Approaches to Enable Demand Response by Industrial Loads for Ancillary Services Provision

Submitted in partial fulfillment of the requirements for
the degree of

Doctor of Philosophy
in
Electrical and Computer Engineering

Xiao Zhang

B. S., Electrical Engineering, Tsinghua University

Carnegie Mellon University
Pittsburgh, PA
January, 2017

This dissertation is dedicated to my parents and my wife.

Abstract

Demand response has gained significant attention in recent years as it demonstrates potentials to enhance the power system's operational flexibility in a cost-effective way. Industrial loads such as aluminum smelters, steel manufacturers, cement plants, and air separation units demonstrate advantages in supporting power system operation through demand response programs, because of their intensive power consumption, already existing advanced monitoring and control infrastructure, and the strong economic incentive due to the high energy costs. In this thesis, we study the three aforementioned manufacturing processes each with its own capabilities and constraints. We provide approaches to efficiently integrate each of these types of manufacturing processes as demand response resources.

Aluminum smelting is an energy-intensive electrolytic process that is widely used to produce aluminum. The electricity cost thereby constitutes a significant portion of the total operation cost. At the same time, the smelting process is able to change its power consumption both accurately and quickly by controlling the pots' DC voltage, without affecting the production quality. Hence, an aluminum smelter has both the motivation and the ability to participate in demand response. First, we focus on determining the optimal regulation capacity that such a manufacturing plant should provide to maximize the combined profit from producing aluminum and providing regulation. The approach is based on stochastic programming and the stochastic variable is the regulation signal sent to the smelter. Next, we focus on determining the optimal bidding strategy in the day-ahead energy and spinning reserve markets for an aluminum smelter. By bidding into the electricity market, the smelter provides flexibility to the power system operator and gets compensation which reduces the overall electricity cost. Finally, we focus on industrial loads which provide both energy and regulation service and study the optimal bidding strategy to maximize their revenues in the day-ahead markets. The approach is based on stochastic programming with a set of possible price curves as input, which represents the possible price scenarios for the next day. We also use a recently proposed method called the Multiple Quantile Graphical Model to represent the price distributions and use Gibbs sampling to sample the price curves.

Electric arc furnaces (EAFs) in steel manufacturing consume a large amount of electric energy, and the energy cost constitutes a significant proportion of the total costs of producing steel. However, a steel plant can take advantage of time-based electricity prices by optimally arranging energy-consuming activities to avoid peak hours. Besides, the EAFs' power rate can be adjusted by switching transformers' taps, which offers additional flexibility for arranging energy consumption and minimizing the cost of electricity. We first propose scheduling methods that incorporate the EAFs' flexibilities to reduce the electricity cost. However, the scheduling of steel plants is very

complex and the involved computations are intense. Hence, we then focus on these difficulties and propose methods such as adding additional constraints as cuts and implementing an application-specific branch and bound algorithm to make the computations more tractable. Finally, since the steel plants are very flexible in terms of adjusting their power consumption rate through switching the transformer tap position, we extend the scheduling formulations to enable the provision of spinning reserve.

As typical industrial loads, cement plants are able to quickly adjust their power consumption rate by switching on/off the crushers. However, in the cement plant as well as other industrial loads, switching on/off the loading units only achieves discrete power changes, which restricts the load from offering valuable ancillary services such as regulation and load following, as continuous power changes are required for these services. We propose methods that enable these loads to provide regulation or load following with the support of an on-site energy storage device. Secondly, we propose scheduling methods for the industrial plant to determine the optimal regulation capacity for each operating hour to maximize its daily revenue. The scheduling approach takes into account the revenues from market participation, the cost of regulation provision, as well as the coupling of the crushing process with other processing stages within the industrial plant.

As demonstrated by the case studies presented in this thesis, the proposed approaches are effective and can generate practical production instructions for the industrial loads. This thesis not only provides methods to enable demand response by industrial loads but also potentially encourages industrial loads to be active in electricity markets.

Acknowledgments

This dissertation comprises the results I have obtained during my Ph.D. research at Carnegie Mellon University. My Ph.D. research is sponsored by ABB, a pioneering technology leader in industrial digitalization. I want to first express my gratitude to CMU, for providing such a great research atmosphere, and to ABB, for sponsoring the project.

The research presented in this thesis would not have been possible without the contributions and support of my advisors, Prof. Gabriela Hug and Prof. J. Zico Kolter. I want to gratefully thank Gabriela for providing me such a valuable opportunity, for always holding me up to high standards, for her precise logic and critical thinking, for her permanent support, guidance, and trust. I want to gratefully thank Zico for sharing his excellent machine learning knowledge, for his perceptive insights and profound perspectives in approaching the problems, for his constant help, patience, and encouragement.

I am also very grateful for my thesis committee members, Dr. Iiro Harjunkoski, Prof. Ignacio E. Grossmann, and Prof. Anthony Rowe. To Iiro, for constantly providing inputs and suggestions for my research throughout all these years, and for bringing in so much knowledge and practical experience from the industrial operations. To Ignacio, for being a great lecturer from whom I have learned a lot about industrial scheduling and integer programming, and for being a great researcher who has devoted decades of diligent work and has obtained enormous achievements which always inspire me to work hard. To Anthony, for providing the opportunity to participate in the micro-grid course project in which my power system knowledge has a chance to help people living in the rural areas, for all those interesting discussions and the cool ideas in your lab which broaden my mindset. I also want to say thank you to Dr. Alf Isaksson, Dr. Xiaoming Feng, and Dr. Marija Zima from ABB for the insights and expert opinions throughout our quarterly project meetings all these years.

Thank you to Prof. Marija D. Ilić, Prof. John M. Dolan, Prof. Ozan Tonguz, Prof. Bruno Sinopoli, Prof. Xin Li, Prof. Soummya Kar, Prof. Aarti Singh, Prof. Barnabs Pczos, Prof. Larry Wasserman, Prof. Ryan Tibshirani for organizing the great courses at CMU and for all those warm discussions on study or research. To Claire E. Bauerle, for her helpful assistance over the years and taking care of so many errands ranging from room reservations to travel reimbursement.

I have also spent a great time with the other members in Gabriela's group and Zico's group. Thank you to Rui Yang, Dinghuan Zhu, Kyri Baker, Javad Mohammadi, Andrew Hamann, Hameed Safiullah, Chenye Wu, Junyao Guo, Dmitry Shchetinin, Jesse Thornburg, Mohsen Rahmani, Amin Kargarian, for keeping up the good research, for helping me with my qualification exam and grammar checking, for driving to conferences together and having fun there, for working together to assist Gabriela's optimization course. Thank you to Matt Wytock, Alnur Ali, Po-Wei Wang, Eric

Wong, Brandon Amos, for the inspiration to explore machine learning, for discussing the details of theories and methods, for sharing good papers and reading materials, for all those interesting chats.

I also want to say thank you to many friends whom I met with in this nice city of Pittsburgh. To Prof. Jimmy Hsia, for all of your support and trust. To Tao Cui, Qixing Liu, Yang Weng, Yilin Mo, Liangyan Gui, Xiaoqi Yin, Xia Miao, Yaoqing Yang, Jonathan Donadee, Jhi-Young Joo, Max Buevich, Aurora Schmidt, Joya Deri, Stephen Kruzick, Subhro Das, June Zhang, Kevin Bachovchin, Andrew Hsu, Shanghang Zhang, and all others from the Porter Hall community for spending so much time together, for keeping the community lively and friendly. To Thom Popovici, Richard Veras, Marie Nguyen, Yu Wang, Guanglin Xu, Jiyuan Zhang, Milda Zizyte, John Filleau, Zhipeng Zhao, Jing Huang and all other Hamerschlag Hall A-300 residents, for keeping everything in order, for staying up late together and all those happy chats. To Dr. Pedro Castro, Qi Zhang, Xue Yang, Jun Shi, Yajun Wang, Hubert Hadera, Dragoljub Gajic, Raymond Todd, Brian Helms, and all those friends from the Chemical Engineering field, for helping me better understand the industrial processings, for all those enjoyable discussions. To Wenda Xu, Yi Zhou, Yang Hu, Fa Wang, Yuxiong Wang, Fan Tong, Tianyu Gu, Chiyu Dong, Danyang Li, Wenlu Hu, Yuanyuan Feng, Qi Guo, Weiguang Mao, Song Luan, Jiaxing Cui, and all other friends, for going to restaurants together, for celebrating festivals together, for organizing events together, for all those happy and warm moments in Pittsburgh which I will always keep in my memory.

Looking back, I also feel grateful for the people that I have worked with at Tsinghua University. To Prof. Qiang Lu, Prof. Shengwei Mei, Prof. Chen Shen, Prof. Guangyu He, Prof. Feng Liu, Prof. Ying Chen, Prof. Xuemin Zhang, Dr. Feng Gao, Shufeng Dong, Xinyi Su, Shaowei Huang, Laijun Chen, Yin Xu, Wei Wei, Guanqun Wang, Ying Li, Bin Wang, De Zhang, Wentao Guo, Min Zhao, and all others, for all the guidance and help, for the happy moments in our lab trips and festival celebrations, for working hard together, for all those support and understanding. Thank you to all my classmates and friends at Tsinghua, for all those memories that warm my heart, and I wish I could have spent more time with you. Thank you to Peng Liu, Zhen Geng, Kaicheng Liu, Xian Qin, and all other friends from my high school, for our friendship over the decade. Thank you to Xijun Gu, Bo Li, Sichao Zhu, Wensheng Sui, Laiqing Huang, Cheng Hong, and all of my teachers in my middle and high schools, for your teaching and your confidence in me.

Finally, I own the greatest gratitude to my family. To my mom, Sheng Dong, my dad, Xinhua Zhang, and my brother, Yang Zhang, I want to thank you for all your support and love. To my beloved wife, Yiqun Cao, thank you for your unconditional love, support, and your companion in Pittsburgh and beyond. I cannot image going through all these years without you by my side.

Thanks to all of you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The main goal of this dissertation is to provide methods for industrial loads to support power grid operation through demand response, especially by providing ancillary services such as spinning reserve, regulation and load following/tracking; at the same time, these methods are able to reduce the electricity cost of the industrial loads, supporting the economic development of the manufacturing industry. Section 1.1 presents the motivation for the thesis, explaining the basic idea of the research topic and why this topic is important and interesting. Section 1.2 formally states the problems that we are investigating and the expected research results from this dissertation. The contributions of this dissertation are summarized in Section 1.3, and the thesis outline is presented in Section 1.4.

1.1 Motivation

Increased operational flexibility is an inherent characteristic of what is commonly referred to as the smart grid. This is because a large share of renewable generation resources such as wind and solar generation are expected to be deployed to enable a sustainable energy future. However, the power output of these renewable resources is intermittent and uncertain which requires significant amounts of balancing resources to increase the operational flexibility of the grid. Traditionally, the power system relies on generators to provide such flexibility, but it is not economical for generators to frequently change their output. Realizing the balancing potential from the demand side, electricity providers start offering economic incentives to encourage demands to change their electricity usage behavior, and thereby help maintaining the supply-demand balance. Hence, demand response has gained significant attention in recent years as it demonstrates potentials to enhance the power system's operational flexibility in a cost-effective way [1–5]. Additional benefits include that demand response is beneficial for the power grid as it helps mitigating daily supply and transmission

bottlenecks and slowing down the needs for constructing more generation capacity.

There have been intensive discussions and promising solutions for demand response provided by electric vehicles [6–8], residential areas [9–11], buildings [12–18], data centers[19, 20], aluminum smelters [21, 22], and steel plants [23–25]. However, for many potential players, especially the smaller loads in the residential and commercial areas, it may not be profitable to participate in demand response, as the payments are usually not high enough for them to justify the investment in implementing the platforms for participation. The investment includes the infrastructure to enable the control, monitor, and communication of the appliances. Besides, it is difficult for the electricity consumers in these areas to sacrifice their living or working convenience either. Meanwhile, industrial loads, given their large energy consumption, are ideal candidates for providing demand response as they have both the ability to provide demand response as well as the motivation to do so [26–29]:

- many industrial loads are able to offer large, fast, and accurate adjustments in their power consumption;
- most industrial plants are already equipped with the infrastructures for control, measurement, and communications that are required for demand response providers;
- the demand response programs are also financially appealing to the industrial plants, especially to those energy-intensive plants who treat demand response as an opportunity to increase their profits by making full use of their assets.

The range of industrial loads that can support the operation of the electric power system include aluminum smelting pots, steel melting furnaces, cement crushers, fans, freezers, pumps, etc.

On the other hand, there are still challenges and difficulties for industrial loads to participate as demand response resources:

- First of all, there are critical production constraints that need to be respected and satisfied to ensure operation feasibility when providing demand response, especially when participating in an electricity market with a lot of uncertainties such as market prices and instantaneous control commands. For example, the temperature of the aluminum smelting pots must be maintained within given bounds to ensure production efficiency and equipment safety, which needs to be satisfied no matter what the electricity price is.
- Secondly, the industrial production process is usually very complex with multiple process stages and parallel equipment, and consequently the computations associated with the demand response provision by industrial loads are often very intense: the associated computation is usually a large scale optimization problem with a lot of integer variables that is very

difficult to solve. For instance, it is recognized that steel manufacturing is one of the most difficult industrial processes for scheduling [30], and the daily scheduling of a typical steel plant for demand response provision may need several hours of computation.

