the amount of energy to be sold on the market. We assume $C_{max} \geq W_{max}$, $C_{min} \geq W_{min}$ and $C(t) > W(t)$ for all t. That is, the MG cannot make profit by greedily purchasing energy from the market and then sell it back to the market at a higher price simultaneously.

3.2.2 Problem Formulation

Given the above models, a control policy $\mathbb{A}(t) = \{Q(t), S(t), R_k(t), D_k(t), p_n(t)\}$ is designed to minimize the operation cost of the MG and guarantee the QoSE of the residents. We formulate the electricity scheduling problem as

$$\begin{aligned} \text{minimize:} \quad & \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{Q(\tau)C(\tau) - S(\tau)W(\tau)\} \\ \text{s.t.} \quad & \text{(3.1), (3.3), (3.4), (3.5), (3.6), (3.7), (3.8)} \\ & \text{battery queue stability constraints.} \end{aligned} \tag{3.9}$$

Problem (3.9) is a stochastic programming problem, where the utility prices, utility generation of DRERs, and utility consumption of residents are all random. The solution also depends on the evolution of battery states. It is challenging since the supply, demand, and price are all general processes.

Virtual Queues

We first adopt a *battery virtual queue* $X_k(t)$ that tracks the charge level of each battery k:

$$X_k(t) = E_k(t) - D_k^{max} - E_k^{min} - VC_{max}, \ \ \forall\, k, t, \tag{3.10}$$

where $0 < V \leq V_{max} = \min_k \left\{ \frac{E_k^{max} - E_k^{min} - R_k^{max} - D_k^{max}}{C_{max} - W_{min}} \right\}$ is a constant for the trade-off between algorithm performance and ensuring the battery constraints. This constant V_{max} is carefully selected to ensure the evolutions of the battery levels always satisfy the battery constraints (3.1), which will be examined in detail in Section 3.3.3.The virtual queue can be deemed as a shifted version of the battery dynamics in (3.2) as

$$X_k(t+1) = X_k(t) - D_k(t) + R_k(t), \quad \forall\, k, t. \tag{3.11}$$

These queues are "virtual" because they are maintained by the MGCC control algorithm. Unlike an actual queue, the virtual queue backlog $X_k(t)$ may take negative values.

We next introduce a conceptual *QoSE virtual queue* $Z_n(t)$, whose dynamics are governed by the system equation as

$$Z_n(t+1) = [Z_n(t) - \delta_n \cdot \alpha_n(t)]^+ + I_n(t), \quad \forall\, n, t. \tag{3.12}$$

where $[x]^+ = \max\{0, x\}$.

Theorem 3.1. *If an MGCC control policy stabilizes the QoSE virtual queue $Z_n(t)$, the outage quality usage of resident n will be stabilized at the average QoSE rate $\rho_n \leq \delta_n \cdot \lambda_n$.*

Proof. See Appendix B.1. □

Problem Reformulation

With Theorem 3.1, we can transform the original problem (3.9) into a queue stability problem with respect to the QoSE virtual queue and the battery virtual queues, which leads to a system

stability design from the control theoretic point of view. We have a reformulated stochastic programming problem as follows.

$$\begin{aligned} \text{minimize:} \quad & \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{Q(\tau)C(\tau) - S(\tau)W(\tau)\} \\ \text{s.t.} \quad & \text{(3.3), (3.4), (3.6), (3.7), (3.8)} \\ & \text{battery and QoSE virtual queue stability} \\ & \text{constraints.} \end{aligned} \tag{3.13}$$

Theorem 3.1 indicates that QoSE provisioning is equivalent to stabilizing the QoSE virtual queue $Z_n(t)$, while stabilizing the virtual queues (3.11) ensures that the battery constraints (3.1) are satisfied. We then apply *Lyapunov optimization* to develop an adaptive electricity scheduling policy for problem (3.13), in which the policy greedily minimize Lyapunov drift in every slot t to push the system toward the stability.

3.2.3 Lyapunov Optimization

We define the *Lyapunov function* for system state $\vec{\Theta}(t) = [\vec{X}(t), \vec{Z}(t)]^T$ with dimension $(N + K) \times 1$ as follows, in which $\vec{X}(t) = [X_1(t) \cdots X_K(t)]^T$ and $\vec{Z}(t) = [Z_1(t) \cdots Z_N(t)]^T$.

$$L(\vec{\Theta}(t)) = \frac{1}{2} \sum_{k=1}^{K} [X_k(t)]^2 + \frac{1}{2} \sum_{n=1}^{N} [Z_n(t)]^2, \tag{3.14}$$

which is positive definite, since $L(\vec{\Theta}(t)) > 0$ when $\vec{\Theta}(t) \neq \vec{\mathbf{0}}$ and $L(\vec{\Theta}(t)) = 0 \Leftrightarrow \vec{\Theta}(t) = \vec{\mathbf{0}}$. We then define the conditional one slot *Lyapunov drift* as

$$\Delta(\vec{\Theta}(t)) = \mathbb{E}\{L(\vec{\Theta}(t+1)) - L(\vec{\Theta}(t))|\vec{\Theta}(t)\}. \tag{3.15}$$

With the drift defined as in (3.15), it can be shown that

$$\begin{aligned}\Delta(\vec{\Theta}(t)) &= \frac{1}{2}\mathbb{E}\left\{\sum_{k=1}^{K}[(X_k(t+1))^2-(X_k(t))^2|X_k(t)]+\right.\\ &\qquad\left.\sum_{n=1}^{N}[(Z_n(t+1))^2-(Z_n(t))^2|Z_n(t)]\right\}\\ &\leq B+\sum_{n=1}^{N}\mathbb{E}\{Z_n(t)(1-\delta_n)\alpha_n(t)|Z_n(t)\}+\\ &\quad\sum_{k=1}^{K}\mathbb{E}\{X_k(t)(R_k(t)-D_k(t))|X_k(t)\}-\\ &\quad\sum_{n=1}^{N}\mathbb{E}\{(Z_n(t)+\alpha_n(t))p_n(t)|Z_n(t)\}, \end{aligned} \tag{3.16}$$

where $B = \frac{1}{2}\sum_{k=1}^{K}(\max\{D_k^{max}, R_k^{max}\})^2 + \frac{1}{2}\sum_{n=1}^{N}(2+\delta_n^2)(\alpha_n^{max})^2$ is a constant. See Applendix B.2 for the derivation of (3.16).

To minimize the operation cost of the MG, we adopt the *drift-plus-penalty method* [112]. Specifically, we select the control policy $\mathbb{A}(t) = \{Q(t), S(t), R_k(t), D_k(t), p_n(t)\}$ to minimize the bound on the drift-plus-penalty as:

$$\begin{aligned}&\Delta(\vec{\Theta}(t)) + V\mathbb{E}\{Q(t)C(t) - S(t)W(t)|\vec{\Theta}(t)\}\\ \leq\ &\text{right-hand-side of (3.16)} +\\ &V\mathbb{E}\{Q(t)C(t) - S(t)W(t)|\vec{\Theta}(t)\},\end{aligned} \tag{3.17}$$

where $0 < V \leq V_{max}$ is defined in Section 3.2.2 for the trade-off between stability performance and operation cost minimization. Given the current virtual queue states $X_k(t)$ and $Z_n(t)$, market prices $S(t)$ and $W(t)$, available DRERs energy $P(t)$, and the resident quality usage request $\alpha_n(t)$,

the optimal policy is the solution to the following problem.

$$\begin{aligned} \text{minimize:} \quad & B + \sum_{n=1}^{N}[Z_n(t)(1-\delta_n)\alpha_n(t)] + \\ & V[Q(t)C(t) - S(t)W(t)] + \\ & \sum_{k=1}^{K}[X_k(t)(R_k(t) - D_k(t))] - \\ & \sum_{n=1}^{N}[(Z_n(t) + \alpha_n(t))p_n(t)] \\ \text{s.t.} \quad & \text{(3.3), (3.4), (3.6), (3.7), (3.8) .} \end{aligned} \tag{3.18}$$

Since the control policy $\mathbb{A}(t)$ is only applied to the last three terms of (3.18), we can further simplify problem (3.18) as

$$\begin{aligned} \text{minimize:} \ & V[Q(t)C(t) - S(t)W(t)] + \sum_{k=1}^{K}[X_k(t)(R_k(t)- \\ & D_k(t))] - \sum_{n=1}^{N}[(Z_n(t) + \alpha_n(t))p_n(t)] \\ \text{s.t.} \ & \text{(3.3), (3.4), (3.6), (3.7), (3.8) ,} \end{aligned} \tag{3.19}$$

which can be solved based on observations of the current system state $\{X_k(t), Z_n(t), C(t), W(t), P(t), \alpha_n(t)\}$.

3.3 Optimal Electricity Scheduling

3.3.1 Properties of Optimal Scheduling

With the Lyapunov penalty-and-drift method, we transform problem (3.13) to problem (3.19) to be solved for each time slot. The solution only depends on the current system state; there is no need for the statistics of the supply, demand and price processes and no need for any future information. The solution algorithm to this problem is thus an *online algorithm*. We have the following properties for the optimal scheduling.

Lemma 3.1. *The optimal solution to problem (3.19) has the following properties:*

1. *If $Q(t) > 0$, we have $S(t) = 0$,*

   (a) If $X_k(t) > -VC(t)$, the optimal solution always selects $R_k(t) = 0$; if $X_k(t) < -VC(t)$, the optimal solution always selects $D_k(t) = 0$.

   (b) If $Z_n(t) > VC(t) - \alpha_n(t)$, the optimal solution always selects $p_n(t) \geq (1 - \delta_n)\alpha_n(t)$; if $Z_n(t) < VC(t) - \alpha_n(t)$, the optimal solution always selects $p_n(t) = 0$.

2. *When $Q(t) = 0$, we have $S(t) > 0$,*

   (a) If $X_k(t) > -VW(t)$, the optimal solution always selects $R_k(t) = 0$; if $X_k(t) < -VW(t)$, the optimal solution always selects $D_k(t) = 0$.

   (b) If $Z_n(t) > VW(t) - \alpha_n(t)$, the optimal solution always selects $p_n(t) \geq (1 - \delta_n)\alpha_n(t)$; if $Z_n(t) < VW(t) - \alpha_n(t)$, the optimal solution always selects $p_n(t) = 0$.

Proof. The proof of Lemma 3.1 is given in Appendix B.3. □

Lemma 3.2. *The optimal solution to the battery management problem has the following properties:*

1. *If $X_k(t) > -VW_{min}$, the optimal solution always selects $R_k(t) = 0$.*

2. *If $X_k(t) < -VC_{max}$, the optimal solution always selects $D_k(t) = 0$.*

Proof. The proof of Lemma 3.2 is given in Appendix B.4. □

Lemma 3.3. *The optimal solution to the QoSE provisioning problem has the following properties:*

1. *If $Z_n(t) > VC_{max}$, the optimal solution always selects $p_n(t) \geq (1 - \delta_n)\alpha_n(t)$.*

2. *If $Z_n(t) < VW_{min} - \alpha_{max}$, the optimal solution always selects $p_n(t) = 0$.*

Proof. The proof of Lemma 3.3 is given in Appendix B.5. □

Lemma 3.1 provides useful insights for simplifying the algorithm design, which will be discussed in Section 3.3.2. The intuition behind these lemmas is two-fold. On the ESS management side, if either the purchasing price $C(t)$ or the selling price $W(t)$ is low, the MG prefers to recharge the ESS's to store excess electricity for future use. On the other hand, if either $C(t)$ or $W(t)$ is high, the MG is more likely to discharge the ESS's to reduce the amount of energy to purchase or sell more stored energy back to the macrogrid. On the QoSE provisioning side, if either $C(t)$ or $W(t)$ is high and the quality usage $\alpha_n(t)$ is low, the MG is apt to decline the quality usage for lower operation cost. On the other hand, if either $C(t)$ or $W(t)$ is low and $\alpha_n(t)$ is high, the quality usage are more likely to be granted by purchasing more energy or limiting the sell of energy.

3.3.2 MG Optimal Scheduling Algorithm

In this section, we present the MG control policy $\mathbb{A}(t)$ to solve problem (3.19). Given the current virtual queue state $\{X_k(t), Z_n(t)\}$, market prices $C(t)$ and $W(t)$, quality usage $\alpha_n(t)$ and available energy $P(t)$ from the DRERS for serving quality usage, problem (3.19) can be decomposed into the following two linear programming (LP) sub-problems (since one of $S(t)$ and $Q(t)$ must be zero, see (3.7)).

$$\begin{aligned}\text{minimize: } & VQ(t)C(t) + \sum_{k=1}^{K}[X_k(t)(R_k(t) - D_k(t))] - \\ & \sum_{n=1}^{N}((Z_n(t) + \alpha_n(t))p_n(t)) \\ \text{s.t. } & S(t) = 0, \text{(3.3), (3.4), (3.6), (3.8).}\end{aligned} \tag{3.20}$$

$$\begin{aligned}\text{minimize: } & -VS(t)W(t) + \sum_{k=1}^{K}[X_k(t)(R_k(t) - D_k(t))] - \\ & \sum_{n=1}^{N}((Z_n(t) + \alpha_n(t))p_n(t)) \\ \text{s.t. } & Q(t) = 0, \text{(3.3), (3.4), (3.6), (3.8).}\end{aligned} \tag{3.21}$$

Algorithm 3: Adaptive Electricity Scheduling Algorithm

MGCC initializes the QoSE target to δ_n and the virtual queues backlogs $Z_n(t)$ and $X_k(t)$, for all n and k ;
while *TRUE* **do**
    Residents send usage request (with basic and quality usage) to MGCC via the information network ;
    MGCC solves LPs (3.20) and (3.21) ;
    MGCC selects the optimal solution $\mathbb{A}(t)$ comparing the solutions to (3.20) and (3.21) ;
    MGCC updates the virtual queues $X_k(t)$ and $Z_n(t)$ according to (3.11) and (3.12), for all n and k ;
end

In sub-problem (3.20), we set $R_k(t) = 0$ if $X_k(t) > -VC(t)$, and $D_k(t) = 0$ if $X_k(t) < -VC(t)$ according to Lemma 3.1. Also, if $Z_n(t) < VC(t) - \alpha_n(t)$, we set $p_n(t) = 0$; otherwise, we reset constraint (3.6) to a smaller search space of $(1 - \delta_n)\alpha_n(t) \leq p_n(t) \leq \alpha_n(t)$. We take a similar approach for solving sub-problem (3.21) by replacing $C(t)$ with $W(t)$. Then we compare the objective values of the two sub-problems and select the more competitive solution as the MG control policy. The complete algorithm is presented in Algorithm 3.

3.3.3 Performance Analysis

The proposed scheduling algorithm dynamically balances cost minimization and QoSE provisioning. It only requires current system state information (i.e., as an online algorithm) and requires no statistic information about the random supply, demand, and price processes. The algorithm is also robust to non-i.i.d. and non-ergodic behaviors of the processes [113, 114].

Theorem 3.2. *The constraint on the ESS battery level $E_k(t)$, $E_k^{min} \leq E_k(t) \leq E_k^{max}$, is always satisfied for all k and t.*

Proof. The proof of Theorem 3.2 is given in Appendix B.6.

□

Theorem 3.3. *The worst-case backlogs of the QoSE virtual queue for each resident n is bounded by $Z_n(t) \leq Z_n^{max} = VC_{max} + \alpha_n^{max}$, for all n, t. Moreover, the worst-case average amount of outage of quality usage for resident n in a period T is upper bounded by $Z_n^{max} + T\delta_n\alpha_n^{max}$.*

Proof. The proof of Theorem 3.3 is given in Appendix B.7.

□

Theorem 3.4. *The average MG operation cost under the adaptive electricity scheduling algorithm in Algorithm3, y^*, is bounded as $y^* \leq y_{opt} + \hat{B}/V$, where $\hat{B} = B + \sum_{n=1}^{N} Z_n^{max}(1-\delta_n)\alpha_n^{max}$.*

Proof. The proof of Theorem 3.4 is given in Appendix B.8. □

It is worth noting that the choice of V controls the optimality of the proposed algorithm. Specifically, a larger V leads to a tighter optimality gap. However, from the proof of Theorem 3.2, V is limited by V_{max}, which ensures the feasibility of the battery constraints. This is actually a similar phenomenon to the so-called *performance-congestion trade-off* [115]. Through the definition of V_{max} (see Section 3.2.2), it can be seen that if we invest more on the individual storage components for a larger ESS capacity, the proposed algorithm can achieve a better performance (i.e., a smaller optimality gap).

It is also worth noting that all the performance bounds of the proposed algorithm are deterministic, which provide "hard" guarantees for the performance of the proposed adaptive scheduling policy in every time slot. Unlike probabilistic approaches, the proposed method provides useful guidelines for the MG design, while guaranteeing the MG operation cost, grid stability, and the usage quality of residents.

3.4 Simulation Study

We demonstrate the performance of the proposed adaptive MG electricity scheduling algorithm through extensive simulations. We simulated an MG with 500 residents, where the electricity from DRERs is supplied by a wind turbine plant. We use the renewable energy supply data from the

Western Wind Resources Dataset published by the National Renewable Energy Laboratory [116]. The ESS's consists of 100 PHEV Li-ion battery packs, each of which has a maximum capacity of 16 kWh and the minimum energy level is 0. The battery can be fully charged or discharged within 2 hours [117].

The residents' pre-agreed power demand is uniformly distributed in [2 kW, 25 kW], and the quality usage power is uniformly distributed in [0, 10 kW]. The MG works in the grid-connected mode and may purchase/sell electricity from/to the macrogrid. The utility prices in the macrogrid are obtained from [118] and are time-varying. We assume the sell price by the broker is random and below the purchasing price in each time slot. The time slot duration is 15 minutes. The MGCC serves a certain level of quality usage according to the adaptive electricity scheduling policy. The QoSE target is set to $\delta_n = 0.07$ for all residents. The control parameter is $V = V_{max}$, unless otherwise specified.

3.4.1 Algorithm Performance

We first investigate the average QoSEs and total MG operation cost with default settings for a five-day period. We use MATLAB LP solver for solving the sub-problems (3.20) and (3.21). For better illustration, we only show the QoSEs of three randomly chosen users in Fig. 3.3. It can be seen that all the average QoSEs converge to the neighborhood of 0.08 within 200 slots, which is close to the MG requested criteria $\delta_n = 0.07$. In fact the proposed scheme converges exponentially, due to the inherent exponential convergence property in Lyapunov stability based design [110].

We also plot the MG operation traces from this simulation in Fig. 3.6. The energy for serving quality usage from the DEREs are plotted in Fig. 3.6(A). It can be seen that the DRERs generate excessive electricity from slot 150 to 200, which is more than enough for the residents. Thus, the MGCC sells more electricity back to the macrogrid and obtains significant cost compensation accordingly. In Fig. 3.6(B), we plot the traces of electricity trading, where the positive values are the purchased electricity (marked as brown bars), and the negative values represent the sold electricity (marked as dark blue bars). The MG operation costs are plotted in Fig. 3.6(C). The

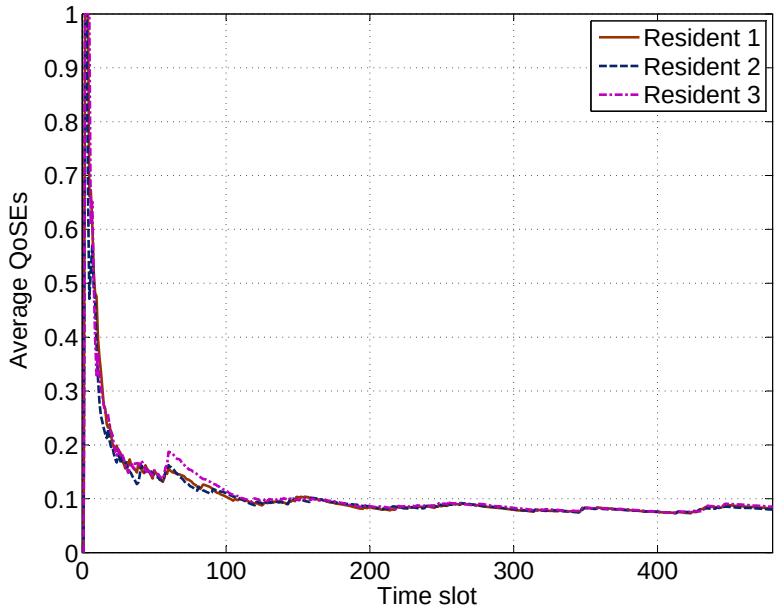


Figure 3.3: Average QoSEs of three residents ($V = V_{max}$).

curve rises when the MG purchases electricity and falls when the MG sells electricity. From slot 150 to 200, the operation cost drops significantly due to profits of selling excess electricity from the DEREs. The operation cost is $\$418.10$ by the end of the period, which means the net spending of the MG is $\$418.10$ on the utility market.

We then examine the energy levels of the batteries in Fig. 3.4. We only plot the levels of three batteries in the first 50 time slots for clarity. The proposed control policy charges and discharges the batteries in the range of 0 to 16 kWh, which falls strictly within the battery capacity limit. It can be seen that the amount of energy for charging or discharging in one slot is limited by 2 kWh in the figure, due to the short time slots comparing to the 2-hour fully charge/discharge periods. For longer time slot durations and batteries with faster charge/discharge speeds, the variation of the energy level in Fig. 3.4 could be higher. However, Theorem 3.2 indicates that the feasibility of the battery management constraint is always ensured, if the control parameter V satisfies $0 < V \leq V_{max}$.

We next evaluate the performance of the proposed adaptive control algorithm under different values of control parameter V. For different values $V = \{V_{max}, V_{max}/2, V_{max}/4\}$, the QoSEs are stabilized at 0.081, 0.061, and 0.055, and the total operation cost are $\$418.10$, $\$625.69$, and

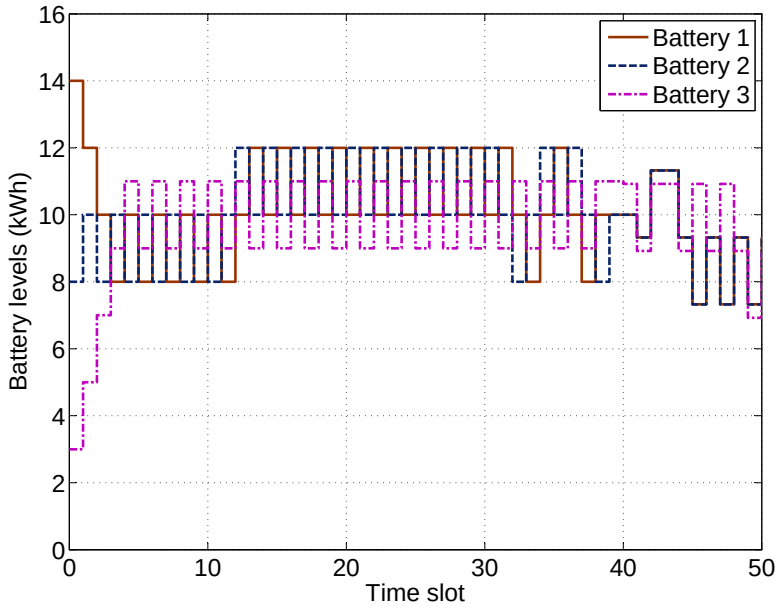


Figure 3.4: Energy levels of three Li-ion batteries ($V = V_{max}$).

$717.75, respectively. We find the QoSE decreases from 0.081 to 0.055, while the total operation cost is increased from $418.10 to $717.75, as V_{max} is decreased. This demonstrates the performance-congestion trade-off as in Theorem 3.4: a larger V leads to a smaller objective value (i.e., the operating cost), but the system is also penalized by a larger virtual queue backlog, which corresponds to a higher QoSE. On the contrary, a smaller V favors the resident quality usage, but increases the total operation cost. In practice, we can select a proper value for this parameter based on the MG design specifications.

It would be interesting to examine the case where the residents require different QoSEs. We assume 5 residents with a service contract for lower QoSEs. We plot the average QoSEs of three residents with $V = V_{max}/2$ in Fig. 3.5. Resident 1 prefers an outage probability $\delta_1 = 0.02$, while residents 2 and 3 require an outage probability $\delta_2 = \delta_3 = 0.07$. It can be seen in Fig. 3.5 that resident 1's QoSE converges to 0.015, while the other two residents' QoSEs remains around 0.063.

3.4.2 Comparison with a Benchmark

We compare the performance of the proposed scheme with a heuristic MG electricity control policy (MECP), which serves as a benchmark. In MECP, the MGCC blocks quality usage requests

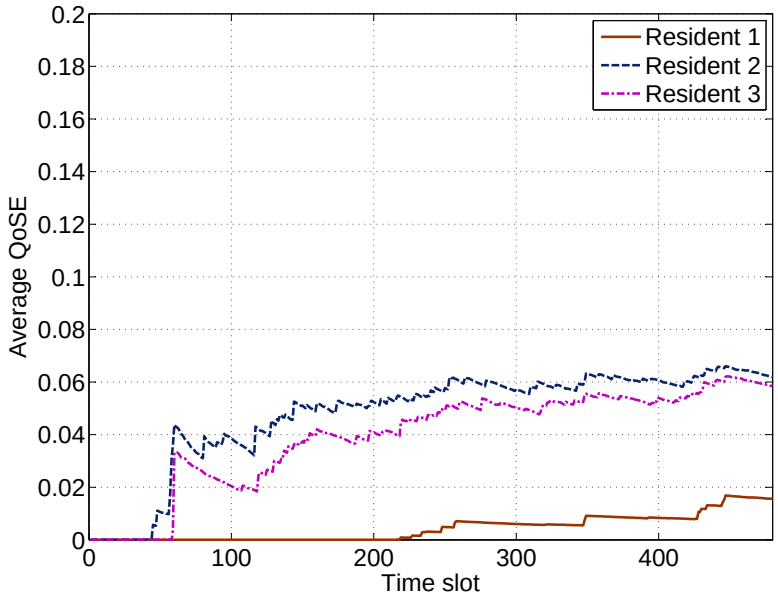


Figure 3.5: QoSEs for three residents with different service contracts ($V = V_{max}/2$).

simply by tossing a coin with the target probability. We use $\delta_n = 0.03$ in the following simulations. If there is sufficient electricity from the DRERs, all the quality usage requests will be granted and the excess energy will be stored in the ESS's. If there is still any surplus energy, the MGCC will sell it to the macrogrid. If there is insufficient electricity from the DRERs, the ESS's will be discharged to serve the quality usage requests. The MGCC will purchase electricity from the macrogrid if even more electricity is required. Finally, with a predefined probability, e.g., 0.5 in the following simulation, the MG purchases as much energy as possible to charge the ESS's.

We run 100 simulations with different random seeds for a seven-day period. We assume in the first five days the resident behavior is the same as previous default settings. In the last two days, we assume the residents are apt to request more electricity (e.g., more activities in weekends) We assume in the last two days the resident pre-agreed basic usage power demand is uniformly distributed from 5 kW to 35 kW. The quality usage power is uniformly distributed from 0 to 20 kW.

We find that the proposed algorithm earns $\$947.27$ from the utility market (with 95% confidence interval $[950.65, 943.89]$). The profit mainly comes from the abundant DRER generation in the last two days, as shown in Fig. 3.7. MECP only earns $\$379.74$ from the market (with 95%

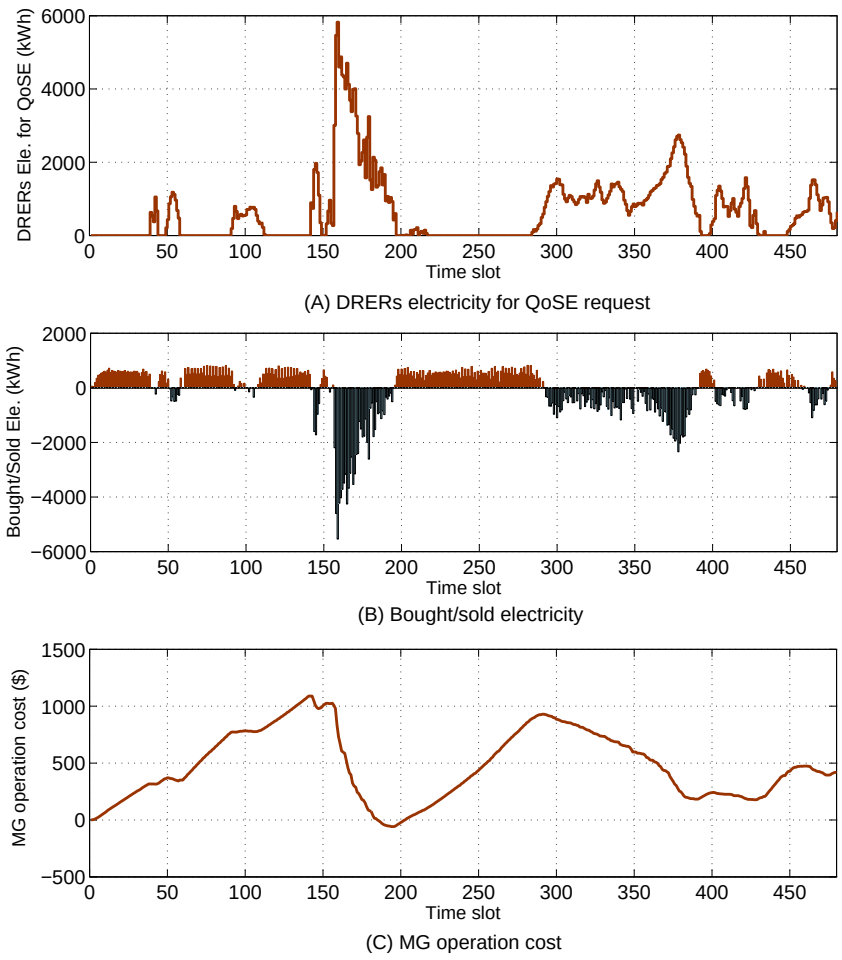


Figure 3.6: MG operation traces of the proposed algorithm for the 5-day period.

confidence interval $[387.96, 371.52]$), which is 60% lower than that of the proposed control policy. We also find that the QoSEs under the proposed control policy remains about 0.025, which is lower than the criteria $\delta_n = 0.03$. This is because there are a sudden price jump from $\$27$/MWh to $\$356$/MWh in the afternoon of the last day. This sharp increment increases C_{max} eight times and decreases the value of V_{max}. Due to the performance-congestion trade-off, the QoSEs become smaller (lower than MECP's 0.03 level).

3.5 Related Work

SG is regarded as the next generation power grid with two-way flows of electricity and information. Several comprehensive reviews of SG technologies can be found in [3, 4]. Recently,

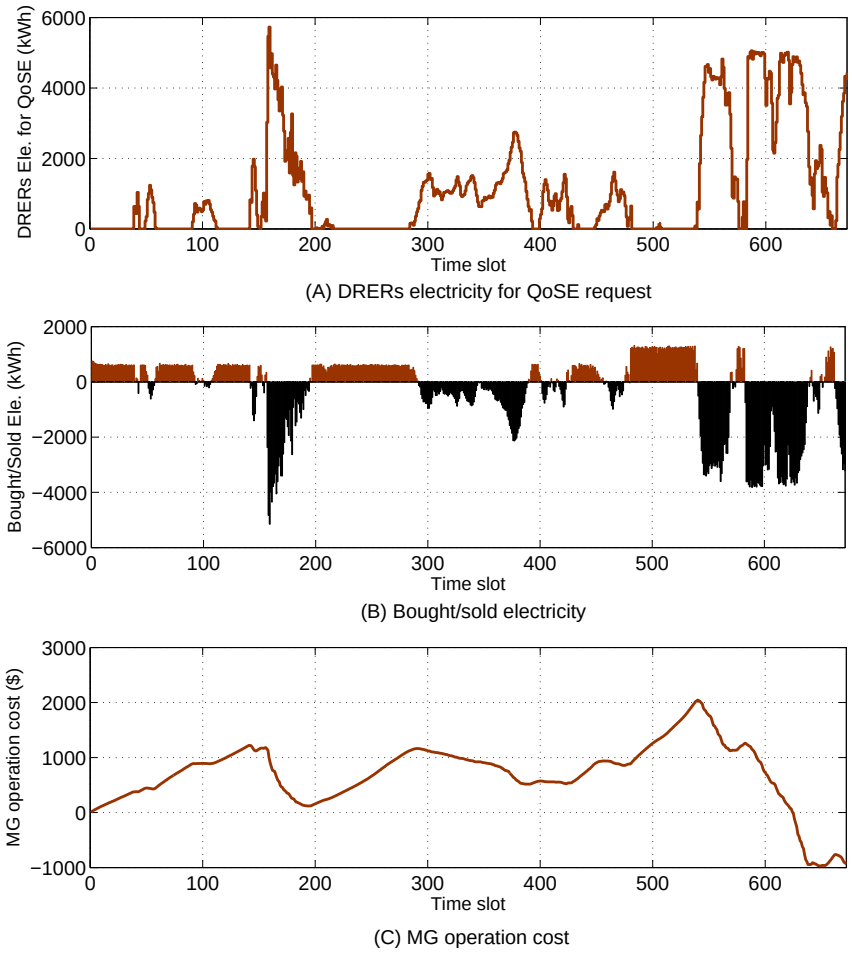


Figure 3.7: MG operation traces of proposed algorithm for the 7-day period.

SG research is attracting considerable interest from the networking and communications communities [119–124]. For example, the design of wireless communication systems in SG is studied in [120]. The authors of [121, 122] explore the important wireless communication security issues in smart grid. The energy management and power flow control in the grid is investigated in [119] to reach system-wide reliability under uncertainties. The frequency oscillation in power networks is studied in [123] by epidemic propagation and a social network based approach. The electric power management with PHEVs are examined in [124].

Microgrid is a new grid structure to group DRERs and local residents loads, which provides a promising way for the future SG. In [108], the authors review the MG structure with distributed energy resources. In [125], the integration of random wind power generation into grids for cost

effective operation is investigated. In [126], the authors propose a useful online method to discover all available DRERs within the islanded mode mircogrid and compute a DRER access strategy. The problem of optimal residential demand management is studied in [92], aiming to adapt to time-varying energy generation and prices, and maximize user benefit. In [127], the authors investigate energy storage management with a dynamic programming approach. The size of the ESS's for MG energy storage is explored in [128].

Lyapunov optimization is a useful stochastic optimization method [12]. It integrates the Lyapunov stability concept of control theory with optimization and provides an efficient framework for solving schedule and control problems. It has been widely used and extended in the communications and networking areas [12, 112]. In two recent work [114, 129], the Lyapunov optimization method is applied to jointly optimize power procurement and dynamic pricing. In [114], the authors investigate the problem of profit maximization for delay tolerant consumers. In [129], the authors study electricity storage management for data centers, aiming to meet the workload requirement. Both of the work are designed based on a *single* energy consumption entity model.

In this chapter, we investigate a novel smart energy management system for MGs based on the concept of QoSE, which is different from above work. By jointly considering *multiple* residents, ESS's and utility market participation, the adaptive electricity scheduling policy is designed with Lyapunov optimization for ensuring the quality of service of the electricity usage and minimizing the MG operation cost.

3.6 Conclusion

In this chapter, we developed an online adaptive electricity scheduling algorithm for smart energy management in MGs by jointly considering renewable energy penetration, ESS management, residential demand management, and utility market participation. We introduced a QoSE model by taking into account minimization of the MG operation cost, while maintaining the outage probabilities of resident quality usage. We transformed the QoSE control problem and ESS management problem into queue stability problems by introducing the QoSE virtual queues and

battery virtual queues. The Lyapunov optimization method was applied to solve the problem with an efficient online electricity scheduling algorithm, which has deterministic performance bounds. Our simulation study validated the superior performance of the proposed approach.