- Thirdly, even though the magnitude of power consumption by an industrial plant and the change in power it can provide are generally very large, its granularity is usually not fine enough which restricts it from fully contributing its potential flexibility. Many industrial loads are able to provide very fast change of power in both directions, qualifying them for regulation and load following. For example, the crushers or mills in the cement industry can be switched on and off very rapidly. However, most of these industrial loads can only provide power changes in a discrete manner, e.g. the power change is several MWs at a time. This coarse granularity hinders those industrial loads from providing the most valuable ancillary services, as the regulation and load following in the current electricity markets require a continuous change of power.

Realizing the opportunities brought by demand response, the power and energy communities have been investigating demand response and have proposed a range of promising solutions. For example, utilities are developing all kinds of demand response programs to encourage electricity consumers to actively participate in the supply/demand balancing [31, 32]; the Southern California Edison has promoted programs such as the Agricultural and Pumping Interruptible Program to allow the utility to temporarily suspend electricity to the pumping equipment, the Automated Demand Response to automatically reduce energy usage during demand response events, etc. [33]; the Midcontinent Independent System Operator encourages its loads to take part in Demand Response Resources (DRRs), Load Modifying Resources (LMRs), Emergency Demand Response (EDR), Aggregators of Retail Customers (ARCs), Energy Efficiency (EE), and Price Responsive Demand (PRD) [34]; [6] and [35] investigate the coordination of electric vehicle charging with other controllable loads and co-generation units; [18] and [17] discuss the demand response provided by commercial buildings; [36] and [37] investigate the interaction and coordination between utilities and customers. The integration of these demand resources into power systems and their interactions have been intensively studied. However, most of these works focus on the residential and commercial loads such as electric vehicles and buildings, but neglect the industrial loads which constitute a great portion of the total loads and have a lot of potential as demand response resources.

On the other hand, the chemical engineering community has also been investigating the potentials from industrial plants as demand response resources. The industrial Demand Side Management (DSM) has been an active research topic which has been recognized as an effective approach to improving consumer benefits and supporting power system operations [38–40]. The industrial DSM

studies both how to improve the energy efficiency within the plant and how to enable demand response for the power system; in this thesis, we focus more on the latter. For example, scheduling methods for industrial plants taking into account the impacts of electricity markets and varying electricity prices are proposed in [41] and [42]; for the steel manufacturing industry, how to track a pre-specified energy curve is investigated in [43] and [44], and the optimal scheduling of production activities are studied by resource-task network models in [45]; for the cement crushing industry, the energy cost minimization for cement plants under time-based electricity prices are investigated in [46]. Nevertheless, most of these works only focus on the electric energy markets, but do not consider the ancillary service markets[1] which require more flexibilities from the loads and more interactions with the power system. Taking both sides into account, we have identified a research gap which is the integration of industrial demand response resources into the power grid, especially the participation of industrial loads in the ancillary service markets.

With the opportunities and challenges discussed above, in this thesis, we focus on developing tools and methods that help industrial loads to fully achieve their potential as demand response participants, especially in the ancillary service markets. For each of the three aforementioned challenges, we select one representative industry as an example to investigate: the aluminum smelting, the steel manufacturing, and the cement crushing. We study these three industries because they are representative of the industrial loads which have both the ability and the motivation to provide demand response. Note that the approaches developed for these three industries also apply to other industries. For instance, the Chloralkali process is another electrolysis process that is similar to aluminum smelting, which has also been identified as a potential candidate for demand response resource in the ancillary service markets [47]; the air separating industry consumes large amounts of energy but the process is very complex to schedule [42], which shares similar opportunities and challenges with steel manufacturing; similar to the crushers and mills in the cement industry, the refiners in the thermal mechanical pulping process can also be switched on and off very rapidly, which sees great potential in the ancillary service market [47, 48]. With the approaches developed in this dissertation, a lot of industrial loads will be able to utilize their potentials and contribute to the power system operation through demand response.

1.2 Problem Statement

In the previous section, we have described the three challenges for industrial loads to provide demand response. In this section, we further discuss these challenges within the context of the representative industries, and formally state the problems that we are interested in.

[1]The ancillary services are introduced described in Section 2.1.

1.2.1 Account for Uncertainty - Aluminum Smelter

Aluminum smelting is the electrolytic process that transforms alumina to aluminum - the most widely used non-ferrous metal that is used anywhere from making cars to packaging cans. In the smelting plant, the electrolytic process takes place in the so-called cell and is enabled by a DC electric current that passes through the cell. A typical cell is displayed in Fig 1.1. In addition to the main material alumina, several other elements are added to facilitate the chemical reaction. The cells, or pots, are connected in series to form a potline of hundreds of pots. The total power consumption of a potline can be hundreds of MWs. Typically, there are several potlines in an aluminum smelter. Aluminum smelting is an energy-intensive electrolytic process and electricity cost thereby constitutes a significant portion of the total operation cost.

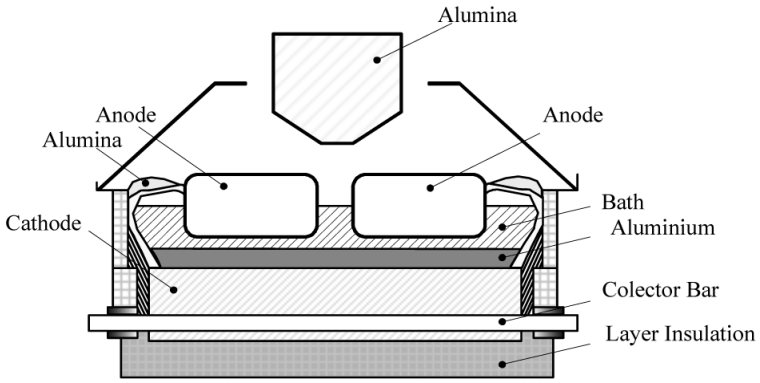


Figure 1.1: Scheme of a prebaked anode cell [49].

Though the chemical relationship within electrolysis is fairly complex [49] and the traditional view for over almost a century has been that keeping current and voltage stable was critical for a stable and efficient aluminum production, it has been demonstrated that the voltage of each potline can be changed very frequently at Alcoa's Warrick Operation, providing regulation service while carrying on electrolysis production at the same time [50]. The power consumption of a potline can be manipulated by adjusting the voltage at the output of the rectifier that supports the required DC current to the potline. By doing so, the power consumption rate can be adjusted very quickly and accurately, e.g. a potline can change its power consumption by about 1 MW within seconds. Another way to achieve power consumption flexibility is to shut down an entire potline totally by switching the breaker, which results in a larger amount of power change within a short amount of time. The smelter's flexibility enabled in this way makes aluminum smelting an ideal demand response resource (DRR). In fact, Alcoa Warrick Operation is actively participating in the MISO electricity market[50], providing both energy and ancillary services to MISO as a *DRR-Type-2* resource [51]; it is also reported that Trimet Aluminium, Germany's largest producer of aluminum,

is experimenting to support power system operation by using its vast pools of molten metal as storage batteries [52].

However, when providing flexibilities by either controlling the rectifier or switching the breaker, the thermal balance of the pots must be maintained to ensure safe operation, i.e. the temperature of the pots must be maintained within given bounds, to ensure production efficiency and equipment safety. If the thermal balance is severely violated, the potline will get damaged which is an expensive loss to the plants. On the other hand, there are a lot of uncertainties for demand response participation in the electricity markets, such as the energy prices, the spinning reserve and regulation prices, as well as the reserve dispatch commands and the regulation and load balancing signals. Taking these concerns into account, we intend to solve the following problem:

- *Problem 1 (uncertainty)*: how to optimally participate in the electricity market with a lot of uncertainties, while still ensuring that the critical operational feasibility constraints are satisfied?

In particular, we will take the aluminum smelter as an example and study how to optimize its participation in the electricity market. Specifically, we will investigate what is the optimal regulation capacity to provide and how to optimally bid in the spinning reserve market. Stochastic programming is the main approach adopted to handle the uncertainties in the market, in which we optimize over a set of scenarios that represent the possible regulation commands or market prices.

1.2.2 Handle Complexity of Process - Steel Manufacturer

The Iron and Steel industrial sector refers to the manufacturing of steel into basic shapes and forms that then can be used by other industries. In the United States, this sector consists of two basic types of production, i.e. Basic Oxygen Furnace (BOF) and Electric Arc Furnace (EAF), that each represents approximately half of domestic production. The BOF technology corresponds to the integrated steel mills that use the blast furnace for the making of iron, while the EAF technology corresponds to the "mini-mills" that produce steel from metal scrap without operations for coking or ironmaking [53].

The typical process of steel production by EAF is illustrated in Fig. 1.2. Solid metal scrap (from recycled steel such as discarded cars) is first molten in the electric arc furnace (EAF), then further processed in the argon oxygen decarburization unit (AOD) to reduce the carbon content. The molten steel is then refined in the ladle furnace (LF) and finally transported in ladles to the continuous casters (CC) to be casted into slabs - the final products of the steel manufacturing process. The steel can be characterized by grade, slab width, and thickness. Different kinds of products require different chemical ingredients and different casting procedures.

The first three processing stages operate in batch mode which means that a specified amount of metal is processed at a time. Each such amount of metal is called a *heat*. Meanwhile, the casting stage operates continuously and has some critical processing constraints. Due to the extreme conditions in the caster, it can only process a limited number of heats, after which it needs maintenance such as changing the caster mold and tundish before further operation. Several heats sharing the same or very similar grade characteristics and shape requirements form a campaign (a group of heats) and are casted sequentially. The method for forming casting campaigns is proposed and discussed in [30], and in this thesis we assume the campaigns have already been formed. The casting order should follow certain rules and the casting sequence for the heats within one campaign must not be interrupted.

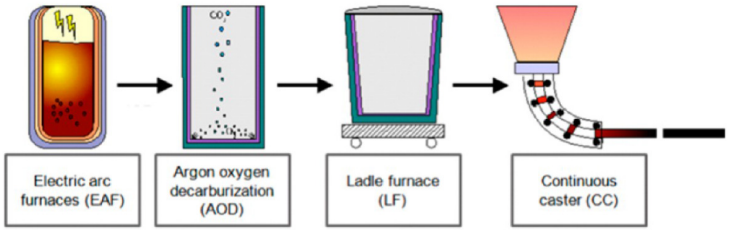


Figure 1.2: Production process of steel manufacturing[45].

Electric arc furnaces (EAFs) in steel manufacturing plants are identified as having great potential for demand response, since these furnaces not only consume large amounts of electric energy, but they operate in batch mode and are also fairly flexible in terms of changing their power consumption rate [47]. The EAFs are powered by transformers and their power consumption rate can be changed very quickly by adjusting the settings of the on-load tapchangers (OLTC). The general objective in industrial plant scheduling is to optimize the allocation of available resources (raw materials, equipment, workforce, utilities) to tasks over a specific time horizon, so as to achieve operational and economic benefits. The objective for the optimal scheduling of steel plants has traditionally been to minimize the make-span, i.e. to maximize the throughput and therefore to exploit the capacities of the heavily invested facilities. In recent years, the participation of steel plants in demand response has been studied with a variety of emphases, such as peak load management [54], prespecified energy curve tracking [43][44], and electricity fee minimization [45].

Meanwhile, it is recognized that steel plant scheduling is one of the most difficult industrial processes for scheduling, as steel manufacturing is a large-scale, multi-stage, multi-product batch process which involves parallel equipment and critical production-related constraints: most equipment can only process one job at a time; the final products have to be delivered at certain due times; in multistage production, the previous stage of a job needs to be completed before proceed-

ing to the next stage; some jobs may be scheduled to wait before a certain processing stage during equipment maintenance; and most intermediate products have to be processed by the following stage in time to prevent adverse cooling effects. In order to enable the steel plants to optimally provide demand response and to quickly adapt their decisions in the changing markets, we intend to solve the following problem:

- Problem 2 (*complexity*): how to model and solve the complex industrial demand response problems within acceptable computation time?

The problems are usually mixed integer linear programming problems that are difficult to solve. We propose to address the computation problems with both efficient modeling and tailored algorithms. On one hand, we plan to exploit the resource task network modeling framework to efficiently model the steel plant scheduling which enables controllable transformers and spinning reserve provision. Our premise is that efficient modeling can provide more flexibilities and more demand response options without largely increasing the formulation complexity, so that the resulting mixed integer programming problem can be solved by commercial solvers within reasonable time. On the other hand, we will explore the special structures in the steel plant scheduling problem and design tailored algorithms to make the computation more tractable. For example, the sequence of several production activities can be fixed to reduce the searching space in the branch and bound algorithm.

1.2.3 Overcome Granularity Restriction - Cement Plant

Cement crushing is an energy-intensive process that uses electricity intensively where electricity costs account for up to 20% of the total production costs [55]. The process starts with the mining of limestone, which serves as the major raw material for the production of cement. Limestone is then crushed and blended with several other raw materials such as shale and iron ore to form a mixture. The mixture is then grounded and fed into the kiln (a rotating, cylindrical, high-temperature oven filled with burning fuels like coal or oil) to be burned into cement clinker. Finally, the clinker is grounded with gypsum and limestone in a cement mill to form cement. The clinker burning process mostly consumes heat energy from the burning of the fuels, while electric energy is consumed by the crushing and grinding processes and various auxiliaries [56]. Obviously, the crushers and millers are the largest consumers of electric energy and also have the greatest potentials to provide demand response. The crushers and millers can be switched on and off rapidly while production goes on uninterrupted by using the stockpiled material which has been already crushed.

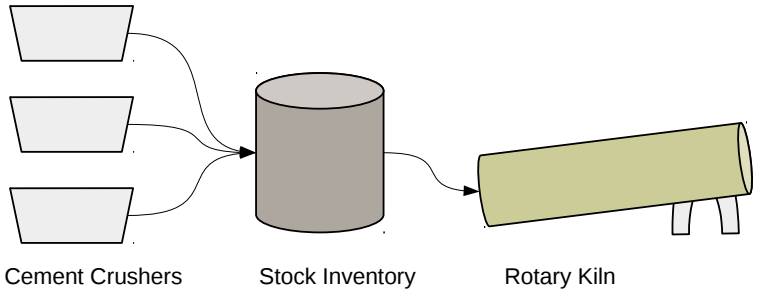


Figure 1.3: Cement plant illustration.

The coupling between processes needs to be considered in scheduling industrial loads. Usually, there are multiple processing stages in an industrial plant, and successive stages interact with each other through the generation and consumption of intermediate products. Taking the cement plant in Fig. 1.3 as an example, the crushing machines belong to the first stage which breaks the raw material (e.g. limestone, clay) into finer particles. In the second stage, the kiln heats up these finer particles to a higher temperature for further processing in its following stages. In other words, the crushing machines generate the finer particles and the kiln consumes these finer particles, therefore these two stages are coupled through the intermediate product. In scheduling the cement plant, we want to keep the kiln operating at a constant consumption rate, because it is very expensive to turn on/off the kiln due to its large thermal capacity.

There has been research studying the cement plants' participation in the electricity markets. For instance, the case study of a typical cement plant in South Africa presented in [57] demonstrates the benefits obtained by shifting the crushing processes to off-peak periods; the scheduling of continuous single-stage multi-product plants (i.e. cement plants) with parallel units and shared storage tanks over a weekly horizon is investigated in [41], in which the energy constraints related to time-dependent electricity pricing and power availability are considered, and efficient modeling and computing methods are proposed; problem instances on cement plants are solved in [58] to demonstrate that practical schedules with lower electricity costs and limited number of changeovers can be obtained very quickly by using the modeling framework of operating modes; in [48], the problem of calculating flexible schedules is tackled with a robust optimization based approach, in which two cement milling machines are scheduled to provide spinning reserve in the ancillary service markets.

Most of the above research focuses on the energy side, and there has been only a few papers discussing spinning reserve provision. Different from spinning reserve which only involves the reduction of power consumption, regulation and load following require a much faster response

of power change, both up and down, and hence are much more valuable in the ancillary service markets. Lots of industrial loads are able to provide very fast change of power in both directions, qualifying them for regulation and load following. For example, the crushers or mills in the cement industry can be switched on and off very rapidly [48, 59]. However, most of these industrial loads can only provide power changes in a discrete manner, e.g. the power change is several MWs at a time. The power capacity for a typical rock crusher is around 2 MW, and the consumption is not varied in a continuous way, but by merely switching on and off the rock crushers. This coarse granularity hinders those industrial loads from providing the most valuable ancillary services, as the regulation and load following in the current electricity markets require a continuous change of power. Consequently, these balancing resources with fast consumption changing capability are not utilized to the full extent. In this dissertation, we intend to overcome the granularity restriction for those industrial loads, e.g. cement plants, to provide regulation or load following service, as stated in:

- Problem 3 (*granularity*): how can we integrate industrial loads with poor granularity to fully exploit their demand response potentials?

In particular, we propose a coordination framework in which the industrial loads provide the regulation or load following services with the help from an on-site energy storage: the industrial loads provide a large but discrete power change which constitutes as the main body for the service, while the energy storage provides a fine and continuous power change which ensures that the combination of the two accurately follows the desired power signal. We use a model predictive control (MPC) approach which incorporates the prediction of the upcoming signals of the services into the decision making to coordinate these two parts. We also plan to study how to optimally decide the regulation capacity and baseline for the combination of industrial loads and energy storage, and how to schedule the entire industrial plant such that providing regulation will not interrupt the processing flows within the plant.

1.3 Contributions

The contributions of this dissertation are as follows:

- **Development of Bidding Strategies to Account for Uncertainty:** We have developed the optimal bidding strategies for the aluminum smelter, an ideal demand response resource, in the day-ahead markets for the provision of energy and ancillary service (spinning reserve or regulation). We propose stochastic programming models that generate the day-ahead bidding strategies for the smelters. The inputs to the models are the smelting plant parameters and

the price scenarios that represent future price trends. The output of the models are the energy bidding curves and the optimal spinning reserve or regulation provision as well as the power consumption levels of the potlines. We also utilize nonparametric Multiple Quantile Graphical Model to represent the distributions of hourly prices and use the Gibbs sampling approach to sample the price curves, which serve as the price scenarios for the stochastic programming model. These models are mixed-integer linear programming problems which can be solved by commercial solvers very quickly. The effectiveness of the models are demonstrated by case studies presented in Chapter 3. The strategies can take advantage of the future price trends and arrange the smelting activities to make profits from both electricity markets participation and aluminum production. The developed bidding methods can also be applied for the electricity market bidding by other industrial loads including the steel manufacturers and cement plants.

- **Development of Scheduling Approaches to Handle Complexity of Process:** We have developed methods for the efficient scheduling of steel manufacturing which has great demand response potentials but is also recognized as one of the most complex chemical processes to schedule. We investigate the mathematical modeling and propose an optimization algorithm for steel plant scheduling. From the mathematical modeling aspect, we develop efficient approaches to model the controllable transformers which supply power for the furnaces, in order to achieve more flexibility in providing demand response; we also provide methods to enable spinning reserve provision. From the optimization algorithm aspect, we propose a tailored branch and bound algorithm which utilizes the knowledge from steel manufacturing to speed up the computation. The proposed Multiple Melting Modes model provides a good trade-off between enabling the exploitation of the flexibilities given by the OLTCs and computational complexity. And as demonstrated through numerical studies in Chapter 4, the proposed tailored branch and bound algorithm shows potentials in reducing the computation time and iteration number of the problem considered. These ideas developed in the modeling and the algorithm for steel plant scheduling are also promising to help solve the difficulties in other complex industrial loads, e.g., the scheduling of cement plants.

- **Development of Coordination Framework to Overcome Granularity Restriction:** We have developed a coordination framework to overcome the granularity restriction in which the industrial loads provide the regulation or load following services with the help from an on-site energy storage device. The industrial loads provide a large but discrete power change which constitutes the main body for the service, while the energy storage provides a fine and continuous power change which ensures that the combination of the two accurately follows the

desired power signal. These two parts are coordinated by a model predictive control (MPC) approach which incorporates the prediction of the upcoming signals of the services into the decision making. The MPC approach is aimed for the real-time operation in each hour. We also study the daily operation and provide the tool to determine the optimal quantity of regulation provision for each hour. The coupling between successive processing stages are considered in our daily scheduling method. We demonstrate through case studies in Chapter 5 that, thanks to the coordination framework, the whole is greater than the sum of its parts. Note that the energy storage device can also be supplied from a third-party service. Also note that the approaches proposed in this thesis apply to a variety of loads, e.g. cement crushing, paper milling, buildings, and can enable both load following and regulation service. The proposed framework also has other potential applications outside of demand response, e.g. the coordination among fast and slow generators and storage.

1.4 Thesis Outline

The chapters that comprise this thesis are outlined as follows:

- **Chapter 2: Background and Methods** reviews the basic concepts and methods used in this thesis, including ancillary services, demand response, stochastic programming, ARMA models, linear regression, and mixed integer linear programming. The intention of this chapter is to help the readers better understand this thesis.

- **Chapter 3: Accounting for Uncertainty** takes the aluminum smelter as an example to study how to address the aforementioned *Problem 1 (uncertainty)*. In particular, we will study how to optimally provide regulation for the real-time operation within each hour, and how to bid in the day-ahead electricity markets with energy, spinning reserve, or regulation.

- **Chapter 4: Handling Complexity of Process** intends to solve the *Problem 2 (complexity)* by focusing on the steel plant scheduling. We first introduce the resource task network modeling framework, then propose modeling methods to take advantage of the controllable transformers to offer more flexibilities in demand response participation. We then explore approaches to make the computation more tractable. Finally, we extend the formulations to enable spinning reserve provision by the steel manufacturers.

- **Chapter 5: Overcoming Granularity Restriction** focuses to address the *Problem 3 (granularity)* for industrial loads such as cement plants to provide regulation or load following service. An coordination framework with the help from an on-site energy storage device is

proposed for the hourly operation, and an optimal scheduling method which considers the coupling between processing stages is proposed for the daily operation.

- **Chapter 6: Conclusion and Future Work** concludes the dissertation and discusses potential directions for future work in this area.

Chapter 2

Background and Methods

The focus of this dissertation is on demand response by industrial loads. This chapter reviews some basic concepts about power system and electricity market including ancillary services and demand response. The basic concepts of the specific industrial loads are already introduced in Section 1.2. This chapter also reviews some existing methods which are utilized in this thesis.

2.1 Power System and Electricity Market

An electric power system is a network of electrical components for the generation, transmission, distribution, and consumption of electricity. The power system is known as the grid and can be broadly divided into generators that supply the power, transmission system that delivers the power from the generating centers to the load centers, and a distribution system that feeds power to nearby residents, commercial buildings, and industries.

Traditionally, utilities are responsible for the operations and management of the power system, and also for providing electricity to the consumers. These utilities are typically vertically integrated, i.e., they own almost everything to serve electricity including the generation, transmission as well as distribution systems. In economic terms, electricity (both energy and power) is a commodity capable of being bought, sold, and traded. An electricity market is a system for electricity purchases, sales, and trades. In order to encourage competition and improve efficiency, the electricity markets worldwide have been gradually evolving from monopoly markets into liberalized markets since the 1980s. For example, the Federal Energy Regulatory Commission (FERC) in the United States has promoted to form the independent system operators (ISOs) to operate the transmission systems independently such that all market participants have open access to the transmission systems and hence can compete for electricity generation. Along with this revolution, the research community has contributed many studies on various topics of electricity markets, e.g.

market modeling [60], market derivatives and risk [61], reactive power market [62], ancillary service market [63], transmission planning [64], decision making [65], etc. Nowadays, the electric energy is usually traded through bilateral transactions and power pool agreements, and most of the markets are bidding-based in which buyers and sellers can bid for or offer generation.

Beyond electric energy, ancillary services are also bought, sold, and traded in the markets. Ancillary services refer to those services which are necessary to support the reliable operation of the bulk transmission system. Generally speaking, ancillary services can be classified as regulating reserve, operating reserve, reactive power support, and black start capability.

2.1.1 Ancillary Services

The active power ancillary services include regulating reserve, load following, spinning reserve, and non-spinning reserve. Regulating reserve, or regulation, is an online resource that can rapidly respond to system-operator requests for up-and-down movements, which is used to track the fast fluctuations in system load and generation. Load following or load tracking is similar to regulation but is slower, which serves as the bridge between the regulation service and the hourly energy market. Operating reserve includes spinning reserve and non-spinning reserve. Spinning reserve refers to online generation (or load) that is synchronized to the grid and can immediately increase power output (or reduce power consumption) in response to the outage of a major generator or transmission line. Non-spinning reserve is the same as spinning reserve except that it does not need to be online and its response can be slower. The comparison among these active power ancillary services over response speed, duration, cycle time, and market cycle is presented in Table 2.1 [50, 51]. Since our focus is on active power ancillary services, we will mainly study how industrial loads provide regulation and spinning reserve.

Table 2.1: Active Power Ancillary Services [50, 51]

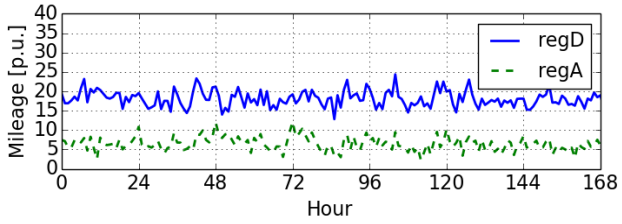
Service	Response Speed	Duration	Cycle Time	Market Cycle
Regulation	seconds to $\leq$ 1 min	minutes	minutes	hourly
Load Following	$\leq$ 10 min	10 min to hours	10 min to hours	hourly
Spinning Reserve	seconds to $\leq$ 10 min	10 min to 120 min	hours to days	hourly
Non-Spinning Reserve	$\leq$ 10 min	10 min to 120 min	hours to days	hourly

Regulation

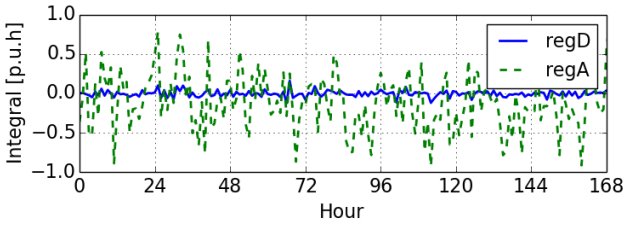
A power grid requires that generation and load exactly balance each other at all time. As large scale storage is still not economical, frequent adjustments to the output of generators are

necessary to balance the change of load. With the increasing penetration of renewable generation such as wind turbines and solar panels, the power grid needs more balancing power due to the volatile nature of these renewable resources. In an electric power system, automatic generation control (AGC) is used to adjust the power output of multiple generators at different power plants, in response to fluctuations in system load and generation. If a generator clears a certain amount of regulation capacity in this market, then this generator is obligated to follow the AGC command within that capacity.