Table 3.1: Notation for Chapter 3

Symbol	*Description*
N	total number of residents
K	total number of batteries
T	total number of slots
$E_k(t)$	energy level for battery k at time slot t
$R_k(t)$	recharging energy for battery k at time slot t
$D_k(t)$	discharging energy for battery k at time slot t
E_k^{max}	maximum battery energy level for battery k
E_k^{min}	minimum battery energy level for battery k
R_k^{max}	maximum supported recharging energy for batter k in a slot
D_k^{max}	maximum supported discharging energy for battery k in a slot
λ_n	average quality usage arrival rate for resident n
ρ_n	average outage rate of quality usage for resident n in MG
δ_n	target QoSE for resident n in MG
$\alpha_n(t)$	quality usage of residents n in time slot t
α_n^{max}	maximum quality usage of resident n in a single slot
$\alpha_n^b(t)$	basic electricity usage of resident n in time slot t
$P(t)$	available electricity from DRERs to supply quality usage in time slot t
$U(t)$	electricity generated from DRERs in time slot t
$Q(t)$	electricity purchased from macrogrid in time slot t
$S(t)$	electricity sold on the market in time slot t
$p_n(t)$	electricity to the resident n
$C(t)$	purchasing price on the utility market in time slot t
$W(t)$	selling price ob the utility market in time slot t
$I_n(t)$	indicator function for outage events of quality usage of resident n in time slot t
C_{min}	minimum purchasing price of utility from macrogrid
C_{max}	maximum purchasing price of utility from macrogrid
W_{min}	minimum selling price of utility to macrogrid
W_{max}	maximum selling price of utility to macrogrid
$X_k(t)$	battery virtual queue for the battery k
$Z_n(t)$	QoSE virtual queue for the resident n
$\vec{\Theta}(t)$	states of the virtual queues $X_k(t)$ and $Z_n(t)$
$L(\cdot)$	Lyapunov function
$\Delta(t)$	Lyapunov one step drift
$\mathbb{A}(t)$	proposed scheduling policy including $Q(t), S(t), R_k(t)$, $D_k(t)$ and $p_n(t)$
y^*	optimal objective value of problem (3.19)
$\hat{\mathbb{A}}(t)$	relaxed scheduling policy for problem (B.6)
$\hat{y}$	optimal objective value of problem (B.6)
y_{opt}	optimal objective value of problem (3.9)

Chapter 4

Overview of Green Video Streaming over Celluar Networks and Variable Bit Rate Video

4.1 Green Video Streaming over Celluar Networks

Besides the redesign of the electricity delivery networks, it is equally important to study the energy efficiency at the demand side. The rapid proliferation of *information and communications technology* (ICT) infrastructures continuously contribute to the overall carbon footprint and bring the intensity of "green" communications to the research community. Among various green communication technologies, we focus on the energy efficiency of *base stations* (BS's) for downlink video streaming. This is due to the expected surge in wireless video data, as well as the drastic increase in the deployment of BS's. It is reported that, in a typical cellular network, more than 50% of the total power consumption is directly attributed to BS equipment [11]. At every year, $120,000$ BS's are added, catering to the 300 million to 400 million new mobile phone users adopting mobile services around the world [130]. Furthermore, Many wireless operators have launched femtocell service recently, such as AT&T, Sprint, Verizon, and Vodafone. The wide adoption of femtocells will greatly intensify the proliferation of BS's. Therefore, any small improvement in the energy efficiency of video coding or wireless video streaming system will be amplified by the huge volume of wireless video data and number of BS's deployed, and will result in considerable environmental impact. Considerable savings on electrical bills could be achieved for wireless operators when the power of BS's is minimized for video streaming. The reduced electricity consumption will also bring about important improvement in the overall carbon footprint of the wireless industry and achieve the goal of "green" communications.

In the following chapters, we consider the challenging problem of streaming multiuser *variable bit rate* (VBR) videos in the downlink of cellular networks. This is motivated by the fact that VBR video offers stable and superior quality over *constant bit rate* (CBR) videos. Furthermore,

(a) Frame 1, VBR (b) Frame 29, VBR (c) Frame 1, CBR (d) Frame 29, CBR

Figure 4.1: Perceived quality of VBR and CBR videos: *Football* video coded with an H.264 codec.

many stored video content are coded in the VBR format. It is important to support such stored VBR video over existing wireless networks without the need for transcoding.

VBR video has stable visual quality for the frames, but at the cost of large variations in the bit rate, while CBR video maintains a stable bit rate, but the frames have large variations in visual quality. This is illustrated in Fig. 4.1, where the 15 fps *Football* sequence is encoded using an H.264 codec. Both VBR and CBR videos are encoded at the approximately same bit rate (250 kb/s). It can be seen that although the two Frame 1s have similar visual quality, CBR Frame 29 looks worse than VBR Frame 29 when there is high motion. However, we may also observe that the sizes of frames of VBR video have much larger variation than those of CBR video in Fig. 4.2. It can be noted that the frame size varies in CBR from frame 42 to 45. That is because that the content on the video switches from the high motion players to the static field, and this simple scene is kept from frame 42 to frame 43, which allows the rate control algorithm to choose smaller frame sizes. Starting from frame 44, the players come back to the scene, thus the larger frame sizes are selected to compensate the rate decrease in previous frames to keep the average rate constant.

The following chapters distinguish themselves from other energy efficient designs over wireless networks in the following aspects: First, instead of power-aware mobile video devices, we focus on the BS power efficiency when transmitting multiuser videos. As mentioned before, the BS equipment consumes more than 50% of the total power in a typical cellular network. Thus, it is important to improve the BS energy efficiency to achieve the goal of green communications. Second, we explicitly investigate streaming of multiple VBR videos. VBR videos can offer constant and better QoE over CBR videos with the same bit budget. However, VBR videos are notoriously difficult to schedule and control in wireless networks, due to the high variability and the

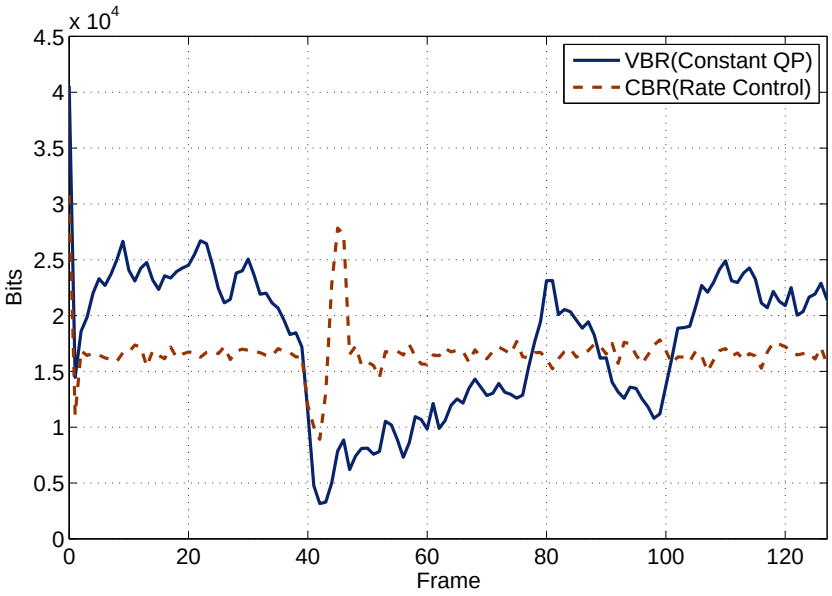


Figure 4.2: Frame size of VBR and CBR videos: *Football* video coded with an H.264 codec.

complex autocorrelation structure [131–133]. Third, we adopt the control and optimization approach to optimize the BS energy allocation as well as the *quality of experience* (QoE) of users, by jointly considering power control, wireless channel condition, playout buffer constraints, and playout deadlines.

4.2 VBR Video System Model

It is a challenging problem to support VBR video traffic, which is found to exhibit both strong long-range and short-range-dependence [131, 132]. It is nontrivial to develop parsimonious traffic models that can accurately capture the auto-correlation structure. The large frame size variations may cause frequent playout buffer underflow or overflow [134]. To address this issue, we adopt a *deterministic traffic model* for stored VBR video, which considers frame size, frame rate, and playout buffers [54, 95, 135, 136]. Unlike prior work that is focused on a single video session over a given CBR or VBR channel, we exploit power control, a unique capability in wireless networks, to adjust the downlink capacities based on prior knowledge of frame sizes and playout schedules. Usually large frames are rarely transmitted simultaneously. Thus jointly optimizing

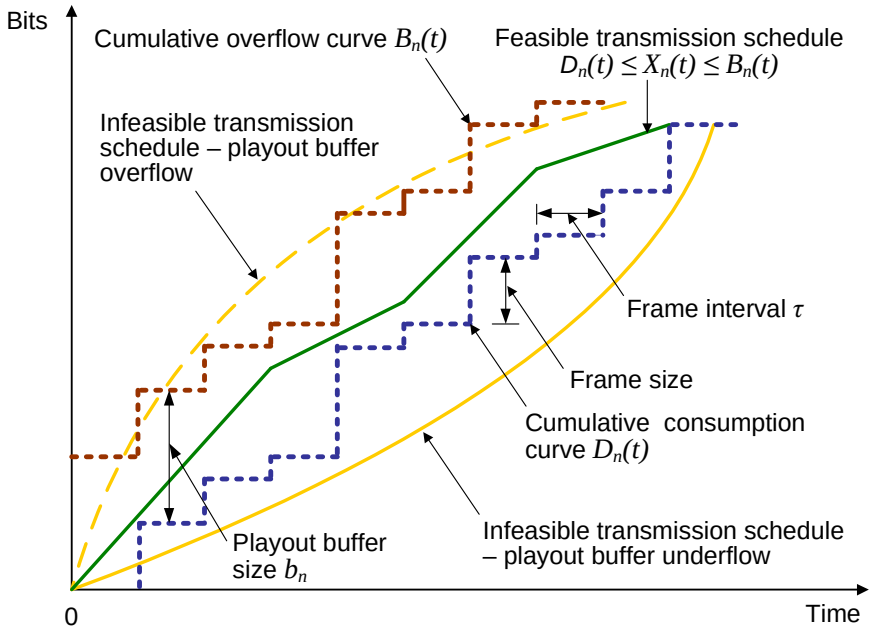


Figure 4.3: Transmission schedules for VBR video session n.

the BS transmit powers is, in some sense, analogous to statistical multiplexing VBR videos in the cellular networks.

A stochastic model capturing the auto-correlation structure often requires a large number of parameters, and is thus hard to be incorporated for scheduling real-time video data. To this end, we adopt a *deterministic model* that considers frame sizes, playout buffers, and schedule [95]. Let $D_i(t)$ denote the *cumulative consumption curve* of the i-th user, representing the cumulative amount of bits consumed by the decoder at time t. The cumulative consumption curve is determined by video characteristics such as frame sizes and rates, and playout schedule. Assume user i has a playout buffer of size b_i bits and its video has L_i frames. We can derive a *cumulative overflow curve* for user i as

$$B_i(t) = \min\{D_i(t-1) + b_i, D_i(L_i)\},\ 0 \le t \le L_i. \tag{4.1}$$

$B_i(t)$ is the maximum number of cumulative received bits at time t without overflowing user i's playout buffer. Finally we define *cumulative transmission curve* $X_i(t)$ as the cumulative amount of bits transmitted to user i at time t. To simplify notation, we assume the video sessions have

identical frame rate and the frame intervals are synchronized. Thus a time slot t is equal to the t-th frame interval, denoted as τ, for $0 \leq t \leq \max_i\{L_i\}$.[1]

Since $D_i(t)$, $B_i(t)$ and $X_i(t)$ are cumulative curves, they are all nondecreasing functions of time. The three curves for user n are illustrated in Fig. 4.3. A feasible transmission schedule will produce a cumulative transmission curve $X_i(t)$ that lies within $D_i(t)$ and $B_i(t)$, i.e., causing neither underflow nor overflow at the playout buffer. In practice, $D_i(t)$'s are known for stored videos and are delivered to the BS's (or a centralized video scheduler) during the session setup phase, and $B_i(t)$'s are then derived as in (4.1).

This deterministic VBR video model will be adopted in following chapters.

[1]This assumption can be relaxed for more general cases. For example, if the frame rates are different, we can use a time slot duration that is equal to the greatest common divisor of all the frame intervals (if not too small). If the frame intervals are not synchronized, a time slot can be a fraction of a frame interval within which the $D_i(t)$'s of all the videos remain constant. In fact, the time slot duration could be arbitrary as in [50] (i.e., equal to multiple frame intervals). Since the cumulative overflow and consumption curves are known, we can still determine the upper and lower bounds for the transmission rate in each time slot. The problem formulation and proposed solution procedures to be discussed in the following sections apply to these cases.

Chapter 5

Downlink Power Allocation for Stored Variable-Bit-Rate Video in Cellular Network

5.1 Introduction

In this chapter, we present a downlink power control framework for streaming multiple *variable bit rate* (VBR) videos in a cellular network with intracell interference. With the deterministic VBR video traffic model 4.2, we formulate an optimization problem that jointly considers donwlink power control, intra-cell interference, VBR video traffic characteristics, playout buffer underflow and overflow constraints, and base station (BS) peak power constraint. The objective is to maximize the total throughput, which can achieve high playout buffer utilization. As a result, playout buffer underflow or overflow events can be minimized. We analyze the convex/concave regions of the formulated problem and develop a two-step downlink power allocation algorithm for solving the problem. We also develop a distributed algorithm based on the dual decomposition technique from convex optimization, in order to reduce the control and computation overhead at the BS. We evaluate the performance of the proposed distributed algorithm with simulations using VBR video traces. Our simulation results verify the accuracy of the analysis and demonstrate the efficacy of the proposed algorithms.

The remainder of this chapter is organized as follows. The deterministic VBR video model is introduced in Section 4.2. The system model is presented in Section 5.2. We develop a two-step algorithm to solve the power allocation problem in Section 5.3, and a distributed algorithm based on dual decomposition in Section 5.4. Simulation results are presented in Section 5.5 and related work are discussed in Section 5.6. Section 5.7 concludes this chapter.

The notation used in this chapter are summarized in Table 5.1.

Table 5.1: Notation for Chapter 5

Symbol	*Description*
N	total number of users in a cell
L	processing gain
β	interference proportion
$\mathcal{U}$	set of users sharing the same channel
T_n	total number of frames for user n video
b_n	playout buffer size of user n
$D_n(t)$	cumulative consumption curve at user n
$X_n(t)$	cumulative transmission curve at user n
$B_n(t)$	cumulative overflow curve at user i
$\vec{P}(t)$	BS transmit power vector in time slot t
$\vec{P}_{max}$	max. power allocation vector without overflow
$\vec{P}_{min}$	min. power allocation vector without underflow
$\bar{P}$	peak power constraint for the BS's
$\bar{P}_{min}$	sum of the elements in $\vec{P}_{min}$
$\vec{P}^*$	inflection power vector
$\hat{\vec{P}}$	optimal power vector
G_n	path gain from BS to user n
B_w	channel bandwidth
τ	duration of a time slot
η_n	noise power at user n
C_n	capacity from the base station to user n
B_w	Channel bandwidth
κ	Constant for the proof of Lemma 5.2
$\vec{P}'(t), \vec{P}''(t)$	Auxiliary power allocation in the Lemma 5.2 proof
A_n	Ratio of noise power and channel gain of user n
P_n^{th}	Minimum between P_n^{max} and P_n^*
$\theta(l)$	Stepsize of step l in (5.26)
$\alpha_\lambda(l), \alpha_\mu(l), \alpha_\nu(l)$	Stepsize of step l in (5.29)
$\gamma_n(t)$	SINR at user un_n in time slot t
$\gamma_n^{min}(t)$	minimum SINR corresponding to $C_n^{min}(t)$
$\gamma_n^{max}(t)$	max. SINR for user un_n without overflow
γ_n^{th}	receiver sensitivity at user n
$\mathbf{F}$	$N \times N$ matrix defined in (5.13)
λ, μ, ν	Lagrange Multipliers
$\mathcal{L}$	Lagrange function

5.2 System Model and Problem Formation

We consider the downlink of a cellular network. In the cell, a BS streams multiple VBR videos simultaneously to mobile users in the cell, which share the downlink bandwidth. We assume the

last-hop wireless link is the bottleneck, while the wired segment of a session path is reliable with sufficient bandwidth. Thus the corresponding video data is always available at the BS before the scheduled transmission time. The *deterministic VBR video model* is adopted as indicated in Section 4.2.

We consider N subscribers in the cell and let $\mathcal{U}$ denote the set of users. In each time slot t, the BS transmits to each user n with power $P_n(t)$ and the *power allocation* is $\vec{P}(t) = [P_1(t), \cdots, P_n(t)]^T$. We also consider a *maximum transmit power* constraint $\bar{P}$, i.e., $\sum_{n\in\mathcal{U}} P_n(t) \leq \bar{P}$, for all t. When the power allocation $\vec{P}(t)$ is determined, the *Signal to Interference-plus-Noise Ratio* (SINR) at user n can be written as [50, 137]

$$\gamma_n(\vec{P}(t)) = \frac{L_n G_n P_n(t)}{\beta \sum_{k\neq n} G_n P_k(t) + \eta_n}, \tag{5.1}$$

where P_n is the power allocated to user n, G_n is the path gain between the BS and user n, η_n is the noise power at user n, L_n is a constant for user n (e.g., processing gain), and β denotes the orthogonality factor, with $0 \leq \beta \leq 1$. In this chapter, we consider the case $\beta = 1$, where the SINR of a user not only depends on its own power allocation but also the power allocations of other users.

We assume slow-fading channels such that the path gains do not change within each time slot [50]. The downlink capacity $C_n(t)$ depends on the SINR at user n, the channel bandwidth B_w, and the transceiver design, such as modulation and channel coding. Without loss of generality, we use the upper bound as predicted by Shannon's Theorem:

$$C_n(\vec{P}(t)) = B_w \log\left(1 + \gamma_n(\vec{P}(t))\right). \tag{5.2}$$

In time slot t, $C_n(t)\tau$ bits of video data will be delivered to user n. The cumulative transmission curve $X_n(t)$ is

$$X_n(0) = 0; \quad X_n(t) = X_n(t-1) + C_n(t)\tau. \tag{5.3}$$

For a feasible power allocation, the cumulative transmission curves should satisfy

$$D_n(t) \leq X_n(t) \leq B_n(t), \text{ for all } n, t, \tag{5.4}$$

i.e., without causing playout buffer underflow or overflow.

From (5.2)$\sim$(5.4), the lower and upper limit on the feasible SINR at user n can be derived as

$$\begin{cases} \gamma_n^{min}(t) = \max\left\{\exp\left\{\frac{\max\{0, D_n(t)-X_n(t-1)\}}{B_w\tau}\right\}, \gamma_n^{th}\right\} \\ \gamma_n^{max}(t) = \exp\left\{\frac{B_n(t)-X_n(t-1)}{B_w\tau}\right\}, \end{cases} \tag{5.5}$$

where γ_n^{th} is the minimum SINR requirement imposed by the transceiver design. $\gamma_n^{min}(t)$ is the SINR that the just empties the buffer at the end of time slot t, without causing underflow; $\gamma_n^{max}(t)$ is the SINR that just fills up the buffer at the end of time slot t, without causing overflow.

Generally, feasible power allocation $\vec{P}(t)$ is not unique for a given set of VBR video sessions. Among the set of feasible solutions, a schedule that transmits more data is more desirable since it provides more flexibility for optimizing future power allocations. We formulate the problem of optimal downlink power control for VBR videos, termed problem **A**, as

(A) maximize
$$\sum_{n\in\mathcal{U}} \log(1+\gamma_n(t)) \tag{5.6}$$

subject to:

$$\gamma_n(t) = \frac{L_n G_n P_n(t)}{\sum_{k\neq n} G_n P_k(t) + \eta_n}, \text{ for all } n \tag{5.7}$$

$$\gamma_n^{min}(t) \leq \gamma_n(t) \leq \gamma_n^{max}(t), \text{ for all } n \tag{5.8}$$

$$\sum_{n\in\mathcal{U}} P_n \leq \bar{P}. \tag{5.9}$$

In problem **A**, the objective is to achieve the maximum buffer uitilization at the users, under playout buffer underflow and overflow constraints and BS maximum transmit power constraints. This is a nonlinear nonconvex problem, to which traditional convex optimization techniques cannot directly apply. Due to the large variability of VBR traffic, the SINRs may assume values ranging

from very low to very high, to avoid playout buffer underflow and overflow. Thus the existing high SINR approximation [59] and low SINR approximation [138] techniques cannot be directly applied.

5.3 Two-Step Downlink Power Allocation

In problem **A**, we consider an interference-limited system, where the capacity of downlink n depends on the power allocations for all the users. In the following, we first derive conditions for the optimal solution, and then present a two-step power allocation algorithm for solving problem **A**.

Lemma 5.1. *If there exists a feasible power allocation $\vec{P}(t)$ that achieves $\gamma_n^{max}(t)$ for all n, the solution is optimal.*

Proof. See Appendix C.1. □

Lemma 5.2. *If the upper limit $\gamma_n^{max}(t)$ cannot be achieved for every user n, then the optimal power allocation $\vec{P}(t)$ satisfies $\sum_{n\in\mathcal{U}} P_n(t) = \bar{P}$.*

Proof. See Appendix C.2.

□

We have the following result for the optimal solution of problem **A**, which directly follows Lemmas 5.1 and 5.2.

Theorem 5.1. *A solution to problem* **A** *is optimal if (i) it achieves the maximum SINR $\gamma_n^{max}(t)$ for all n; or (ii) its total transmit power is $\bar{P}$.*

Proof. By *Lemma* 5.1 and *Lemma* 5.2, it is straightforward to obtain the result. □

Theorem 5.1 implies that we can examine the SINR (or buffer) constraints and the peak power constraint separately. In the rest of this section, we present a two-step power allocation algorithm

for solving problem **A**. We first examine problem **A** under condition (i) in Theorem 5.1, to obtain problem **B** as

$$\textbf{(B)} \qquad \gamma_n^{max}(t) = \frac{L_n G_n P_n(t)}{\sum_{k \neq n} G_n P_k(t) + \eta_n}, \text{ for all } n, \tag{5.10}$$

subject to:

$$\sum_{n \in \mathcal{U}} P_n \leq \bar{P}. \tag{5.11}$$

In problem **B**, (5.10) is a system of linear equations of power allocation $\vec{P}(t)$. Rearranging the terms, we can rewrite (5.10) in the matrix form as:

$$(\mathbf{I} - \mathbf{F})\, \vec{P}(t) = \vec{u}, \;\; \text{for } \vec{P}(t) \succ \vec{0}, \tag{5.12}$$

where **I** is the *identity matrix*, **F** is a $N \times N$ matrix with

$$F_{nm} = \begin{cases} 0, & \text{if } n = m \\ \gamma_n^{max}/L_n, & \text{otherwise}, \end{cases} \tag{5.13}$$

and $\vec{u} = [\eta_1 \gamma_1^{max}/L_n G_1, \eta_2 \gamma_2^{max}/L_n G_2, \cdots, \eta_N \gamma_N^{max}/L_n G_N]^T$.

Since all the variables are nonnegative, **F** is a non-negative matrix. According to the Perron-Frobenius Theorem, we have the following equivalent statements [47]:

Fact 5.1. *The following statements are equivalent: (i) there exits a feasible power allocation satisfying (5.12); (ii) the spectrum radius of* **F** *is less than 1; (iii) the reciprocal matrix* $(\mathbf{I} - \mathbf{F})^{-1} = \sum_{k=0}^{\infty} (\mathbf{F})^k$ *exists and is component-wise positive.*

Based on Theorem 5.1 and Fact 5.1, we derive the *first step* of the two-step power allocation algorithm, as given in Algorithm 4. If problem **B** is solvable, the Step I algorithm in Algorithm 4 produces the optimal solution for problem **A** according to Theorem 5.1. Otherwise, we derive

Algorithm 4: Two-Step Power Allocation Algorithm: Step I

BS obtains b_n, D_n, and B_n, and computes γ_n^{max} for all user n;
BS tests the existence of feasible solutions using (5.12);
if *(5.12) is solvable* **then**
    Compute its solution $\vec{P}(t)$;
else
    Go to Step II of the algorithm, as given in Algorithm 5;
end
if $\sum_{n\in\mathcal{U}} P_n(t) \leq \bar{P}$ **then**
    Stop with the optimal solution $\vec{P}(t)$;
else
    Go to Step II of the algorithm, as given in Algorithm 5;
end

problem **C** by applying Lemma 5.2, as

$$\textbf{(C) maximize} \quad \sum_{n\in\mathcal{U}} \log(1+\gamma_n(t)) \tag{5.14}$$

subject to:

$$\gamma_n(t) = \frac{L_n P_n(t)}{\bar{P} - P_n(t) + A_n}, \text{ for all } n \tag{5.15}$$

$$P_n^{min}(t) \leq P_n(t) \leq P_n^{max}(t), \text{for all } n \tag{5.16}$$

$$\sum_{n\in\mathcal{U}} P_n(t) = \bar{P}, \tag{5.17}$$

where $A_n = \eta_n/G_n$ is the ratio of noise power and channel gain, representing the quality of the user n downlink channel. $P_n^{min}(t)$ and $P_n^{max}(t)$ are solved from (5.8) and (5.15), as

$$\begin{cases} P_n^{min}(t) = \gamma_n^{min}(\bar{P} + A_n)/(L_n + \gamma_n^{min}) \\ P_n^{max}(t) = \gamma_n^{max}(\bar{P} + A_n)/(L_n + \gamma_n^{max}). \end{cases} \tag{5.18}$$

Since the total transmit power is $\bar{P}$, the objective value in (5.14) and the SINR in (5.15) for each user only depends on its own power. Note that all the constraints are now linear. To solve problem **C**, we examine the objective function to see if it is convex. We omit time index t in the following for brevity.

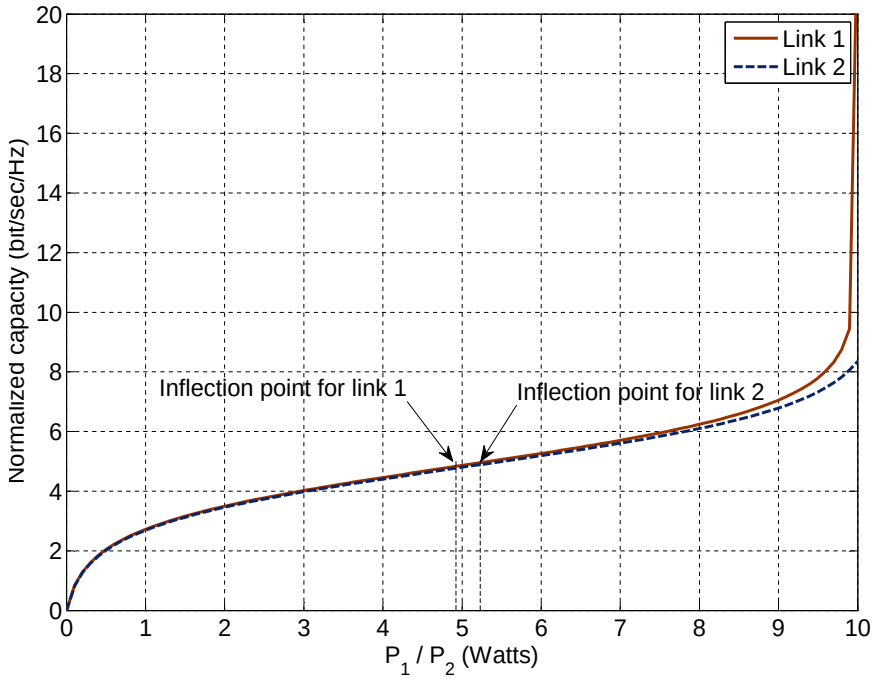


Figure 5.1: Normalized capacity curves and inflection points for a two-user system, where link 1 has better quality than link 2, i.e. $A_1 < A_2$.

Lemma 5.3. *The capacity of each user n, C_n, has one* inflection point P_n^**: when $P_n < P_n^*$, C_n is in concave; when $P_n > P_n^*$, C_n is convex.*

Proof. See Appendix C.3. □

The normalized capacities for a two-user system is plotted in Fig. 5.1, with the inflection points marked. It can been observed that the curves are concave on the left hand side of the inflection points and convex on the right hand side of the inflection points. The processing gain is usually large for practical systems (e.g., $L_n = 128$ in IS-95 CDMA). We assume $L_n \gg 1$ in the following analysis.

Theorem 5.2. *For problem* **C***, there can be at most two links operating in the convex region if* $L_n \geq (4\bar{P} + 6A_n)/(\bar{P} + 3A_n)$.

Proof. See Appendix C.4. □

For a clean channel where $A_n \approx 0$, $L_n \geq 4$ will guarantee at most two links operating in the convex region. The following results are on the impact of channel quality $A_n = \eta_n/G_n$.

Theorem 5.3. *For a given L_n, the inflection point P_n^* is an increasing function of A_n. For two links i and j with the same transmit power P, if $A_i < A_j$, we have $C_i(P, A_i) > C_j(P, A_j)$ and $\frac{\partial C_i(P_i,A_i)}{\partial P_i}|_{P_i=P} > \frac{\partial C_j(P_j,A_j)}{\partial P_j}|_{P_j=P} > 0$.*

Proof. See Appendix C.5. □

Theorem 5.3 shows that, for two links in the convex region with the same initial power P, allocating more power to the link with better quality can achieve larger objective value than alternative ways of splitting the power between the two links (i.e., achieving the multi-user diversity gain). Based on the above analysis, we develop the *second step* of the power allocation algorithm for solving problem **C**, as given in Algorithm 5. In Algorithm 5, Lines $3 \sim 4$ tests the feasibility of the power allocation. If the sum of the total minimum required power is larger than the BS peak power, there is no feasible power allocation and there will be buffer underflow. In this case, we select users with "good" channels for transmission and suspend the users with "bad" channels.

The Step II algorithm checks the three possible solution scenarios for problem **C** depending on the network status and video parameters:

- All links operate in the convex region;
- One link operates in the convex region and the remaining links operate in the concave region
- Two links operate in the convex region and the remaining links operate in the concave region.

Each of the three phases in Algorithm 5 considers the optimality condition for one of the three scenarios. In particular, Phase 1 first optimizes the power allocation in the concave region and then allocates the remaining power to the links that could be moved to the convex region. Phase 2 allocates as much power as possible to the link with the best quality, which could work in the convex region. Phase 3 attempts to move the second best link to the convex region if the total power constraint is not violated. Usually when L_n and n are large, Phase 3 will rarely occur due to the peak power constraint.

Algorithm 5: Two-Step Power Allocation Algorithm: Step II

1 Initialization;
2 BS obtains b_n, D_n, and B_n for all user n;
3 BS computes γ_n^{max}, γ_n^{min}, and P_n^*, for all n;
4 BS computes the minimum required sum power $\bar{P}_{min} = \sum_{n\in\mathcal{U}} P_n^{min}$ and gap $\Delta_P = \bar{P} - \bar{P}_{min}$;
5 **if** $\bar{P}_{min} > \bar{P}$ **then**
6 | Remove links from $\mathcal{U}$, according to descending order of A_n, until $\bar{P}_{min} \leq \bar{P}$;
7 **end**
8 Compute $R_n = \frac{C_n(\min\{P_n^{max}, P_n^{min}+\Delta_P\}) - C_n(P_n^{min})}{\min\{P_n^{max}, P_n^{min}+\Delta_P\} - P_n^{min}}$, for all $P_n^{max} > P_n^*$;
9 Phase 1;
10 Select all the users satisfying $P_n^{min} < P_n^*$ as a set $\mathcal{U}' \subseteq \mathcal{U}$;
11 Solve problem **C** under constraints $P_n^{min} \leq P_n \leq \min(P_n^{max}, P_n^*)$ and $\sum_{n\in\mathcal{U}'} P_n \leq \bar{P}' = \bar{P} - \sum_{n\in\bar{\mathcal{U}}'} P_n^{min}$, where $\bar{\mathcal{U}}'$ is the complementary set of $\mathcal{U}'$, and obtain solution $\vec{P}_1$;
12 Calculate R_n by updating P_n^{min} to the solution in Line 11 and assign the remaining power to the nodes in set $\mathcal{U}$, in descending order of R_n;
13 Obtain the Phase 1 solution, $\vec{P}_{p_1}$, and objective value f_{p_1};
14 Phase 2;
15 Select the link with the maximum R_n, and assign all the available power $\bar{P} - \bar{P}_{min}$ to the link, until either all the power is assigned or the link attains power P_n^{max};
16 **if** *there is still power to allocate* **then**
17 | Select all the nodes in set $\mathcal{U}\backslash n$ and repeat Lines $8 \sim 12$;
18 **end**
19 Obtain the Phase 2 solution, $\vec{P}_{p_2}$, and objective value f_{p_2};
20 Phase 3;
21 Select the first 2 links with the largest R_n's, and assign all the available power $\bar{P} - \bar{P}_{min}$ to the links, until all the power is assigned or the links attains power P_n^{max}, and repeat Lines $16 \sim 18$;
22 Obtain the Phase 3 solution, $\vec{P}_{p_3}$, and objective value f_{p_3};
23 Decision;
24 Choose the largest objective value among f_{p_1}, f_{p_2} and f_{p_3}, and stop with the corresponding power assignment;

In Algorithm 5, Line 7 presents a convex optimization component, for which several effective solution techniques can be applied. In the following section, we describe a distributed algorithm for Line 7 based on dual decomposition.

5.4 Distributed Algorithm

As discussed in Section 5.3, the core of the Step II algorithm is to solve problem **C** in the concave region (see Fig. 5.1). In this section, we present a distributed algorithm for this purpose,

where the users are involved in power allocation to reduce the control and computation overhead on the BS. In the concave region, we have problem **D** as

$$\textbf{(D) maximize} \quad \sum_{n\in\mathcal{U}} \log(1+\gamma_n(t)) \tag{5.19}$$

subject to:

$$\gamma_n(t) = \frac{L_n P_n(t)}{\bar{P} - P_n(t) + A_n}, \text{ for all } n \tag{5.20}$$

$$P_n^{min}(t) \le P_n(t) \le \min\{P_n^{max}, P_n^*\}, \text{for all } n \tag{5.21}$$

$$\sum_{n\in\mathcal{U}} P_n(t) \le P_{tot}, \tag{5.22}$$

where $P_{tot} \le \bar{P}$ is the total power budget for the links in the concave region. For brevity, we define $P_n^{th} = \min\{P_n^{max}, P_n^*\}$ and drop the time slot index t in the following analysis.

Introducing non-negative Lagrange multipliers λ_n, μ_n, and ν for constraints (5.21) and (5.22), respectively, we obtain the Lagrange function as

$$\begin{aligned} &\mathcal{L}(\vec{P}, \vec{\lambda}, \vec{\mu}, \nu) \\ &= \sum_{n\in\mathcal{U}} \left[\log\left(1 + \frac{L_n P_n}{\bar{P} - P_n + A_n}\right) + \lambda_n (P_n - P_n^{min}) \right] + \\ &\quad \sum_{n\in\mathcal{U}} \left[\mu_n (P_n^{th} - P_n)\right] + \nu \left(P_{tot} - \sum_{n\in\mathcal{U}} P_n \right) \\ &= \sum_{n\in\mathcal{U}} \left[\mathcal{L}_n(P_n, \lambda_n, \mu_n, \nu) + (\mu_n P_n^{th} - \lambda_n P_n^{min}) \right] + \nu P_{tot}, \end{aligned} \tag{5.23}$$

where

$$\mathcal{L}_n(P_n, \lambda_n, \mu_n, \nu) = \log\left(1 + \frac{L_n P_n}{\bar{P} - P_n + A_n}\right) + (\lambda_n - \mu_n - \nu) P_n. \tag{5.24}$$

Since $\mathcal{L}_n$ only depends on user n's own parameters, we have the dual decomposition for each user n. For given Lagrange multipliers (or, prices) $\hat{\lambda_n}$, $\hat{\mu_n}$, and $\hat{\nu}$, we have the following subproblem for each user n.

$$\hat{P}_n(\hat{\lambda_n}, \hat{\mu_n}, \hat{\nu}) = [P_n^{min} \le P_n \le P_n^{th}] \arg\max \mathcal{L}_n(P_n, \hat{\lambda_n}, \hat{\mu_n}, \hat{\nu}), \text{ for all } n. \tag{5.25}$$

Subproblem (5.25) has a unique optimal solution due to the strict concavity of $\mathcal{L}_n$. We use the gradient method [14] to solve (5.25), where user n iteratively updates its power P_n as:

$$\begin{aligned} & P_n(l+1) \\ &= [P_n(l) + \theta(l)\nabla_n \mathcal{L}_n(P_n)]^* \\ &= \left[P_n(l) + \theta(l)\frac{L_n(\bar{P}+A_n)}{(\bar{P}-P_n+A_n)(\bar{P}+(L_n-1)P_n+A_n)} + \theta(l)(\lambda_n - \mu_n - \nu)\right]^*, \end{aligned} \tag{5.26}$$

where $[\cdot]^*$ denotes the projection onto the range of $[P_n^{min}, P_n^{th}]$. The update stepsize $\theta(l)$ varies in each step l and is determine by the Armijo Rule [14]. Due to the strict concavity of $\mathcal{L}_n$, the series $\{P_n(1), P_n(2), \cdots\}$ will converge to the optimal solution $\hat{P}_n$ as $l \to \infty$.

For a given optimal solution for problem (5.25), $\vec{\hat{P}} = [\hat{P}_1, \cdots, \hat{P}_N]^T$, the master dual problem is as follows:

$$\textbf{minimize} \qquad \mathcal{L}(\vec{\hat{P}}, \vec{\lambda}, \vec{\mu}, \nu) \tag{5.27}$$

$$\textbf{subject to:} \qquad \lambda_n, \mu_n, \nu \geq 0, \text{ for all } n. \tag{5.28}$$

Since the objective function (5.27) is differentiable, we also apply the gradient method to solve the master dual problem [14], where the Lagrange multipliers are iteratively updated as

$$\begin{cases} \lambda_n(l+1) = [\lambda_n(l) - \alpha_\lambda(l) \cdot \frac{\partial \mathcal{L}(\vec{\lambda},\vec{\mu},\nu)}{\partial \lambda_n}]^+, \text{ for all } n \\ \mu_n(l+1) = [\mu_n(l) - \alpha_\mu(l) \cdot \frac{\partial \mathcal{L}(\vec{\lambda},\vec{\mu},\nu)}{\partial \mu_n}]^+, \text{ for all } n \\ \nu(l+1) = [\nu(l) - \alpha_\nu(l) \cdot \frac{\partial \mathcal{L}(\vec{\lambda},\vec{\mu},\nu)}{\partial \nu}]^+, \end{cases} \tag{5.29}$$

where $[\cdot]^+$ denotes the projection onto the nonnegative axis. The update stepsizes are also determined by the Armijo Rule [14]. As the dual variables $\vec{\lambda}(l), \vec{\mu}(l), \nu(l)$ converge to their stable values as $l \to \infty$, the primal variables $\vec{\hat{P}}$ will also converge to the optimal solution [100].

Algorithm 6: Distributed Power Control Algorithm

1 BS sets $l = 0$ and prices $\lambda_n(l), \mu_n(l), \nu(l)$ equal to some nonnegative initial values for all n;
2 BS broadcasts the prices to the selected users;
3 Each user locally solves problem (5.25) as in (5.26) to obtain its requested power;
4 Each user sends its requested power to the BS;
5 BS updates prices $\lambda_n(l), \mu_n(l), \nu(l)$ as in (5.29) and broadcasts new prices $\lambda_n(l+1), \mu_n(l+1), \nu(l+1)$ for all n;
6 Set $l = l + 1$ and go to Step 3, until the solution converges;

The distributed algorithm is given in Algorithm 6, where the above procedures are repeated iteratively. The BS first broadcasts Lagrange multipliers to the users. Each user updates its requested power as in (5.26), using local information P_n^{min}, P_n^{max}, P_n^*, A_n, L_n, and BS peak power $\bar{P}$. Each user then sends its requested power back to the BS, and the BS will updates the Lagrange multipliers as in (5.29). And so forth, until the optimal solution is obtained.

5.5 Simulation Results

We evaluate the proposed algorithms with MATLAB simulations, where the deterministic VBR traffic model and the optimization solution algorithms are implemented. We use a cellular network with 20 users;[1] the network topology is illustrated in Fig. 5.2. The downlink bandwidth is 1 MHz. The path gain averages are $G_n = d_n^{-4}$, where d_n is the physical distance from the BS to user n. The downlink channel is modeled as log-normal block fading with zero mean and variance 8 dB [50]. The processing gains are set to $L_n = 128$ for all n. The distance d_n is uniformly distributed in [100m, 1000m]. The device temperature is $T_0 = 290$ Kelvin and the equivalent noise bandwidth is $B_w = 1MHz$. The BS peak power constraints is set to $\bar{P} = 10$ Watts. We use three VBR movies traces, *Star Wars*, *NBC News*, and *Tokyo Olympics*, from the Video Trace Library maintained at Arizona State University [139]. We plot the sizes of the first 100 frames of the *NBC News* video sequence in Fig. 5.3, to illustrate the high variation of VBR video frame sizes, which makes it very challenging to develop accurate mathematical models. Each playout buffer is set to 1.5 times of the largest frame size in the requested VBR video.

[1]The number of users/links in the cellular network is chosen according to the resource specified in the simulation: bandwidth and the total BS power limit.

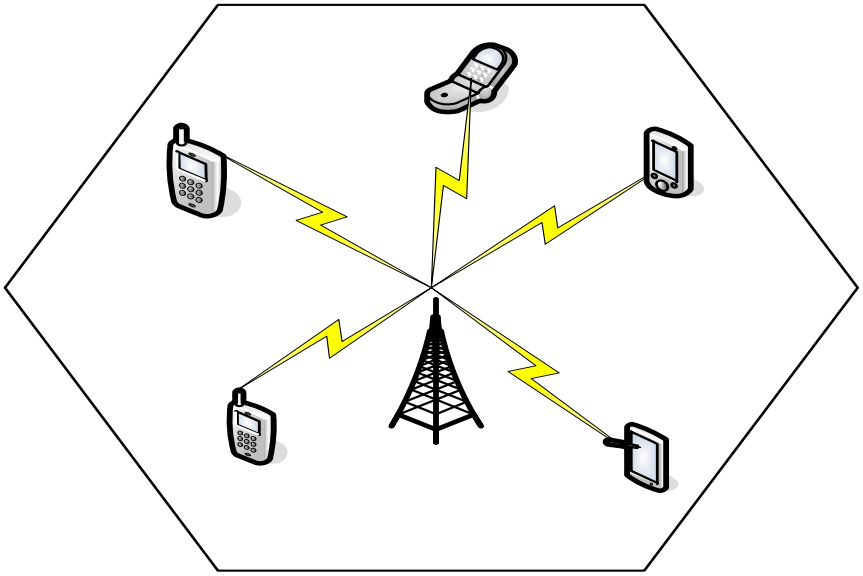


Figure 5.2: Topology of the cellular network.

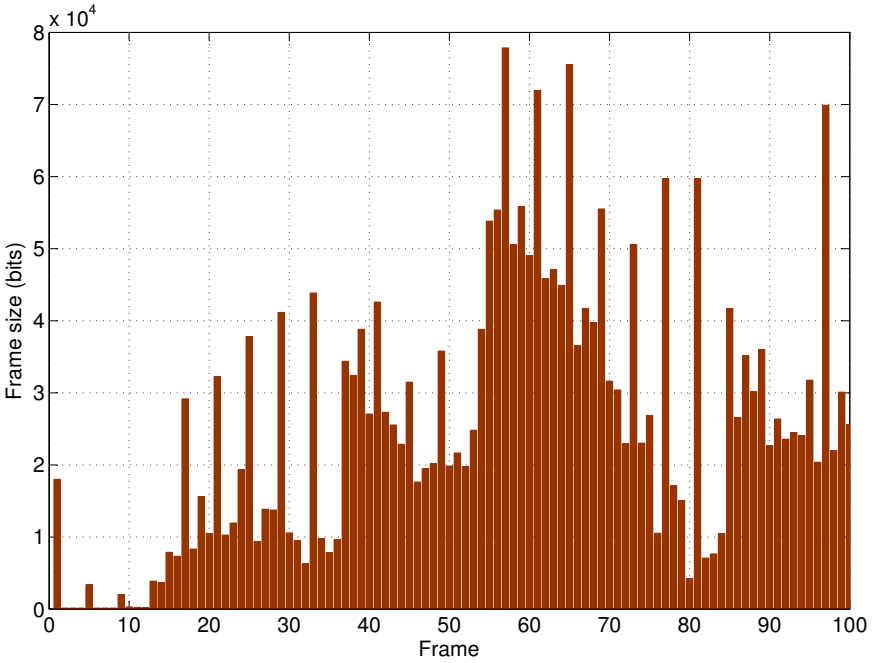


Figure 5.3: The sizes of the first 100 frames of the *NBC News* sequence.

In the simulations, we have 7 user streaming NBC news, 7 users streaming Star Wars, and 6 users streaming Tokyo Olympics. The proposed power allocation algorithm is executed at the beginning of each time slot. In Fig. 5.4, we plot the cumulative consumption, overflow and transmission curves for *NBC News* transmitted to user 2. The top sub-figure is the overview of $10,000$

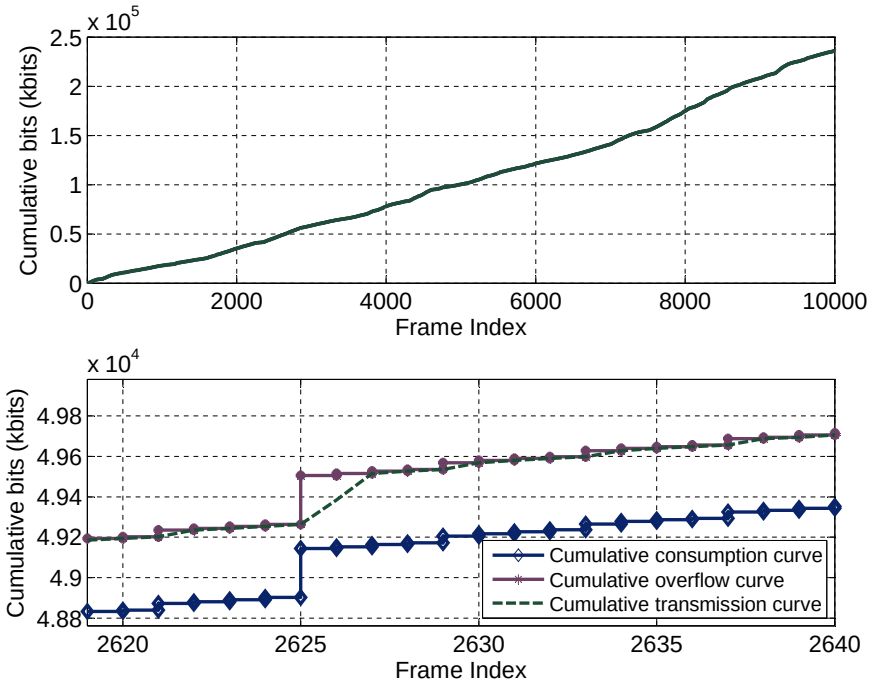


Figure 5.4: Transmission schedule for video *NBC News* to user 2.

frames. We also plot the curves from frame $2,620$ to $2,640$ in the bottom sub-figure. We observe that the cumulative transmission curve $X(t)$ is very close to the cumulative overflow curve $B(t)$, indicating that the algorithm always aim to maximize the transmission rate as allowed by the buffer and power constraints. The playout buffers are almost fully utilized most of the time. There is no playout buffer overflow and underflow for the entire range of $10,000$ frames. Among the *NBC News* frames, frame $2,625$ is the largest frame. We let seven out of the 20 links playout this largest frame simultaneously at time slot $2,625$ in the simulation. There is no buffer underflow under such heavy load.

In Fig. 5.5, we plot the power allocation and price updates for all the 20 links in one of the 10,000 time slots. The power and prices converges in around 70 steps. The converged power vector is $\vec{P}$ = [0.0022, 1.396, 0.0356, 0.0024, 1.396, 0.0351, 0.0016, 1.396, 0.0356, 0.0026, 1.396, 0.0356, 0.0023, 1.396, 0.0356, 0.0018, 1.396, 0.0356, 0.0034, 1.394] Watts. Note that with the distributed algorithm, the computation in each iteration only consisting updating power or price as in (5.26) and (5.29), which takes only a negligible amount of time. The 70-step convergence time is very small comparing to the power control in cellular standards (e.g., 1500 Hz for UMTS power

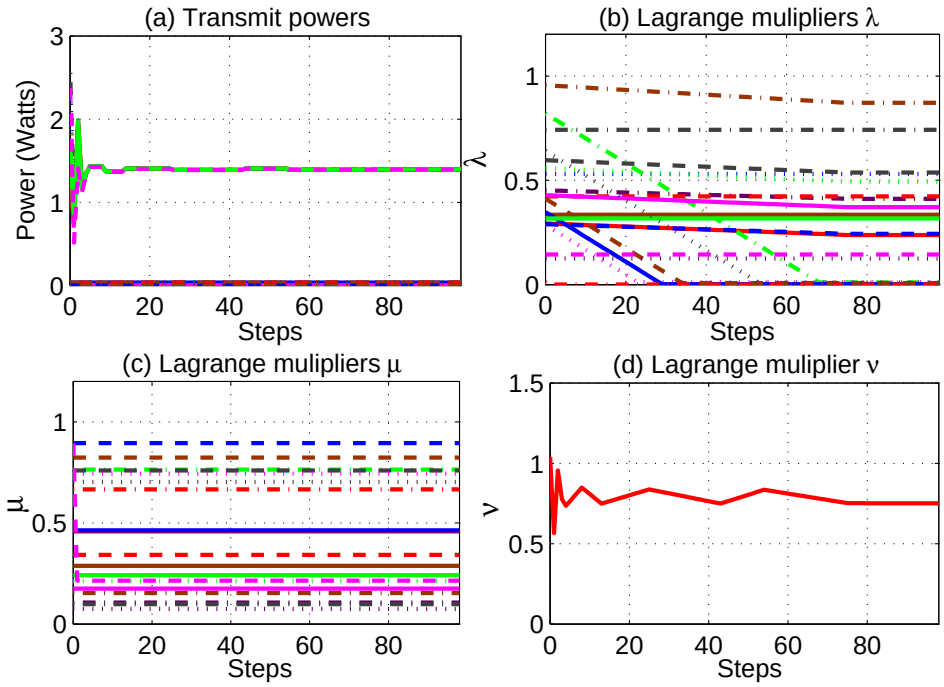


Figure 5.5: Convergence of power allocation and Lagrange multipliers.

control [140]). Since the gradient method is used, the convergence of the algorithm is dependent on the gradients, which further depend on the system parameters such as L_n and A_n. Another main factor for the convergence speed is the choice of the stepsize. As discussed, we use Armijo Rule to determine step size, in which the stepsize evolves according to the difference of the target values between steps.

Finally, we compare the proposed algorithm with a diversity-aware power allocation scheme, where the BS allocates power according to channel quality. With this scheme, the best channel n will be assigned power to achieve its maximum required power $P_n^{max}(t)$. Then the second best channel will be allocated power until its maximum required power is achieved, and so forth until all of $\bar{P}$ is allocated. In this simulation, we increase the number of users to 50 to stress the capacity of the cellular network, such that the system is close to saturate. The purpose is to show the performance of the algorithms under a nearly congested scenario, which is more interesting in performance analysis than an under-load scenario.

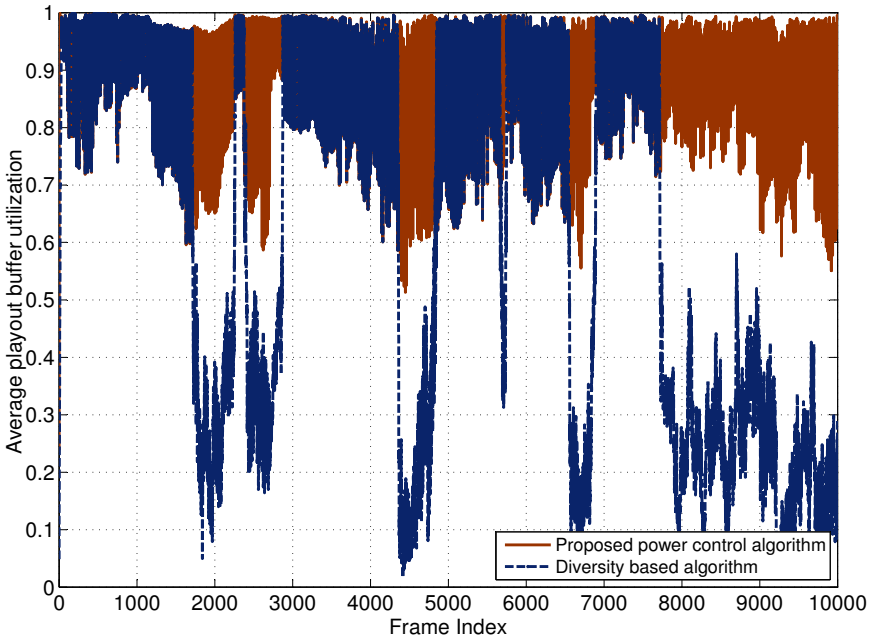


Figure 5.6: Average playout buffer utilization for the entire video sequence (10000 frames).

We compare the algorithms by their average playout buffer utilization. In Fig. 5.6, we plot the average buffer utilizations achieved by the proposed scheme and the diversity-aware scheme for the entire video sequence. A zoomed in version is presented in Fig. 5.7 from frames ranging from $2,000$ to $2,500$. It can be seen that the proposed algorithm consistently achieves high buffer utilization, ranging from 60% to 100%. The diversity scheme achieves buffer utilization lower than 50% for frames from $2,000$ to $2,250$. Such considerably higher buffer utilization translates to better video quality: there is no buffer overflow or underflow for proposed algorithm, while there is buffer underflow in 17% of the playout frames for the diversity scheme.

5.6 Related Work

Most of the prior work on VBR video streaming consider wired networks, which can be classified according to their traffic models, i.e., *statistical* or *deterministic* models. With the former approach, stochastic models are developed to capture the burstiness in VBR traffic. In [131, 132], the authors observed the *long-range-dependence* in VBR video traffic and modeled the autocorrelation with self-similar processes. This class of work provides valuable insights on the nature of

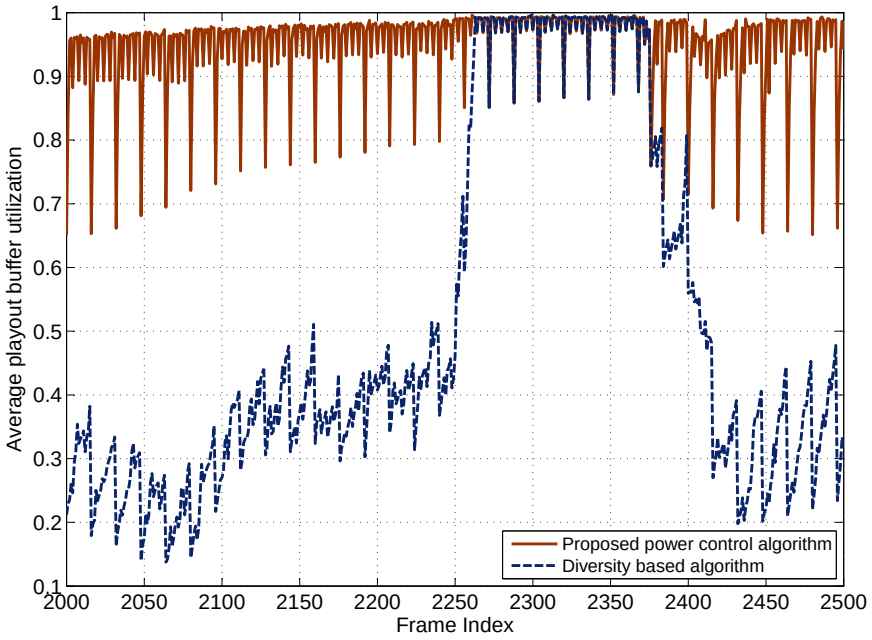


Figure 5.7: Average playout buffer utilization for frames 2000 to 2500.

VBR video traffic. The stochastic models can be incorporated in *quality of service* (QoS) mechanisms for VBR videos, and for traffic synthesizing in simulations [133].

With the deterministic approach, the piecewise-constant-rate transmission and transport (PCRTT) method was used, aiming to optimize one or more objectives while preserving continuous video playout. In [135], Liew and Chan proposed bandwidth allocation schemes for dynamically sharing a CBR channel among multiple VBR video streams, either i) to minimize the total receiver buffer size, or ii) to avoid underflow and overflow for a given playout buffer size. In [95], Salehi *et al.* considered smoothing VBR video over a CBR link and developed an effective algorithm to achieve the greatest smoothness in rate. In [141], McManus and Ross introduced a dynamic programming framework to set PCRTT rates and intervals to optimize different objective functions. These techniques do not directly apply to our problem of VBR over wireless networks, due to the fundamental difference between wireless and wired CBR links.

The downlink power allocation problem was studied in [50, 137], aiming to obtain the power allocation that maximizes a properly defined system utility. A distributed algorithm based on dynamic pricing and partial cooperation was proposed. Deng, Webera, and Ahrens [142] studied

the achievable maximum sum rate of multi-user interference channels. These papers provide the theoretical foundation and effective algorithms for utility maximization of downlink traffic, but the techniques used cannot be directly applied for VBR video over wireless networks with buffer and delay constraints.

In [54, 143], the authors studied the problem of one VBR stream over a given time-varying wireless channel. In [143], it was shown that the separation between a delay jitter buffer and a decoder buffer is in general suboptimal, and several critical system parameters were derived. In [54], the authors studied the frequency of jitters under both network and video system constraint and provided a framework for quantifying the trade-offs among several system parameters. In this chapter, we jointly consider power control in wireless networks, playout buffers, and video frame information, and address the more challenging problem of streaming multiple VBR videos, and present a cross-layer optimization approach that does not depend on any specific channel or video traffic models.

5.7 Conclusions

We developed a downlink power allocation model for streaming multiple VBR videos in a cellular network. The model considers interactions among downlink power control, channel interference, playout buffers, and VBR video traffic characteristics. The formulated problem aims at maximizing the total transmission rate under both peak power and playout buffer overflow/underflow constraints. We presented a two-step approach for solving the problem and a distributed algorithm based on the dual decomposition technique. Our simulation studies validated the efficacy of the proposed algorithms.

Chapter 6

Downlink Power Control for Variable Bit Rate Video over Multicell Wireless Networks

6.1 Introduction

In this chapter, we extend power control in *variable bit rate* (VBR) video streaming to multicell wireless networks scenario. We consider video streaming over a multicell wireless network, a wireless network architecture widely deployed all over the world. We consider the typical case of downlink video transmissions. For the multicell system, generally intra-cell interference can be effectively controlled with precise synchronization or the use of guard times. The capacities of the downlinks are mainly limited by the inter-cell interference due to simultaneous base station (BS) transmissions using the same channel. Therefore, effective downlink power control is necessary to support concurrent videos.

In this chapter, we presented a problem formulation that considers downlink power control, inter-cell interference, VBR video characteristics, and playout buffer requirements. The objective is to achieve high playout buffer utilization, under playout buffer underflow and overflow constraints and peak power constraint. This is a nonlinear nonconvex problem to which traditional convex optimization techniques [59] and low- or high- *Signal to Interference-plus-Noise Ratio* (SINR) approximations [59, 138] do not directly apply.

We first derive the condition of the existence of feasible power assignments, which can achieve downlink capacities to guarantee no buffer underflow and overflow. We then develop a centralized algorithm that can produce solutions with bounded optimality gap. Specifically, we use the Reformulation-Linearization Technique (RLT) to obtain a linear programming (LP) relaxation of the original problem. Solving this LP relaxation yields an upper bound to the original problem. Interestingly, since the constraints are preserved in the relaxation procedure, the upper-bounding

solution is also feasible to the original problem; the corresponding objective value with this solution provides a lower bound to the global optimum. The LP relaxation is then incorporated into the branch-and-bound framework to obtain a centralized algorithm, which can produce a solution within the (1-ϵ) range of the global optimal.

To simplify computation and control, we also develop a distributed algorithm based on distributed constrained power control (DCPC) [48], where each BS iteratively updates transmit power based on feedback of measured SINR at the target receiver. It is shown that with DCPC, the power vector converges to a unique power vector that can achieve the goal of maximizing playout buffer utilization and avoiding playout buffer underflow and overflow. We evaluate the proposed algorithms with simulations using VBR video traces [139] and fading channels. The distributed algorithm is shown to achieve a performance very close to that of the centralized algorithm. Both algorithms are demonstrated to be highly effective for streaming VBR videos over multicell wireless networks.

In the reminder of this chapter, we present the problem formulation in Section 6.2. We describe a centralized algorithm in Section 6.3 and a distributed algorithm in Section 6.4. Simulation results are presented in Section 6.5 and related work is discussed in Section 6.6. Section 6.7 concludes this chapter. The notation used in this chapter are summarized in Table 6.1.

6.2 Problem Statement

6.2.1 Network and Video System Model

We consider the downlinks of an M-cell wireless network as shown in Fig. 6.1. In each cell, a BS streams video to mobile users in the cell, each allocated with a downlink channel. A channel is a spectral resource slot, the nature of which depends on the specific multiple access technique adopted for the multicell network. Without loss of generality, we assume that the downlink channels within a cell are orthogonal (e.g., due to perfect synchronization of spreading codes or use of guard times). The main interference at a user stems from the concurrent downlink transmissions

Table 6.1: Notation Table for Chapter 6

Symbol	*Description*
M	total number of cells (or, BS's)
$\mathcal{U}$	set of users sharing the same channel
L_i	total number of frames for user i video
b_i	playout buffer size of user i
$D_i(t)$	cumulative consumption curve at user i
$X_i(t)$	cumulative transmission curve at user i
$B_i(t)$	cumulative overflow curve at user i
$P_m(t)$	transmit power of BS m in time slot t
$\vec{P}(t)$	BS transmit power vector in time slot t
$\bar{P}$	peak power constraint for the BS's
$\vec{P}^*$	optimal power vector to the LP relaxation
G_k^m	path gain from BS k to user un_m
B_w	channel bandwidth
τ	duration of a time slot
η_m	noise power at user un_m
C_m	capacity of the cell m downlink
$C_m^{min}(t)$	min. rate for user un_m without underflow
$\bar{C}_m^{min}(t)$	the largest value of $C_m^{min}(t)$
$\gamma_m(t)$	SINR at user un_m
$\gamma_m^{min}(t)$	minimum SINR corresponding to $C_m^{min}(t)$
$\bar{\gamma}_m^{min}(t)$	SINR corresponding to $\bar{C}_m^{min}(t)$
$\gamma_m^{max}(t)$	max. SINR for user un_m without overflow
γ_m^{th}	receiver sensitivity at user un_m
$\mathbf{A}$	matrix of path gain ratios defined in (D.5)
$\mathbf{\Gamma}^{min}$	defined as diag$\{\gamma_1^{min}(t), \gamma_2^{min}(t), \cdots, \gamma_M^{min}(t)\}$
$\bar{\mathbf{\Gamma}}^{min}$	defined as diag$\{\bar{\gamma}_1^{min}(t), \bar{\gamma}_2^{min}(t), \cdots, bar\gamma_M^{min}(t)\}$
$\mathbf{\Delta}$	$M \times M$ matrix defined as $(\bar{\Gamma}^{min} - \Gamma^{min})$
γ_m^{tar}	target user un_m SINR for distributed alg.
$\bar{\mathbf{\Gamma}}^{tar}$	defined as diag$\{\gamma_1^{tar}(t), \gamma_2^{tar}(t), \cdots, \gamma_M^{tar}(t)$
ν_m	vector of elements η_m/G_m^m
u_m	RLT substitution variable for logarithm terms
v_{mk}	RLT substitution variable for quadratic terms
α, β	parameters for the distributed algorithm

in neighboring cells that use the same channel. There is a need for the BS's to adopt power control to mitigate such inter-cell interference.

We consider the problem of streaming multiple VBR videos in the multicell network. We assume the wired segment of a video session path is reliable with sufficient bandwidth, while the last-hop wireless link is the bottleneck [144]. Thus the corresponding video data is always

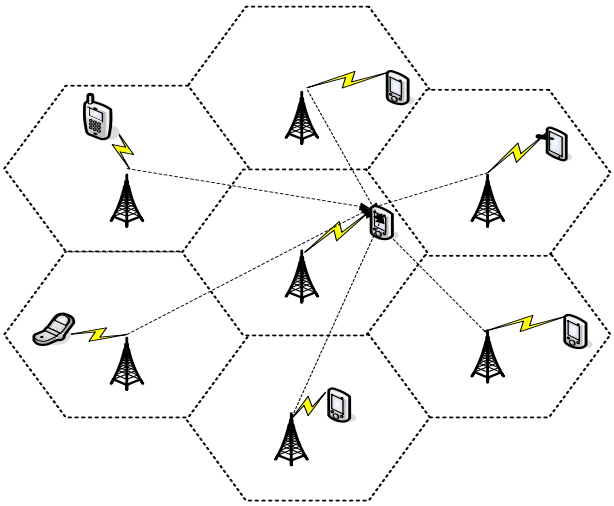


Figure 6.1: A multicell wireless network with concurrent VBR video sessions. The inter-cell interference experienced by the central cell user is illustrated.

available at the BS before the scheduled transmission time. We adopt the *deterministic VBR video model* in 4.2.

6.2.2 Problem Formation

For the multicell wireless video network, consider a specific channel and let $\mathcal{U} = \{un_1, un_2, \cdots, un_M\}$ denote the set of users sharing the channel, where un_m is the user in cell m.[1] Let the BS *transmit power vector* be $\vec{P}(t) = [P_1(t), P_2(t), \cdots, P_M(t)]^T$ in time slot t. The capacity of the downlink from BS m to user un_m, denoted as $C_m(t)$, depends on the SINR at un_m, which can be written as

$$\gamma_m(\vec{P}(t)) = \frac{G_m^m P_m(t)}{\sum_{k \neq m} G_k^m P_k(t) + \eta_m}, \tag{6.1}$$

where G_k^m is the path gain from BS k to user un_m and η_m is the noise power at un_m. We assume slow-fading channels such that the path gains do not change within each time slot [50], but vary over different time slots following a certain distribution. The downlink capacity $C_m(t)$ also depends on the channel bandwidth B_w and the transceiver design, such as modulation and channel

[1] 0-1 index variables can be used to model the case where no user uses the channel in some cells, but are omitted for brevity.

coding. Without loss of generality, we use the upper bound as predicted by Shannon theorem:

$$C_m(\vec{P}(t)) = B_w \log\left(1 + \gamma_m(\vec{P}(t))\right). \tag{6.2}$$

The impact of fading channels is incorporated in the SINR in (6.2). For practical systems, the achievable capacity may be a fraction of $C_m(\vec{P}(t))$, but this part is omitted for brevity.

Once the link capacity is determined, $C_m(t)\tau$ video bits will be delivered to user un_m in that time slot. The cumulative transmission curve $X_m(t)$ can be written as

$$X_m(0) = 0; \;\; X_m(t) = X_m(t-1) + C_m(t)\tau. \tag{6.3}$$

Assume peak power constraint $0 \leq P_m \leq \bar{P}$, for all m. The problem is to determine the transmit power vector $\vec{P}(t)$, for $0 < t \leq \max_i\{L_i\}$, such that the resulting cumulative transmission curves satisfy

$$D_m(t) \leq X_m(t) \leq B_m(t), \text{ for all } m, t, \tag{6.4}$$

i.e., without causing playout buffer underflow or overflow. Since the video frames have variable sizes and the video sessions have random phases, large frames from different sessions are less likely to occur in the same time slot. Jointly considering power control for the downlinks is, in some sense, analogous to statistical multiplexing of VBR video flows.

From (6.2)$\sim$(6.4), the feasible SINR range at user un_m is

$$e^{\frac{\max\{0, D_m(t) - X_m(t-1)\}}{B_w\tau}} - 1 \leq \gamma_m \leq e^{\frac{B_m(t) - X_m(t-1)}{B_w\tau}} - 1. \tag{6.5}$$

In (6.5), the lower bound is the SINR that just empties the buffer without causing underflow. The upper bound is the SINR that just fills up the buffer without causing overflow.

Generally, the feasible transmit power vector $\vec{P}(t)$ is not unique for a given set of VBR video sessions. Among the set of feasible solutions, a schedule that transmits more data is more desirable since it provides a larger search space for optimizing transmit power vectors for future time

slots. Omitting the constant B_w, we formulate the optimal power control problem for VBR videos, termed problem OPT-VBR, as

$$\textbf{maximize} \quad \sum_{m \in \mathcal{U}} \log(1 + \gamma_m(t)) \tag{6.6}$$

$$\textbf{subject to:} \quad \gamma_m(t) = \frac{G_m^m P_m(t)}{\sum_{k \neq m} G_k^m P_k(t) + \eta_m}, \ \forall\, m \tag{6.7}$$

$$\gamma_m^{min}(t) \leq \gamma_m(t) \leq \gamma_m^{max}(t), \ \forall\, m \tag{6.8}$$

$$0 \leq P_m \leq \bar{P}, \ \forall\, m, \tag{6.9}$$

where $\gamma_m^{max}(t)$ is the upper bound in (6.5) and $\gamma_m^{min}(t)$ is the larger one between the lower bound in (6.5) and γ_m^{th}, a minimum SINR requirement imposed by the transceiver design.

In problem OPT-VBR, the total amount of video data delivered in time slot t is maximized, under playout buffer underflow and overflow constraints and peak transmit power constraints. This is a nonlinear nonconvex problem, to which traditional convex optimization techniques do not directly apply. Furthermore, to achieve the objective of avoiding playout buffer underflow and overflow, the SINRs may assume values ranging from very low to very high. Thus the existing high SINR approximation [59] and low SINR approximation [138] techniques cannot be used. In the following, we first prove the existence of feasible solutions. We then derive effective centralized and distributed algorithms to solve problem OPT-VBR in Sections 6.3 and 6.4.

6.2.3 Existence of Feasible Solutions

Due to the wide range of VBR video frame sizes, the corresponding SINR requirements also assume a wide range of values. Under conditions where many video sessions coincidently transmit their large frames in the same time slot, problem OPT-VBR may not have a feasible power assignment to deliver all the frames. In this section, we derive the conditions for the existence of feasible power assignments. We assume a centralized scheduler in the multicell network, which has prior knowledge of all the path gains and the cumulative consumption and overflow curves.

We define the *minimum required rate* for user un_m in time slot t, denoted as $C_m^{min}(t)$, as the bit rate such that the playout buffer is just emptied, but without underflow, at the end of time slot t. We have the following result for $C_m^{min}(t)$.