The AGC commands are sent by the control center based on the real time power system conditions. The regulation provider commits to responding to the commands which are sent every 2s or 4s. The p.u. AGC signal multiplied by the provided regulation capacity is the AGC command that the plant committed to follow. The regulation signal of PJM (started in 1927 and renamed the Pennsylvania-New Jersey-Maryland Interconnection (PJM) in 1956, currently the largest competitive wholesale electricity market in the U.S.) is published online at [66], of which we will make use for our study. PJM distinguishes between two different regulation signals: RegD (Dynamic Regulation Control Signal) and RegA (Regulation Control Signal). RegD is designed for fast reacting resources and is developed specifically for resources with limited storage capabilities such as energy storage devices. The hourly mileage and integral of the PJM regulation signals over a week are plotted in Fig. 2.1 respectively, from which we can tell that the RegD signal moves intensively but is well balanced. We will use RegD for our research by assuming that the industrial loads participate as such fast acting resource, which is true for industrial demand response resources like aluminum smelters.



(a) Hourly regulation mileage.



(b) Hourly regulation integral.

Figure 2.1: Hourly mileage and integral of PJM regulation signals over a week.

In the ancillary service market of regulation, a market participant bids its hourly regulation capacity and the desired price. The bid and the clearing price are determined by the market operator after collecting offers from all market participants [51]. The participant should provide both positive and negative capacities indicating that if a manufacturing plant wants to provide regulation, it needs to operate below its full production capacity to ensure the ability to follow regulation in both directions. This regulation capacity is required by most ISOs to remain the same for the entire hour. In this thesis, we assume that positive and negative capacities are equal, but this could easily be relaxed. The plant also needs to provide a prediction of its energy consumption for the next hour in a resolution of 5 minutes, i.e. 12 predicted values for the hour. These predictions serve as the baseline for the regulation performance evaluation. This power prediction reflects the plant's production schedule if no regulation is provided.

Once the plant is cleared for a certain regulation capacity, the plant receives compensation for the lost opportunity of production by being paid for the provision of regulation capacity. It is obliged to fully respond to AGC commands within that capacity range, otherwise a nonperformance penalty is imposed and its performance scoring is affected [67].

Spinning Reserve

The operating reserve is the generating capacity available to the system operator within a short interval of time to meet demand in case a generator trips offline or another disruption occurs. Most power systems are designed such that, under normal conditions, the operating reserve equals

the capacity of the largest generator plus some fraction of the peak load. The operating reserve is made up of the spinning reserve as well as the non-spinning (or supplemental) reserve. The spinning reserve is the extra generating capacity that is available by increasing the power output of generators that are already connected to the power system. The non-spinning reserve or supplemental reserve is the extra generating capacity that is not currently connected to the system but can be brought online within minutes. We will mainly focus on spinning reserve in this thesis.

In terms of bidding rules in the day-ahead markets, the participant must submit its offers of energy and spinning reserve in the day-ahead market before both prices are known. Energy offers should take the form of a price curve (either a block offer or a slope offer), and up to ten Price/MW pairs can be submitted for each hour of the next operating day. While for spinning reserve, only one Price/MW pair can be offered for each hour of the next operating day. In terms of offering strategy for spinning reserve, it is advantageous if as much as possible of the available spinning reserve from the smelter is cleared. This is because the power system control center seldom dispatches spinning reserve and the smelter can make impressive profits simply by standing by. For example, the spinning reserve deployment rate at Alcoa Warrick Operation is less than 0.5% according to [50]. Thus, it is wise for the smelter to ask for a relatively low price to sell its maximum spinning reserve amount. After all, the spinning reserve is sold at the market clearing price which will not be affected by the smelter's offer, as the smelter's capacity is too small compared to the system capacity.

2.1.2 Demand Response

With the increased penetration of advanced technologies like smart meters and distributed generators, the loads are becoming more active in providing demand response. From a system perspective, an increase in active power production from a generator, in response to a positive regulation signal r, is equivalent to a decrease in active power consumption from a load. According to Federal Energy Regulatory Commission, demand response (DR) is defined as: Changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized. "DR includes all intentional modifications to consumption patterns of electricity of end use customers that are intended to alter the timing, level of instantaneous demand, or the total electricity consumption" [68]. Utilities are developing demand response programs to encourage electricity consumers' activeness.

Demand response is important because it provides a cost-effective solution to support the power system. Demand response reduces the need for new power plants. To respond to high peak demand, utilities traditionally have been building capital-intensive power plants and lines. However, peak

demand happens just a few times a year, so those assets run at a mere fraction of their capacity. Meanwhile, electric users pay for this idle capacity through the prices they pay for electricity. According to the Demand Response Smart Grid Coalition, 10%-20% of electricity costs in the United States are due to peak demand during only 100 hours of the year [69]. DR is a way for utilities to reduce the need for large capital expenditures, and thus keep rates lower overall. After all, if more loads participate in the market, the cost of the entire system will be reduced due to the increased competition and the improved market efficiency.

Of course, there is a limit to such economic improvement or cost reduction because demand response resources lose the productive or convenience value of the electricity not consumed. Thus, it is misleading to only look at the cost savings that demand response can produce without also considering what the consumer gives up in the process. In this thesis, we consider all these factors and investigate how to optimally participate in the electricity markets as demand response resources for the industrial loads.

2.2 Methods

The existing methods utilized by this dissertation are reviewed here, including stochastic programming, ARMA models, linear regression, and mixed-integer programming. To avoid redundancy and to clearly distinguish the contributions of this thesis from the pre-existing work, the relevant aspects of these existing methods are discussed here, separate from the methods developed in this thesis.

2.2.1 Stochastic Programming

In Chapter 3, stochastic programming is utilized to generate the optimal bidding strategies for the industrial load, an aluminum smelter, participating in the day-ahead electricity market. Stochastic programming is a technique for modeling optimization problems that involve uncertainty, which minimizes the total cost over a chosen number of scenarios while accounting for uncertainties in the problem [70]. Deterministic optimization problems are formulated with known parameters, whereas stochastic programming use a set of scenarios to represent the possible outcomes of uncertain parameters which are assumed to be subject to certain distributions. Stochastic programming has seen applications in optimal design, planning and scheduling, as well as operations under uncertainties [71–80].

In two-stage stochastic programming [81], there are two groups of decision variables: the first stage variables which are common to all scenarios, and the second stage variables which depend on the first stage variables and are specific to each scenario. A standard two-stage stochastic problem

is formulated as follows:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) + E\left[Q(x,\xi)\right] \\ \text{subject to} \quad & g(x) \leq 0 \\ & h(x) = 0 \end{aligned} \tag{2.1}$$

in which variables x denote the first-stage variables, $E\left[\cdot\right]$ represents expectation and $Q(x,\xi)$ denotes the optimal solution of the second stage problem, which also depends on the values of the second-stage variables y; the vector ξ is composed of a finite number of realizations $\xi_1, \ldots, \xi_k$ of the random parameters, with respective probabilities $p_1, \ldots, p_k$. This set of realizations $\xi_1, \ldots, \xi_k$ is called scenarios, which represent the possible outcomes of the uncertain parameters. The scenarios are usually obtained from Monte Carlo sampling or Gibbs sampling, combined with scenario reduction techniques [82–84]. By considering the random parameter space as a set of discrete events, the stochastic programming is then formed as its deterministic equivalent in the following:

$$E[Q(x,\xi)] = \sum_{k=1}^{K} p_k Q(x,\xi) \tag{2.2}$$

Hence, the stochastic programming problem can be solved to find the optimal solution for all variables while taking into account the probabilities for each of the scenarios.

Robust optimization is another approach that takes parameter uncertainty into consideration [85–87]. In optimization under uncertainty, one should decide whether to adopt the framework of robust optimization or stochastic programming [88]. Generally speaking, robust optimization tends to generate more conservative solutions and is easier to formulate, which is more suitable for short-term scheduling [89]; meanwhile, stochastic programming relies on historical data and is more suitable for long-term planning. We choose stochastic programming for this thesis because it is natural to generate the bidding curves for the market participation by examining the optimal variables in each scenario. Besides, stochastic programming is suitable to develop strategic decisions as it does not fix all decisions at the initial point, and our market participant can adapt its strategy by making decisions in the real-time market. Moreover, our stochastic variables are the electricity prices which have great historical data support.

2.2.2 ARMA Models and Linear Regression

ARMA stands for auto-regressive moving-average models for time series analysis, which represents the time series as a (weakly) stationary stochastic process composed of two polynomials, one for the auto-regression and the other for the moving-average. The model is typically referred to as the ARMA(p,q) model, where p stands for the order of the auto-regressive (AR) part, and q stands for the order of the moving-average (MA) part. The AR part regresses the variable on its own past

values. The MA part models the error term as a linear combination of error terms at various times in the past. For instance, the ARMA (2,1) model is described by:

$$\omega_t = \phi_1 \omega_{t-1} + \phi_2 \omega_{t-2} + \theta_1 \epsilon_{t-1} + \epsilon_t \tag{2.3}$$

in which ω_t stands for the time series variable, ϵ_t stands for the white noise. The auto-regressive parameters are ϕ_1, ϕ_2 and the moving-average parameter is θ_1. ARMA models can be estimated following the BoxJenkins approach [90]. There are also many existing software packages for ARMA analysis, e.g., the Time Series Analysis package in Python (*statsmodels.tsa*).

Linear regression is an approach for modeling the relationship between a dependent variable y and one or more explanatory variables x. If there are more than one dependent variable, it is usually referred to as multivariate linear regression. The relationship between y and x is modeled as:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.4}$$

with

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

for a given date set $\{y_i, x_{i1}, \cdots, x_{ip}\}_{i=1}^{n}$ of n observation samples with p exogenous (independent) features (variables); $\boldsymbol{\beta}$ is the p-dimensional parameter vector and ε_i is the error term. The relationship between output and input variables is assumed to be linear and is represented by the $\boldsymbol{\beta}$ vector, which can be estimated by ordinary least squares (OLS). The OLS method minimizes the sum of squared residuals for the training data set, and leads to a closed-form expression for the estimated parameter vector:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}\boldsymbol{y} \tag{2.5}$$

Note that including transformations of input variables enables representing a more complex relationship in linear regression. Ridge regression [91] and Lasso [92] can help with regression shrinkage and selection by adding a penalty term of $\boldsymbol{\beta}$ (l_2 norm or l_1 norm) to the OLS objective function.

We utilize ARMA models for the prediction of the AGC signal in Chapter 5 in the MPC coordination framework. We also utilize linear regression to predict electricity prices in the market bidding case studies in Chapter 3. Both methods can also be used to extend our scheduling work to predict the price signal for the plant.

2.2.3 Mixed-Integer Linear Programming

Mixed-Integer Linear Programming (MILP) is an extension to the Linear Programming problem where a subset of the variables are restricted to take integer values - mostly commonly binary values.

The general form of the MILP problem is given as follows:

$$\begin{aligned} \underset{x,\,y}{\text{minimize}} \quad & f = a^T x + c^T y \\ \text{subject to} \quad & Ax + By \leq b \\ & y \in \{0,1\} \end{aligned} \tag{2.6}$$

where y corresponds to a vector of variables restricted to be binary. The MILP problem is very useful for modeling discrete decisions in industrial plant operations. Several typical examples are given in the following [93, 94]:

- Multiple choice constraints:
  Select only one item: $\sum_i y_i = 1$
  Select at least one item: $\sum_i y_i \geq 1$
  Select at most one item: $\sum_i y_i \leq 1$
- Implication constraints:
  If item k is selected, then item j must be selected: $y_k - y_j \leq 0$
  If a binary y is zero, then a continuous variable x $(0 \leq x \leq U)$ must be zero: $x - Uy \leq 0$
- Either-or constraints (disjunctive constraints):
  Either constraint $g_1(x) \leq 0$ or constraint $g_2(x) \leq 0$ must hold, with M being a big number:
  $$g_1(x) - My \leq 0, \quad g_2(x) - M(1-y) \leq 0$$

One standard method to solve MILP problems is the branch and bound algorithm. The algorithm starts by first solving the relaxed LP problem. If the x variables in the solution take integer values, then we stop, as we have already found the best feasible solution. Otherwise, the relaxed solution provides a lower bound f^{lo} for the objective function, then we select one binary variable and restrict it to be either one or zero, which gives us two branches. The global optimal solution must lie in one of the branches. For each branch, we successively solve the relaxed LP, then restrict another binary variable and generate two more branches. Each time we solve the relaxed LP, we obtain a lower bound, and f^{lo} keeps increasing as the feasible region keeps shrinking. Along the process, we also round the relaxed solution to obtain integer solutions, and each of these feasible solutions provides an upper bound f^{up} for the objective function. As $f^{lo} \leq f^* \leq f^{up}$, once f^{lo} is close enough to f^{up}, we have found the global optimal solution. A detailed presentation of the algorithm is presented in Chapter 4. Note that methods such as the cutting plane algorithm are also very helpful in solving the MILP problems [95–98].

Many of the problems we study in bidding, scheduling, and coordination are MILP problems, as discussed in Chapter 3, 4, and 5. Most of the problems are solved by using existing solver CPLEX. We have also developed a tailored branch and bound algorithm in Chapter 4.

Chapter 3

Accounting for Uncertainty

Industrial loads such as aluminum smelters are ideal candidates to support power system operation through demand response programs. The aluminum smelters are qualified to participate in energy market, spinning reserve market, and regulation market. However, there are uncertainties in all of these markets such as the uncertain regulation signal, the uncertain dispatch call of spinning reserve, the uncertain energy price, etc. How to account for the uncertainties is an interesting and important problem that needs to be solved.

We have proposed an optimal regulation capacity provision model with given regulation prices for electrolysis processing plants such as the aluminum smelters in [21], which is presented in Section 3.1. We have also studied the demand response for aluminum smelters that participate in both energy and spinning reserve day-ahead markets in [22], as described in Section 3.2. Additionally, we have studied the bidding problem for demand response resources such as aluminum smelters with uncertain prices in the markets, as presented in Section 3.3.