Lemma 6.1. *The largest value for the minimum required rate $C_m^{min}(t)$ is $\bar{C}_m^{min}(t) = [D_m(t) - D_m(t-1)]/\tau$.*

Proof. See Appendix D.1. □

We have the following condition for the existence of a feasible power assignment for problem OPT-VBR.

Theorem 6.1. *There exits a feasible power assignment for problem OPT-VBR for time slot t, if there exits a feasible power assignment that can achieve the rate vector $\left[\bar{C}_1^{min}(t), \bar{C}_2^{min}(t), \cdots, \bar{C}_M^{min}(t)\right]$.*

Proof. See Appendix D.2.

□

Theorem 6.1 allows us to evaluate, for a given set of videos, if there is a feasible power assignment for each time slot. There is no need to consider the transmission schedules and playout buffer occupancies in previous time slots. At the beginning of time slot t, we obtain $\bar{\gamma}_m^{min}(t)$ from the cumulative consumption curve $D(t)$ and channel gains. If the linear system (D.4) is solvable and the resulting $\vec{P}$ satisfies constraint (6.9), then there is a feasible power assignment for problem OPT-VBR for this time slot. The following fact from [51] can be used for the feasibility test.

Fact 6.1. *The following statements are equivalent: (i) there exits a feasible power assignment satisfying (D.4); (ii) the maximum modulus eigenvalue of $\left(\bar{\mathbf{\Gamma}}^{min}\mathbf{A}\right)$ is less than 1; (iii) the reciprocal matrix $(\mathbf{I} - \bar{\mathbf{\Gamma}}^{min}\mathbf{A})^{-1} = \sum_{k=0}^{\infty}\left(\bar{\mathbf{\Gamma}}^{min}\mathbf{A}\right)^k$ exists and is positive component-wise.*

6.2.4 Comparison with a Lazy Scheme

A "lazy" scheme is proposed in [136] for VBR video transmission over a wired network. This is an ON-OFF scheme and it transmits a video frame as late as possible before its playout deadline at the maximum link speed, which minimizes the required client buffer size. In multicell multi-user wireless VBR video streaming, the maximum link speed varies from time to time due to interference and channel fading. Thus, the original lazy scheme cannot be applied directly.

We enhance the lazy scheme to support multicell multi-user VBR video streaming, termed W-Lazy, where every BS transmits a frame that is needed for playout in the next time slot. Then we can determine the rate vector (and the transmit powers) as given in Theorem 6.1. We use W-Lazy as a benchmark for comparison and evaluation of the proposed algorithms. We have the following results for W-Lazy.

Corollary 6.1.1. *Problem OPT-VBR has a larger solution space than the W-Lazy scheme.*

Proof. This result directly follows Theorem 6.1. □

Corollary 6.1.2. *If $\vec{C}^*(t) = [C_1^*(t), ..., C_n^*(t)]$ is the solution to problem VBR-OPT, then any other vector $\vec{C}(t)$ that is element-wise smaller than $\vec{C}^*(t)$ has a smaller solution space.*

Proof. This result also follows a similar process as in the proof of Theorem 6.1. □

6.3 Centralized Algorithm

As discussed, problem OPT-VBR is a nonlinear nonconvex problem, to which traditional convex optimization techniques do not directly apply. In this section, we present a centralized algorithm to provide solutions with bounded optimality gap. We first use RLT to obtain a linear programming (LP) relaxation of problem OPT-VBR [145]. We then incorporate the linear relaxation into a branch-and-bound framework, which can produce (1-ϵ)-optimal solutions.

6.3.1 Reformulation and Linearization

We first apply *polyhedral outer approximation* for the logarithm functions in problem OPT-VBR to obtain a Polynomial Programming Problem OPT-VBR(p) [146]. We then use *RLT bound-factor product constraints* to relax the quadratic terms to obtain an LP relaxation OPT-VBR(l). The time slot index (t) is dropped in the following to simplify notation.

We first process the logarithm functions in the objective function. Letting $u_m = \log(1+\gamma_m)$, we obtain a linear objective function $\sum_{m\in\mathcal{U}} u_m$ and new constraints $u_m = \log(1+\gamma_m)$. We deal with the new constraints using polyhedral outer approximation. Since $\gamma_m^{min} \le \gamma_m \le \gamma_m^{max}$, we choose H points, denoted as $\{\gamma_m^h\}$, within this range as

$$\gamma_m^h = (1+\gamma_m^{min})\left(\frac{1+\gamma_m^{max}}{1+\gamma_m^{min}}\right)^{\frac{h}{H-1}} - 1, h = 0, \cdots, H-1, \tag{6.10}$$

where $\gamma_m^0 = \gamma_m^{min}$ and $\gamma_m^{H-1} = \gamma_m^{max}$. We can obtain a *convex envelop* for the logarithm function in [γ_m^{min}, γ_m^{max}], which consists of H tangent lines at the H points given in (6.10) and the line segment connecting the two end points. We relax the logarithm constraint by using its convex envelop, represented by the following new linear constraints:

$$\begin{cases} u_m \ge \frac{\log(1+\gamma_m^{min})}{\gamma_m^{max}-\gamma_m^{min}}(\gamma_m^{max}-\gamma_m) + \frac{\log(1+\gamma_m^{max})}{\gamma_m^{max}-\gamma_m^{min}}(\gamma_m-\gamma_m^{min}) \\ u_m \le \log(1+\gamma_m^h) + \frac{\gamma_m-\gamma_m^h}{1+\gamma_m^h},\ h = 0, 1, \cdots, H-1. \end{cases}$$

The first line is for the segment connecting the two end points, and the second line is for the tangent lines at the H points. A four-point approximation is illustrated in Fig. 6.2.

With the polyhedral outer approximation, we obtain a polynomial programming problem OPT-VBR(p), as given in (6.11) $\sim$ (6.18). We can rewrite the last constraint (6.18) as

$$\sum_{k\ne m} G_k^m \gamma_m P_k - G_m^m P_m + \eta_m \gamma_m = 0,$$

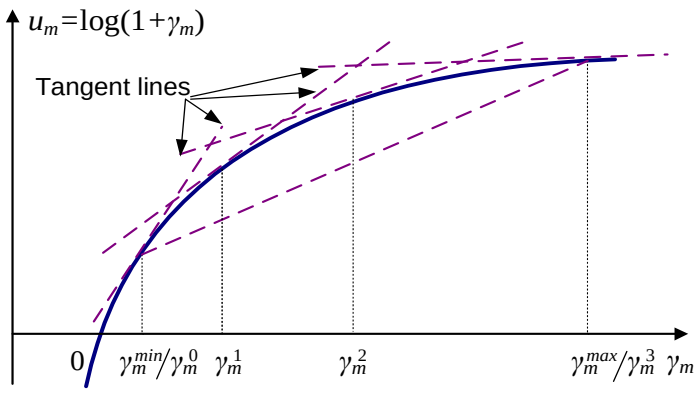


Figure 6.2: Four-point polyhedral outer approximation for $u_m = \log(1+\gamma_m)$, $1 < \gamma_m^{min} \leq \gamma_m \leq \gamma_m^{max}$.

$$\textbf{maximize} \sum_{m \in \mathcal{U}} u_m \tag{6.11}$$

subject to:

$$G_m^m P_m - \left(\sum_{k \neq m} G_k^m P_k + \eta_m \right) \gamma_m^{min} \geq 0, \ \forall\, m \tag{6.12}$$

$$G_m^m P_m - \left(\sum_{k \neq m} G_k^m P_k + \eta_m \right) \gamma_m^{max} \leq 0, \ \forall\, m \tag{6.13}$$

$$0 \leq P_m \leq \bar{P}, \ \forall\, m \tag{6.14}$$

$$u_m \geq \frac{\log\left(1+\gamma_m^{min}\right)}{\gamma_m^{max} - \gamma_m^{min}} (\gamma_m^{max} - \gamma_m) + \frac{\log\left(1+\gamma_m^{max}\right)}{\gamma_m^{max} - \gamma_m^{min}} (\gamma_m - \gamma_m^{min}), \ \forall\, m \tag{6.15}$$

$$u_m \leq \log(1+\gamma_m^h) + \frac{\gamma_m - \gamma_m^h}{1+\gamma_m^h}, \ \forall\, m, h \tag{6.16}$$

$$\gamma_m^h = (1+\gamma_m^{min}) \left(\frac{1+\gamma_m^{max}}{1+\gamma_m^{min}} \right)^{\frac{h}{H-1}} - 1, \forall\, m, h \tag{6.17}$$

$$\gamma_m = \frac{G_m^m P_m}{\sum_{k \neq m} G_k^m P_k + \eta_m}, \ \forall\, m. \tag{6.18}$$

---

which contains quadratic terms in the form of $\gamma_m P_k$. We next introduce RLT bound-factor product constraints to remove such terms and to obtain an LP relaxation.

Define substitution variables $v_{mk} = \gamma_m P_k$, for all m, k. Since γ_m and P_k are bounded by their respective lower and upper bounds as $\gamma_m^{min} \leq \gamma_m \leq \gamma_m^{max}$ and $0 \leq P_k \leq \bar{P}$, we obtain the

following RLT bound-factor product constraints

$$\begin{cases} (\gamma_m - \gamma_m^{min}) \cdot (P_k - 0) \geq 0 \\ (\gamma_m^{max} - \gamma_m) \cdot (P_k - 0) \geq 0 \\ (\gamma_m - \gamma_m^{min}) \cdot (\bar{P} - P_k) \geq 0 \\ (\gamma_m^{max} - \gamma_m) \cdot (\bar{P} - P_k) \geq 0. \end{cases}$$

Substituting $\gamma_m P_k = v_{mk}$, we obtain the following four linear constraints for v_{mk}:

$$\begin{cases} v_{mk} - \gamma_m^{min} P_k \geq 0 \\ \gamma_m^{max} P_k - v_{mk} \geq 0 \\ \gamma_m \bar{P} - v_{mk} - \gamma_m^{min} \bar{P} + \gamma_m^{min} P_k \geq 0 \\ \gamma_m^{max} \bar{P} - \gamma_m^{max} P_k - \gamma_m \bar{P} + v_{mk} \geq 0. \end{cases}$$

The quadratic terms $P_k \gamma_m$ are thus replaced with v_{mk} with the above linear RLT bound-factor constraints, and an LP relaxation OPT-VBR(l) is obtained as given in (6.19) $\sim$ (6.30).

The LP relaxation OPT-VBR(l) can be effectively solved with an LP solver in polynomial time. The optimal solution to the LP relaxation consists of $\{\vec{P}^*, \vec{u}^*, \vec{\gamma}^*, \vec{v}^*\}$. It is worth noting that during the reformulation and linearization procedure, we mainly relax the logarithm function in the objective function of OPT-VBR. The original constraints of OPT-VBR are preserved in OPT-VBR(l). Therefore, we have the following theorem regarding the feasibility of the solution, which greatly simplifies the *local search* procedure of the branch-and-bound algorithm to be presented in Section 6.3.2.

Theorem 6.2. *The optimal transmit power vector $\vec{P}^*$ to the LP relaxation OPT-VBR(l) is a feasible solution to the original problem OPT-VBR.*

6.3.2 Branch-and-Bound Algorithm

According to Theorem 6.2, we can substitute the optimal power assignment $\vec{P}^*$ for the LP relaxation into problem OPT-VBR to obtain a lower bound, while the LP solution itself provides

maximize $\sum_{m \in \mathcal{U}} u_m$ (6.19)

subject to:

$$G_m^m P_m - \left(\sum_{k \neq m} G_k^m P_k + \eta_m \right) \gamma_m^{min} \geq 0, \ \forall \ m \tag{6.20}$$

$$G_m^m P_m - \left(\sum_{k \neq m} G_k^m P_k + \eta_m \right) \gamma_m^{max} \leq 0, \ \forall \ m \tag{6.21}$$

$$0 \leq P_m \leq \bar{P}, \ \forall \ m \tag{6.22}$$

$$u_m \geq \frac{\log\left(1+\gamma_m^{min}\right)}{\gamma_m^{max}-\gamma_m^{min}}(\gamma_m^{max}-\gamma_m) + \frac{\log\left(1+\gamma_m^{max}\right)}{\gamma_m^{max}-\gamma_m^{min}}(\gamma_m-\gamma_m^{min}), \ \forall \ m \tag{6.23}$$

$$u_m \leq \log(1+\gamma_m^h) + \frac{\gamma_m - \gamma_m^h}{1+\gamma_m^h}, \ \forall \ m, h \tag{6.24}$$

$$\gamma_m^h = (1+\gamma_m^{min}) \left(\frac{1+\gamma_m^{max}}{1+\gamma_m^{min}} \right)^{\frac{h}{H-1}} - 1, \forall \ m, h \tag{6.25}$$

$$v_{mk} - \gamma_m^{min} P_k \geq 0, \ \forall \ m, k \neq m \tag{6.26}$$

$$(\gamma_m - \gamma_m^{min})\bar{P} - v_{mk} + \gamma_m^{min} P_k \geq 0, \ \forall \ m, k \neq m \tag{6.27}$$

$$\gamma_m^{max} P_k - v_{mk} \geq 0, \ \forall \ m, k \neq m \tag{6.28}$$

$$(\gamma_m^{max} - \gamma_m)\bar{P} - \gamma_m^{max} P_k + v_{mk} \geq 0, \ \forall \ m, k \neq m \tag{6.29}$$

$$\sum_{k \neq m} v_{mk} G_k^m - G_m^m P_m + \eta_m \gamma_m = 0, \ \forall \ m. \tag{6.30}$$

an upper bound. We next incorporate the LP relaxation into a branch-and-bound framework to obtain an algorithm that can produce (1-ϵ)-optimal solutions.

Branch-and-bound is an iterative method for solving optimization problems, especially for discrete and combinatorial problems. A branch-and-bound procedure has two key components. The first one, called *branching*, is to partition a problem into subproblems. The procedure is repeated recursively to each of the subproblems and all produced subproblems naturally form a tree structure, i.e., the *branch-and-bound tree*. Its nodes are the constructed subproblems. The leaves of the tree is also call the *Problem List*. The other component is *bounding*, which is a fast way of finding upper and lower bounds for the optimal solution for each subproblem. For a maximization problem, an infeasible upper bound (UB) can be found by solving a relaxed problem. A *local*

search algorithm is then used to explore the neighborhood, to find a feasible lower-bounding solution (LB). As discussed, we can easily derive upper and lower bounds by solving the LP relaxation (no need for local search). The core of the approach is an observation that, for a maximization task, if the upper bound for a subproblem l_1 is smaller than the lower bound for any other subproblem l_2, then l_1 and the branch rooted at l_1 can be safely discarded from the tree, such that the computational complexity can be reduced. This procedure is called *pruning*.

The algorithm terminates when the upper bound reaches $(1+\epsilon)$ of the lower bound. Let the optimal object value be $O \leq UB$, we have $LB \geq \frac{1}{1+\epsilon}UB \geq \frac{1}{1+\epsilon}O = (1-\epsilon+\epsilon^2-\epsilon^3+\cdots)O \approx (1-\epsilon)O$, for $0 \leq \epsilon \ll 1$. The pseudo code for the branch-and-bound algorithm is given in Algorithm 7.

6.3.3 Enhancement

In this section, we further introduce a heuristic to accelerate the convergence of the branch-and-bound algorithm. At the beginning of time slot t, if the playout buffer occupancy is above a certain threshold, say, 80%, and $X_m(t-1) \geq D_m(t)$ at user m, we set $P_m(t) = 0$ and remove the link from the optimization process.

Generally the playout buffer size should at least be greater than the largest frame size. Given the large variations in VBR frame sizes, there could be multiple frames stored when the buffer is close to full. When the above conditions are satisfied, there is little chance of buffer underflow at the end of time slot t even if we do not transmit anything to user m. On the other hand, if we schedule a non-zero power $P_m(t)$ for this link, only a small amount of bits can be transmitted due to the buffer overflow constraint, but at the cost of reduced SINRs at all other links. Excluding such links from transmission not only greatly speeds up the convergence of the branch-and-bound algorithm, but also increases the SINR and capacity of other active links.

Algorithm 7: Branch-and-Bound Algorithm

Initialization ;
Obtain LP relaxation OPT-VBR(l) as Prob 1 ;
Set optimal solution $sol = \phi$, Problem list $\mathcal{S} = \{\text{Prob 1}\}, UB = \infty$, and $LB = 0$;
Solve Prob 1 for solution $\{\vec{P}', \vec{u}', \vec{\gamma}', \vec{v}'\}$ and upper bound UB_1 ;
Use $\vec{P}'$, (6.6), and (6.7) to get lower bound LB_1 ;
Set $UB = UB_1$ and $LB = LB_1$;
Iteration & pruning ;
Select Prob l with the largest UB_l in $\mathcal{S}$ and set $UB = UB_l$;
if $LB_l > LB$ **then**
    Set $sol = \vec{P}_l'$ and $LB = LB_l$;
    if $UB \leq (1+\epsilon)LB$ **then**
        stop with solution sol ;
    else
        remove all probs k in $\mathcal{S}$ with $UB_k \leq (1+\epsilon)LB$;
    end
end
Partition ;
For Prob l, find the maximum relaxation error among all RLT variables, e.g., $\max_{m,k}\{|\gamma_m P_k - v_{mk}|\}$;
Evaluate the following condition: $(\gamma_m^{max} - \gamma_m^{min}) \cdot \min\{\gamma_m' - \gamma_m^{min}, \gamma_m^{max} - \gamma_m'\} \geq (P_m^{max} - P_m^{min}) \cdot \min\{P_m' - P_m^{min}, P_m^{max} - P_m'\}$;
if *true* **then**
    partition $[\gamma_m^{min}, \gamma_m^{max}]$ into $[\gamma_m^{min}, \gamma_m']$ and $[\gamma_m', \gamma_m^{max}]$;
else
    partition $[P_m^{min}, P_m^{max}]$ into $[P_m^{min}, P_m']$ and $[P_m', P_m^{max}]$;
end
Bounding ;
Solve the partitioned probs l_1 and l_2 to get solutions sol_{l_1}, sol_{l_2} and bounds UB_{l_1}, UB_{l_2}, LB_{l_1}, LB_{l_2} ;
Remove Prob l from $\mathcal{S}$;
if $(1+\epsilon)LB < UB_{l_1}$ **then**
    add Prob l_1 into $\mathcal{S}$;
end
if $(1+\epsilon)LB < UB_{l_2}$ **then**
    add Prob l_2 into $\mathcal{S}$;
end
if $\mathcal{S} = \phi$ **then**
    stop ;
else
    go to Step 8 ;
end

6.4 Distributed Algorithm

Although the RLT-based branch-and-bound algorithm can provide a $(1-\epsilon)$-optimal solution, it requires a centralized implementation. A centralized controller is needed to collect network, link and video related information, and to update transmit power for each downlink. In this section, we develop a distributed algorithm for problem OPT-VBR that can be implemented in each BS and operate with local information.

We assume each BS obtains video cumulative consumption curves and playout buffer sizes for its users during the video session initiation phase. At the beginning of time slot t, each BS m computes for user un_m the minimum rate as $[D_m(t) - X_m(t-1)]/\tau$, i.e., the data rate that empties the playout buffer at the end of time slot t but without underflow, and the maximum rate as $[B_m(t) - X_m(t-1)]/\tau$, i.e., the data rate that makes the playout buffer full at the end of time slot t but without overflow. BS m then translates the minimum and maximum rates to minimum and maximum SINRs, i.e., $\gamma_m^{min}(t)$ and $\gamma_m^{max}(t)$ as given in (6.5). In the following, we again drop the time slot index (t) to simplify notation.

To maximize objective function (6.6), BS m sets a target SINR as $\gamma_m^{tar} = \gamma_m^{max}$, and tries to achieve the target SINR by adjusting its transmit power. The problem then becomes a *Distributed Constrained Power Control* (DCPC) problem [48]. BS m first randomly sets its initial transmit power as $0 < P_m^0 \leq \bar{P}$. Let γ_m^i be the i-th SINR measurement at user un_m, which is fed back to BS m. BS m then uses the following DCPC algorithm to update its power after receiving the i-th SINR feedback:

$$P_m^i = \min\left\{\bar{P}, \frac{\gamma_m^{tar}}{\gamma_m^i} P_m^{i-1}\right\},\ i = 1, 2, \cdots . \tag{6.31}$$

If the γ_m^{tar}'s are feasible (see Section 6.2.3), the power vector series $\{\vec{P}^0, \vec{P}^1, \cdots, \vec{P}^i, \cdots\}$ is proved to converge to a unique positive power vector satisfying the following equation [48]

$$\vec{P} = \min\left\{\vec{\bar{P}}, \mathbf{\Gamma}^{tar}(\mathbf{A}\vec{P} + \vec{\nu})\right\}, \tag{6.32}$$

Algorithm 8: DCPC Algorithm

1 Initialization ;
2 BS m obtains b_m, D_m, and B_m for user un_m ;
3 BS m computes SINR bounds γ_m^{max} and γ_m^{min} ;
4 BS m sets $\gamma_m^{tar} = \gamma_m^{max}$ and $P_m(0) \in (0, \bar{P}]$;
5 Iteration ;
6 BS m receives SINR feedback γ_m^i and updates its power as: $P_m^i = \min\left\{\bar{P}, (\gamma_m^{tar}/\gamma_m^i)\, P_m^{i-1}\right\}$;
7 **if** *($P_m^i = \bar{P}$ for β iterations) & ($\gamma_m^i \neq \gamma_m^{tar}$)* **then**
8 reset the target SINR as: $\gamma_m^{tar} = \gamma_m^{min} + \alpha \cdot (\gamma_m^{tar} - \gamma_m^{min})$;
9 **end**
10 $i = i + 1$ and go to Step 6 ;

where $\mathbf{\Gamma}^{tar} = \text{diag}\{\vec{\gamma}^{tar}\} = \text{diag}\{\gamma_1^{tar}, \gamma_2^{tar}, \cdots, \gamma_M^{tar}\}$. Furthermore, the converged power vector $\vec{P}^*(t)$ also achieves the target SINR $\gamma_m^{tar}(t)$ for each BS m. The convergence result is summarized as the following fact from [48].

Fact 6.2. *With the DCPC algorithm (6.31), the transmit power vector converges to a unique positive power vector $\vec{P}^*$ satisfying (6.32). After convergence, either $\vec{P}^*$ achieves $\vec{\gamma}^{tar}$ or at least one of the components in $\vec{P}^*$ is equal to $\bar{P}$.*

The pseudo code for the distributed DCPC algorithm is given in Algorithm 8, where α is a fraction in (0,1) and β is a positive integer. If BS m's transmit power remains at the maximum power $\bar{P}$ for β iterations, while the target SINR γ_m^{tar} is still not achieved, we reset the target SINR as $\gamma_m^{tar} = \gamma_m^{min} + \alpha \cdot (\gamma_m^{tar} - \gamma_m^{min})$ and restart the iterative update process. We choose $\alpha = 0.618$, the reciprocal of the *golden ratio*, and β from 2 to 5 in our simulations.

In practice, the path gains vary over time due to channel fading. It is possible that during some time slot, the transmission is not feasible even for the minimum required rate. It is nontrivial to test the feasibility of the target SINR vector $\vec{\gamma}^{tar}$ in a distributed manner with only local information. In fact, if the target SINR vector is infeasible, the problem of finding the largest set of links that can be supported at the given SINRs is proved to be NP-Complete [147]. Therefore, we adopt the following heuristic strategies to handle the case when the target SINR vector cannot be achieved by a feasible power assignment due to deep fading channels.

i) In the first time slot, if the DCPC algorithm does not converge in a certain number of steps, suspend the transmission of the video with the largest frame size for sometime and retry the algorithm.

ii) Adopt the acceleration enhancement as in the centralized algorithm, which is described in Section 6.3.3.

iii) If the DCPC algorithm does not converge for the reduced γ_m^{tar} (see Line 5 in Algorithm 8), further reduce the target SINR as $\gamma_m^{tar} = \gamma_m^{min} + \alpha \cdot (\gamma_m^{tar} - \gamma_m^{min})$. If still no convergence when $\gamma_m^{tar} = (1+\epsilon) \cdot \gamma_m^{min}$, for $0 < \epsilon \ll 1$, all the links whose buffer will not be empty in the next time slot will pause their transmissions. Since the algorithm always tries to transmit as more data as possible (i.e., by setting a high target SINR whenever possible), it is highly likely that such links won't have buffer underflow in the following time slots.

iv) If all the above steps fail, the BS suspends its transmission and the user freezes the playout precess until the next time slot.

6.5 Simulation Results

To evaluate the performance of the proposed algorithms, we simulate streaming VBR videos in a 7-cell wireless network. We assume the channels within a cell are orthogonal and inter-cell interference is the major limiting factor. The channel bandwidth is $B_w = 1$ MHz. The path gain averages are set to $\bar{G}_k^m = d_{km}^{-4}$, where d_{km} is the physical distance from BS k to user un_m. We assume Rayleigh fading channels in all the simulations, where the normalized path gain is exponentially distributed as $f(G_k^m) = \exp\{-G_k^m/\bar{G}_k^m\}$ for $G_k^m \geq 0$. The distance from a user to its corresponding BS is uniformly distributed from 100 m to 1000 m and the inter-cell BS distance is from 1600 m to 2000 m. The temperature is $T_0 = 290$ Kelvin and the equivalent noise bandwidth is also 1 MHz. The peak power constraint is $\bar{P} = 1$ Watt.

In each cell, the channel is dedicated to one mobile user for VBR video streaming. We assume BS's 1, 4 and 7 are streaming movie *Star Wars*, BS's 2 and 5 are streaming *NBC News*, and the

remaining links 3 and 6 are transmitting *Tokyo Olympics*. We use the VBR traces for these videos from the Video Trace Library hosted at Arizona State University [139] in all the simulations. The playout buffer size is set to be 1.5 times of the largest frame size in the requested VBR video.

6.5.1 Centralized Algorithm

We implement the branch-and-bound centralized algorithm using MATLAB. We choose $\epsilon = 10\%$ for the simulations. From the VBR video traces, we derive the cumulative consumption and overflow curves. The centralized algorithm computes the optimized power assignment for the BS's at beginning of each time slot. In Fig. 6.5.1, we plot the cumulative consumption, overflow and transmission curves for *Star Wars* transmitted on link 1. The top subfigure is for 10,000 frames. We also plot the curves from frame 1,960 to frame 1,980 in the bottom subfigure, while frame 1,969 has the largest size among the 10,000 frames. We observe that the cumulative transmission curve $X_1(t)$ is very close to the cumulative overflow curve $B_1(t)$, indicating that the centralized algorithm always aims to maximize the transmission rate as allowed by the buffer and power constraints, and the playout buffer is fully utilized for most of the time. There is no playout buffer overflow or underflow for the entire range of the movies.

In Fig. 6.6, we plot the upper and lower bounds for objective function (6.6) for time slot 1. This is the hardest time slot with respect to power control, since all the sessions are transmitting I-frames and all the playout buffers are empty in this time slot in our simulations. We observe the optimality gap between UB and LB is continuously decreased until the $\epsilon = 0.1$ threshold is reached. In other time slots where the frame sizes are not consistently large and the playout buffers are close to full, it usually takes only a few (e.g., 5 or 6) iterations to reach the optimality gap threshold.

We also evaluate the accelerated scheme under the same video and network conditions. The curves for link 1 are plotted in Fig. 6.5.1. It can be seen that during time slots 1,963, 1,967, and 1,971, there is no transmission on link 1 since the playout buffer is over 80% full. Pausing transmission in these time slots makes it easier for other links to transmit large frames and speeds

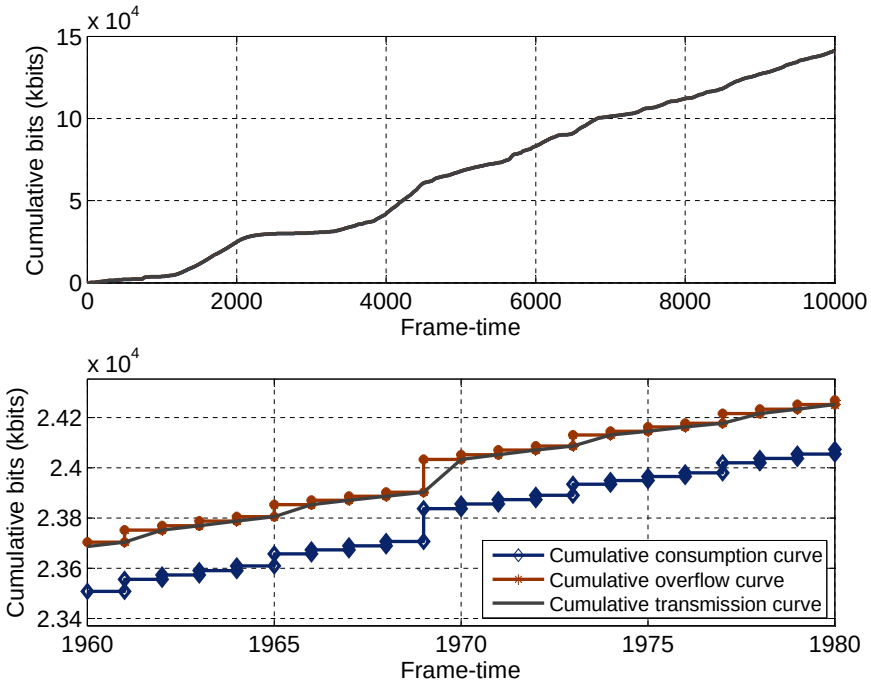


Figure 6.3: The cumulative overflow, transmission, and consumption curves when transmitting *Star Wars* at link 1 with centralized algorithm in the seven-cell network.

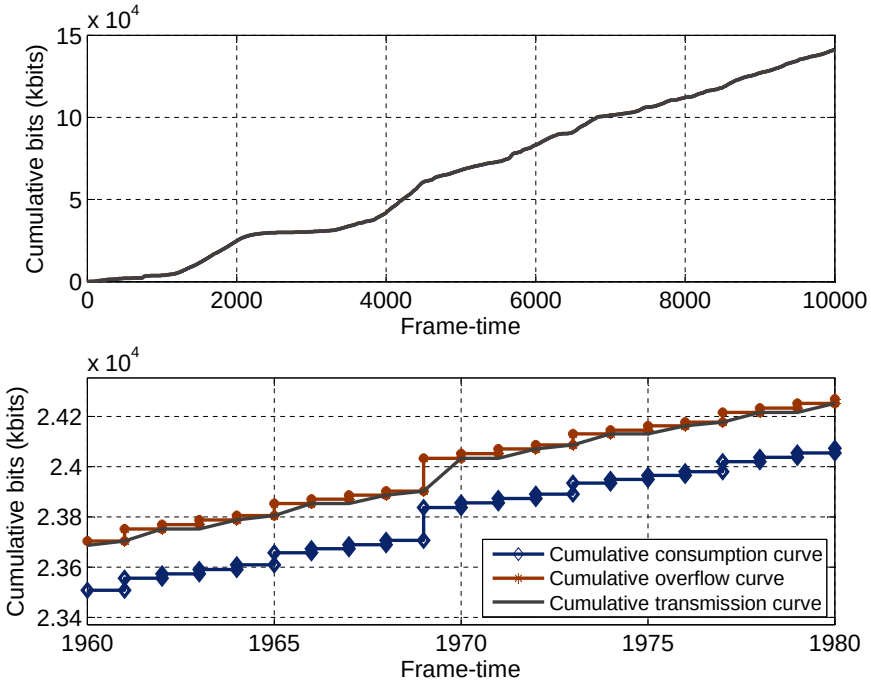


Figure 6.4: The cumulative overflow, transmission, and consumption curves when transmitting *Star Wars* at link 1 with Accelerated centralized algorithm in the seven-cell network.

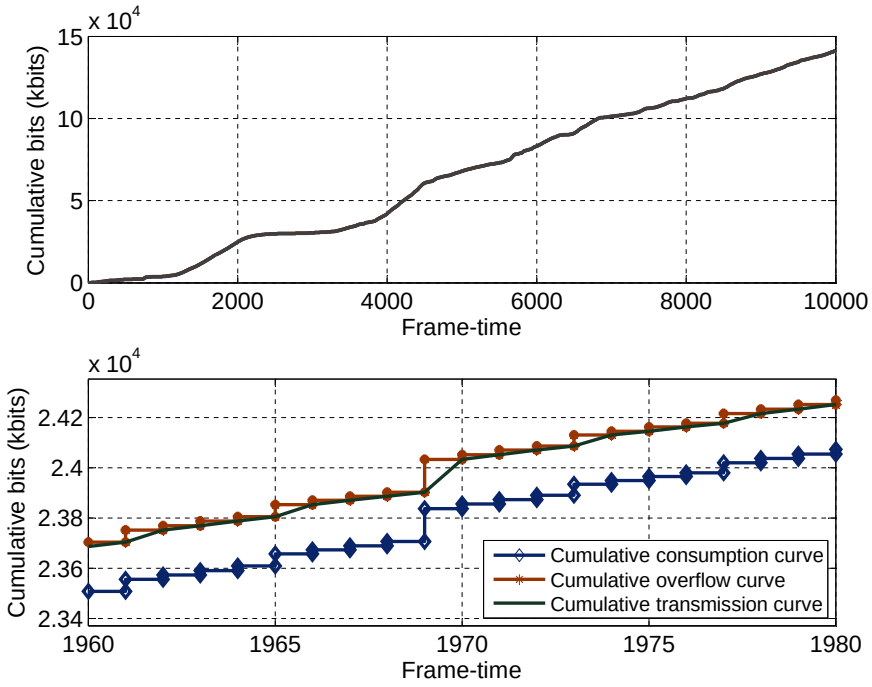


Figure 6.5: The cumulative overflow, transmission, and consumption curves when transmitting *Star Wars* at link 1 with DCPC in the seven-cell network.

up the convergence of the algorithm, while causing no buffer underflow at link 1. Since usually large frames rarely occur in the same time slot (except for time slot 1), this is analogous to statistical multiplexing of VBR videos. We find in the simulation, a link can pause in over 60% of the time slots with the acceleration heuristic, resulting in significant reduction in computation time.

6.5.2 Distributed Algorithm

We next examine the performance of DCPC. The network and video setups are the same as those in the centralized algorithm simulations. The cumulative overflow, transmission, and consumption curves obtained by DCPC are plotted in Fig. 6.5.1 for *Star Wars* transmitted on link 1. We observe very similar performance as in the case of the centralized algorithm shown in Fig. 6.5.1. The cumulative transmission curve is again very close to $B_m(t)$, and there is neither buffer overflow nor underflow during the transmission of 10,000 frames.

To compare the distributed and centralized algorithms, we compute the sum of the bit rates of all the links in each time slot. The acceleration scheme is not used for both algorithms in this

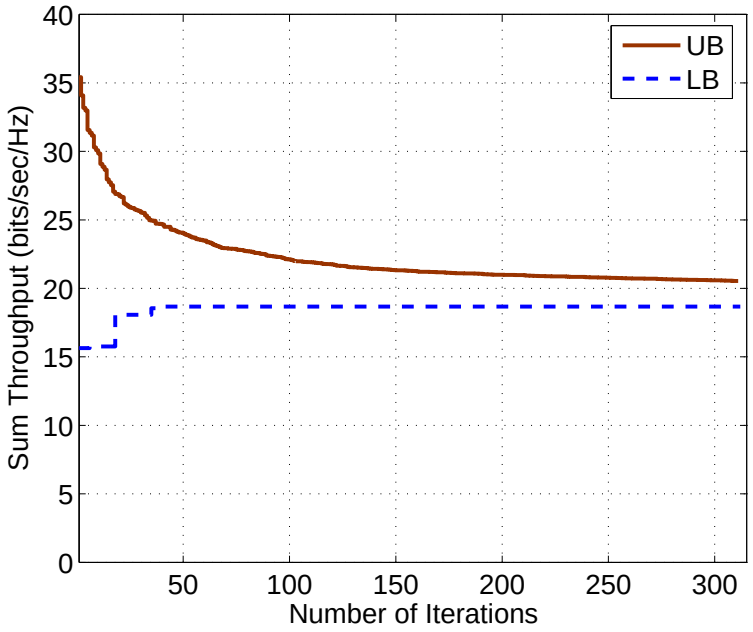


Figure 6.6: Convergence of the branch-and-bound algorithm in time slot 1 when all the I-frames are transmitted and all the buffers are empty (i.e., the worst case scenario).

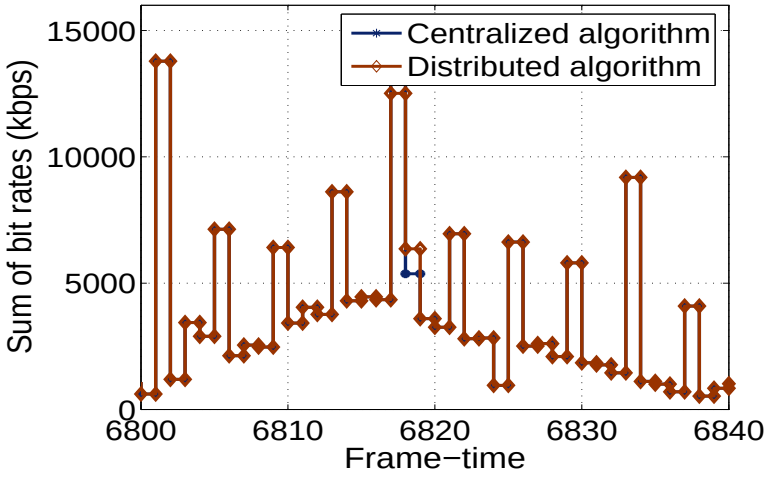


Figure 6.7: Rate sums with the algorithms with a seven-link network.

simulation. The rate sums are plotted in Fig. 6.5.2 from time slot 6,800 to 6,840. We observe that the sum rates achieved by the centralized algorithm and that by the distributed algorithm are identical for most part of this interval. Examining the rate sums for the entire 10,000 time slots, we find that the rate sum achieved by the DCPC algorithm is within 99% of the corresponding rate sum achieved by the centralized algorithm in over 97% of the time slots.

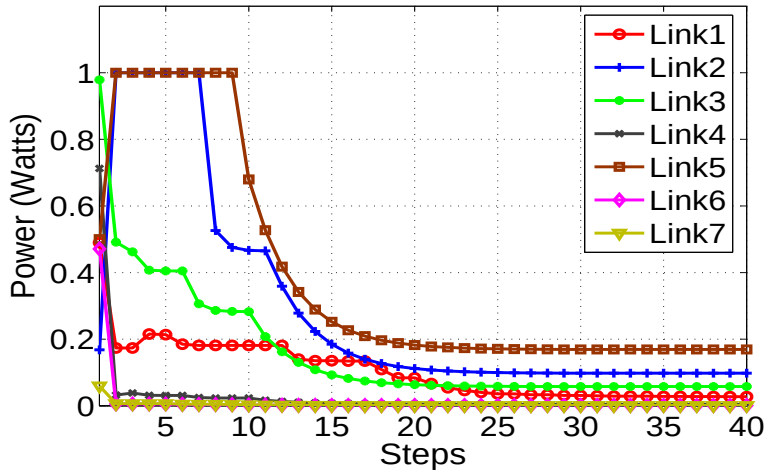


Figure 6.8: Convergence of transmit powers of DCPC with a seven-link network.

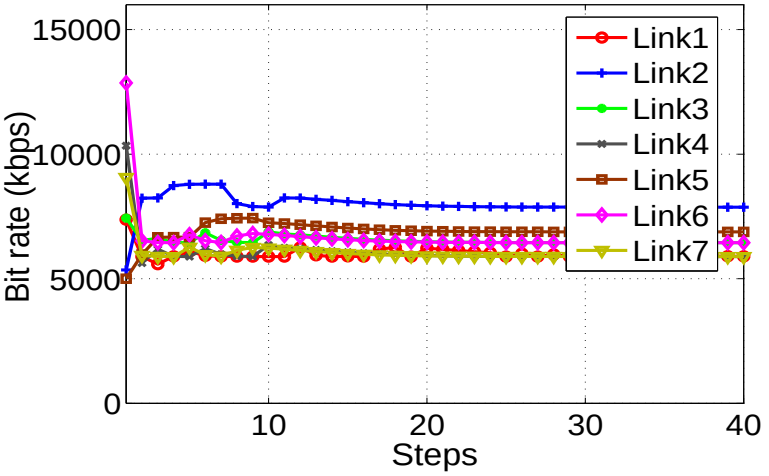


Figure 6.9: Convergence of bit rates of DCPC with a seven-link network.

The convergence of the distributed DCPC algorithm is plotted in Figs. 6.5.2 and 6.5.2 for one of the time slots. The accelerated scheme is incorporated with DCPC, such that a link m may pause its transmission if its buffer is over 80% full and $X_m(t-1) > D_m(t)$. The evolution of the BS transmit powers are plotted in Fig. 6.5.2, where after 23 steps, all the transmit powers converges to a value between 0 and $\bar{P} = 1$ Watt. The converged power vector is $\vec{P}^*$ = [0.0023, 0.208, 0.185, 0.0013, 0.163, 7.1×10^{-04}, 0.188] Watt. The evolution of the bit rates are plotted in Fig. 6.5.2. It is interesting to see the data rates converge faster than the transmit powers in this case. All the data rates reach stable values after a few steps.

6.5.3 Empirical Performance Evaluation

We evaluate the performance of the proposed schemes by comparing them with the following two schemes.

- A round-robin scheme where the BS allocates power in a *quality of service* (QoS) based round-robin fashion, which favors the session that would suffer buffer starvation if no transmission is scheduled. When a specific BS is selected for transmission, it transmits the video with maximum power without overflowing the client buffer, and all its neighbors remain silent in the same frame-time slot.

- W-Lazy, as described in Section 6.2.3.

First, we investigate the average buffer utilization at end of each time slot. When underflow happens, the missing frame is discarded, and the next frame will be scheduled for the transmission in the next time slot. We observe that the proposed RLT and DCPC schemes achieve higher average buffer utilization than the other two schemes. Fig. 6.10 shows the average buffer utilization from frame-time slot $1,600$ to $1,700$. We find that the buffer utilization of RLT and DCPC fluctuate around 90% mostly, while the utilization of the Round-robin scheme is in the range of 50% to 80%. We also find that the W-Lazy scheme always achieves a zero buffer utilization,since it only transmits each frame as late as possible in each time slot. At the end of a time slot, all the data will be comsumed by the user and the buffer is left empty.

We then compare the average number of underflow events in Table 6.2. we find RLT achieves underflow free transmission, while the number of underflow events for DCPC is negligible in the simulations. This is because both schemes aim to transmit as much video data as possible under the feasible condition in each frame-time slot. The extra video data transmitted will be in the playout buffer to provide a cushion to future large frames or network dynamics. On the other hand, both Round-robin and W-Lazy suffers a large number of underflow events. We also illustrate the buffer underflow events in the period from $1,680$ to $1,700$ in Fig. 6.11. The red dot circles indicate the buffer underflow. It can be seen that the cumulative transmission curve lies below the cumulative

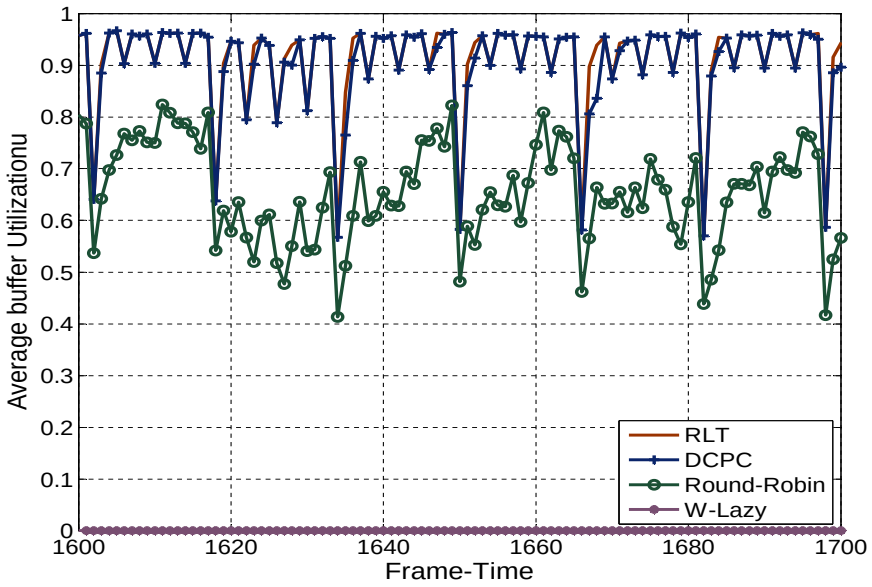


Figure 6.10: Average buffer utilization of the four schemes.

Table 6.2: Number of underflow events

	RLT	DCPC	Round-robin	W-Lazy
Mean	0	0.2	516	1076
Conf. Int.	$[0, 0]$	$[-0.355, 0.755]$	$[442, 591]$	$[514, 1637]$

consumption curve when buffer underflow events occur. This results in an infeasible transmission schedule, which causes frozen playout.

The average power consumption of the schemes are shown in Fig 6.12. W-Lazy has the lowest power consumption. Due to the variation of frame size and network condition, the transmission of W-Lazy are infeasible in many time slots. To prevent the divergence of power allocation, some video sessions should be paused and the power savings of W-Lazy are achieved by pausing video transmissions. However, this is at the cost of significantly more buffer underflow events, which are undesirable for user experience. The Round-robin scheme tries to transmit as much video data as possible. However, it chooses a session greedily and pauses other unselected video sessions. This also causes many underflow events for the unselected sessions. Also, due to the round robin fashion and limited buffer size, when the unselected session become selected, its low buffer utilization will lead to a larger power consumption in order to fill the buffer, especially when it misses the previous good channel condition and the channel condition is worse at the current time slot. Thus, the

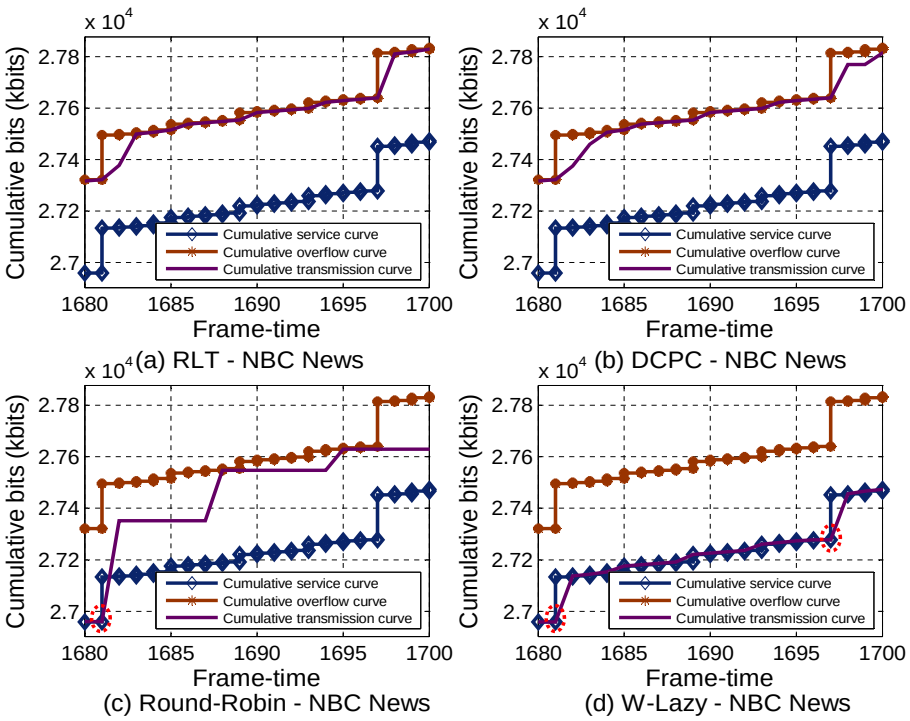


Figure 6.11: Illustration of underflow events.

proposed algorithms achieve the better balance between the energy consumption and the quality of experience of the video streaming.

6.6 Related Work

A thorough review in VBR video model and VBR video streaming over wired networks has been explored in 5.6. Due to the fundamental difference between wireless and wired CBR links, these techniques do not directly apply to our problem of VBR over the multicellular wireless networks. In this chapter, we take advantage of power control in wireless networks to adjust the capacity of wireless links based on video frame size information, such that we can jointly optimize the transmission of *multiple* VBR video sessions over *multiple* VBR channels. Our approach does not depend on any channel or video traffic models, and can be adopted for CBR video as well.

Power control is an important problem for interference-limited wireless networks. Most prior work focuses on maximizing network utility in the forms of SINR or bit rate [48, 59, 138]. In [48], Grandhi, Zander, and Yates presented centralized and distributed power control algorithms for

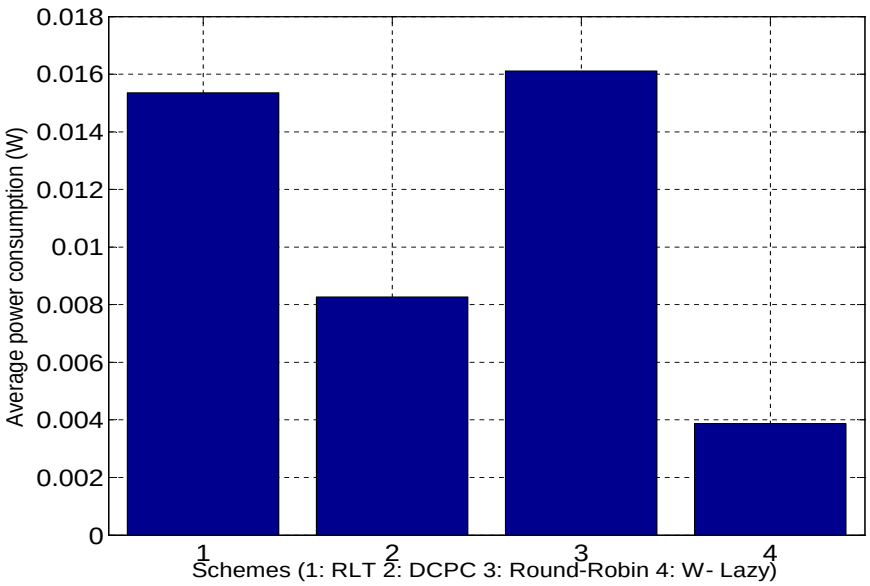


Figure 6.12: Average power consumption.

achieving target SINRs in a cellular network. In [59], Chiang studied the problem of joint power control and congestion control, aming to maximize the throughput of TCP-Vegas over an ad hoc network. Gjendemsj *et al.* [138] presented centralized binary power control algorithms for maximizing the sum rate over multiple interfering links. Although laid out the theoretical foundation and developed effective algorithms, these techniques cannot be directly applied for VBR video over multicell wireless networks with buffer and delay constraints.

6.7 Conclusions

We studied downlink power control for VBR video streaming in multicell wireless networks. The problem formulation considers downlink power control, inter-cell interference, VBR video characteristics, and playout buffer requirements. We developed a centralized algorithm that can provide (1-ϵ)-optimal solutions, and a fast distributed algorithm that only needs local information. The algorithms are evaluated with extensive simulations with VBR video traces and fading channels, and are demonstrated to be effective for streaming VBR videos over multicell wireless networks.

Chapter 7

Energy Efficiency on Downlink Multiuser VBR Video Streaming

7.1 Introduction

In this chapter, we present a power efficient downlink power control framework in wireless system with orthogonal channels for *variable bit rate* (VBR) streaming with focus on minimizing the power consumption for the total streaming period. We consider the problem of optimal power allocation for multiuser VBR video streaming in the downlink of a cellular network with orthogonal channels. We assume the wireline segment of a video session path is reliable with sufficient bandwidth, while the last-hop wireless link is the bottleneck. Thus the corresponding video data is always available at the BS before the scheduled transmission time. We adopt a deterministic model for VBR video traffic that incorporates video frame and playout buffer characteristics. The BS allocates a transmit power to each user in each time slot. The problem is to find the optimal power control schedule to stream the requested VBR video data to users, such that the total transmit power consumption can be minimized, while minimizing the buffer underflow and overflow events.

The problem is formulated as a constrained stochastic optimization problem. We show that the problem fits well with majorization theory, which concerns with partial ordering of real vectors and order-preserving functions. It answers the question of how to order vectors with nonnegative real components and its order-preserving functions [13]. A majorization-based solution framework is developed to tackle the problem. First, we prove that the objective function of the formulated problem is Schur-convex with the order-preserving property [13]. Second, we investigate the case of a single VBR video session with relaxed peak power constraint. We develop a majorization-based power optimal algorithm with low complexity, and prove the power optimality of the proposed

algorithm and the uniqueness of the global optimum. We also demonstrate that the proposed algorithm is smoothness optimal as well. Third, we investigate the case of multiuser VBR streaming, where power allocations for the users are coupled with the BS peak power constraint. We develop a heuristic algorithm that selectively suspends some video sessions, which will not incur underflow in the next time slot, when the peak power constraint is violated. Finally, the proposed algorithms are evaluated with trace-driven simulations [139], and are shown to achieve considerable power savings and improved video quality over a conventional "lazy" scheme [136]. The rest of this chapter is organized as follows. The system model and problem statement are presented in Section 7.2. We transform the problem into a majorization problem in Section 7.3. The proposed algorithms are described in Section 7.4 and simulation results are presented in Section 7.5. We review related work in Section 7.6. Section 7.7 concludes this chapter. The notation used in this chapter is summarized in Table 7.1.

7.2 System Model

7.2.1 Network and Video Source Model

We consider the downlink in a cellular network, as shown in Fig. 7.1. There are N active mobile users in a set $\mathcal{U} = \{1, 2, \cdots, N\}$ in the cell that subscribe to the video service. A BS transmits multiple VBR videos to the mobile users. Each user occupies a downlink channel, which is a spectral/time resource slot, the nature of which depends on the specific multiple access technique adopted. We assume that the downlink channels within a cell are orthogonal, due to perfect synchronization of the spreading codes or the use of guard times or frequencies. We further assume the wireline segment of a video session path is reliable with sufficient bandwidth, while the last-hop wireless link is the bottleneck. Thus the corresponding video data is always available at the BS before the scheduled transmission time. We adopt a *deterministic VBR video model*, which was presented in the Section 4.2.

The BS allocates a transmit power to each user in each time slot. Let $\vec{P}(t) = [P_1(t), \cdots, P_N(t)]$ be the power allocation in time slot t. The *Signal to Interference-plus-Noise Ratio* (SINR) at user

Table 7.1: Notation Table for Chapter 7

Symbol	*Description*
N	number of mobile users in the cell
$\mathcal{U}$	set of users
b_n	playout buffer size at user n
T_n	total number of frames of the user n video
Ω_n	total number of bits of the use n video
$D_n(t)$	cumulative consumption curve of user n
$B_n(t)$	cumulative overflow curve of user n
$X_n(t)$	cumulative transmission curve of user n
$P_n(t)$	transmit power of user n in time slot t
$\vec{P}(t)$	power allocation in time slot t
$\gamma_n(t)$	SINR at user n in time slot t
$G_n(t)$	path gain from BS to user n in time slot t
$\eta_n(t)$	noise power at user n in time slot t
$c_n(t)$	downlink data rate of user n in time slot t
B_w	channel bandwdith
κ	a transceiver dependent constant
$\bar{P}$	peak power constraint
$\vec{C}_n^i$	the i-th feasible transmission schedule
$\vec{C}_n^*$	the optimal solution to (7.9)
$\vec{C}_n^{opt}$	an evenly distributed rate vector
$\vec{C}_n^1$	an auxiliary schedule used in Theorem 7.2 proof
$\Phi(\cdot)$	a mapping function $\mathcal{R}^{T_n} \rightarrow \mathcal{R}$ defined in (7.12)
$\vec{X}, \vec{Y}, \vec{Z}$	n-dimensional nonnegative vectors
$C_{max}(t), C_{min}(t)$	rate of probe lines
$U(\vec{C})$	smoothness utility function

n in time slot t can be written as

$$\gamma_n(t) = G_n(t)P_n(t)/\eta_n(t), \tag{7.1}$$

where $G_n(t)$ is the path gain from BS to user n and $\eta_n(t)$ is the noise power at user n in time slot t. We assume block fading channels, where the $G_n(t)$'s are i.i.d. random variables with a certain distribution, for $t = 1, \cdots, T_n$ [50]. The downlink data rate can be written as $c_n(t) = B_w \log(1 + \kappa\gamma_n(P_n(t)))$, where B_w is the channel bandwidth and κ depends on the transceiver design, such as modulation and channel coding. Without loss of generality, we use the Shannon

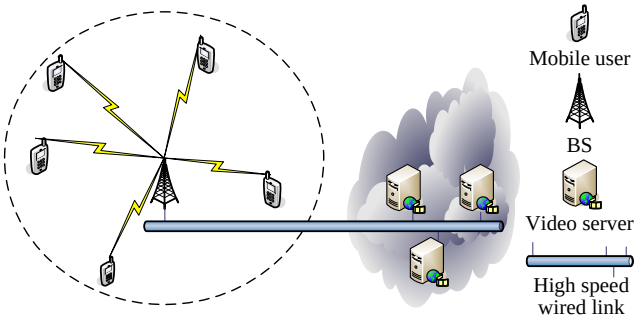


Figure 7.1: Cellular network and video streaming system model.

capacity as an upper bounding approximation:

$$c_n(t) = B_w \log\left(1 + \gamma_n(P_n(t))\right). \tag{7.2}$$

Once the link capacity is determined, $c_n(t)\tau$ bits of video data will be delivered to user n in that time slot. The cumulative transmission curve $X_n(t)$ can be written as

$$X_n(0) = 0; \quad X_n(t) = X_n(t-1) + c_n(t)\tau. \tag{7.3}$$

A feasible transmission schedule should cause neither playout buffer underflow nor overflow, i.e., satisfying

$$D_n(t) \leq X_n(t) \leq B_n(t), \text{ for all } t, n. \tag{7.4}$$

7.2.2 Power-Aware Transmission Scheduling

As discussed, we jointly consider the traffic source model in the application layer and power allocation in the physical layer. We adopt cross-layer design to compute the optimal feasible transmission schedule $\{X_n(t), 0 < t \leq T_n\}$, for all users $n \in \mathcal{U}$, such that the total transmit power consumption can be minimized. From (7.2), the required transmit power for user n is

$$P_n(t) = (2^{c_n(t)/B_w} - 1)\eta_n(t)/G_n(t). \tag{7.5}$$

A peak power constraint may be applied at the BS, i.e., $\sum_{n\in\mathcal{U}} P_n(t) \leq \bar{P}$, for all t. We then formulate the following *constrained stochastic optimization problem*, aiming to minimize the expected total transmit power.

$$\text{minimize } \sum_{n\in\mathcal{U}}\sum_{t=1}^{T_n} \mathbb{E}[P_n(t)] \tag{7.6}$$

$$\text{subject to: } D_n(t) \leq X_n(t) \leq B_n(t), \text{for all } n, t \tag{7.7}$$

$$\sum_{n\in\mathcal{U}} P_n(t) \leq \bar{P}, \text{for all } t. \tag{7.8}$$

Due to orthogonal channels, transmission in one channel does not interfere with those in other channels. We first relax the peak power constraint (7.8) (i.e., the case when $\bar{P}$ is large). Then, problem (7.6) can be decomposed into N sub-problems, each minimizing the transmit power of a video session.

$$\text{minimize } \sum_{t=1}^{T_n} \mathbb{E}[P_n(t)], \text{for all } n \in \mathcal{U} \tag{7.9}$$

$$\text{subject to: } D_n(t) \leq X_n(t) \leq B_n(t), \text{ for all } n, t. \tag{7.10}$$

For given $B_n(t)$ and $D_n(t)$, the feasible transmission schedule satisfying (7.10) is not unique. The i-th feasible transmission schedule is a piece-wise linear curve that can be represented as a vector $\vec{C}_n^i = [c_n^i(1), \cdots, c_n^i(T_n)]$, where $c_n^i(t) \geq 0$ is the data rate in time slot t, for all t. Let $\vec{C}_n^* = [c_n^*(1), \cdots, c_n^*(T_n)]$ be the optimal solution to (7.9). For a given VBR video, all the feasible transmission schedules transmit the same amount of video data, i.e.,

$$\sum_{t=1}^{T_n} c_n^i(t) = \sum_{t=1}^{T_n} c_n^*(t) = \Omega_n, \text{ for all } i, n. \tag{7.11}$$

Furthermore, the total transmit power for a feasible schedule can be viewed as a mapping function $\Phi : \mathcal{R}^{T_n} \rightarrow \mathcal{R}$ with

$$\Phi(\vec{C}_n^i) = \sum_{t=1}^{T_n} (2^{c_n^i(t)/B_w} - 1)\eta_n(t)/G_n(t). \tag{7.12}$$

Given such an interpretation of the relaxed problem (7.9), the objective is to find an optimal feasible vector $\vec{C}_n^*$, such that its total power P_n^*, obtained through the mapping $\Phi(\cdot)$, is the minimum among all feasible vectors $\vec{C}_n^i$. This interpretation fits well with the *majorization* theory introduced in Chapter 2, which provides useful order preserving results for inequality problems. Applying these results, we design an optimal algorithm for solving the decomposed sub-problem (7.9) in Section 7.4.1. Then we will examine the case of multiuser VBR video streaming coupled with the peak power constraint in Section 7.4.3.

7.3 Problem Reformulation

7.3.1 Majorization Preliminaries

The preliminaries of majorization have been given in Section 2.3 of Chapter 2.

7.3.2 Schur-convexity of Problem (7.9)

As discussed, problem (7.9) fits well with majorization theory with a mapping function (7.12). To solve the problem, we need to find the optimal rate vector $\vec{C}_n^*$ that is majorized by all other feasible transmission rate vectors $\vec{C}_n^i$, as $\vec{C}_n^* \prec \vec{C}_n^i$, for all i. If the mapping (7.12) is Schur-convex, then the total transmit power to achieve $\vec{C}_n^*$ will also be dominated by those of other feasible transmission rate vectors. That is, the minimum power is found for problem (7.9). Due to random path gains and noise powers, stochastic majorization (rather than ordinary majorization) should be used, which investigates the inequality properties related to random variables [13]. We have the following theorem for the mapping (7.12) in problem (7.9).

Theorem 7.1. *The objective function of (7.9) is an increasing Schur-convex function.*

Proof. See Appendix E.1. □

With Theorem 7.1, solving problem (7.9) is equivalent to finding the optimal rate vector $\vec{C}_n^*$, such that $\vec{C}_n^* \prec \vec{C}_n^i$, for all i. Then the total power associated with $\vec{C}_n^*$ is the minimum since the

mapping (7.12) is order-conserving. The feasible rate vector that is closest to equal distribution will be majorized by all other feasible rate vectors. Therefore, we transform problem (7.9) to finding a transmission schedule with the most evenly distributed rates for all the time slots.

7.4 Power Allocation Algorithms

Based on the stochastic majorization interpretation of problem (7.9) and the Schur-convex property of its objective function, we first develop a power minimization algorithm (PMA) for the case of relaxed peak power constraint. We prove its optimality and the uniqueness of the global optimal, as well as the equivalence of power optimal and smoothness optimal. We then describe a heuristic algorithm for the case of multiple videos coupled with the peak power constraint.

7.4.1 Power Minimization Algorithm

From Section 7.3, an evenly distributed rate vector $\vec{C}_n^{opt} = [\Omega_n/(T_n\tau), \cdots, \Omega_n/(T_n\tau)]$ is majorized by all feasible schedules, i.e., $\vec{C}_n^{opt} \prec \vec{C}_n^i$, for all i. However, due to the high variability of VBR video frames, limited playout buffer size, random path gains and noise powers, $\vec{C}_n^{opt}$ may not always be feasible. In general, each feasible schedule is piece-wise linear with a set of rate change points, where the rate is increased or decreased to prevent buffer underflow or overflow. $\vec{C}_n^{opt}$ is a special case with no such rate change points.

The algorithm in Algorithm 9, termed PMA, can generate a piece-wise linear schedule, while keeping each piece as long as possible and rate variation as small as possible. The operation of the algorithm is illustrated in Fig. 7.2. Starting from t_{start} (e.g., h_1 in Fig. 7.2), PMA first computes two probe lines:

- One through the starting point and one of the future corner points of $B_n(t)$, which can go the furthest into the future without causing buffer underflow or overflow (e.g., lines h_1h_2 in Case 1 and h_5h_6 in Case 2 of Fig. 7.2). The rate of this probe line is $C_{max}(t) = \frac{B_n(t)-X_n(t_{start})}{t-t_{start}}$.

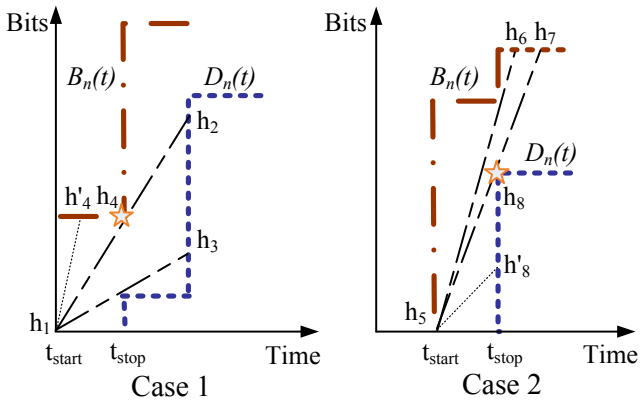


Figure 7.2: Two cases for determine the next transmission rate.

- The other through the starting point and one of the future corner points of $D_n(t)$, which can go the furthest into the future without causing buffer underflow or overflow (e.g., lines h_1h_3 in Case 1 and h_5h_7 in Case 2 of Fig. 7.2). The rate of this probe line is $C_{min}(t) = \frac{D_n(t)-X_n(t_{start})}{t-t_{start}}$.

All feasible transmission curves should lie in between these two probe lines in order to go that far. Furthermore, when the probe lines end, they hit *both* on either $B_n(t)$ or $D_n(t)$. Otherwise, we can always adjust one of the probe lines to make them go even further into the future. For example, see lines h_1h_3 and $h_1h'_4$ in Case 1 of Fig. 7.2. We can use line h_1h_2, which goes further into the future, to replace line $h_1h'_4$, and both probe lines hit $D_n(t)$ eventually (also see lines h_5h_6 and $h_5h'_8$ in Case 2).

If both probe lines hit $D_n(t)$ (i.e., Case 1 in Fig. 7.2), any feasible schedule for this interval will also hit $D_n(t)$, since they must lie in between the two probe lines. We then trace back the upper probe line (i.e., line h_1h_2) to find the latest time when the buffer is full (i.e., point h_4 at time t_{stop}). Then segment h_1h_4 will be chosen as the transmission schedule for this interval, with rate $\frac{B_n(t_{stop})-X_n(t_{start})}{t_{stop}-t_{start}}$.

If both probe lines hit $B_n(t)$ (i.e., Case 2 in Fig. 7.2), any feasible schedule for this interval will also hit $B_n(t)$. We then trace back the lower probe line (i.e., line h_5h_7) to find the latest time when the buffer is empty (i.e., point h_8 at time t_{stop}). Then segment h_5h_8 will be chosen as the transmission schedule for this interval, with rate $\frac{D_n(t_{stop})-X_n(t_{start})}{t_{stop}-t_{start}}$.

Algorithm 9: Power Minimization Algorithm (PMA-1)

1 BS obtains b_n, D_n, B_n, and B_w for all user n ;
2 Set $t = 1, t_{start} = 0, t_{stop} = t_{c_1} = t_{c_2} = 1, C_{min} = 0, C_{max} = \infty$;
3 **while** *some time slots are not assigned a rate* **do**
4 Calculate $C_{max}(t)$ and $C_{min}(t)$ over interval $[t_{start}, t]$;
5 **if** $C_{min} \leq C_{min}(t)$ & $C_{min}(t) \leq \min\{C_{max}, C_{max}(t)\}$ **then**
6 $C_{min} = C_{min}(t)$ and $t_{c_1} = t$;
7 **end**
8 **if** $C_{max} \geq C_{max}(t)$ & $C_{max}(t) \geq \max\{C_{min}, C_{min}(t)\}$ **then**
9 $C_{max} = C_{max}(t)$ and $t_{c_2} = t$;
10 **end**
11 **if** $C_{min} > \min\{C_{max}, C_{max}(t)\}$ **then**
12 Select C_{min} from t_{start} to $t_{stop} = t_{c_1}$;
13 **else if** $C_{max} < \max\{C_{min}, C_{min}(t)\}$ **then**
14 Select C_{max} from t_{start} to $t_{stop} = t_{c_2}$;
15 **else**
16 $t++$;
17 CONTINUE ;
18 **end**
19 $t_{start} = t_{stop}, t_{stop} = t_{c_1} = t_{c_2} = t_{start} + 1, t = t_{start} + 1, C_{min} = 0, C_{max} = \infty$;
20 **end**
21 **while** *more video frames to transmit* **do**
22 Measure the channel gain of the time slot, calculate power using (7.5), and transmit the video data ;
23 **end**

After the transmission schedule for $[t_{start}, t_{stop})$ is determined, we set $t_{start} = t_{stop}$ and repeat the above procedure to find the schedule for the next time interval. In Algorithm 9, the algorithm probes for the longest feasible rate starting from t_{start} in Steps 4–10. In Steps 11–14, the transmission rate for the interval $[t_{start}, t_{stop})$ is determined depending on which of the two cases it is as illustrated in Fig. 7.2. Steps 16–17 are for the case that the rate does not change in the time slot. Step 19 resets the variables to start the computation for the next segment of $X_n(t)$. Finally, Steps 21–23 transmit the frames following the computed schedule.

7.4.2 Optimality Proof

We next show that the algorithm given in Algorithm 9 computes the optimal solution to problem (7.9).

Theorem 7.2. *The power minimization algorithm PMA is optimal to problem (7.9).*

Corollary 7.2.1. *The power optimal transmission scheme $\vec{C}_n^*$ is unique for given $B_n(t)$ and $D_n(t)$.*

Corollary 7.2.2. *The computational complexity of Algorithm PMA is $\mathcal{O}(T_n^2)$.*

The proofs of above conclusions are similar to the proofs of in Chapter 2 and thus omitted for brevity.

Note that Algorithm PMA is executed during the session setup time. It only incurs a small initialization delay. In our simulations with VBR video traces, we find the execution time is usually negligible. When the channel statistics are changed (i.e., due to handoff), the schedule will be recomputed for the remaining video frames.

Corollary 7.2.3. *The power optimal transmission schedule $\vec{C}_n^*$ is also the smoothest one among all feasible schedules.*

Proof. The proof of Corollary 7.2.3 is given in Appendix E.2.

□

7.4.3 Multiuser Video Transmissions

We now consider problem (7.6) to compute transmission schedules for N VBR video sessions, which are coupled by the peak power constraint (7.8). Due to the peak power constraint and random channel gains, the individually calculated transmit powers may violate (7.8) in some time slots. The problem is further complicated because of the random channel gains, which is not available a priori (except for the statistics of the channels).