3.1 Optimizing Regulation Capacity

In this section, we focus on determining the optimal regulation capacity that such an aluminum smelting plant should provide to maximize the combined profit from producing aluminum and providing regulation. The approach is based on stochastic programming and the stochastic variable is the regulation signal sent to the smelter.

Using linear approximations of the high-resolution regulation signal as scenarios, we can reduce the computational burden to solve the associated optimization problem significantly. Simulations for a specific aluminum smelting plant provide insights into the optimal regulation provision by smelters with various cost and price parameters.

To simplify the problem, we made the following assumptions for the considered aluminum

smelting process:

Assumption 1: The production rate is proportional to the power consumption and the power consumption is proportional to the voltage at the rectifier which controls the electrolysis process. The voltage on the other hand is linearly dependent on the rectifier tap changer position.

Assumption 2: The regulation revenue is proportional to the cleared capacity, the production revenue is proportional to the energy consumed, the control cost is proportional to the tap movement and the performance penalty is proportional to the difference between regulation command and mileage. All those prices are assumed to be parameters.

3.1.1 AGC Signal Simplification

Automatic generation control regulates the power output of generators to balance generation and load momentarily. It is expected that in the future a greater contribution of such regulation is provided by storage and flexible demand. The AGC commands are sent by the control center based on the real time power system conditions. As discussed in Chapter 2, the regulation provider commits to responding to the commands which are sent every 2s or 4s and the p.u. AGC signal multiplied by the provided regulation capacity is the AGC command that the plant committed to follow.

Given the high resolution of the AGC signal it is very difficult if not impossible to predict the signal for a longer time frame. However, as we intend to use stochastic programming to determine the optimal regulation capacity, we need to find a way to generate scenarios and capture the expected regulation provision without having an accurate prediction of the AGC signal. Hence, we propose a linear simplification of AGC signals by local extreme points and then create scenarios based on such extreme points. First, we apply a low pass filter to the original AGC signal to remove high frequency noise. Local extreme points are then identified by numerical derivatives along the curve, i.e. whenever the derivative changes its sign, a local extreme point has been detected. Successive local extreme points that fall within short distances both in time and magnitude, e.g. time gaps less than 5s and value gaps less than 0.025p.u., are further aggregated as one single point. Hence, local extreme points which will be called AGC events in the remainder of the section are identified and used for scenario generation. Piecewise segments characterized by those AGC events serve as a linear simplification of the AGC signal in between extreme points as shown in Fig. 3.1.

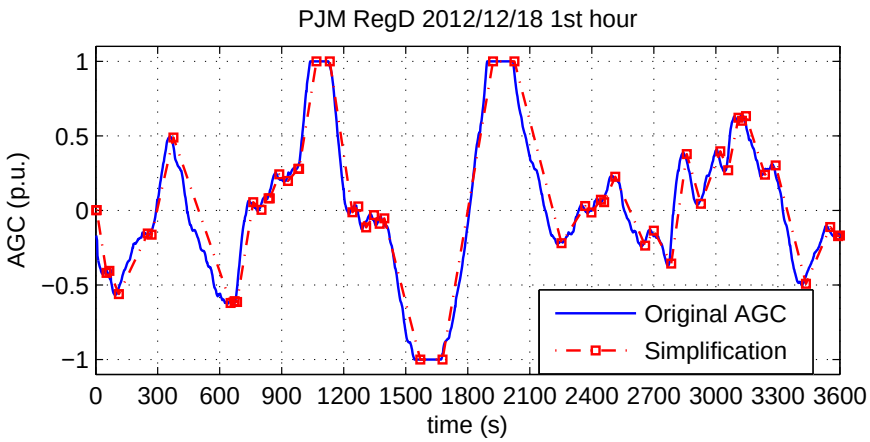


Figure 3.1: Simplification of AGC Signals. The red square dashed line represents the linear simplification. The blue solid line represents the original PJM's RegD signals of the first hour on 12/18/2012 [66].

This simplification approximates the real trajectory of the AGC signal by piecewise linear segments. The statistics of the AGC events' magnitude and time interval are provided in Fig. 3.2 and Fig. 3.3. Typically around 40 - 60 AGC events are presented in an hour's worth of regulation D signal. Given the original scenarios of AGC signals, we apply the simplification and obtain the simplified scenarios. Then, the proposed model only responds to AGC events of the simplified scenarios and tracks the piecewise linear trajectory. The approximation here is reasonably sound and is able to reduce the computational burden in the stochastic programming significantly.

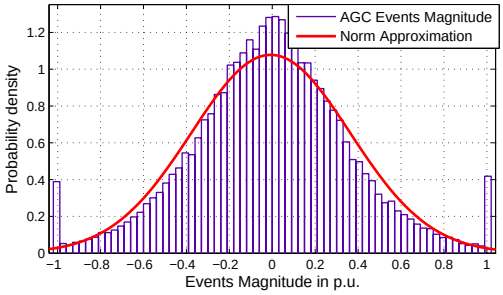


Figure 3.2: Histogram of AGC Events' Magnitude. The AGC events are simplifications of PJM's RegD Signals from 12/18/2012 to 01/18/2013 [66].

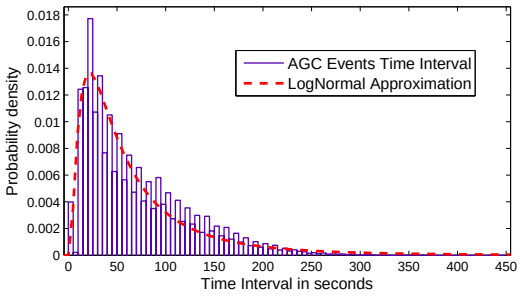


Figure 3.3: Histogram of AGC Events' Time Interval - the time between two successive AGC events.

3.1.2 Mathematical Formulations

In this section, the proposed optimal capacity provision model is presented by which the plant decides its hourly regulation capacity V_h that would be optimal for the plant for the upcoming operating period. We focus on the regulation provision from the industrial electrolysis process and optimize its participation. The following sections will focus on building a bidding curve which the plant should bid into the market. As of now, we determine the capacity that the plant ideally is cleared for, assuming that the price it will receive has been predicted reasonably well. Power flow constraints are not considered, as we do not study the overall system here. The plant first needs to submit its predicted power consumption for every 5 minutes of the next hour. These values serve as a baseline for the regulation provision and basically correspond to the scheduled rectifier tap changer positions $z_{l,i}$, which correspond to the first-stage variables in the stochastic programming. The regulation command it needs to follow is equal to the multiplication of V_h and $X_{s,k}$ where s indicates the scenario and k the k-th AGC event. In order to follow this command, the tap position of the rectifier of production line l moves by $m_{l,s,k}$ steps from its original position $z_{l,i}$. The rectifier tap positions evolve in a discrete manner, hence, $z_{l,i}$ and $m_{l,s,k}$ are integer variables. However, since the number of $m_{l,s,k}$ grows with the number of scenarios and trajectory simplification is adopted, $m_{l,s,k}$ is relaxed and assumed to be continuous in order to simplify the calculations.

Constraints

Rectifier Tap Changer. Rectifier tap changers are designed to operate within a certain range. In addition, changing the tap position is a mechanical process which means that there is a limit on how much change can be achieved within a certain amount of time. Consequently, the following constraints need to hold for $\forall l \in L, i \in I, s \in S, k \in K_s$:

$$z_l^{lo} \leq z_{l,\mathcal{I}(k)} + m_{l,s,k} \leq z_l^{up} \tag{3.1}$$

$$dz_{l,s,k} = |z_{l,\mathcal{I}(k+1)} + m_{l,s,k+1} - z_{l,\mathcal{I}(k)} - m_{l,s,k}| \tag{3.2}$$

$$dz_{l,s,k} \cdot \tau_l \leq t_{s,k+1} - t_{s,k} \tag{3.3}$$

where $\mathcal{I}(\cdot)$ is a function mapping event k to its corresponding 5min interval i.

Regulation Capacity. The maximum regulation capacity is the summation of available capacities over all production lines. However, this capacity needs to remain the same for the entire hour which is limited by the available capacity within every 5min interval. For $\forall l \in L, i \in i(h)$, the following constraints need to hold:

$$V_{\mathcal{H}(i)} \leq \sum_{l \in L} V_{l,i}^{+} \tag{3.4}$$

$$V_{\mathcal{H}(i)} \leq \sum_{l \in L} V_{l,i}^{-} \tag{3.5}$$

$$V_{l,i}^{+} \leq \theta_l \cdot (z_l^{up} - z_{l,i}) \tag{3.6}$$

$$V_{l,i}^{-} \leq \theta_l \cdot (z_{l,i} - z_l^{lo}) \tag{3.7}$$

where $\mathcal{H}(\cdot)$ is a function mapping the 5min intervals i to its corresponding operating hour h.

Plant Scheduling. The main purpose of the plant is to carry out electrolysis and produce aluminium. Hence, we impose a constraint which ensures that production targets are met. Since production is proportional to electricity consumption, the constraint can be formulated as a lower constraint on the energy consumption of the plant. Hence, for $\forall j \in J, s \in S$:

$$\sum_{l \in L} \sum_{k \in \mathcal{K}(j)} ((z_{l,\mathcal{I}(k)} + m_{l,s,k}) \cdot \theta_l + P_l^0) \cdot \mathcal{T}(k) \geq W_j \tag{3.8}$$

where $\mathcal{K}(j)$ stands for all AGC events that occur during production task j. $\mathcal{T}(\cdot)$ is a function that calculates the time duration for each AGC event k. For event k in scenario s, $\mathcal{T}(\cdot)$ returns $(t_{s,k+1} - t_{s,k-1})/2$, i.e. the time duration it affects the energy consumption. This constraint should hold for every scenario making sure the plant's minimum production level is ensured.

Objective

The overall objective of the optimization problem is a sum of multiple cost and revenue terms.

Energy Consumption. The value λ_h^E denotes the per unit profits made by consuming electricity and producing aluminum, which includes the revenue from selling aluminum and the combined costs of producing aluminum. That part of the objective function is given by

$$EProfit = \sum_{s \in S} p_s \sum_{l \in L} \sum_{k \in K_s} \theta_l (z_{l,i(k)} + m_{l,s,k}) \mathcal{T}(k) \lambda_h^E \tag{3.9}$$

Control Action. There is a limit on how much the tap of a rectifier could be changed because intensive movement leads to frequent replacement of the rectifiers. The price λ_l^{Tap} reflects the average cost per movement resulting in the following contribution to the overall objective

$$ACost = \sum_{s \in S} p_s \sum_{l \in L} \sum_{k \in K_s} dz_{l,s,k} \lambda_l^{Tap} \tag{3.10}$$

Deployment Error. As a demand resource, the regulation deployment is the negative summation of all tap changers multiplied by the change in energy consumption per tap step, thus the regulation error $\delta_{s,k}$ is equal to:

$$\delta_{s,k} = |X_{t,s} V_h + \sum_{l \in L} \theta_l m_{l,s,k}|$$

and the resulting penalty is given by:

$$Penalty = \sum_{s \in S} p_s \sum_{k \in K_s} \delta_{s,k} \lambda^P_{\mathcal{H}(\mathcal{I}(k))} \tag{3.11}$$

For absolute values, we can use additional variables and constraints to represent them in the final formulations to solve the optimization problems. For example, the above equation involving absolute terms can be transformed as:

$$\delta_{s,k} \geq X_{t,s} V_h + \sum_{l \in L} \theta_l m_{l,s,k}$$
$$\delta_{s,k} \geq -X_{t,s} V_h - \sum_{l \in L} \theta_l m_{l,s,k}$$

Regulation Capacity. The electrolysis plant receives a payment for the provision of its regulation capacity V_h, i.e. the plant's revenue from regulation provision is given by:

$$VRevenue = \sum_{h \in H} V_h \cdot \lambda_h^V \tag{3.12}$$

The overall objective function is given by the sum over all of these objectives resulting in:

$$\min \quad -EProfit + ACost + Penalty - VRevenue \tag{3.13}$$

The resulting overall problem formulation is a mixed-integer linear programming problem.

3.1.3 Case Study

Simulations for a specific aluminum smelting plant provide insights into the optimal regulation provision by smelters with various cost and price parameters. We consider an aluminum smelter

which consists of two potlines for which the parameters are listed in Table 3.1. Ten AGC scenarios are considered and it is assumed that the probability of occurrence of each scenario is the same. The linearized evolution of the AGC signal in these scenarios is shown in Fig. 3.4. There's one production task to complete within the considered hour and the minimum production is set to 95% of the base production which corresponds to the production achieved if the tap position is zero. The minimum regulation capacity required by the ISO is assumed to be 1MW.

Table 3.1: Potline Parameters

l	θ_l/MW	τ_l/s	z_l^{up}	z_l^{lo}	P_l^0/MW
1	0.8	5	10	-10	100
2	1.2	5	9	-9	100

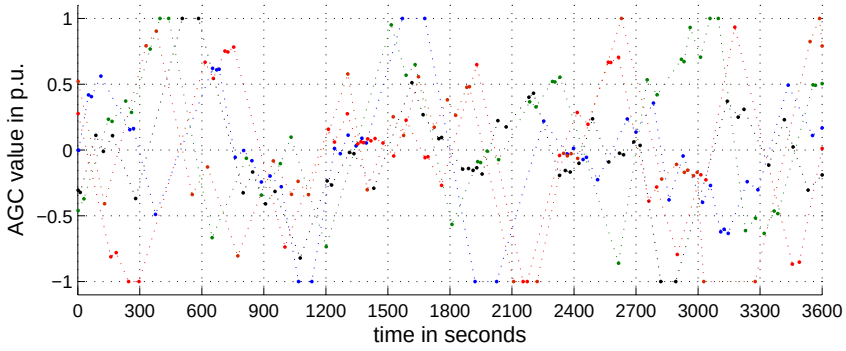


Figure 3.4: Simplified AGC Signal Scenarios. Five scenarios are taken from the first hour's PJM Regulation-D signals from 2012/12/18 to 2012/12/22 [66]. These five scenarios and their negative constitute the original ten scenario AGC set that is positive and negative well balanced.

We carry out a three hour simulation with 10 different scenarios of the AGC signal. The simulated price parameters encourage regulation participation in the 1st and 3rd hour while encouraging increased production in the 2nd hour. The results for the optimal regulation capacities are shown in Fig. 3.5. We also simulate with different penalties on the regulation provision mismatches. When increasing the penalty, the regulation error decreases as shown in Fig. 3.6.

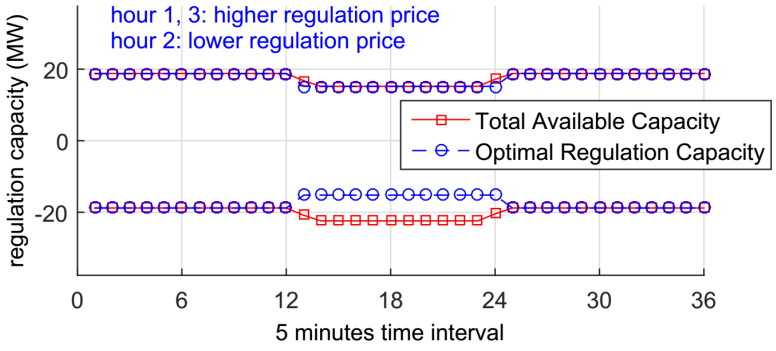
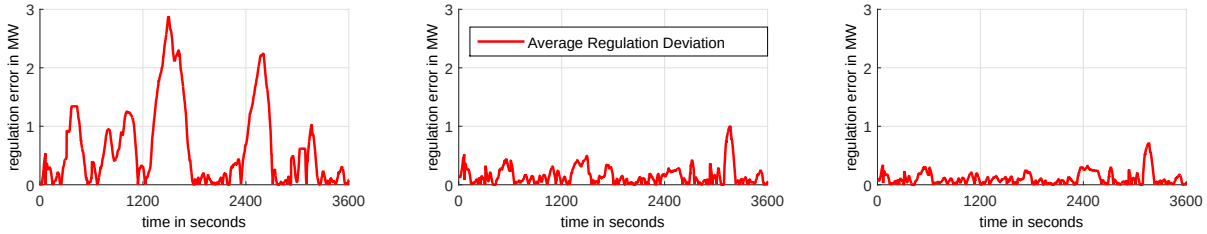


Figure 3.5: Regulation capacity for 3 hour simulation.



(a) λ^P=1 (b) λ^P=10 (c) λ^P=20

Figure 3.6: Regulation performance with changing penalties on mismatch.

The proposed model takes into account the impact of a variety of benefits and costs on the decision of regulation capacity provision, and is able to help the plant operators to determine the optimal regulation capacity to provide.

3.2 Optimal Bidding of Energy and Spinning Reserve

In this section, we design the optimal bidding strategy for energy and spinning reserve for an aluminum smelter in the day-ahead electricity market. We assume that the smelter's flexibility is realized by controlling the rectifiers, and the smelter's power consumption is adjustable within a given range. We only consider controlling rectifiers to achieve flexibility because turning off an entire potline causes significantly more interruption to the plant operation. It is assumed that the smelter has a long-term energy contract with the electricity utility, and the smelter can sell energy back to the market if the actual amount of energy usage is less than the contracted amount. In addition, the smelter can provide spinning reserve to the power system when its power consumption is higher than its lower bound. In that case, the difference between the current loading level and

the minimum loading level is the maximum available spinning reserve.

In the following, we focus on determining the optimal bidding strategy in the day-ahead energy and spinning reserve markets for an aluminum smelter. The approach is based on stochastic programming in which the market prices are the stochastic variables. Case studies demonstrate the effectiveness of the approach and provide insights into the demand response from industrial plants.

3.2.1 Bidding Strategy

The market prices are treated as stochastic variables where $\lambda_{s,h}$ represents the energy price and $\rho_{s,h}$ stands for the spinning reserve price. The subscript h denotes the hour of the day and s denotes the scenario index. As previously mentioned, the values for these stochastic variables can be obtained by price prediction techniques. The potline's power consumption level is $P_{l,s,h}$, where l stands for the potline index. The decision variables are the smelter's offers for the day-ahead market, i.e. the energy offer Price/MW pairs and the spinning reserve offer MW/Price pair. For energy offers, we use $E_{s,h}$ to represent the energy to sell in scenario s at hour h. After the optimal values for $E_{s,h}, s = \{1, ..., S\}$ is obtained from the proposed model, the hourly energy offering curve is constructed by connecting the MW/Price pairs ($E_{s,h}$, $\lambda_{s,h}$) from different scenarios in the same hour. For spinning reserve offering, since only one Price/MW pair can be submitted for each hour, we denote this offered capacity as V_h. Note that there is no subscript s for the spinning reserve offer. Once cleared, the smelter needs to make sure that the committed amount of spinning reserve is available for any possible scenario.

3.2.2 Mathematical Formulations

The potline's power consumption is bounded by parameters P_l^{min} and P_l^{max}, which are given by the maximum achievable flexibility of the plant and the limitations to ensure a safe operation of the plant. The power consumption of $P_{l,s,h}$ is modeled by piece-wise linear segments. This is because we model the aluminum production efficiency by a piece-wise linear approximation. The number of segments is n_l and the ascending parameters $\{a_{l,1}, ..., a_{l,n_l+1}\}$ represent the segments. Note that $a_{l,1}, a_{l,n_l+1}$ equal P_l^{min}, P_l^{max}, respectively. The binary variable $N_{l,s,h,i}$ denotes whether the power $P_{l,s,h}$ is within the i-th segment, and its summation over i should be one. The continuous variable $\Delta P_{l,s,h,i}$ denotes the excess value of $P_{l,s,h}$ over the i-th segment. This results in the following set of equations:

$$P_{l,s,h} = \sum_{i=1}^{n_l} (a_{l,i} N_{l,s,h,i} + \Delta P_{l,s,h,i}) \quad \forall l, s, h \tag{3.14}$$

$$0 \leq \Delta P_{l,s,h,i} \leq (a_{l,i+1} - a_{l,i}) N_{l,s,h,i} \quad \forall l, s, h, i \tag{3.15}$$

$$\sum_{i=1}^{n_l} N_{l,s,h,i} = 1 \quad \forall l, s, h \tag{3.16}$$

For simplicity, we assume $\sum_l P_l^{max}$ equals the contracted power consumption in the long term energy contract[1]. Thus the smelter can sell energy to the market if its potlines are operating below P_l^{max}, i.e. consuming less than the contracted amount. Hence, the energy to sell $E_{s,h}$ is modeled as:

$$E_{s,h} = \sum_l (P_l^{max} - P_{l,s,h}) \qquad \forall s, h \tag{3.17}$$

The available spinning reserve is limited by the smelter's ability to further reduce its power consumption. Consequently, we require that the offered spinning reserve, once cleared in the market, should be available in every scenario. Thus the spinning reserve availability is modeled by

$$V_h \leq \min_s \sum_l (P_{l,s,h} - P_l^{min}) \qquad \forall h \tag{3.18}$$

which indicates that V_h needs to be less than the available amount in any of the considered scenarios; V_h correspond to the first stage variables in the stochastic programming.

As mentioned before, the thermal balance is the most critical issue in providing flexibility, and the potlines' temperature should be kept within a certain range to ensure high smelting efficiency as well as operation safety. This means that the energy consumption for every successive τ_l hours should be greater than E_l^τ, as in

$$\sum_{h'=h}^{h+\tau_l-1} (P_{l,s,h'} - V_{h'}) \geq E_l^\tau \qquad \forall l, s, h \tag{3.19}$$

where E_l^τ is the minimum input energy required for τ_l hours to sustain the temperature. The spinning reserve should be committed to last for at least one hour in most electricity markets. Note that (3.19) states that the temperature should also be sustained even if the spinning reserve is called and dispatched by the system operator. The impact of spinning reserve dispatch is considered in both (3.18) and (3.19), as these constraints are related to the potlines' operation safety.

Furthermore, there is daily aluminum production scheduled by a certain higher-level longer-horizon plant planning. It is also assumed that the plant has storage capability, meaning that there is some flexibility in terms of when the aluminum production takes place. Thus the total energy consumption during the operating day, which is proportional to the aluminum production quantity, is bounded according to

$$E_d^{min} \leq \sum_{h,l} P_{l,s,h} \leq E_d^{max} \qquad \forall s \tag{3.20}$$

[1]Note that relaxing this assumption is straightforward.

where E_d^{min} and E_d^{max} are the daily minimum and maximum energy consumption, which is proportional to the minimum and maximum aluminum production amount.

In order to get a monotonous bidding curve, we require the following constraint to hold:

$$E_{s,h} - E_{s',h} \leq 0 \quad \forall h, s, s' : O_h(s) + 1 = O_h(s') \tag{3.21}$$

where $O_h(s)$ denotes the order of the energy price for each scenario in hour h. The scenarios are ordered in each hour in an ascending order. For example, if s is the scenario with the lowest price in hour h, then $O_h(s)$ equals 1; if s' is the scenario with the highest price in hour h, then $O_h(s')$ equals the total number of scenarios.

The revenues from electricity market participation are calculated as

$$R = \sum_s p_s \cdot \sum_h \lambda_{s,h}(E_{s,h} + \rho_{h,s} V_h) \tag{3.22}$$

in which p_s stands for the probability of scenario s. Note that we do not consider the economics of the actual dispatch of spinning reserve due to the low dispatch rate. The economics analysis on spinning reserve dispatch should be conducted for longer-horizon scheduling, e.g. weekly scheduling or quarterly scheduling.

As mentioned before, the profit P from producing aluminum is approximated by piece-wise linear functions, as illustrated in Fig. 3.7, in which we assume that the marginal production profit is constant within each segment. The production profit is approximated by

$$P = \sum_s p_s \sum_h \sum_l \sum_{i=1}^{n_l} (c_{l,i} N_{l,s,h,i} + b_{l,i} \Delta P_{l,s,h,i}) \tag{3.23}$$

in which we assume that the potline l's marginal production profit is $b_{l,i}$ in its i-th segment, and the total value of production profit at the segment's left boundary (i.e. when $P_{l,s,h} = a_{l,i}$) is $c_{l,i}$. In this way, we can model the differences in smelting efficiency when the potline is operating at different loading levels. Generally, the production efficiency is higher when the loading level is higher, as the potline is originally designed to produce with full capacity.

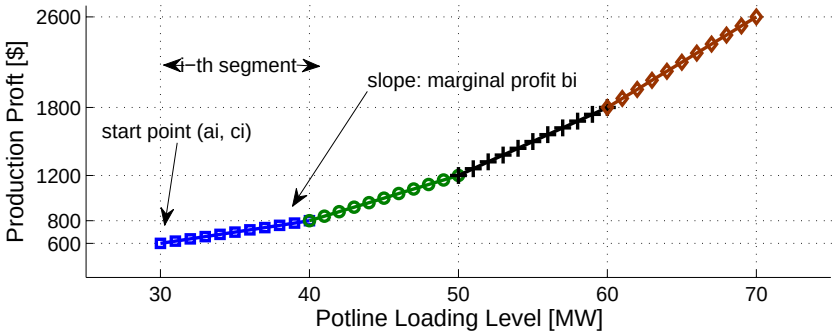


Figure 3.7: The illustration for piece-wise linear approximation of the smelter's production profit.

The optimization objective of the daily bidding is to maximize the revenues from electricity market and the profit of producing aluminum, i.e.

$$\max \qquad R + P \tag{3.24}$$

Consequently, the overall problem is a mixed-integer linear programming problem.

3.2.3 Case Study

Simulation Setup

We consider an aluminum smelting plant with two potlines. The potlines' parameters are listed in Table 3.2. We approximate the production profit by 4 piece-wise linear segments, and the corresponding parameters are listed in Table 3.3.