Algorithm 10: Power Minimization Algorithm for Multiuser Videos (PMA-m)

```
1  Execute power minimization algorithm PMA to compute transmission schedules for all
   active users ;
2  while there are more video frames to transmit do
3      Measure channel gains of the current time slot and calculate the transmit powers using
       (7.5) ;
4      if peak power constraint is violated then
5          Select the users who won't have underflow even without transmission in this time
           slot ;
6          Sort the selected users in decreasing order of powers ;
7          while peak power constraint is not satisfied do
8              Decrease the power of the selected users by the order ;
9          end
10     end
11     Transmit the videos and recalculate the optimal transmission scheme for the paused
       mobile users for the next time interval ;
12 end
```

To solve problem (7.6), we develop a heuristic algorithm, termed PMA-m, as presented in Algorithm 10. The PMA-m algorithm uses PMA to compute transmission schedules for all active users. Then based on current channel state information, it computes the power needed to achieved the rate for each user, and checks the peak power constraint $\sum_{n\in\mathcal{U}_n} P_n(t) \leq \bar{P}$. If the constraint is not violated, each user's video data will be transmitted at the computed power. Otherwise, as in Steps 4–10, PMA-m selects those users who will not have buffer underflow if their transmissions are suspended in the following time slot, and sort them in the deceasing order of their required powers. Starting with the first user, PMA-m decreases the powers of the users in the list; if the first user's power reaches 0 W but the peak power constraint is till not satisfied, PMA-m starts to reduce the power of the second user in the list; and so forth until the peak power constraint is satisfied.

In some extremely severe channel conditions, the total power $\bar{P}$ cannot even support the minimum required bit rate for all the users. Some users have to be paused and the current frames be discarded. The corresponding playout of such a user will be frozen until the next time slot. Finally, the transmission schedules for the suspended users will be recomputed using PMA as in Step 11 and the above procedure is repeated.

7.4.4 Application to Interactive Video Streaming

The ubiquitous spread of mobile devices and trend of multimedia applications require the interactive service be supported [148–150]. Thus, it is necessary to investigate how to apply the proposed schemes to the case of interactive video streaming. Interactive video is a relatively new and still evolving technology with a broad scope. We focus on the three interactive video streaming related typical scenarios in the following, and show that the proposed schemes are applicable for these scenarios for improved performance.

First, for quick response to user inputs, many interactive video streaming applications have *tight delay requirements*. Such stringent delay requirements have two implications: (i) unlike stored video, not all the future frame sizes are known now; only the frames sizes for a short *look-ahead period* (LAP) are known. (ii) the playout buffer sizes cannot be large, since a large buffer usually introduces large delay. Clearly, the proposed schemes can be applied to the look-ahead time period for which the future frame sizes are known, to compute a schedule for the near future. Furthermore, given the small playout buffer size, usually a chosen transmission rate won't last very long into the future before it hits either the cumulative overflow curve or the cumulative consumption curve (see Fig.4). Therefore the impact of limited look-ahead period would be small or moderate at best.

Second, many interactive video applications support *VCR controls* [151]. For example, a user may slide the progress bar of the video player to replay or skip a part of the video. This case is equivalent to a change in the cumulative overflow and consumption curves. The proposed algorithms will seek to the new start frame that the user requires and recalculate the transmission schedules for the following frames.

Third, in both *"exploratory"* online interactive videos (where a user can move through different locations in a space or view an object from different angles) and *video click throughs* [152] (where a user can click objects in the video that are linked to other contents), new data will be transmitted after each user input. These are equivalent to the case of VCR controls. A new set

Table 7.2: VBR Video Trace Statistics

Video Trace	*Average Rate*	*Frame Rate*	*Average PSNR*
Star Wars	331,681 b/s	30 f/s	44.62 dB
NBC News	784,840 b/s	30 f/s	38.80 dB
Tokyo Olympics	509,191 b/s	30 f/s	41.46 dB
Terminator 2	5,085,453 b/s	30 f/s	43.92 dB
From Mars to China	4,849,711 b/s	30 f/s	39.26 dB
Sony Demo	5,803,650 b/s	30 f/s	44.07 dB

of cumulative overflow and consumption curves will be delivered (derived from the new data requested) and new schedules computed.

In Section 7.5.1, we evaluate the performance of the proposed schemes under the above interactive video streaming scenarios. Our simulation results show that the proposed schemes still achieve considerable power savings and better video quality than a conventional scheme for interactive videos.

7.5 Performance Evaluation

We demonstrate the performance of the proposed optimal power control algorithm through trace-driven simulations. We simulate the downlink of a cell with 1 mile radius. The channels are assumed to be orthogonal, each with $B_w = 1$ MHz bandwidth. We assume that bit errors can be corrected by error correction codes. The path gain averages are $\bar{G}_n = d_n^{-4}$, where d_n is the physical distance from the BS to user n. We assume log-normal fading with zero mean and 8 dB standard deviation. The device temperature is 290 Kelvin and the equivalent noise bandwidth is $B_w = 1$ MHz. The BS streams three movies *Star Wars*, *NBC News*, and *Tokyo Olympics* to active users. The video traces are obtained from the Video Trace Library at Arizona State University [139]. The statistics of the three video traces are summarized in Table 7.2.

We first investigate the performance of the power optimal algorithm. In the simulation, the BS streams $3,000$ frames of a video sequence to each mobile user located at different distances to the BS. The cumulative consumption, overflow and transmission curves of the *Star wars* video session

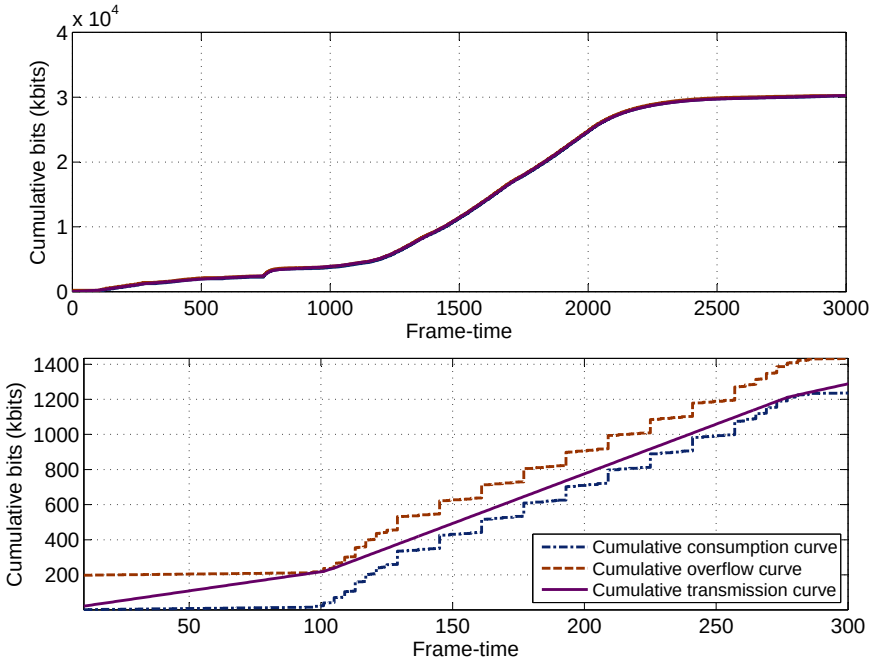


Figure 7.3: Simulation results: transmission curves of *Star wars*.

are plotted in Fig. 7.3. It can be seen that the transmission schedule always lie in between the cumulative consumption and overflow curves, indicating that there is no playout buffer underflow or overflow events in this simulation.

We next compare the optimal power algorithm with a conventional transmission scheme with respect to the average power consumption at the BS. In each time slot, the conventional scheme only transmits the video data that is needed by the decoder at the end of the time slot. It achieves a cumulative transmission curve that connects all the corner points of $D_n(t)$ (also called the "*lazy*" scheme). Intuitively, such lazy approach should be energy efficient since it always transmits the minimal amount of data as needed. However, we will see that the proposed algorithm outperform this lazy approach in the simulations.

In Fig. 7.4, we plot the average power consumption achieved by the two schemes for increased distance to the BS. Each point in the figure is the average of $10,000$ simulation runs. The 95% confidence intervals are plotted as vertical bars in the figure, which are all very small.

It can be seen that the proposed algorithm outperforms the conventional scheme for the entire range of distances examined. When the distance to BS is small, both schemes use small transmit

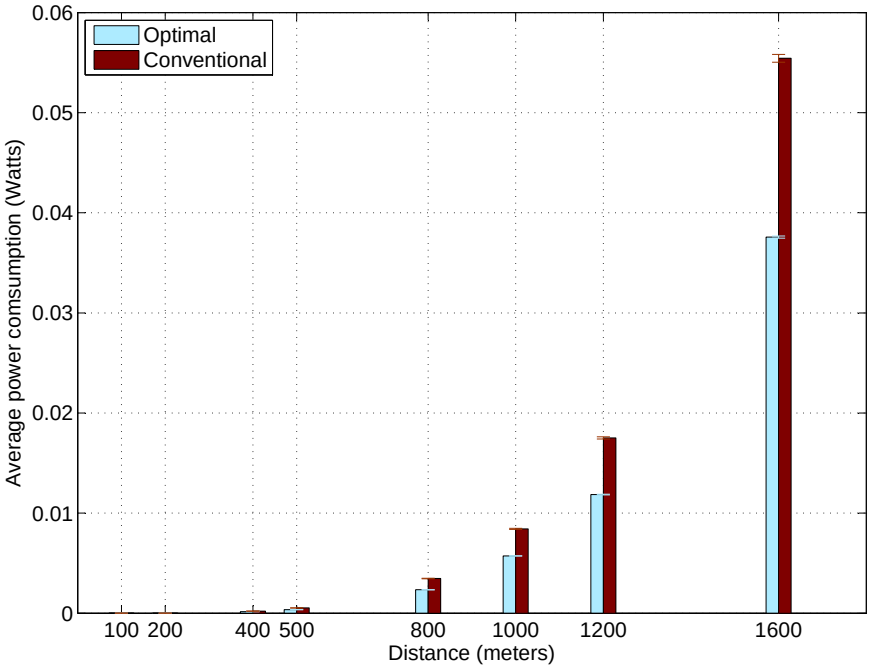


Figure 7.4: Simulation results: average power consumption.

powers and the power savings are not very big. However, when the distance is increased, channel fading has a bigger impact on interference and channel capacity. The proposed algorithm achieves considerable power savings than the conventional scheme. When the distance is 1,600 m, the total power of the proposed scheme is only 46.62% of that of the conventional approach, corresponding to a 54.34% normalized improvement.

We further investigate in more detail the difference of the transmissions between the two schemes. The position of a mobile node is set to $1,000$ m from the BS. The first $3,000$ frames of the *Star wars* movie are transmitted to the node using the PMA-1 scheme and conventional scheme, respectively. Fig. 7.5 shows the cumulative power consumption for the first $1,000$ frames, while the energy consumption for each video frame is plotted in Fig. 7.6 for frames in [200, 250]. We observe that at the beginning, the "*lazy*" scheme archives smaller power consumption than the PMA scheme, due to the fact that it only transmits the minimum amount of required frames in each time slot. However, the transmission of frame 241 of the conventional scheme generates a sharp power increase, because it encounters a large frame as well as bad channel condition, as indicated in Fig. 7.6. The transmission curves of the two schemes are plotted in Fig. 7.7 for the

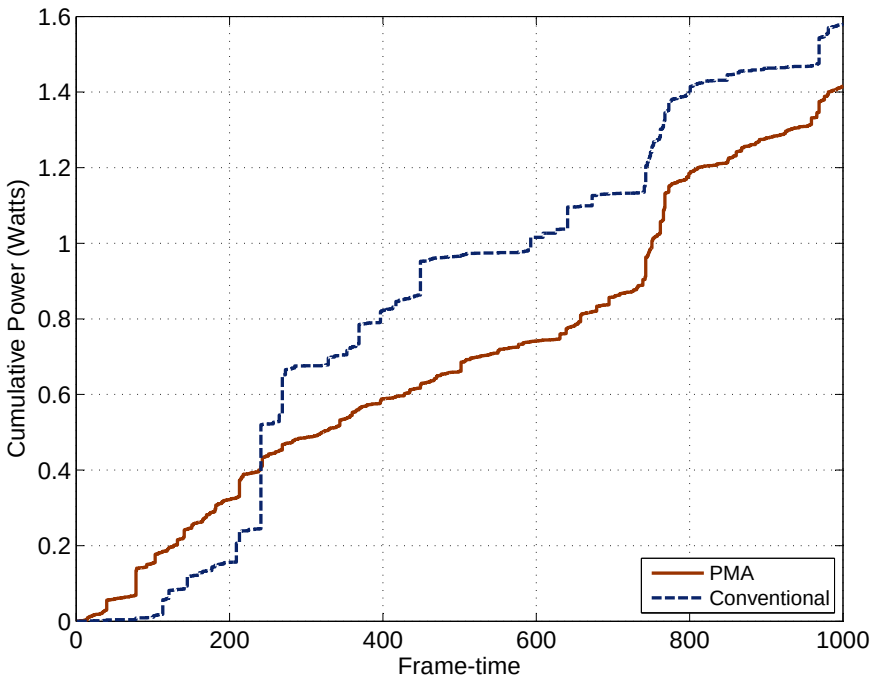


Figure 7.5: Cumulative power consumptions achieved by PMA-1 and the conventional scheme.

first 250 frames for the two schemes. Although the conventional scheme archives smaller power consumption frame by frame for about 80% of the $3,000$ frames, the average power consumption of the proposed scheme during the entire period (0.0055 W) is much smaller than that of the conventioanal scheme (0.0141 W). In summary, the "*lazy*" scheme only uses the current video and channel status, and transmits only the minimum amount of required video data, It does not effectively utilize the playout buffer capacity. Thus during the entire transmission period, the cumulative power may increased due to some large frames and bad channel conditions. On the contrary, the proposed scheme aims at minimizing the total average transmission power during the entire period. Thus, it achieves considerable power savings comparing to the conventional scheme.

We also obtain the average execution time of the proposed algorithm, under the same setting but for $20,000$ *Star Wars* frames. We find that the average execution time is about 0.06 s on an IBM Laptop with Intel T2400 1.83 GHz processor and 2 GB RAM.

Finally, we examine the buffer underflow events. The following scenarios are simulated:

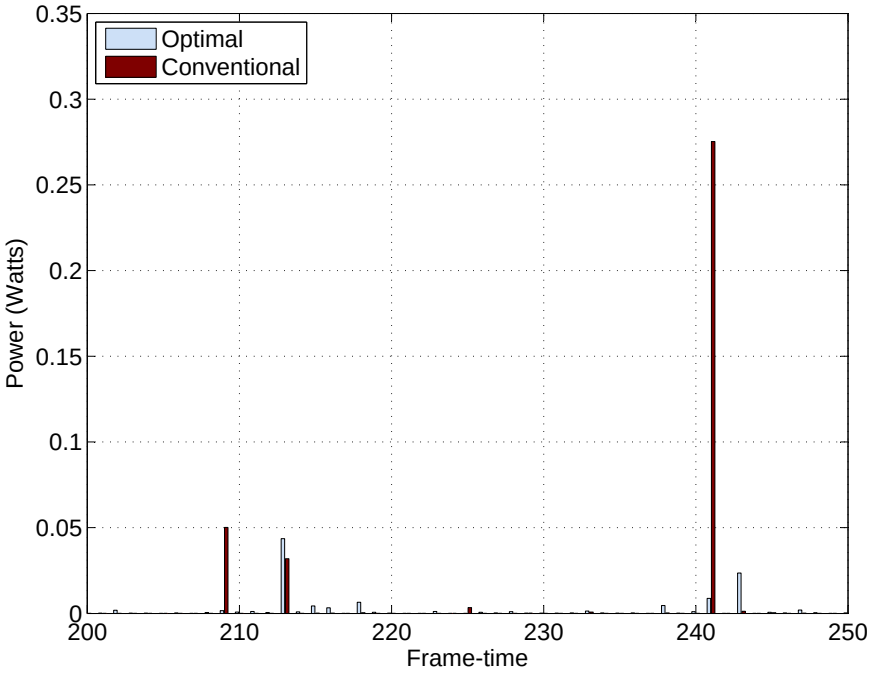


Figure 7.6: Power consumption for each frame achieved by PMA-1 and the conventional scheme.

- *Scenario 1*: $\bar{P} = 1$ W; the movies are *Star Wars*, *NBC News*, and *Tokyo Olympics*; 50 mobile users; $B_w = 1$ MHz;
- *Scenario 2*: The same setting as i) except that $B_w = 125$ KHz;
- *Scenario 3*: $\bar{P} = 10$ W; the HD movies are *Terminator 2*, *From Mars to China*, and *Sony Demo*; 20 mobile users; $B_w = 1$ MHz.

The HD movies have larger frame sizes and higher variability in frame sizes. The buffer underflow rates are presented in Table 7.3, each being the ratio of the number of underflow frames over the total number of frames. PMA-m achieves considerably lower underflow rates for all the three scenarios. The PMA-m underflow rates are 0.056%, 13.60%, and 14.26% of that of the conventional scheme. Therefore, PMA-m achieves not only considerable energy savings, but also much better video quality for the mobile users.

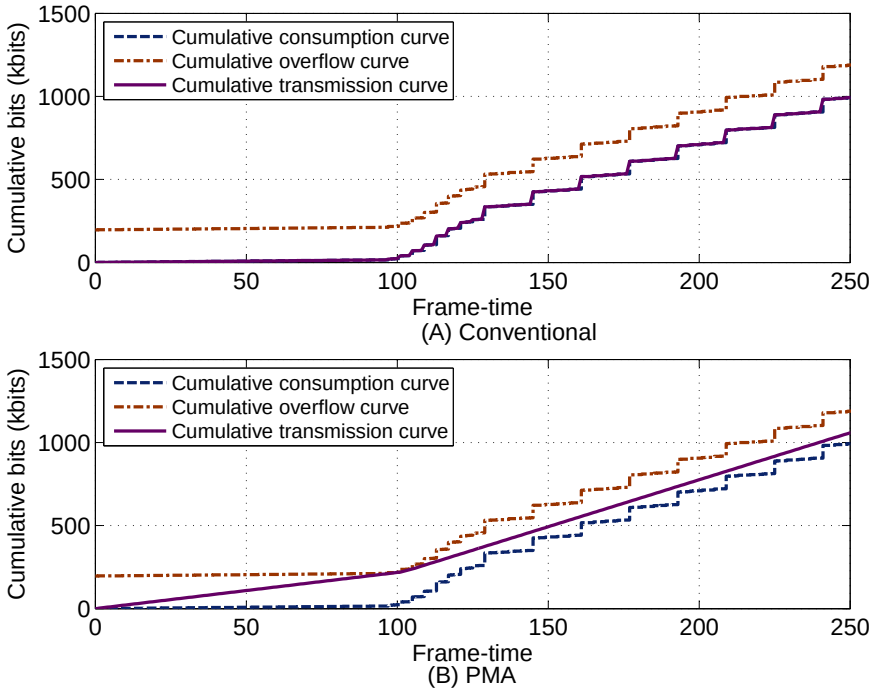


Figure 7.7: Transmission curves achieved by PMA-1 and the conventional scheme.

Table 7.3: Simulation Results: Playout Buffer Underflow Rates

	Scenario 1	*Scenario 2*	*Scenario 3*
PMA-m	0.0005%	1.8%	1.66%
Conventional	0.89%	13.24%	11.64%

7.5.1 Simulation Results for Interactive Video Streaming

In this section, we study the performance of the proposed algorithms for interactive video streaming. First, we simulate the interactive real-time video streaming with stringent delay requirements and small playout buffer sizes. The same simulation settings are used. All the positions of mobile nodes are randomly generated in the cell. We apply the proposed PMA-m, but only assume only the frame sizes in a small LAP are known.

The transmission curves of VBR movie *NBC News* are plotted in Fig. 7.8 for the first 100 frames, where the length of the LAP is 16 frames. We may observe that the proposed algorithm PMA-m is executed piecewisely for each block of LAP frames, where LAP=16. For this range of frames, there is neither playout buffer overflow nor playout buffer underflow occurs. We then

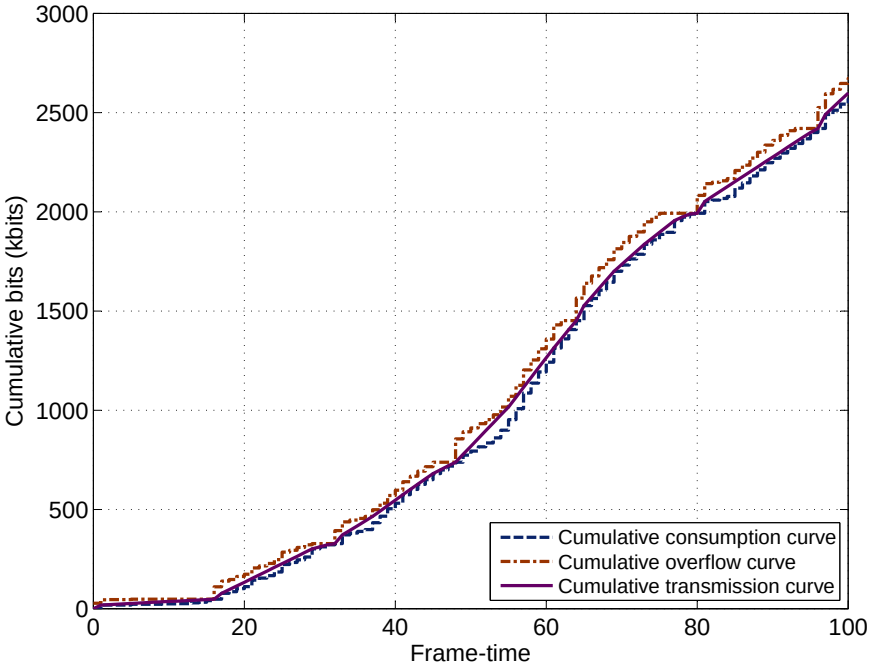


Figure 7.8: Transmission curves for real-time interactive video streaming: *NBC News* and LAP = 16.

Table 7.4: Playout Buffer Underflow Rates for Interactive VBR Videos with Different LAPs

	Scenario I	*Scenario II*	*Scenario III*
	LAP=16		
PMA-m	0.387%	5.969%	4.766%
Conventional	0.711%	12.446%	12.495%
	LAP=8		
PMA-m	0.377%	5.690%	4.894%
Conventional	0.620%	11.409%	9.645%

compute the number of the underflow frames over the total number of frames, as presented in Table 7.4. The PMA-m still archive the underflow rate that are 54.56%, 47.96%, 38.15% of those of the conventional scheme in this case.

To illustrate the impact of the length of LAP, we further decrease it to 8 frames and then run the simulations with random deployed mobile nodes. The playout buffer underflow rates are presented in Table 7.4. For the halved delay requirement, naturally the proposed scheme's performance is slightly degraded due to limited information about the video frame sizes. However, the PMA-m scheme still archives underflow rates that are 60.90%, 49.88%, 50.74% of those of the

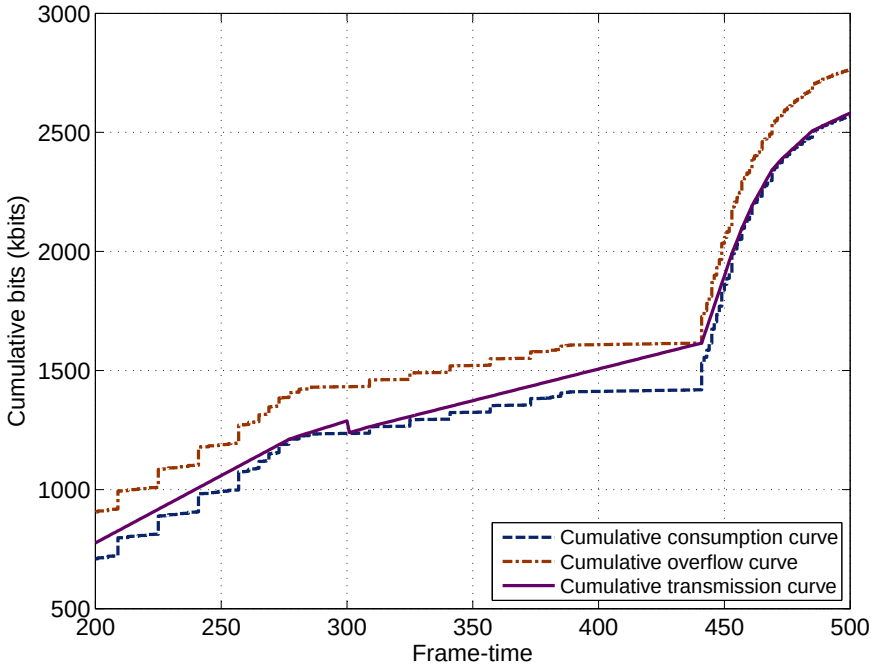


Figure 7.9: Transmission curves after the user skips the frames in [10s, 20s]: *Star Wars*.

conventional scheme, indicating considerably superior viewer performance over the conventional "lazy" approach.

Finally, we demonstrate the application of the proposed schemes to VCR control in interactive video streaming. We assume that after 10 seconds of streaming the VBR video (i.e, corresponding to 300 frames), the user skips the next 10 seconds of the video, and then resumes the video playout from 20 second. We plot the dynamics in the transmission/schedule curves of the *Star Wars* video in Fig. 7.9. Comparing the curves with the original non-skipped transmissions in Fig. 7.10, we observe that after playing out the 300^{th} frame, the frame from 301 to 600 are skipped by user's operation. Then the frame 601 is moved to the time-slot 301 and a new transmission schedule is computed for the following frames.

7.6 Related Work

A thorough review in VBR video model and VBR video streaming over wired networks has been explored in 5.6. In an interesting work [95], Salehi, et al. applied majorization to VBR video smoothing and developed a smoothness optimal algorithm. The proof of Theorem 7.2 follows a

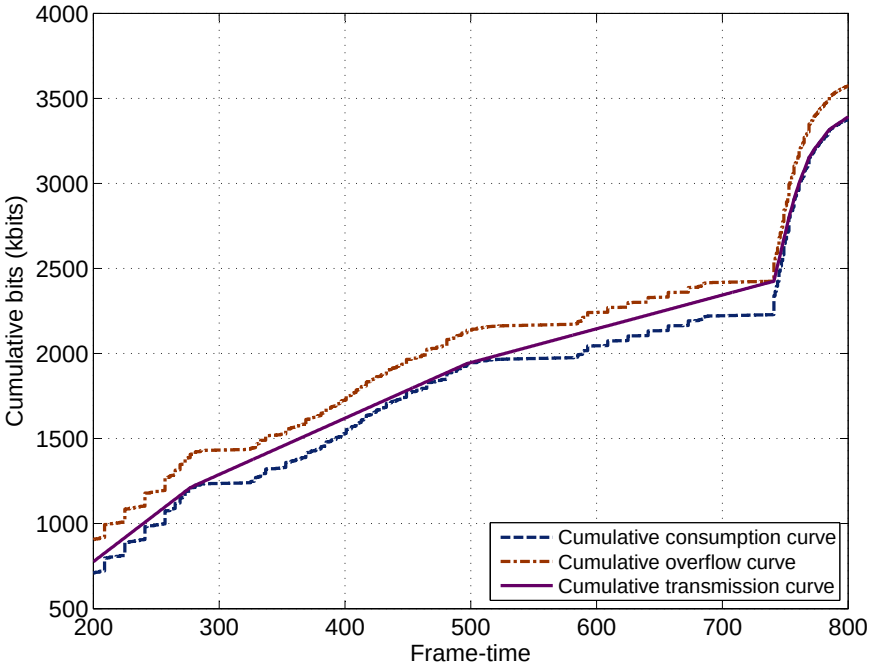


Figure 7.10: Transmission curves for the original non-skipped video streaming: *Star Wars*.

similar approach as in [95]. These prior work are based on the assumptions of a single video session and constant rate channels, which may not be directly applied to the case of wireless networks.

In this chapter, we consider multiuser VBR video streaming within a cellular network with orthogonal channels. We jointly consider power control, video traffic, and video palyout information for power minimization. Our stochastic majorization theory based approach is quite different from the prior works [54, 143], which allow us to develop effective algorithms with low complexity and proven optimality.

7.7 Conclusion

In this chapter, we studied the problem of downlink multiuser VBR video streaming in cellular networks. Our formulation takes into account the interactions among power control, fading channels, VBR video traffic and playout characteristics. We formulate a constrained stochastic optimization problem aiming to minimize the BS power consumption and to avoid playout buffer

overflow or underflow. We developed majorization-based algorithms to solve the formulated problem. For the case of large peak power constraints, we prove the optimality of the proposed algorithm and the uniqueness of the global optimal, as well as the equivalence of power optimal and smoothness optimal. For the case of multiple videos coupled with the peak power constraint, we develop an effective heuristic algorithm that selectively suspends some video sessions when the peak power constraint is violated. The superior performance of the proposed algorithms over a conventional scheme is validated with trace-driven simulations.

Chapter 8

Summary and Future Work

8.1 Summary

In the previous chapters, we proposed frameworks for energy efficient designs in Cyber-physical systems. We investigated the problems by a control and optimization approach, which contains Lyapunov optimization [12], majorization [13], nonlinear and convex optimization [14]. The synergy of these advanced mathematical tools produces new visions for the energy efficient solutions to alleviate energy resource depletion, decrease greenhouse gases emission and air pollution, which evolves a green world in the future.

In Chapter 2, we presented the electric power scheduling policies for smoothing the demand profile in power distribution networks. We introduced a *deterministic electricity supply/demand model* that takes into account time-varying demands and their deadlines. We formulated a constraint nonlinear optimization problem and incorporated the theory of majorization to develop algorithms that can compute smooth optimal electric power schedules. After the smooth power schedule is obtained, a distributed user benefit maximization load control scheme is used to allocate the scheduled power to individual users, while maximizing their level of satisfaction. The proposed algorithms are highly desirable for grid design and operation, which provide smooth electric power scheduling, minimum peak power and operating cost. The simulation showed that the proposed algorithms can alleviate the peak power up to 45%. This means we can deploy transformers, transmission lines and other electrical devices with much smaller capacity, which save the captital investment in the power grid construction. The generation cost saving is about 5% by the proposed algorithms. Due to the tremendous amount of energy of bulk generation, 5% cost savings could indeed contribute to billions of dollars.

In Chapter 3, we investigated the problem of balancing supply and demand of electric energy in microgrid (MG). We presented a novel framework for smart energy management based on the concept of *quality-of-service in electricity* (QoSE). Specifically, the resident electricity demand is classified into basic usage and quality usage. The basic usage is always guaranteed by the MG, while the quality usage is controlled based on the MG status. The *microgrid control center* (MGCC) aims to minimize the MG operation cost and maintain the outage probability of quality usage, i.e., QoSE, below a target value, by scheduling electricity among DRERs, ESS's, and macrogrid. The problem is formulated as a constrained stochastic programming problem. The Lyapunov optimization technique is then applied to derive an adaptive electricity scheduling algorithm by introducing the QoSE virtual queues and energy storage virtual queues. The proposed algorithm is an online algorithm since it does not require any statistics and future knowledge of the electricity supply, demand and price processes. We derive several "hard" performance bounds for the proposed algorithm. and evaluate its performance with trace-driven simulations. The proposed electricity scheduling algorithm enables an efficient integration of DRERs, ESS's, and residential power quality management into the smart grid by plug-and-play interfaces, and provides a promising paradigm for smart energy management systems in smart grid. The simulation showed that the MG operation cost can be saved up to 60%, while keeping the power quality of the users.

In Chapter 5, we studied the problem of power allocation of *base station* (BS) for streaming multiple *variable bit rate* (VBR) videos in the downlink of a wireless cellular network with intracell interference. We considered a *deterministic model* for VBR video traffic and finite playout buffer at the mobile users. The objective is to derive the optimal downlink power allocation for the VBR video sessions, such that the video data can be delivered in a timely fashion without causing playout buffer overflow and underflow. The formulated problem is a nonlinear nonconvex optimization problem. We analyzed the convexity conditions for the formulated problem and proposed a two-step greedy approach to solve the problem. We also developed a distributed algorithm based on the dual decomposition technique. The proposed algorithms effectively allocated the power in BS's

to stream the VBR video in cellular networks, while preserving the *quality of experience* (QoE) requirement.

In Chapter 6 we further extended the problem of downlink power control for streaming multiple VBR videos in a multicell wireless networks, where downlink capacities are limited by inter-cell interference. We adopted a deterministic model for VBR traffic that considers video frame sizes and playout buffers at the mobile users. The problem is to find the optimal transmit powers for the BS's, such that VBR video data can be delivered to mobile users without causing playout buffer underflow or overflow. We formulated a nonlinear nonconvex optimization problem and proved the condition for the existence of feasible solutions. We then developed a centralized branch-and-bound algorithm incorporating the Reformulation-Linearization Technique, which can produce $(1-\epsilon)-$optimal solutions. We also proposed a low-complexity distributed algorithm with fast convergence. Numerical results showed that the proposed algorithms is QoE performance bounded and achieve effective usage of the BS's power in streaming VBR videos.

In Chapter 7, we addressed the energy saving for BS in wireless cellular networks with orthogonal channels to achieve energy efficient VBR video streaming. We took into account the interactions among power control, fading channels, VBR video traffic, and playout characteristics. A constrained stochastic optimization problem was formulated aiming to minimize the BS power consumption and to avoid playout buffer overflow or underflow. We then developed majorization-based algorithms to achieve energy efficiency, while preserving the QoE demands, for BS downlink VBR streaming in cellular networks. The simulation showed that the average power consumption can achieve 54% improvement and the proposed algorithms are also compativle to interactive video streaming in wireless cellular networks.

8.2 Future Work

Although there has been considerable advances in energy efficient Cyber-physical systems, many problems still remain open in this interesting area. We briefly extend our discussion on

the control and optimization in distributed smart electric energy management systems and low complexity energy efficient wireless multimedia networks.

Distributed Electric Energy Management Systems

We investigated the efficient electric energy scheduling in the power delivery in previous chapters. The proposed policies are mainly based on the centralized control in the electric power delivery networks and Microgrids, which were largely inherited from the control infrastructure of legacy grid. In the evolution of smart grid, all the components, such as DRERs, ESS's, MGs, and users, will be deployed in an ad-hoc mode and accessed in a plug-and-play fashion. Accordingly, such distributed electric power delivery networks prefer distributed intelligent electric energy management systems.

Compared to the centralized approaches, distributed management systems make decisions based on local information with very limited timely information exchange, which reduce the amount of real time information delivery in the communication networks. The distributed management systems are less susceptible to the information loss due to the impairment of the communication links and also have faster response to the grid status. The security design and cryptographic systems would also be simplified due to reduced amount of information exchange. However, due to the limitation on the accessible information, the distributed management systems usually are not capable of computing a globally optimal decision. Further, analytical models and mathematical tools would be necessary to provide the performance bounded optimal decisions locally.

Low Complexity Energy Efficient Wireless Multimedia Networks

Due to the intrinsic complexity of video sources and the dynamics and uncertainty of wireless systems, we conjecture that a holistic approach that encompasses the parameter space would be necessary and the trade-off between complexity and efficiency of a solution algorithm should be carefully investigated. In particular, we list some interesting problems that may be worth of further investigation in the following.