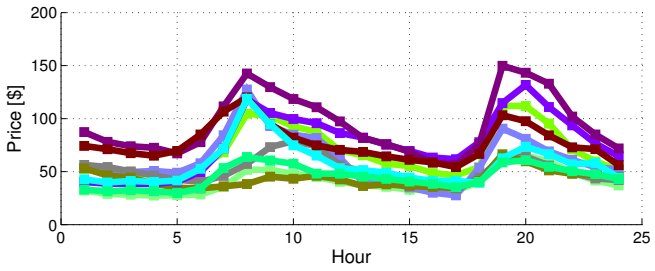
Table 3.2: Smelter Parameters

l	$P_l^{min}[MW]$	$P_l^{max}[MW]$	$\tau_l[h]$	E_l^τ [MWh]
1	30	70	4	180
2	40	60	3	135

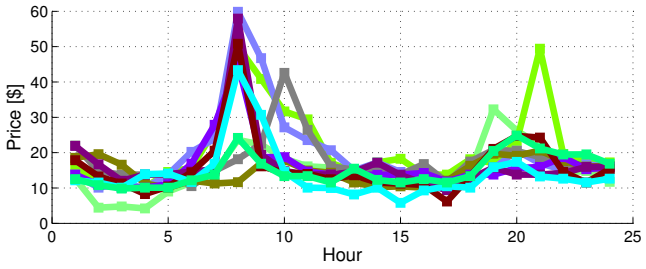
Table 3.3: Piecewise Linear Parameters for Production Profit

$l = 1$	$\{a_i\}[MW]$	{30, 40, 50, 60, 70}
	$\{b_i\}[MW/\$]$	{56, 58, 60, 62}
	$\{c_i\}[\$]$	{1680, 2240, 2820, 3420}
$l = 2$	$\{a_i\}[MW]$	{40, 45, 50, 55, 60}
	$\{b_i\}[MW/\$]$	{66, 68, 70, 72}
	$\{c_i\}[\$]$	{2640, 2970, 3310, 3660}

The scheduling is carried out on a daily basis and we focus on the day-ahead energy and spinning reserve markets. Price prediction techniques such as ARIMA and neural networks can be applied to generate price scenarios for the stochastic programming problem. Scenario reduction methods can be adopted to alleviate the computation burden of the mixed-integer programming by using a small number of representative scenarios. The price prediction and scenario reduction are not the focus of this thesis, so in our case study we use historical MISO price curves (shown in Fig. 3.8) as our scenarios. The price curves correspond to the 10 days of 02/5/2014 to 02/14/2014. Energy and spinning reserve prices taken from the same day serve as one scenario, and are plotted with the same color. The probabilities of all scenarios are assumed to be equal.



(a) Energy Prices

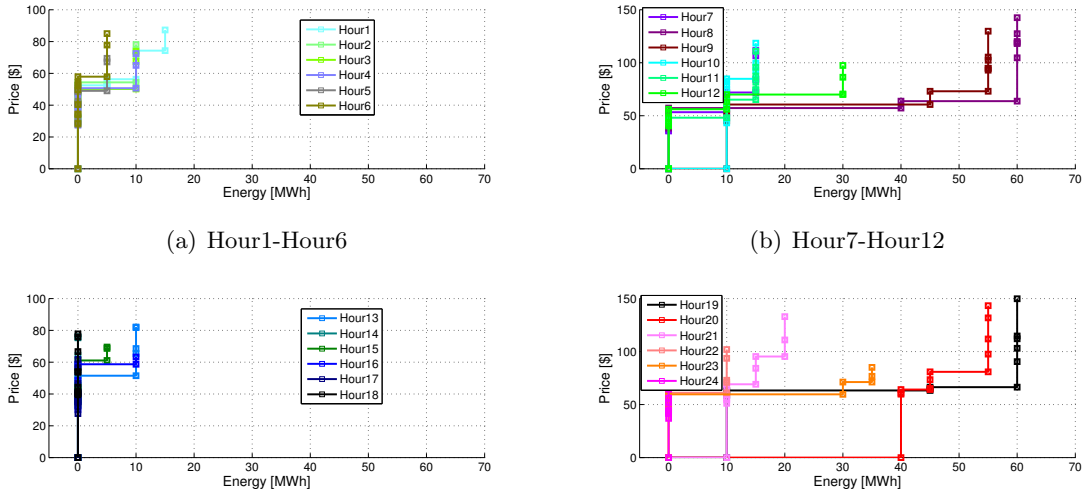


(b) Spinning Reserve Prices

Figure 3.8: Price scenarios taken from MISO's historical data. The spinning reserve prices follow the trend of energy prices. The peak hours for both prices are around hour 8 and 20.

Simulation Results

The bidding model described in Section 3.2.1 is a mixed-integer linear programming problem. We solve this problem in MATLAB by the solver TOMLAB\CPLEX on a 64-bit Linux machine.



(a) Hour1-Hour6

(b) Hour7-Hour12

(c) Hour13-Hour18

(d) Hour19-Hour24

Figure 3.9: The day-ahead energy bidding curves for each hour.

The resulting hourly energy bidding curves given by our model are shown in Fig. 3.9. From the figures, we observe that the bidding curves are more conservative in terms of selling energy in hours 1-6 and hours 13-18, as the smelter asks for a very high price for selling very little energy. The smelter is even reluctant to sell any energy for hours 17 and 18. As seen in Fig. 3.8, the energy prices are relatively lower during these hours, so it is wise of the smelter to focus on producing aluminum and sell little energy during these hours. On the other hand, the bidding curves are more aggressive in hours 7-12 and hours 19-24, in which the smelter bids significant amounts of energy into the market. In particular, the smelter wants to sell around 55 MW of energy in hours 7, 8, 19, and 20. Comparing with Fig. 3.8, we observe that the energy prices during these hours are really high, so it makes sense for the smelter to be aggressive in selling energy.

The spinning reserve provision given by our model is displayed in Fig. 3.10. We can tell that the smelter is willing to provide more reserve when the reserve prices are higher. But the smelter provides little reserve at the exact peak hours 8, 9, 19 and 20. This can be explained by the fact that the smelter is focused on selling energy in these hours, because the energy prices are significantly higher than the spinning reserve prices. Thus, the potlines' loading levels during these peak hours are very low, leaving less space for providing spinning reserve. It should be kept in mind that the available spinning reserve capacity is upper bounded by the difference between the potline's current loading level and its minimum loading level: if the smelter lowers the potlines' loading levels to sell energy, then there is little spinning reserve capacity left. As discussed before, the smelter should bid this optimal spinning reserve schedule by asking a relatively low price.

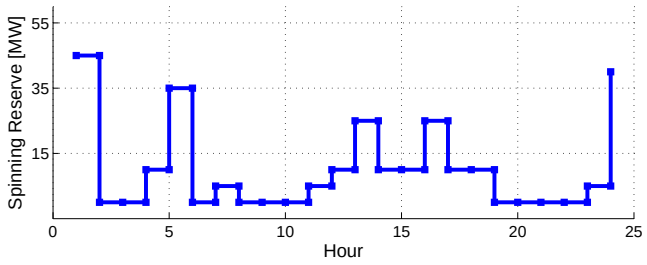


Figure 3.10: Scheduled Spinning Reserve Provision

Besides, we also analyze the power consumption of each potline. The power consumption in scenario 7 of both potlines are compared in Fig. 3.11. As we can see, potline 1 contributes more in providing flexibility while potline 2 concentrates more on smelting. This can be explained by the fact that the marginal production profit of potline 2 is higher than that of potline 1, as seen in Table 3.3.

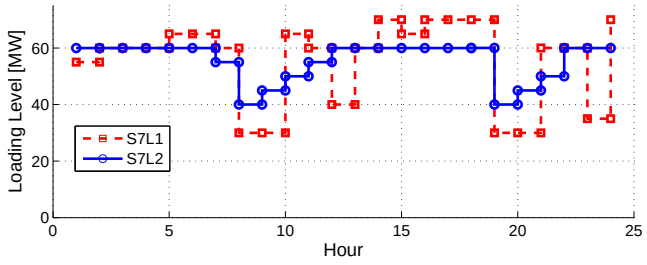


Figure 3.11: Potline Power Consumption

As demonstrated by the case studies, the proposed model can take advantage of the future price trends and arrange the smelting activities to make profits from both electricity markets participation and aluminum production.

3.3 Optimal Bidding of Energy and Regulation

Here, we consider the decision making in the day-ahead markets for industrial demand response resources such as aluminum smelters with energy and regulation provision. Compared with the previous section, here we also study how to sample price scenarios for the bidding process and also study the performance of the bidding strategy. It is assumed that the load has a long-term energy contract with the electricity utility, and it can sell energy back to the market if the actual amount of energy usage is less than the contracted amount. According to the bidding rules, the participant must submit its offers in the day-ahead markets before prices are known. Energy offers should take

the form of a price curve (either a block offer or a slope offer), and up to ten Price/MW pairs can be submitted for each hour of the next operating day. For regulation, only one Price/MW pair can be offered for each hour of the next operating day. Every day at a certain time, the markets for the following day are cleared for each operating hour, and the final market prices are settled, as well as the shares of energy and ancillary service for each participant. Once settled with a certain amount of regulation to provide, the power consumption of the load is obligated to follow the regulation signals sent by the power system control center. The regulation signals are determined according to the real-time power system conditions and are sent every 2s or 4s.

To optimally decide how much regulation to provide for each hour, we need to know the hourly integral and mileage of the regulation signal, in which the integral corresponds to the algebraic summation of the signal and the mileage refers to the absolute length of its moving trace. For example, consider a signal series of $\{0.1, -0.1, 0.1, -0.1\}$, its integral is 0 and its mileage is 0.6. The regulation mileage is strongly relevant to the moving cost of the equipment as well as the mileage compensation from the electricity market. The regulation integral determines the net energy exchange between the regulation provider and the power system, and this net energy exchange is particularly crucial for the aluminum smelter as it is relevant to both the final product yield and the thermal balance inside of the potlines. The hourly mileage and integral of the PJM regulation signals over a week were provided in Fig. 2.1 respectively, and we concluded that the RegD signal moves intensively but is well balanced. To simplify the model, we assume in this section that the hourly mileage of the RegD is a constant number (i.e. historical average of hourly mileage) and its hourly integral is zero.

If we can accurately forecast the future market prices, then it is quite straightforward to develop the optimal bidding strategy. Even though there is a large amount of literature on electricity price prediction which has seen great progress, it is still difficult to provide a reliable point-price prediction. However, by employing the prediction techniques proposed in the literature, we can obtain a distribution of future prices that we can incorporate into a stochastic programming formulation using a set of possible price scenarios. We adopt the non-parametric Multiple Quantile Graphical Model [99] to approximate the price distribution and use Gibbs sampling to generate the price scenarios. The stochastic programming approach is able to hedge the risk from prediction uncertainty, and its objective is to optimize the load's decisions to maximize the average profit (or expected profit) over all scenarios.

In the proposed bidding approach, we assume that the load's bidding/offering into the electricity markets generally will not impact the final market clearing prices. This is reasonable as the load's total power capacity is small compared to the power system's total generation capacity.

3.3.1 Price Scenarios

To optimize the market bidding through stochastic programming, we need a set of price scenarios that represent the possible price curves of the next operating day. One classical method is to first forecast the price curve, then draw Monte Carlo samples based on the forecast, in which the underlying distribution is known. In other words, the Monte Carlo approach assumes that the electricity price is subject to a given probability model, e.g. Gaussian distribution. However, the actual distribution of electricity price may not be accurately represented by parametrized families of probability distributions. The other classical method is to approximate the price trace by ARIMA models or stochastic processes, where the correlation between prices at neighboring hours can be taken into account. However, these models can only consider a constant correlation that does not change along the process.

Here we rely on the Multiple Quantile Graphical Model (MQGM) proposed in [99] to represent the price distributions in a non-parametric way. Let random variable y_i denote the electricity price in the i-th hour with i ranging from 1 to 24, and use the random vector Y (the concatenation of all these y_i) to represent the daily price curve. We are interested to learn the distribution of Y and draw samples from that distribution as price scenarios for our stochastic programming. Furthermore, these random variables y_i impact and depend on each other, and all these variables might also depend on exogenous variables such as the weather and the time (e.g. day in the week). We denote these exogenous variables as a random vector X. The Multiple Quantile Graphical Model is based upon the neighborhood selection idea where each random variable is modeled as a conditional Gaussian distribution giving the remaining variables, but instead use multiple quantile regression to represent the distribution for $y_i|(y_{\neg i}, X)$ in a non-parametric way.

We only briefly present the MQGM here. In modeling the distribution of a random variable, the cumulative distribution function (CDF) gives the *probability* values as a function of the *quantity* values, while the quantile function does the opposite: it gives the *quantity* values as a function of the *probability* values. Either the CDF or the quantile function gives us all the probabilistic properties of a random variable. The MQGM method is able to discover the quantile function by learning of the available data set through quantile regression. Quantile regression estimates the quantile of a random variable y through regression over relevant variables x, hence representing the distribution of y. The basic approach of (single) quantile regression is to directly estimate the conditional α-quantile of a response variable given some inputs. This is achieved by solving the convex optimization problem

$$\underset{\theta}{\text{minimize}} \quad \sum_{i=1}^{m} \psi_\alpha(y^{(i)} - \theta^T x^{(i)})$$

where $\psi_\alpha(z) = \max\{\alpha z, (\alpha - 1)z\}$ is the quantile loss function, α denotes the interested quantile ($0 < \alpha < 1$, e.g. $\alpha = 0.5$ corresponds to estimating the median of y), $y(1), \ldots, y(m)$ are sample outputs, and $x(1), \ldots, x(m)$ are sample inputs. Multiple quantile regression attempts to build models for multiple different quantiles simultaneously. The details of the MQGM can be found in [99].

After learning the non-parametric representation for the conditional distribution of each y_i variable given the remaining variables, the Gibbs sampling procedure can be efficiently adopted thanks to the conditional structure of the distribution representation. We adopt the Gibbs sampling to generate the price scenarios and choose the price forecast value as the starting sample in the iterative procedure of Gibbs sampling. The MQGM approach and the Gibbs sampling have the following advantages: the method is able to accurately represent the distribution of the random variable in a non-parametric way, especially when the real distribution cannot be captured by any parametric model (e.g. Gaussian distribution); the relative dependence among prices from nearby hours can be captured, i.e. each sampled price curve should be smooth in the sense that the hourly price does not jump up and down, and the prices around the peak hour should generally be higher than other hours.