- Complexity-distortion analysis and the design of energy-efficient video codecs require accurate models of video codecs. However, a video codec is a combination of complex functional blocks, which makes accurate mathematic modeling extremely difficult. In addition, the quality of compressed video and the power consumption of the video codec depend on a large number of parameters. A content-based power-aware design may encounter a large search space for optimal solutions. It would be useful to develop an accurate and effective model for video codec that can be incorporated into the mathematical optimization frameworks for both energy efficient codec design and wireless multimedia system design.

- Cross-layer design has been widely adopted for video networking problems. It has been shown that an adaptive strategy with cooperation of several layers can achieve optimal power efficiency for video streaming. Most prior work assume that the wireless channel information and network status are known apriori (e.g., by accurate estimation, measurement, and timely feedback). In practice, this assumption may not be true, because of channel/network uncertainty and dynamics, and delay and congestion in the network. Thus, balancing the achievable performance and the control overhead of the design is still an open problem. Effective schemes that are robust to the channel/network uncertainties would be highly appealing.

Appendices

Appendix A

Proofs in Chapter 2

A.1 Proof of Theorem 2.2

The schedule computed by SEPS-DL, $\vec{P}^*$, could be a straight line or in the general case, consist of one or more convex and concave segments. If $\vec{P}^*$ is a straight line, it is obvious that $\vec{P}^* \prec \vec{P}_k$ for any other $\vec{P}_k$ (see Fig. 2.3) and it is smooth optimal. In the general case, we need to show $\vec{P}^* \prec \vec{P}_k$, for all k in every convex or concave segment. Then according to Lemma 2.1, we have $\vec{P}^* \prec \vec{P}_k$ for all k and it is optimal.

Let $\vec{P}_k$ denote an arbitrary feasible schedule. We introduce an auxiliary schedule $\vec{P}_1$, which intersects with $\vec{P}^*$ at all its power changing points in every convex segment, and with $\vec{P}_k$ at all its power changing points in every concave segment, as shown in Fig. A.1.

First, we prove that $\vec{P}^* \prec \vec{P}_1$. For a convex segment of $\vec{P}^*$, because $\vec{P}_1$ intersects with $\vec{P}^*$ at all the power changing points of $\vec{P}^*$, we have $\vec{P}^* = \vec{P}_1$ in all the convex segments. For a concave segment of $\vec{P}^*$, the endpoints of the concave segment should be the last (first) power changing point of the previous (next) convex segment, where $\vec{P}^*$ intersects with $\vec{P}_1$. The power changing points within the concave segment are all on $W_{min}(t)$, as in SEPS-DL. Therefore, $\vec{P}_1$ is an outer concave curve above $\vec{P}^*$ (or, it is farther away from the straight line A in Fig. 2.3) in this segment. From the discussion of Fig. 2.3, we have $\vec{P}^* \prec \vec{P}_1$ for all the concave segments. It follows that $\vec{P}^* \prec \vec{P}_1$ according to Lemma 2.1.

We next prove that $\vec{P}_1 \prec \vec{P}_k$. For a convex segment of $\vec{P}^*$, the endpoints of the convex segment should be the last (first) power changing point of the previous (next) concave segment, where $\vec{P}_k$ intersects with $\vec{P}_1$. The power changing points of $\vec{P}_1$ (or, of $\vec{P}^*$) in the convex segment are all on $W_{max}(t)$. Therefore, $\vec{P}_k$ is an outer convex curve below $\vec{P}_1$ in this segment. From the discussion of Fig. 2.3, it follows that $\vec{P}_1 \prec \vec{P}_k$ in all the convex segments. In a concave segment, we have either

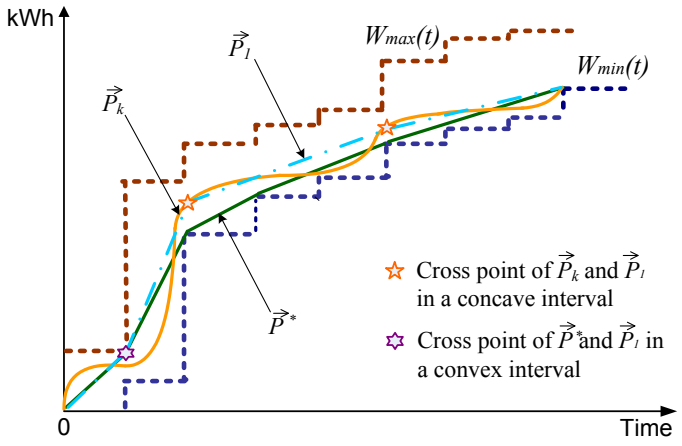


Figure A.1: Illustration for the optimality proof of SEPS-DL algorithm

$\vec{P}_k = \vec{P}_1$ or $\vec{P}_k \prec \vec{P}_1$, because $\vec{P}_1$ intersects with $\vec{P}_k$ at each power changing point. Thus, we obtain $\vec{P}_1 \prec \vec{P}_k$ for all the concave segments, and $\vec{P}_1 \prec \vec{P}_k$ according to Lemma 2.1.

Finally we have $\vec{P}^* \prec \vec{P}_1 \prec \vec{P}_k$. Proposition 2.1 states that problem (2.3) is Schur-convex and order preserving. It follows from Fact 2.1 that $\vec{P}^*$ is optimal to problem (2.3).

A.2 Proof of Theorem 2.5

Without loss of generality, we assume the fuel cost at time slot t is $C(t) = g(P(t), \theta(t))$, which a nondecreasing convex function of the supplied power $P(t)$ [92]. This assumption is generally practical, e.g. classically, the fuel cost for the electric energy generation is usually considered as a quadratic function of its power generation [103]. We also assume the cost $C(t)$ is affected by the random factors $\theta(t)$, which correspond to the cost uncertainty during the period, such as the fuel market price disturbance, and etc. We assume each element in $\theta(t)$ is i.i.d over slots. Thus,

the minimization of the expectation of the total cost over the period L is:

$$\begin{aligned} \text{Minimize:} \quad & \sum_{t=1}^{L} \mathbb{E}\{g(P(t), \vec{\theta}(t))\} \\ \text{s.t.} \quad & W_{min}(t) \leq W(t) \leq W_{max}(t), \forall t \\ & P(t) = W(t)/\tau \\ & \sum_{t=1}^{L} W(t) = \Phi \end{aligned} \tag{A.1}$$

Due to the convexity of the cost function $g(\cdot)$ respective to $P(t)$, problem (A.1) is similar to problem (2.3), except the random variable $\theta(t)$. Thus, we resort to *stochastic majorization* (rather than ordinary majorization) to solve this constraint nonlinear stochastic optimization problem.

Lemma A.1. *The objective function of problem (A.1) is an increasing Schur-convex function.*

Proof. The i.i.d. random variables $\theta(t)$ are exchangeable for all t. The objective function (A.1) can be rewritten as

$$G(\vec{P}) = \mathbb{E}\left\{\sum_{t=1}^{L} g(P(t), \theta(t))\right\},$$

where $g(P(t), \theta(t))$ is convex and increasing with respective to $P(t)$ for each fixed $\theta(t)$, and $\sum_{t=1}^{L} g(P(t), \theta(t))$ is a symmetric, convex and increasing function w.r.t. $P(t)$. According to Proposition 11.B.5 in [13], the expectation $G(\vec{P})$ is symmetric, convex and increasing. Following Fact 2.1, the objective function (A.1) is Schur-convex and increasing. □

By Lemma A.1, the solution of problem (A.1) is equivalent to finding the optimal power vector $\vec{P}^*$, which is majorized by any other feasible power vectors. Thus, the smooth optimal solution $\vec{P}^*$ in Table 1 is also the solution for problem (A.1). Thus, the proposed SEPS-DL achieves fuel cost optimal for energy generation.

Appendix B

Proofs in Chapter 3

B.1 Proof of Theorem 3.1

According to the system equation (3.12), we have

$$\begin{cases} Z_n(1) \geq Z_n(0) - \delta_n \cdot \alpha_n(0) + I_n(0) \\ \dots \\ Z_n(t) \geq Z_n(t-1) - \delta_n \cdot \alpha_n(t-1) + I_n(t-1). \end{cases} \tag{B.1}$$

Summing up the inequalities in (B.1), we have

$$Z_n(t) \geq Z_n(0) - \delta_n \cdot \sum_{\tau=0}^{t-1} \alpha_n(\tau) + \sum_{\tau=0}^{t-1} I_n(\tau). \tag{B.2}$$

Dividing both sides by t and letting t go to infinity, we have

$$\lim_{t\to\infty} \frac{Z_n(t) - Z_n(0)}{t} \geq \lim_{t\to\infty} \frac{1}{t} \left[-\delta_n \sum_{\tau=0}^{t-1} \alpha_n(\tau) + \sum_{\tau=0}^{t-1} I_n(\tau) \right].$$

$Z_n(0)$ is finite. If $Z_n(t)$ is rate stable by a control policy $I_n(t)$, it is finite for all t. We have $\lim_{t\to\infty} \frac{Z_n(t)-Z_n(0)}{t} = 0$, which yields $\rho_n \leq \delta_n \cdot \lambda_n$ due to the definitions of λ_n and $I_n(t)$.

B.2 Derivation of Equation (3.16)

With the drift defined as in (3.15), we have

$$
\begin{aligned}
\Delta(\vec{\Theta}(t)) &= \frac{1}{2}\mathbb{E}\left\{\sum_{k=1}^{K}[(X_k(t+1))^2-(X_k(t))^2|X_k(t)]+\right.\\
&\left.\sum_{n=1}^{N}[(Z_n(t+1))^2-(Z_n(t))^2|Z_n(t)]\right\}\\
&\le \frac{1}{2}\sum_{k=1}^{K}\mathbb{E}\left\{\left[(D_k(t))^2+(R_k(t))^2+2X_k(t)(R_k(t)-\right.\right.\\
&\left.\left.D_k(t))|X_k(t)]\right\}+\frac{1}{2}\sum_{n=1}^{N}\mathbb{E}\left\{\left[I_n(t)^2+\right.\right.\\
&\left.\left.(\delta_n\alpha_n(t))^2+2Z_n(t)(I_n(t)-\delta_n\alpha_n(t))|Z_n(t)\right\}\right.\\
&= \frac{1}{2}\sum_{k=1}^{K}\mathbb{E}\{[(D_k(t))^2+(R_k(t))^2]\}+\\
&\sum_{k=1}^{K}\mathbb{E}\{X_k(t)(R_k(t)-D_k(t))|X_k(t)\}+\\
&\frac{1}{2}\sum_{n=1}^{N}\mathbb{E}\{[(1+(\sigma_n)^2)(\alpha_n(t))^2+(p_n(t))^2]\}+\\
&\frac{1}{2}\sum_{n=1}^{N}\mathbb{E}\{2Z_n(t)(1-\delta_n(t))\alpha_n(t)-\\
&(Z_n(t)+\alpha_n(t))p_n(t)|Z_n(t)\}\\
&\le B+\sum_{n=1}^{N}\mathbb{E}\{Z_n(t)(1-\delta_n)\alpha_n(t)|Z_n(t)\}+\\
&\sum_{k=1}^{K}\mathbb{E}\{X_k(t)(R_k(t)-D_k(t))|X_k(t)\}-\\
&\sum_{n=1}^{N}\mathbb{E}\{(Z_n(t)+\alpha_n(t))p_n(t)|Z_n(t)\}.
\end{aligned}
$$

where $B=\frac{1}{2}\sum_{k=1}^{K}(\max\{D_k^{max},R_k^{max}\})^2+\frac{1}{2}\sum_{n=1}^{N}(2+\delta_n^2)(\alpha_n^{max})^2$ is a constant.

B.3 Proof of Lemma 3.1

In part 1) of Lemma 3.1, if $Q(t) > 0$, we have $S(t) = 0$ according to (3.7). The objective function of problem (3.19) becomes

$$VQ(t)C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) - \sum_{n=1}^{N} (Z_n(t) + \alpha_n(t))p_n(t). \tag{B.3}$$

We first prove Lemma 3.1-1a). If $X_k(t) > -VC(t)$, we assume $R_k(t) > 0$. Then we have $D_k(t) = 0$ according to (3.4). Accordingly, the object function (B.3) is transformed to

$$\begin{aligned}
& VQ(t)C(t) + \sum_{i\neq k} X_i(t)(R_i(t) - D_i(t)) - \\
& \sum_{n=1}^{N} (Z_n(t) + \alpha_n(t))p_n(t) + X_k(t)R_k(t) \\
> \; & VQ(t)C(t) + \sum_{i\neq k} X_i(t)(R_i(t) - D_i(t)) - \\
& \sum_{n=1}^{N} (Z_n(t) + \alpha_n(t))p_n(t) - VC(t)(P(t) + Q(t) - \\
& \sum_{i\neq k}(R_i(t) - D_i(t)) - \sum_{n=1}^{N} p_n(t) \\
= \; & V\left[\sum_{i\neq k}(R_i(t) - D_i(t)) + \sum_{n=1}^{N} p_n(t) - P(t)\right] C(t) + \\
& \sum_{i\neq k} X_i(t)(R_i(t) - D_i(t)) - \sum_{n=1}^{N} (Z_n(t) + \alpha_n(t))p_n(t).
\end{aligned}$$

The above inequality is due to $X_k(t) > -VC(t)$ and $R_k(t) = P(t)+Q(t)-\sum_{i\neq k}(R_i(t)-D_i(t))-\sum_{n=1}^{N} p_n(t) \geq 0$. The last expression shows, given the assumption $R_k(t) > 0$, we may find another feasible electricity allocation scheme $\tilde{Q}(t) = \sum_{i\neq k}(R_i(t) - D_i(t)) + \sum_{n=1}^{N} p_n(t) - P(t)$, which can achieve a smaller objective value by choosing $R_k(t) = 0$ and $D_k(t) = 0$. This contradicts with

the assumption $R_k(t) > 0$. Thus, we prove that $R_k(t) = 0$ when $X_k(t) > -VC(t)$, under the situation $Q(t) > 0, S(t) = 0$.

We then prove the second part of Lemma 3.1-1a). It follows (3.4) that $R_k(t) = 0$ if $D_k(t) > 0$. Then (B.3) becomes

$$\begin{aligned}
& VQ(t)C(t) + \sum_{i\neq k} X_i(t)(R_i(t) - D_i(t)) - \\
& \sum_{n=1}^{N} (Z_n(t) + \alpha_n(t))p_n(t) - X_k(t)D_k(t) \\
> \; & VQ(t)C(t) + \sum_{i\neq k} X_i(t)(R_i(t) - D_i(t)) - \\
& \sum_{n=1}^{N} (Z_n(t) + \alpha_n(t))p_n(t) + VC(t)(\sum_{n=1}^{N} p_n(t) - P(t) - \\
& Q(t) + \sum_{i\neq k}(R_i(t) - D_i(t))) \\
= \; & V\left[\sum_{i\neq k}(R_i(t) - D_i(t)) + \sum_{n=1}^{N} p_n(t) - P(t)\right] C(t) + \\
& \sum_{i\neq k} X_i(t)(R_i(t) - D_i(t)) - \sum_{n=1}^{N} (Z_n(t) + \alpha_n(t))p_n(t).
\end{aligned}$$

The above inequality is due to $X_k(t) < -VC(t) < 0$ and $D_i(t) = -P(t) - Q(t) + \sum_{i\neq k}(R_i(t) + D_k(t)) + \sum_{n=1}^{N} p_n(t) > 0$. The last expression shows, given the assumption $D_k(t) > 0$, we may find another electricity allocation scheme with $\tilde{Q}(t) = \sum_{i\neq k}(R_i(t) - D_i(t)) + \sum_{n=1}^{N} p_n(t) - P(t)$, which can achieve a smaller objective value by choosing $R_k(t) = 0$ and $D_k(t) = 0$. This contradicts with the assumption $D_k(t) > 0$. We thus prove that $D_k(t) = 0$ when $X_k(t) < -VC(t)$, under the situation $Q(t) > 0, S(t) = 0$, which completes the proof of Lemma 3.1-1a).

We next prove Lemma 3.1-1b). For the first part, if $Z_n(t) > VC(t) - \alpha_n(t)$, we assume $0 \leq p_n(t) < (1-\delta_n)\alpha_n(t)$. Following (3.18) and $S(t) = 0$, we have

$$
\begin{aligned}
& B + VQ(t)C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n}(Z_j(t)(1-\delta_j)\alpha_j(t) - (Z_j(t)+\alpha_j(t))p_j(t)) + \\
& Z_n(t)(1-\delta_n)\alpha_n(t) - (Z_n(t)+\alpha_n(t))p_n(t) \\
=& B + VQ(t)C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n}(Z_j(t)(1-\delta_j)\alpha_j(t) - (Z_j(t)+\alpha_j(t))p_j(t)) + \\
& Z_n(t)[(1-\delta_n)\alpha_n(t) - p_n(t)] - \alpha_n(t)p_n(t) \\
>& B + VQ(t)C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n}(Z_j(t)(1-\delta_j)\alpha_j(t) - (Z_j(t)+\alpha_j(t))p_j(t)) + \\
& (VC(t) - \alpha_n(t))[(1-\delta_n)\alpha_n(t) - p_n(t)] - \alpha_n(t)p_n(t) \\
=& B + V[\sum_{k=1}^{K}(R_k(t) - D_k(t)) + \sum_{j\neq n} p_j(t) - P(t) + \\
& (1-\delta_n)\alpha_n(t)]C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n}(Z_j(t)(1-\delta_j)\alpha_j(t) - (Z_j(t)+\alpha_j(t))p_j(t)) + \\
& Z_n(t)(1-\delta_n)\alpha_n(t) - (Z_n(t)+\alpha_n(t))(1-\delta_n)\alpha_n(t).
\end{aligned}
$$

The above inequality is due to $Z_n(t) > VC(t) - \alpha_n(t)$ and the assumption $p_n(t) < (1-\delta_n)\alpha_n(t)$. The last equality shows, given the assumption $p_n(t) < (1-\delta_n)\alpha_n(t)$, we may find another electricity allocation scheme with $p_n(t) = (1-\delta_n)\alpha_n(t)$ and $\tilde{Q}(t) = \sum_{k=1}^{K}(R_k(t)-D_k(t)) + \sum_{j\neq n} p_j(t) - P(t) + (1-\delta_n)\alpha_n(t)$, which can achieve a smaller objective value. This contradicts with the previous assumption. Thus, we have $p_n(t) \geq (1-\delta_n)\alpha_n(t)$.

For the second part of Lemma 3.1-1b), assume $p_n(t) > 0$ for $0 \leq Z_n(t) < VC(t) - \alpha_n(t)$.It follows (3.7) that $S(t) = 0$. The objective function (3.18) can be written as

$$\begin{aligned}
& B + VQ(t)C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n}(Z_j(t)(1-\delta_j)\alpha_j(t) - (Z_j(t) + \alpha_j(t))p_j(t)) + \\
& Z_n(t)(1-\delta_n)\alpha_n(t) - (Z_n(t) + \alpha_n(t))p_n(t) \\
> \; & B + VQ(t)C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n}(Z_j(t)(1-\delta_j)\alpha_j(t) - (Z_j(t) + \alpha_j(t))p_j(t)) + \\
& Z_n(t)(1-\delta_n)\alpha_n(t) - VC(t)p_n(t) \\
\geq \; & B + V[-P(t) + \sum_{k=1}^{K}(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n} p_j(t)]C(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) + \\
& \sum_{j\neq n}(Z_j(t)(1-\delta_j)\alpha_j(t)) - \sum_{j\neq n}((Z_j(t) + \alpha_j(t))p_j(t)).
\end{aligned}$$

The first inequality is due to $0 \leq Z_n(t) < VC(t) - \alpha_n(t)$ and the assumption $p_n(t) > 0$. The second inequality is due to the non-negativity of $Z_n(t)$ and $\alpha_n(t)$. The last equation shows, given the assumption $p_n(t) > 0$, we may find another electricity allocation scheme with $p_n(t) = 0$ and $\tilde{Q}(t) = -P(t) + \sum_{k=1}^{K}(R_k(t) - D_k(t)) + \sum_{j\neq n} p_j(t)$, which can achieve a smaller objective value. This contradicts with the previous assumption. Thus, we have $p_n(t) = 0$, which completes the proof of Lemma 3.1-1b).

In part 2) of Lemma 3.1, if $S(t) > 0$, we have $Q(t) = 0$ according to (3.7). The objective function (3.19) becomes

$$\begin{aligned}
& -VS(t)W(t) + \sum_{k=1}^{K} X_k(t)(R_k(t) - D_k(t)) - \\
& \qquad \sum_{n=1}^{N}(Z_n(t) + \alpha_n(t))p_n(t). \qquad \text{(B.4)}
\end{aligned}$$

We can prove part 2) with a similar approach as in the case of part 3.1. The detailed proof is omitted for brevity.

B.4 Proof of Lemma 3.2

Since $0 \leq C_{min} \leq C(t) \leq C_{max}$ and $V > 0$, we have $R_k(t) = 0$ when $X_k(t) < -VC_{max}$, and $D_k(t) = 0$ when $X_k(t) > -VC_{min}$ according to Lemma 3.1-3.1). Similarly, since $0 \leq W_{min} \leq W(t) \leq W_{max}$ and $V > 0$, we obtain $R_k(t) = 0$ when $X_k(t) < -VW_{max}$, and $D_k(t) = 0$ when$X_k(t) > -VW_{min}$ according to Lemma 3.1-2)

Since $C_{max} > W_{max}$ and $C_{min} > W_{min}$, we conclude that if $X_k(t) > -VW_{min}$, the optimal solution always select $R_k(t) = 0$. If $X_k(t) < -VC_{max}$, the optimal solution always select $D_k(t) = 0$. The proof is completed.

B.5 Proof of Lemma 3.3

The proof directly follows Lemma 3.1 and is similar to the proof of Lemma 3.2. We omit the details for brevity.

B.6 Proof of Theorem 3.2

From the battery virtual queue definition (3.10), the constraint $E_k^{min} \leq E_k(t) \leq E_k^{max}$ is equivalent to

$$-VC_{max} - D_k^{max} \leq X_k(t) \leq E_k^{max} - VC_{max} - D_k^{max} - E_k^{min}.$$

We assume all the batteries satisfy the battery capacity constraint at the initial time $t = 0$, i.e., $E_k^{min} \leq E_k(0) \leq E_k^{max}$, for all k. Supposing the inequalities hold true for time t, we then show the inequalities still hold true for time $t + 1$.

First, we show $X_k(t + 1) \leq E_k^{max} - VC_{max} - D_k^{max} - E_k^{min}$. If $-VW_{min} < X_k(t) \leq E_k^{max} - VC_{max} - D_k^{max} - E_k^{min}$, then with $X_k(t) > -VW_{min} \Rightarrow R_k(t) = 0$ from Lemma 3.2, we

have $X_k(t+1) = X_k(t) - D_k(t) \leq X_k(t) \leq E_k^{max} - VC_{max} - D_k^{max} - E_k^{min}$. If $X_k(t) \leq -VW_{min}$, then the largest value is $X_k(t+1) = -VW_{min} + R_k^{max}$. For any $0 < V \leq V_{max}$, we have

$$\begin{aligned} & E_k^{max} - VC_{max} - D_k^{max} - E_k^{min} \\ \geq\ & E_k^{max} - \min_k \left\{ \frac{E_k^{max} - E_k^{min} - R_k^{max} - D_k^{max}}{C_{max} - W_{min}} \right\} C_{max} \\ & -D_k^{max} - E_k^{min} \geq R_k^{max} \geq X_k(t+1). \end{aligned}$$

It follows that $X_k(t+1) \leq E_k^{max} - VC_{max} - D_k^{max} - E_k^{min}$.

Next, we show $X_k(t+1) \geq -VC_{max} - D_k^{max}$. Assuming $-VC_{max} - D_k^{max} \leq X_k(t) \leq -VC_{max}$, then from Lemma 3.2, we have $X_k(t) \leq -VC_{max} \Rightarrow D_k(t) = 0$. It follows that

$$X_k(t+1) = X_k(t) + R_k(t) \geq X_k(t) \geq -VC_{max} - D_k^{max}.$$

If $X_k(t) \geq -VC_{max}$, following (3.10), we have

$$\begin{aligned} X_k(t+1) &= X_k(t) - D_k(t) + R_k(t) \geq X_k(t) - D_k^{max} \\ &\geq -VC_{max} - D_k^{max}. \end{aligned}$$

Therefore, we have $X_k(t+1) \geq -VC_{max} - D_k^{max}$. Thus the inequalities also hold true for time $t+1$.

It follows that $E_k^{min} \leq E_k(t) \leq E_k^{max}$ is satisfied under the optimal scheduling algorithm for all k, t.

B.7 Proof of Theorem 3.3

(i) We first prove the upper bound Z_n^{max}. Initially, we have $Z_n(0) = 0 \leq VC_{max} + \alpha_n^{max}$. Assume that in time slot t the backlog of the QoSE virtual queue of resident n satisfies $Z_n(t) \leq Z_n^{max} = VC_{max} + \alpha_n^{max}$. We then check the backlog at time $t+1$ and show the bound still holds true.

If $Z_n(t) > VC_{max}$, following Lemma 3.3, the optimal scheduling for the quality usage of resident n satisfies $p_n(t) \geq (1-\delta_n)\alpha_n(t)$. From the virtual queue dynamics (3.12), we have

$$Z_n(t+1) \leq [Z_n(t) - \delta_n \alpha_n(t)]^+ + \delta_n \alpha_n(t).$$

If $Z_n(t) \geq \delta_n \alpha_n(t)$, we have $Z_n(t+1) \leq Z_n(t) \leq VC_{max} + \alpha_n^{max}$; otherwise, it follows that $Z_n(t+1) \leq \delta_n \alpha_n(t) < VC_{max} + \alpha_n^{max}$.

If $Z_n(t) \leq VC_{max}$, we have $Z_n(t+1) \leq [Z_n(t) - \delta_n \alpha_n(t)]^+ + \alpha_n^{max}$. If $Z_n(t) \geq \delta_n \alpha_n(t)$, we have $Z_n(t+1) \leq Z_n(t) - \delta_n \alpha_n(t) + \alpha_n^{max} \leq VC_{max} + \alpha_n^{max}$; otherwise, we have $Z_n(t+1) \leq \alpha_n^{max} \leq VC_{max} + \alpha_n^{max}$.

Thus we have $Z_n(t+1) \leq Z_n^{max} = VC_{max} + \alpha_n^{max}$. The proof of the QoSE virtual queue backlog bound is completed.

(ii) Consider an interval $[t_1, t_2]$ with length of $T = t_2 - t_1$. Summing (3.12) from t_1 to t_2, we have $Z_n(t_2+1) \geq Z_n(t_1) - \delta_n \sum_{\tau=t_1}^{t_2} \alpha_n(\tau) + \sum_{\tau=t_1}^{t_2} [\alpha_n(\tau) - p_n(\tau)] \geq \sum_{\tau=t_1}^{t_2} [\alpha_n(\tau) - p_n(\tau)] - T\delta_n \alpha_n^{max}$. It follows that

$$\sum_{\tau=t_1}^{t_2} [\alpha_n(\tau) - p_n(\tau)] \leq Z_n^{max} + T\delta_n \alpha_n^{max}.$$

B.8 Proof of Theorem 3.4

From Theorem 3.2, the battery capacity constraints is met in each time slot with the adaptive control policy. Take expectation on (3.2) and sum it over the period $[0, t-1]$:

$$\mathbb{E}\{E_k(t)\} - \mathbb{E}\{E_k(0)\} = \sum_{\tau=0}^{t-1} [\mathbb{E}\{R_k(t)\} - \mathbb{E}\{D_k(t)\}], \ \forall \ k.$$

Since $E_k^{min} \leq E_k(t) \leq E_k^{max}$, we divide both sides by t and let t go to infinity, to obtain

$$\lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{R_k(t)\} = \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{D_k(t)\}, \ \forall \ k. \tag{B.5}$$

Consider the following relaxed version of problem (3.9).

$$\begin{aligned} \text{minimize: } & \lim_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}\{Q(\tau)C(\tau)-S(\tau)W(\tau)\} \\ \text{s.t. } & \text{(3.3), (3.4), (3.5), (3.6), (3.7), (3.8), and (B.5).} \end{aligned} \tag{B.6}$$

Since the constraints in problem (B.6) are relaxed from that in problem (3.9), the optimal solution to problem (3.9) is also feasible for problem (B.6). The solution of (B.6) is a stationary and randomized policy does not depend on battery energy levels [115, 129]. Let the optimal solution for problem (B.6) be $\hat{\mathbb{A}}(t)=\{\hat{Q}(t),\hat{S}(t),\hat{R}_k(t),\hat{D}_k(t),\hat{p}_n(t)\}$ and the corresponding object value is $\hat{y}\leq y_{opt}$. According to the properties of optimality of stationary and randomized policies [115], the optimal solution $\hat{\mathbb{A}}(t)$ satisfies $\mathbb{E}\{\hat{R}_k(t)-\hat{D}_k(t)\}=0$ and $\hat{y}=\mathbb{E}\{\hat{Q}(\tau)C(\tau)-\hat{S}(\tau)W(\tau)\}$.

We substitute solution $\hat{\mathbb{A}}(t)$ into the right-hand-side of the drift-and-penalty (3.17). Since our proposed policy minimizes the right-hand-side of (3.17), we have

$$\begin{aligned} & \Delta(\vec{\Theta}(t))+V\mathbb{E}\{Q(t)C(t)-S(t)W(t)|\vec{\Theta}(t)\} \\ \leq\ & B+\sum_{n=1}^{N}\mathbb{E}\{Z_n(t)(1-\delta_n)\alpha_n(t)|Z_n(t)\}+ \\ & \sum_{k=1}^{K}X_k(t)\mathbb{E}\{\hat{R}_k(t)-\hat{D}_k(t)|X_k(t)\}- \\ & \sum_{n=1}^{N}(Z_n(t)+\alpha_n(t))\mathbb{E}\{\hat{p}_n(t)|Z_n(t)\}+ \\ & V\mathbb{E}\{\hat{Q}(t)C(t)-\hat{S}(t)W(t)|\vec{\Theta}(t)\} \\ \leq\ & B+\sum_{n=1}^{N}Z_n^{max}(1-\delta_n)\alpha_n^{max}+V\cdot y_{opt}. \end{aligned}$$

The second inequality is due to the properties of stationary and randomized policy and $\hat{y} \leq y_{opt}$. Taking expectation and sum up from 0 to $T-1$, we obtain

$$\begin{aligned} & \sum_{t=0}^{T-1} V\mathbb{E}\{Q(t)C(t) - S(t)W(t)\} \\ \leq \; & T \cdot \hat{B} + T \cdot V \cdot y_{opt} - \mathbb{E}\{L(\vec{\Theta}(T))\} + \mathbb{E}\{L(\vec{\Theta}(0))\} \\ \leq \; & T \cdot \hat{B} + T \cdot V \cdot y_{opt} + \mathbb{E}\{L(\vec{\Theta}(0))\}. \end{aligned}$$

The second inequality is due to the nonnegative property of Lyapunove functions. Divide both sides by $V \cdot T$ and let T go to infinity. Since the initial system state $\vec{\Theta}(0)$ is finite, we have $\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} V\mathbb{E}\{Q(t)C(t) - S(t)W(t)\} = y^* \leq y_{opt} + \frac{\hat{B}}{V}$.

Appendix C

Proofs in Chapter 5

C.1 Proof of Lemma 5.1

If the feasible power allocation $\vec{P}(t)$ achieves $\gamma_n^{max}(t)$ for all n, then all the user buffers are full at the end of the time slot, according to (5.5). The objective value (5.6) cannot be further improved without causing buffer overflow. Thus the solution is optimal.

C.2 Proof of Lemma 5.2

Consider a feasible power allocation $\vec{P}'(t) = [P_1'(t), P_2'(t), \cdots, P_N'(t)]^T$ and $\sum_{n\in\mathcal{U}} P_n'(t) < \bar{P}$. We can construct another feasible power allocation $\vec{P}''(t) = [P_1''(t), P_2''(t), \cdots, P_N''(t)]^T$, such that $P_n''(t) = \kappa \cdot P_n'(t)$, for all n, and $\kappa \cdot \sum_{n\in\mathcal{U}} P_n'(t) = \sum_{n\in\mathcal{U}} P_n''(t) \leq \bar{P}$, where $\kappa > 1$. For the SINR at user n, we have

$$\begin{aligned}
\gamma_n(\vec{P}''(t)) &= \frac{L_n G_n P_n''(t)}{\sum_{k\neq n} G_n P_k''(t) + \eta_n} \\
&= \frac{\kappa L_n G_n P_n'(t)}{\sum_{k\neq n} \kappa G_n P_k'(t) + \eta_n} \\
&> \frac{\kappa L_n G_n P_n'(t)}{\sum_{k\neq n} \kappa G_n P_k'(t) + \kappa\eta_n} \\
&= \gamma_n(\vec{P}'(t)).
\end{aligned}$$

It follows that $\sum_{n\in\mathcal{U}} \log(1+\gamma_n(\vec{P}''(t))) > \sum_{n\in\mathcal{U}} \log(1+\gamma_n(\vec{P}'(t)))$, since $\log(1+x)$ is an increasing function of x.

Choosing $\kappa = \bar{P}/\sum_{n\in\mathcal{U}} P_n'(t)$, we can construct a feasible solution $\vec{P}'''(t) = \kappa \cdot \vec{P}'(t)$, such that $\sum_{n\in\mathcal{U}} P_n'''(t) = \bar{P}$. Then we have $\gamma_n(\vec{P}'''(t)) > \gamma_n(\vec{P}'(t))$ and $\sum_{n\in\mathcal{U}} \log(1+\gamma_n(\vec{P}'''(t))) > \sum_{n\in\mathcal{U}} \log(1+\gamma_n(\vec{P}'(t)))$. That is, any feasible solution with $\sum_{n\in\mathcal{U}} P_n'(t) < \bar{P}$ will be dominated

by feasible solutions with $\sum_{n\in\mathcal{U}} P_n'''(t) = \bar{P}$. We conclude that the optimal solution $\vec{P}(t)$ must satisfy $\sum_{n\in\mathcal{U}} P_n(t) = \bar{P}$.

C.3 Proof of Lemma 5.3

Taking the first and second derivatives of the objective function (5.14) with respect to P_n, we have

$$\frac{\partial C_n(P_n)}{\partial P_n} = \frac{L_n(\bar{P}+A_n)}{(\bar{P}-P_n+A_n)[\bar{P}+(L_n-1)P_n+A_n]} \tag{C.1}$$

$$\frac{\partial^2 C_n(P_n)}{\partial {P_n}^2} = \frac{-L_n[(L_n-2)(\bar{P}+A_n)+2(1-L_n)P_n](\bar{P}+A_n)}{[(\bar{P}-P_n+A_n)^2+L_nP_n(\bar{P}-P_n+A_n)]^2}. \tag{C.2}$$

Since $P_n \leq \bar{P}$ and $A_n > 0$, both the first and second derivatives exist. Letting $\frac{\partial^2 C_n(P_n)}{\partial {P_n}^2} = 0$, we derive the unique inflection point

$$P_n^* = \frac{L_n-2}{2(L_n-1)}(\bar{P}+A_n). \tag{C.3}$$

When $P_n < P_n^*$, it can be shown that $\frac{\partial^2 C_n(P_n)}{\partial {P_n}^2} < 0$; when $P_n > P_n^*$, it can be shown that $\frac{\partial^2 C_n(P_n)}{\partial {P_n}^2} > 0$.

C.4 Proof of Theorem 5.2

The reflection point is $P_n^* = \frac{L_n-2}{2(L_n-1)}(\bar{P}+A_n)$. As $L_n \to \infty$, we have $P_n^* = 0.5(\bar{P}+A_n)$. Only one link can operate in the convex region due to constraint (5.17). Since $\frac{\partial P_n^*}{\partial L_n} > 0$, P_n^* is an increasing function of L_n. When $1 \ll L_n < \infty$, we have $P_n^* < 0.5 \cdot (\bar{P}+A_n)$. Letting $3P_n^* = \bar{P}$, we have $L_n = (4\bar{P}+6A_n)/(\bar{P}+3A_n)$.

C.5 Proof of Theorem 5.3

The first part can be easily shown by the first derivative of P_n^* with respect to A_n, which is $\frac{\partial P_n^*}{\partial A_n} = \frac{L_n - 2}{2(L_n - 1)} > 0$, for $L_n > 2$. The second part can be easily shown by evaluating (5.14), (5.15), and (C.1).

Appendix D

Proofs in Chapter 6

D.1 Proof of Lemma 6.1

According to the definition of $X_m(t)$ in (6.3), we have $C_m(t) = [X_m(t) - X_m(t-1)]/\tau$. From the definition of $C_m^{min}(t)$, the playout buffer is emptied at the end of time slot t, i.e., $X_m(t) = D_m(t)$. Therefore, we can derive the minimum required rate as

$$C_m^{min}(t) = \max\{0, D_m(t) - X_m(t-1)\}/\tau. \tag{D.1}$$

From the feasibility condition (6.4), we have $X_m(t-1) \geq D_m(t-1)$. Substituting it into (D.1), we have

$$C_m^{min}(t) \leq [D_m(t) - D_m(t-1)]/\tau \equiv \bar{C}_m^{min}(t). \tag{D.2}$$

Rate $\bar{C}_m^{min}(t)$ occurs when the playout buffer is empty at both the beginning and end of time slot t, but without buffer overflow during the entire time slot.

D.2 Proof of Theorem 6.1

Recall that γ_m^{min} is the SINR corresponding to the minimum required rate $C_m^{min}(t)$. Let $\bar{\gamma}_m^{min}(t)$ be the SINR corresponding to $\bar{C}_m^{min}(t)$. Since (6.2) is a monotonically increasing function, we have $0 \leq \gamma_m^{min}(t) \leq \bar{\gamma}_m^{min}(t)$.

We now consider the power assignment that achieves rates $\bar{C}_m^{min}(t)$, or, the corresponding SINRs $\bar{\gamma}_m^{min}(t)$. From (6.7) and (6.8), the minimum SINR constraint is:

$$\gamma_m(t) = \frac{G_m^m P_m(t)}{\sum_{k \neq m} G_k^m P_k(t) + \eta_m} \geq \bar{\gamma}_m^{min}(t), \ \forall\ m. \tag{D.3}$$

Eqn. (D.3) is a system of linear equations of the power vector $\vec{P}(t)$, which can be written in the matrix form as:

$$\left(\mathbf{I} - \bar{\mathbf{\Gamma}}^{min}\mathbf{A}\right)\vec{P}(t) \succeq \bar{\mathbf{\Gamma}}^{min}\vec{\nu}, \tag{D.4}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{A}$ is an $M \times M$ matrix with

$$A_{mk} = \begin{cases} 0, & m = k \\ G_k^m/G_m^m, & m \neq k, \end{cases} \tag{D.5}$$

$\bar{\mathbf{\Gamma}}^{min} = \text{diag}\{\bar{\gamma}_1^{min}(t), \bar{\gamma}_2^{min}(t), \cdots, \bar{\gamma}_M^{min}(t)\}$ is a diagonal matrix, and $\vec{\nu} = [\eta_1/G_1^1, \eta_2/G_2^2, \cdots, \eta_M/G_M^M]^T$.

Define $\mathbf{\Gamma}^{min} = \text{diag}\{\gamma_1^{min}(t), \gamma_2^{min}(t), \cdots, \gamma_M^{min}(t)\}$ and $\mathbf{\Delta} = \bar{\mathbf{\Gamma}}^{min} - \mathbf{\Gamma}^{min} \succeq \mathbf{0}$. Assume $\vec{P}$ is a power assignment that achieves $\bar{\gamma}_m^{min}(t)$ for all m, which satisfies (D.4). Substituting $\bar{\mathbf{\Gamma}}^{min} = \mathbf{\Delta} + \mathbf{\Gamma}^{min}$ into (D.4), we have

$$\left(\mathbf{I} - \mathbf{\Gamma}^{min}\mathbf{A}\right)\vec{P} \succeq \mathbf{\Gamma}^{min}\vec{\nu} + \mathbf{\Delta}\left(\vec{\nu} + \mathbf{A}\vec{P}\right).$$

Since $\mathbf{\Delta}$, $\vec{\nu}$, $\mathbf{A}$ and $\vec{P}$ all have non-negative elements, we have $\mathbf{\Delta}\left(\vec{\nu} + \mathbf{A}\vec{P}\right) \succeq \mathbf{0}$, and therefore,

$$\left(\mathbf{I} - \mathbf{\Gamma}^{min}\mathbf{A}\right)\vec{P} \succeq \mathbf{\Gamma}^{min}\vec{\nu}.$$

That is, $\vec{P}$ can also achieve $\gamma_m^{min}(t)$ for all m and it satisfies the minimum SINR constraint in (6.8).

Once the minimum SINR constraint in (6.8) (i.e., no buffer underflow) is satisfied, the maximum SINR constraint in (6.8) (i.e., no buffer overflow) can be satisfied since BS m can stop transmission when the playout buffer at user un_m is full.

Appendix E

Proofs in Chapter 7

E.1 Proof of Theorem 7.1

Due to i.i.d. channel gains and noise powers, the random variables $\eta_n(t)/G_n(t)$'s are *exchangeable*, for all t. Define $w(g,\eta,c) = (2^{c/B_w}-1)\eta/g$, which is convex and increasing with c, for all $g \geq 0$ and $\eta \geq 0$. Let $\psi(\vec{C}) = \mathbb{E}[\Phi(\vec{C})] = \mathbb{E}[\sum_{t=1}^{T} w(g(t),\eta(t),c(t))]$. $\Phi(\vec{C})$ is a symmetric, convex and increasing function in $\vec{C}$ for each fixed $\vec{G}$ and $\vec{\eta}$. According to Proposition 11.B.5 in [13], $\psi(\vec{C})$ is symmetric, convex and increasing. Following Fact 2.1, the objective function (7.9) is Schur-convex and increasing.

E.2 Proof of Corollary 7.2.3

To evaluate the smoothness of a transmission schedule $\vec{C}$, the following smoothness utility function can be used:

$$U(\vec{C}) = \sum_{t=1}^{T_n}([c(t)-\bar{c}]/T_n), \tag{E.1}$$

where $\bar{c} = \sum_{t=1}^{T_n} c(t)/T_n$ is the average rate. This is a continuous symmetric convex function $U : \mathcal{R}^{T_n} \rightarrow \mathcal{R}$. From Fact 2.1, U is Schur-convex and order preserving. The optimal power transmission schedule $\vec{C}_n^*$ satisfies $\vec{C}_n^* \prec \vec{C}_n^i$ for all i. Therefore, it also achieves the minimum value for $U(\cdot)$.

Appendix F

Acronyms

AMI	Automated Meter Infrastructure
BS	Base Station
CBR	Constant Bit Rate
CDMA	Code Division Multiple Access
CPS	Cyber-Physical System
CSMA	Carrier Sense Multiple Access
DCC	Distribution Control Center
DCPC	Distributed Constrained Power Control
DCT	Discrete Cosine Transform
DR	Demand Response
DRER	Distributed Renewable Energy Resource
DUBMLC	Distributed User Benefit Maximization Load Control
DVS	Dynamic Voltage Scaling
ESS	Energy Storage System
GSEPS	General Smooth Electric Power Scheduling
FDMA	Frequency-Division Multiple Access
HAN	Home Area Network
ICT	Information and Communications Technology
IDCT	Inverse DCT
LAN	Local Area Network
LB	Lower Bound
LP	Linear Programming

MB	Macro Block
MG	Microgrid
MGCC	MG Central Controller
PMA	Power Minimization Algorithm
PMU	Phasor Measurement Unit
PHEV	Plug-in Hybrid Electric Vehicles
PLC	Power Line Communication
QoE	Quality of Experience
QoS	Quality of Service
QoSE	Quality of Service in Electricity
RLT	Reformulation-Linearization Technique
RTP	Real-time Transport Protocol
SEPS-DL	Smooth Electric Power Scheduling for Deferrable Load
SINR	Signal to Interference-plus-Noise Ratio
SG	Smart Grid
SST	Solid State Transformer
SUDP	Supply Until Deadline Policy
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
UB	Upper Bound
UDP	User Datagram Protocol
UMRP	Utility Maximization Real-time Pricing
V2G	Vehicle-to Gird
VBR	Variable Bit Rate
VPP	Virtual Power Plant
VSN	Visual Sensor Networks
WAN	Wide Area Network

Bibliography

[1] N. Wu and X. Li, *RFID Applications in Cyber-Physical System*. InTech, 2011, ch. 16, [online] Available: http://www.intechopen.com/books/deploying-rfid-challenges-solutions-and-open-issues/rfid-applications-in-cyber-physical-system.

[2] K. Kim and P. R. Kumar, "Cyber-physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. 3, pp. 1287–1308, May 2012.

[3] H. Farhangi, "The path of the smart grid," *IEEE Power and Energy Mag.*, vol. 8, no. 1, pp. 18–28, Jan.-Feb. 2010.

[4] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid - the new and improved power grid: A survey," *IEEE Commun. Surveys & Tutorials*, vol. PP, no. 99, pp. 1–37, Dec. 2011.

[5] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang, "Smart transmission grid: Vision and framework," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 168–177, Sept. 2010.

[6] C. W. Clark, *The Smart Grid: Enabling Energy Efficiency and Demand Response*. Lilburn, GA: Fairmont Press, 2009.

[7] Energy.gov, "How the smart grid promotes a greener future," [online] Available: http://energy.gov/oe/downloads/how-smart-grid-promotes-greener-future.

[8] G. Fettweis and E. Zimmermann, "Ict energy consumption - trends and challenges," in *Proc. WPMC'08*, Lapland, Finland, Sep. 2008, pp. 1–4.

[9] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015," Feb. 2011, [online] Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/\\ns705/ns827/white_paper_c11-520862.html.

[10] C. Schaefer, C. Weber, and A. Voss, "Energy usage of mobile telephone services in Germany," *Elsevier Energy*, vol. 28, no. 5, pp. 411–420, Apr. 2003.

[11] S. Vadgama, "Trends in green wireless access," *FUJITSU Scientific Technical Journal*, vol. 45, no. 4, 2009.

[12] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[13] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York, NY: Academic Press, 1979.

[14] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.

[15] Nationalacademies.org, "Greatest engineering achievements of the 20th century," [online] Available: http://www.nationalacademies.org/greatachievements/.

[16] Whitehouse.gov, "Battery and electric vehicle report," Jul. 2010, [online] Available: http://www.whitehouse.gov/files/documents/Battery-and-Electric-Vehicle-Report-FINAL.pdf.

[17] G. Iyer, P. Agrawal, E. Monnerie, and R. S. Cardozo, "Performance analysis of wireless mesh routing protocols for smart utility networks," in *IEEE SmartGridComm'11*, Oct. 2011, pp. 114–119.

[18] Z. Zhong, C. Xu, B. J. Billian, L. Zhang, S.-J. Tsai, R. W. Conners, V. A. Centeno, A. G. Phadke, and Y. Liu, "Power system frequency monitoring network (FNET) implementation," *IEEE Trans. Power Systems*, vol. 20, no. 4, pp. 1914–1921, Nov. 2005.

[19] S. Xu, A. Q. Huang, S. Lukic, and M. E. Baran, "On integration of solid-state transformer with zonal DC microgrid," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 975–985, Jun. 2012.

[20] J. A. P. Lopes, F. J. Soares, and P. M. R. Almeida, "Integration of electric vehicles in the electric power system," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 168–183, Jan. 2011.

[21] J. Kumagai, "Virtual power plants, real power," *Spectrum IEEE*, vol. 49, no. 3, pp. 13–14, Mar. 2012.

[22] A. Q. Huang, M. L. Crow, G. T. Heydt, J. P. Zheng, and S. J. Dale, "The future renewable electric energy delivery and management (FREEDM) system: the energy internet," *Proceedings of IEEE*, vol. 99, no. 1, pp. 133–148, Jan. 2011.

[23] J. Huang, Z. Li, M. Chiang, and A. Katsaggelos, "Joint source adaptation and resource allocation for multi-user wireless video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 582–595, May. 2008.

[24] A. Brooks, E. Lu, D. Reicher, C. Spirakis, and B. Weihl, "Demand dispatch," *IEEE Power and Energy Mag.*, vol. 8, no. 3, pp. 20–29, May/Jun. 2010.

[25] A. N. Netravali and J. O. Limb, "On the performance of distributed polling service-based medium access control," *Proceedings of the IEEE*, vol. 68, no. 3, pp. 366–406, Mar. 1980.

[26] A. N. Netravali and B. G. Haskell, *Digital pictures: representation and compression.* New York, NY: Plenum Press, 1988.

[27] C. J. Sreenan, J. Chen, P. Agrawal, and B. Narendran, "Delay reduction techniques for playout buffering," *IEEE Trans. Multimedia*, vol. 2, no. 2, pp. 88–100, Jun. 2000.

[28] T. Zhang, E. van den Berg, J. Chennikara, P. Agrawal, J. Chen, and T. Kodama, "Local predictive resource reservation for handoff in multimedia wireless ip networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 1931–1941, Oct. 2001.

[29] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 7, pp. 1270–1283, Jul. 1999.

[30] P. Agrawal, E. Hyden, P. Krzyzanowski, P. Mishra, M. B. Srivastava, and J. A. Trotter, "SWAN: a mobile multimedia wireless network," *IEEE Personal Communications*, vol. 3, no. 2, pp. 18–33, Apr. 1996.

[31] S. Misra, M. Reisslein, and X. Guoliang, "A survey of multimedia streaming in wireless sensor networks," *IEEE COMMUNICATION Surveys & Tutorials*, vol. 10, pp. 18–39, 2008.

[32] I. F. Akyildiz, T. Melodia, and K. R. Chowdury, "Wireless multimedia sensor networks: A survey," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 32–39, Dec. 2007.

[33] P. Agrawal, J. Yeh, J. Chen, and T. Zhang, "Ip multimedia subsystems in 3GPP and 3GPP2: overview and scalability issues," *IEEE Commun. Mag.*, vol. 46, no. 1, pp. 138–145, Jan. 2008.

[34] A. K. Katsaggelos, F. Zhai, Y. Eisenberg, and R. Berry, "Energy-efficient wireless video coding and delivery," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 24–30, Aug. 2005.

[35] P. Agrawal, J. Chen, S. Kishore, P. Ramanathan, and K. Sivalingam, "Battery power sensitive video processing in wireless networks," in *Proc. IEEE Personal, Indoor and Mobile Radio Communications'98*, Sep. 1998, pp. 116–120.

[36] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 5, pp. 645–658, May 2005.

[37] D. N. Kwon, P. F. Driessen, A. Basso, and P. Agathoklis, "Performance and computational complexity optimization in configurable hybrid video coding system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 31–42, Jan. 2006.

[38] H. Cheng and L. Dung, "A content-based methodology for power-aware motion estimation architecture," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 52, no. 10, pp. 631–635, Oct. 2005.

[39] N. J. August and D. S. Ha, "Low power design of dct and idct for low bit rate video codecs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 414–422, Jun. 2004.

[40] Z. Cao, B. Foo, L. He, and M. van der Schaar, "Optimality and improvement of dynamic voltage scaling algorithms for multimedia applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 57, no. 3, pp. 681–690, Mar. 2010.

[41] H. M. Wang, H. S. Choi, and J. T. Kim, "Workload-based dynamic voltage scaling with the qos for streaming video," in *Proc. IEEE Electronic Design, Test and Applications'08*, Jan. 2008, pp. 236–239.

[42] M. Li, Z. Guo, R. Y. Yao, and W. Zhu, "A novel penalty controllable dynamic voltage scaling scheme for mobile multimedia applications," *IEEE Trans. Mobile Computing*, vol. 5, no. 12, pp. 1719–1733, Dec. 2006.

[43] A. Goldsmith, *Wireless Communications*. New York, NY: Cambridge University Press, 2005.

[44] IEEE, "Part 11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)," July 2003, ANSI/IEEE Std 802.11e, Draft 5.0.

[45] Y. Huang, P. Walsh, Y. Li, and S. Mao, "A gnu radio testbed for distributed polling service-based medium access control," in *Proc. IEEE MILCOM'11*, Baltimore, MD, Nov. 2011.

[46] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.

[47] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio system," in *Proc. 4th Winlab Workshop on Third Generation Wireless Information Networks*, New Brunswick, NJ, Oct. 1993, pp. 249–257.

[48] S. Grandhi, J. Zander, and R. Yates, "Constrained power control," *Int. J. Wireless Personal Commun.*, vol. 1, no. 4, pp. 257–270, Apr. 1995.

[49] P. Hande, S. Rangan, M. Chiang, and X. Wu, "Distributed uplink power control for optimal sir assignment in cellular data networks," *IEEE/ACM Trans. Networking.*, vol. 16, no. 6, pp. 1420–1433, Dec. 2008.

[50] J. Lee, R. Mazumdar, and N. Shroff, "Downlink power allocation for multi-class wireless systems," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 854–867, Aug. 2005.

[51] N. Bambos, S. C. Chen, and G. J. Pottie, "Radio link admission algorithm for wireless networks with power control and active link quality protection," in *Proc. IEEE INFOCOM'95*, Boston, MA, Apr. 1995, pp. 97–104.

[52] M. Xiao, N. B. Shroff, and E. K. P. Chong, "Distributed admission control for power-controlled cellular wireless systems," *IEEE/ACM Trans. Networking*, vol. 9, no. 6, pp. 790–800, Dec. 2001.

[53] C. Comaniciu and H. V. Poor, "Jointly optimal power and admission control for delay sensitive traffic in cdma networks with lmmse receivers," *IEEE Trans. Signal Processing*, vol. 51, no. 8, pp. 2031–2042, Aug. 2003.

[54] G. Liang and B. Liang, "Balancing interruption frequency and buffering penalties in VBR video streaming," in *Proc. IEEE INFOCOM'07*, Anchorage, AK, May 2007, pp. 1406–1414.

[55] Y. Huang, S. Mao, and Y. Li, "Downlink power allocation for stored variable-bit-rate videos," in *Proc. ICST QShine'10*, Houston, TX, Nov. 2010, pp. 423–438.

[56] Y. Huang and S. Mao, "Downlink power control for variable bit rate video over multicell wireless networks," in *Proc. IEEE INFOCOM'11*, Shanghai, China, Apr. 2011, pp. 2561–2569.

[57] Y. Huang, S. Mao, and Y. Li, "Downlink power control for VBR video streaming in cellular networks: A majorization approach," in *Proc. IEEE Globalcom'11*, Houston, TX, Dec. 2011, pp. 1–6.

[58] ——, "On downlink power allocation for multiuser variable-bit-rate video streaming," *Wiley Security and Communication Networks*, to appear.

[59] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: jointly optimal congestion control and power control," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.

[60] S. Mao, S. Lin, Y. Wang, S. S. Panwar, and Y. Li, "Multipath video transport over wireless ad hoc networks," *IEEE Wireless Commun.*, vol. 12, no. 4, pp. 42–49, Aug. 2005.

[61] S. Mao, Y. T. Hou, H. D. Sherali, and S. F. Midkiff, "Multimedia-centric routing for multiple description video in wireless mesh networks," *IEEE Network*, vol. 22, no. 1, pp. 19–24, Jan./Feb. 2008.

[62] H. Y. Shutoy, D. Gunduz, E. Erkip, and Y. Wang, "Cooperative source and channel coding for wireless multimedia communications," *IEEE J. Sel. Signal Process.*, vol. 1, no. 2, pp. 295–307, Aug. 2007.

[63] T. Koponen, T. Koponen, B. Chun, A. Ermolinskiy, K. H. Kim, and S. Shenker, "A data-oriented (and beyond) network architecture," in *ACM SIGCOMM'07*, Kyoto, Japan, Agu. 2007, pp. 181–192.

[64] X. Lu, E. Erkip, Y. Wang, and D. Goodman, "Power efficient multimedia communication over wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1738–1751, Dec. 2003.

[65] Y. Eisenberg, C. E. Luna, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 411–424, Jun. 2002.

[66] Y. Li, M. Reisslein, and C. Chakrabarti, "Energy-efficient video transmission over a wireless link," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1229–1244, Mar. 2009.

[67] Z. Li, F. Zhai, and A. K. Katsaggelos, "Joint video summarization and transmission adaptation for energy-efficient wireless video streaming," *EURASIP J. on Advances in Signal Processing*, vol. 2008, pp. 1–11, Jan. 2008.

[68] D. Wu, S. Ci, and H. Wang, "Cross-layer optimization for video summary transmission over wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 841–850, May 2007.

[69] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Advances in Multimedia*, vol. 2009, 2009, article ID 640386, 21 pages, doi:10.1155/2009/640386.

[70] A. Seema and M. Reisslein, "Towards efficient wireless video sensor networks: A survey of existing node architectures and proposal for a flexi-wvsnp design," *IEEE Commun. Surveys Tutorials*, vol. 13, no. 3, pp. 462–486, Quarter 2011.

[71] M. Chen, T. Kwon, S. Mao, Y. Yuan, and V. C. M. Leung, "Reliable and energy-efficient routing protocol in dense wireless sensor networks," *International Journal of Sensor Networks*, vol. 4, no. 1/2, pp. 104–117, 2008.

[72] M. Chen, V. C. M. Leung, and S. Mao, "Directional controlled fusion in wireless sensor networks," *ACM/Springer MONET*, vol. 14, no. 2, pp. 220–229, Apr. 2009.

[73] ——, "Directional controlled fusion in wireless sensor networks," in *Proc. QShine'08*, Hong Kong, P.R. China, Jul. 2008, pp. 1–7.

[74] J. Liu and S. Singh, "Atcp: Tcp for mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 7, pp. 1300–1315, Jul. 2001.

[75] V. Tsaoussidis and H. Badr, "Tcp-probing: towards an error control schema with energy and throughput performance gains," in *Proc. Network Protocols'00*, 2000, pp. 12–21.

[76] D. Hu, S. Mao, Y. T. Hou, and J. H. Reed, "Fine grained scalability video multicast in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, Apr. 2010.

[77] D. Hu and S. Mao, "Streaming scalable videos over multi-hop cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3501–3511, Nov. 2010.

[78] Y. Zhao, S. Mao, J. H. Reed, and Y. Huang, "Experimental study of utility function selection for video over cognitive radio networks," in *Proc. TRIDENTCOM 2009*, Washington D.C., Apr. 2009, pp. 1–10.

[79] ——, "Utility function selection for streaming videos with a cognitive engine testbed," *ACM/Springer Mobile Networks and Applications Journal*, vol. 15, no. 3, pp. 446–460, June 2010.

[80] Y. Huang, S. Mao, and R. M. Nelms, "Smooth electric power scheduling in power distribution networks," in *Proc. IEEE GLOBECOM 2012 – Workshop on Smart Grid Communications: Design for Performance*, Anaheim, CA, Dec. 2012, pp. 1–5.

[81] M. LeMay, R. Nelli, G. Gross, and C. A. Gunter, "An integrated architecture for demand response communications and control," in *HICSS '08*, Washington, DC, 2008, pp. 174–183.

[82] Y. Huang, S. Mao, and R. M. Nelms, "Adaptive electricity scheduling in microgrids," in *Proc. IEEE INFOCOM 2013*, Turin, Italy, Apr. 2013.

[83] Y. Huang and S. Mao, "Adaptive electricity scheduling with quality of usage guarantees in microgrids," in *Proc. IEEE GLOBECOM 2012*, Anaheim, CA, Dec. 2012, pp. 1–6.

[84] ——, "On qualify of usage provisioning for electricity scheduling in microgrids," *IEEE Systems Journal, Special Issue on Smart Grind Communications Systems*, 2013, to appear.

[85] S. Mao, Y. Huang, and Y. Li, "Downlink vbr video scheduling in cellular networks with orthogonal channels," *IEEE E-Letter for MMTC*, vol. 6, no. 9, pp. 40–42, Sept. 2011.

[86] Y. Huang, S. Mao, and Y. Li, "A majorization approach to downlink power control for vbr videos over cellular networks," *Elsevier Computer Communications*, vol. 35, no. 15, pp. 1828–1837, Sept. 2012.

[87] ——, *Green video streaming over cellular networks*. New York, NY: CRC Press, 2012, ch. 23, pp. 659–690.

[88] G. T. Heydt, "The next generation of power distribution systems," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 225–235, Dec. 2010.

[89] S. Shao, M. Pipattanasomporn, and S. Rahman, "Demand response as a load shaping tool in an intelligent grid with electric vehicles," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 624–631, Dec. 2011.

[90] A. H. Mohsenian-Rad and A. Leon-Garcia, "Optimal residential load control with price prediction in real-time electricity pricing environments," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 120–133, Sept. 2010.

[91] C. Ibars, M. Navarro, and L. Giupponi, "Distributed demand management in Smart Grid with a congestion game," in *Proc. IEEE SmartGridComm'10*, Gaithersburg, MD, Oct. 2010, pp. 495–500.

[92] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *2011 IEEE PES General Meeting*, Detroit, MI, Jul. 2011, pp. 1–8.

[93] M. G. Kallitsis, G. Michailidis, and M. Devetsikiotis, "Optimal power allocation under communication network externalities," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 162–173, Mar. 2012.

[94] J. Medina, N. Muller, and I. Roytelman, "Demand response and distribution grid operations: Opportunities and challenges," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 193–198, Sept. 2010.

[95] J. Salehi, Z.-L. Zhang, J. Kurose, and D. Towsley, "Supporting stored video: reducing rate variability and end-to-end resource requirements through optimal smoothing," *IEEE/ACM Trans. Networking.*, vol. 6, no. 4, pp. 397–410, Aug. 1998.

[96] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[97] A. Papavasiliou and S. S. Oren, "Supplying renewable energy to deferrable loads: Algorithms and economic analysis," in *Proc. IEEE Power & Energy Society General Meeting'10*, Jul. 2010, pp. 1–8.

[98] B. C. Arnold, *Majorization and the Lorenz Order: A Brief Introduction*. New York, NY: Springer-Verlag, 1987.

[99] A. Keyhani, *Design of Smart Power Grid Renewable Energy Systems*. Hoboken, NJ: Wiley-IEEE Press, 2011.

[100] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[101] M. Fahrioglu and F. L. Alvarado, "Designing incentive compatible contracts for effective demand management," *IEEE Trans. Power Systems*, vol. 15, no. 4, pp. 1255–1260, Nov. 2000.

[102] P. Samadi, A. Mohsenian-Rad, R. Schober, V. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximization for smart grid," in *IEEE SmartGridComm'10*, Oct. 2010, pp. 415–420.

[103] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. New York: John Wiley & Sons, 1984.

[104] F. Rahimi and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 82–88, Jun. 2010.

[105] S. Amin and B. Wollenberg, "Toward a smart grid: power delivery for the 21st century," *IEEE Power and Energy Mag.*, vol. 3, no. 5, pp. 1–37, Sept.-Oct. 2005.

[106] A. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[107] H. K. Nguyen, J. B. Song, and Z. Han, "Demand side management to reduce peak-to-average ratio using game theory in smart grid," in *Proc. IEEE INFOCOM Workshops'12*, Orlando, FL, Mar. 2012, pp. 91–96.

[108] J. Huang, C. Jiang, and R. Xu, "A review on distributed energy resources and microgrid," *ELSEVIER Renewable and Sustainable Energy Reviews*, vol. 12, no. 9, pp. 2472–2483, Dec. 2008.

[109] E. Fumagalli, J. W. Black, I. Vogelsang, and M. Ilic, "Quality of service provision in electric power distribution systems through reliability insurance," *IEEE Trans. Power Systems*, vol. 19, no. 3, pp. 1286–1293, Aug. 2004.

[110] J. Slotine and W. Li, *Applied Nonlinear Control*. Prentice Hall, 1991.

[111] T. T. Kim and H. H. V. Poor, "Scheduling power consumption with price uncertainty," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 519–527, Sept. 2011.

[112] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 89–103, Jan. 2005.

[113] M. J. Neely, "Stock market trading via stochastic network optimization," in *Proc. IEEE CDC'10*, Atlanta, GA, Dec. 2010, pp. 1–8.

[114] M. J. Neely, A. S. Tehrani, and A. G. Dimakis, "Efficient algorithms for renewable energy allocation to delay tolerant consumers," in *IEEE SmartGridComm'10*, Oct. 2010, pp. 549–554.

[115] M. J. Neely and R. Urgaonkar, "Opportunism, backpressure, and stochastic optimization with the wireless broadcast advantage," in *Asilomar Conference on Signals, Systems, and Computers'08*, Pacific Grove, CA, Oct. 2008, pp. 1–7.

[116] The National Renewable Energy Laboratory, "Western wind resources dataset," [online] Available: http://wind.nrel.gov/Web_nrel/.

[117] S. B. Peterson, J. F. Whitacre, and J. Apt, "The economics of using plug-in hybrid electric vehicle battery packs for grid storage," *J. Power Sources*, vol. 195, no. 8, pp. 2377–2384, 2010.

[118] "The Electric Reliability Council of Texas," [online] Available: http://www.ercot.com/.

[119] M. He, S. Murugesan, and J. Zhang, "Multiple timescale dispatch and scheduling for stochastic reliability in smart grids with wind generation integration," in *Proc. IEEE INFOCOM'11*, Shanghai, China, Apr. 2011, pp. 461–465.

[120] B. Karimi and V. Namboodiri, "Capacity analysis of a wireless backhaul for metering in the smart grid," in *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar. 2012, pp. 61–66.

[121] Z. Lu, W. Wang, and C. Wang, "Hiding traffic with camouflage: Minimizing message delay in the smart grid under jamming," in *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar. 2012, pp. 3066–3070.

[122] M. A. Rahman, P. Bera, and E. Al-Shaer, "Smartanalyzer: A noninvasive security threat analyzer for AMI smart grid," in *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar. 2012, pp. 2255–2263.

[123] H. Ma, H. Li, and Z. Han, "A framework of frequency oscillation in power grid: Epidemic propagation over social networks," in *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar. 2012, pp. 67–72.

[124] H. Liang, B. J. Choi, W. Zhuang, and X. Shen, "Towards optimal energy store-carry-and-deliver for PHEVs via V2G system," in *Proc. IEEE INFOCOM'12*, Orlando, FL, Mar. 2012, pp. 1674–1682.

[125] X. Liu, "Economic load dispatch constrained by wind power availability: A wait-and-see approach," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 347–355, Dec. 2010.

[126] X. Fang, D. Yang, and G. Xue, "Online strategizing distributed renewable energy resource access in islanded microgrids," in *IEEE GLOBECOM'11*, Huston, TX, Dec. 2011, pp. 1931–1937.

[127] I. Koutsopoulos, V. Hatzi, and L. Tassiulas, "Optimal energy storage control policies for the smart power grid," in *IEEE SmartGridComm'11*, Oct. 2011, pp. 475–480.

[128] S. X. Chen, H. B. Gooi, and M. Q. Wang, "Sizing of energy storage for microgrids," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 142–151, Mar. 2012.

[129] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal power cost management using stored energy in data centers," in *Proc. ACM SIGMETRICS'11*, San Jose, CA, Jun. 2011, pp. 221–232.

[130] H. Sistek, "Green-tech base stations cut diesel usage by 80 percent," Apr. 2008, [online] Available: http://news.cnet.com/8301-11128_3-9912124-54.html.

[131] M. W. Garrett and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 24, no. 4, pp. 269–280, 1994.

[132] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1566–1579, Feb./Mar./Apr. 1995.

[133] D. P. Heyman and T. V. Lakshmanr, "What are the implications of long-range dependence for VBR-video traffic engineering?" *IEEE/ACM Trans. Networking*, vol. 4, no. 3, pp. 301–317, June 1996.

[134] T. Lakshman, A. Ortega, and A. Reibman, "VBR video: Trade-offs and potentials," *Proc. IEEE*, vol. 86, no. 1, pp. 952–973, May 1998.

[135] S. Liew and H. Chan, "Lossless aggregation: a scheme for transmitting multiple stored VBR video streams over a shared communications channel without loss of image quality," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 6, pp. 1181–1189, Aug. 1997.

[136] S. Sen, D. Towsley, Z. Zhang, and J. K. Dey, "Optimal multicast smoothing of streaming video over the internet," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 7, pp. 1345–1359, Sep. 2002.

[137] J. Lee and J. Kwon, "Utility-based power allocation for multiclass wireless systems," *IEEE Trans. Vehic Tech.*, vol. 58, no. 7, pp. 3813–3819, Sep. 2009.

[138] A. Gjendemsj, D. Gesbert, G. Oien, and S. Kiani, "Binary power control for sum rate maximization over multiple interfering links," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3164–3173, Aug. 2008.

[139] M. Reisslein, "Video trace library," Arizona State University, [online] Available: http://trace.eas.asu.edu/.

[140] UMTS World, *UMTS Power Control*, [online] Available: http://www.umtsworld.com/technology/power.htm.

[141] J. M. Mcmanus and K. W. Ross, "A dynamic programming methodology for managing prerecorded vbr sources in packet-switched networks," in *Proc. SPIE, Performance and Control of Network Systems*, 1997, pp. 140–154.

[142] S. Deng, T. Webera, and A. Ahrens, "Capacity optimizing power allocation in interference channels," *AEU - International Journal of Electronics and Communications*, vol. 63, no. 2, pp. 139–147, Jan. 2009.

[143] T. Stockhammer, H. Jenkac, and G. Kuhn, "Streaming video over variable bit-rate wireless channels," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 268–277, Apr. 2004.

[144] M. Chen and A. Zakhor, "Multiple TFRC connections based rate control for wireless networks," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 1045–1062, Oct. 2006.

[145] H. D. Sherali and W. P. Adams, *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*. Dordrecht, Boston, London: Kluwer Academic Publishers, 1999.

[146] S. Kompella, S. Mao, Y. Hou, and H. Sherali, "On path selection and rate allocation for video in wireless mesh networks," *IEEE/ACM Trans. Networking.*, vol. 17, no. 1, pp. 212–224, Feb. 2009.

[147] M. Andersin, Z. Rosberg, and J. Zander, "Gradual removals in cellular PCS with constrained power control and noise," in *Proc. IEEE PIMRC'95*, Toronto, Canada, Sept. 1995, pp. 56–60.

[148] A. Naman and D. Taubman, "JPEG2000-based scalable interactive video (JSIV)," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1435–1449, May 2011.

[149] W.-K. Liao and Y.-H. Lai, "Type-aware error control for robust interactive video services over time-varying wireless channels," *IEEE Trans. Mobile Comput.*, vol. 10, no. 1, pp. 136–145, Jan. 2011.

[150] G. Cheung, A. Ortega, and N.-M. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 744–761, Mar. 2011.

[151] A. Dan, P. Shahabuddin, D. Sitaram, and D. Towsley, "Channel allocation under batching and vcr control in movie-on-demand servers," *J. Parallel Distrib. Comput.*, vol. 30, pp. 168–179, 1995.

[152] buto.tv, "Feature: Clickable videointeractive, clickable video that anyone can build in minutes," [online] Available: http://get.buto.tv/features/clickable-video/.