To demonstrate the quality of the sampled price scenarios, we applied the MQGM method described above to learn the price distributions for a node in MISO (*ALTE.CC.RIVS1* - a node near Madison, MI). The learning data set consists of 2 years historical price records from January 2014 to December 2016 together with the corresponding weather information. The exogenous variables include the day of a week, the hour of a day, the maximum and minimum temperature of the day. Feature mapping, e.g. radial basis function, is applied to enhance the feature space and improve the representative ability of the variables. The histogram of the real historical prices and the Gibbs samples for different hours are displayed in Fig. 3.12 and Fig. 3.13. From the comparison between sampled histogram and the real histogram of prices, we see that the MQGM method is able to accurately model the price distributions. Besides, we also observe that the distributions of prices from different hours are not the same, and these distributions are not Gaussian.

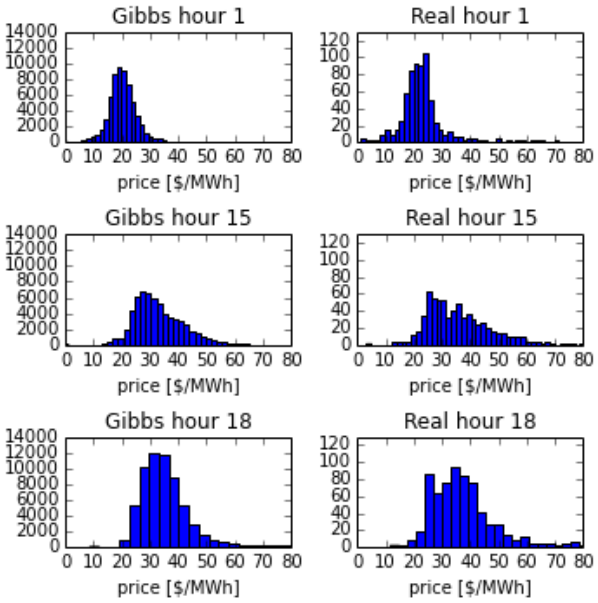


Figure 3.12: Histogram comparison for energy price. Left: counts of Gibbs samples; right: counts of historical observations.

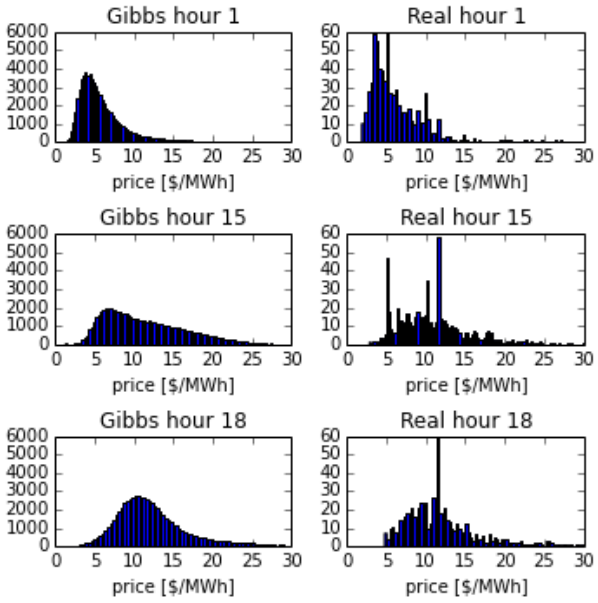
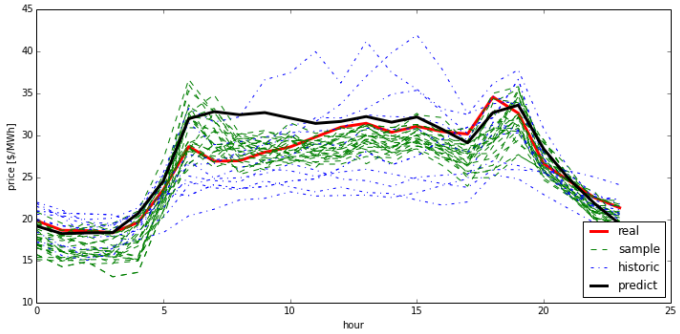


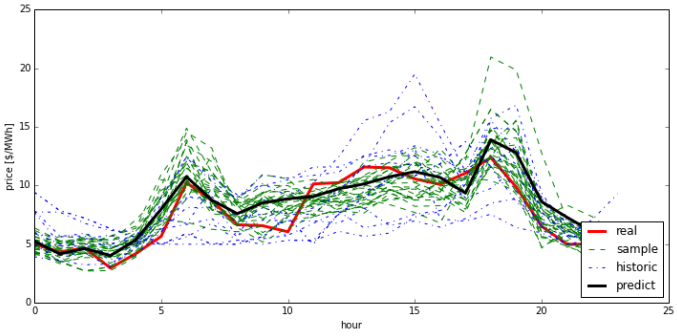
Figure 3.13: Histogram comparison for regulation price. Left: counts of Gibbs samples; right: counts of historical observations.

Examples of the sampled price curves are plotted in Fig. 3.14, in which the red solid line denotes

the real price curve, the black solid line denotes the predicted price curve by linear regression, the green dashed lines are the Gibbs samples which serve as the price scenarios in the stochastic programming, while the blue dashed lines are historical daily price curves up to ten prior days (purely for comparison purpose). From the plots we see that the sampled price curves are pretty smooth and very close to the actual prices; the variance of the samples changes from hour to hour, as the exogenous variables (hour of day and weather information) change along the day; we also observe that the regulation prices follow the trends of the energy prices but are much lower. We will use these generated samples as the price scenarios for the case study in Section 3.3.3.



(a) Energy price scenarios on Oct 4, 2015



(b) Regulation price scenarios on Oct 4, 2015

Figure 3.14: Examples of sampled price scenarios.

3.3.2 Bidding Strategy

In this section, we develop the optimal bidding strategy for the industrial loads in day-ahead markets with energy and regulation provision. The optimal bidding strategy is based on stochastic programming, which hedges the risk from price uncertainties by considering a set of price scenarios. The market prices are treated as stochastic variables where $\lambda_{s,h}$ represents the energy price and $\rho_{s,h}$ stands for the regulation price. The subscript h denotes the hour of the day and s denotes the scenario index.

The decision variables are the power consumption baseline $P_{l,s,h}$ for each potline l and the regulation capacity $R_{l,s,h}$ it provides. Suppose the power consumption quantity in the long-term contract equals to $\sum_l P_l^{max}$, i.e. the industrial plant has already paid a fixed amount of money such that it can consume $\sum_l P_l^{max}$ at any hour [2], where P_l^{max} is the maximum loading level of potline l. We use $E_{s,h}$ to represent the energy to sell in scenario s at hour h. Then we have $E_{s,h} = \sum_l (P_l^{max} - P_{l,s,h})$ for each hour in each scenario.

We want to develop the plant's offering curves for the day-ahead market, i.e. the energy offer MW/Price pairs and the regulation offer MW/Price pair. For energy offers, after the optimal values for $E_{s,h}, s = \{1, ..., S\}$ are obtained from the proposed stochastic programming method, the hourly energy offering curve can be constructed by connecting the MW/Price pairs ($E_{s,h}$, $\lambda_{s,h}$) from different scenarios in the same hour. These MW/Price pairs will form the bidding curve for that hour. For regulation offering, since only one MW/Price pair can be submitted for each hour, we will use the average over all scenarios as the bid. Note that in our stochastic programming model, the regulation capacity R is also sub-indexed by scenario s even though regulation bidding only involves one MW/Price pair while energy bidding requires a curve of up to ten MW/Price pairs; otherwise, the offered regulation capacity would be the same for all the scenarios, which could further lock up the offered energy to be the same among scenarios, as the quantities of available energy and regulation capacity are intertwined with each other.

The constraints and objective for the optimal bidding stochastic programming are presented in the following.

Constraints

Operational Limits. The loading level of the potiline is determined by the decision variables $P_{l,s,h}$ and $R_{l,s,h}$, i.e. its power consumption rate ranges between $P_{l,s,h} - R_{l,s,h}$ and $P_{l,s,h} + R_{l,s,h}$. Obviously, we have the following operational limits,

$$\begin{aligned} P_l^{min} \leq P_{l,s,h} + R_{l,s,h} \leq P_l^{max} \quad \forall l, s, h \\ P_l^{min} \leq P_{l,s,h} - R_{l,s,h} \leq P_l^{max} \quad \forall l, s, h \end{aligned} \tag{3.25}$$

where P_l^{min} and P_l^{max} are the minimum and maximum loading level allowed by the potline, respectively. Additionally, both $P_{l,s,h}$ and $R_{l,s,h}$ should be non-negative.

Thermal Balance. As mentioned before, even though the smelting pots are able to adjust their power consumption fast and accurately, it is crucial to maintain the thermal balance inside of the pots. The temperature within the potlines' should be kept within a certain range to ensure high smelting efficiency as well as operation safety. Since the pots' temperature is dependent on

[2] Note that this assumption can be easily relaxed.

their energy consumption, the energy consumption for every successive τ_l hours need to be greater than E_l^τ, as in

$$\sum_{h'=h}^{h+\tau_l-1} P_{l,s,h'} \geq E_l^\tau \qquad \forall l, s, h \tag{3.26}$$

where E_l^τ is the minimum input energy required for τ_l hours to sustain the temperature. We do not consider regulation capacity in the above constraint because the regulation signal is well balanced and its integral over the hours is almost zero.

Production Schedule. Furthermore, the plant's daily production amount (throughput) is usually scheduled by certain higher-level longer-horizon plant planning. It is also assumed that the plant has storage capability, such that extra products can be stored in case of over-production, and when the production is not enough, the previously stored products can be cashed out for delivery. Hence, the total energy consumption of the plant during the operating day, which is proportional to the aluminum production quantity, is bounded as in

$$E_d^{min} \leq \sum_{h,l} P_{l,s,h} \leq E_d^{max} \qquad \forall s \tag{3.27}$$

where E_d^{min} and E_d^{max} are the daily minimum and maximum energy consumption (in MWh), which is proportional to the minimum and maximum aluminum production amount (in kg). The minimum and maximum aluminum production amounts are determined from the higher-level planning and the storage capacity.

Bidding Curve Monotony. As required by the market, the energy bidding curve for each hour should be monotonous. In other words, a larger amount of energy must be offered at a higher price. In order to get a monotonous bidding curve, we require the following constraint to hold:

$$E_{s,h} - E_{s',h} \leq 0 \quad \forall h, s, s' : O_h(s) + 1 = O_h(s')$$

where $O_h(s)$ denotes the order of the energy price for each scenario in hour h. In the stochastic programming framework, the prices in all scenarios are known, and in each hour, we order the scenarios in an ascending order of price. For example, if s is the scenario with the lowest price in hour h, then $O_h(s)$ equals 1; if s' is the scenario with the highest price in hour h, then $O_h(s')$ equals S - the total number of scenarios. The ordering in different hours could be different. The above equation is equivalent to the following:

$$\sum_l (P_{l,s',h} - P_{l,s,h}) \leq 0 \;\; \forall h, s, s' : O_h(s) + 1 = O_h(s') \tag{3.28}$$

where we instead use the decision variables $P_{l,s,h}$.

Objective

The revenues from aluminum production is assumed to be a quadratic function of its power consumption, i.e.

$$C_P = \sum_s p_s \sum_h \sum_l (a_{2,l} P_{l,s,h}^2 + a_{1,l} P_{l,s,h}) \tag{3.29}$$

in which p_s stands for the probability of scenario s, $a_{1,l}$ and $a_{2,l}$ are the coefficients for the quadratic expression. The revenue function takes a nonlinear form to reflect the fact that the smelting efficiency changes with the loading level, i.e. $P_{l,s,h}$ - the power consumption rate of the potline. Other forms of production revenue expressions can also be considered. The revenues from energy offering C_E and regulation provision C_R are expressed as

$$C_E = \sum_s p_s \sum_h \sum_l -\lambda_{s,h} P_{l,s,h} \tag{3.30}$$

$$C_R = \sum_s p_s \sum_h \sum_l (b_l + \rho_{s,h}) R_{l,s,h} \tag{3.31}$$

respectively. There is a negative sign before the energy price $\lambda_{s,h}$, because we are selling energy to the market which corresponds to the counterpart of power consumption $P_{l,s,h}$. In Eq. (3.31), b_l is the net profit for regulation provision per MW which accounts for both the mileage compensation and the moving cost. Note that the regulation revenue C_R considers both capacity and mileage: the market pays for regulation capacity by price $\rho_{s,h}$, which will be settled after the market clearance of all bids and offers, and we can forecast the price by learning from historical data; the regulation mileage causes equipment degradation and the market also compensates for the mileage, and we use b_l to consider these two effects to simplify the model.

To sum up, the stochastic programming for the optimal bidding strategy is formulated as:

$$\begin{aligned} \underset{P_{l,s,h}, R_{l,s,h}}{\text{maximize}} \quad & C_P + C_E + C_R \\ \text{subject to} \quad & (3.25) - (3.31) \\ & P_{l,s,h} \geq 0, \;\; R_{l,s,h} \geq 0 \quad \forall l, s, h \end{aligned}$$

The overall formulation is a nonlinear optimization problem where the only nonlinear term comes from the aluminum production revenue function in Eq. (3.29). By modeling the revenue function by quadratic expressions, the overall problem is a quadratic programming problem that can be solved efficiently; other forms of nonlinear revenue function can also be solved by general nonlinear optimization solvers such as IPOPT.

3.3.3 Case Study

We consider an aluminum smelting plant with two potlines. The potlines' parameters are the same as in Table 3.2. Parameters P_l^{min} and P_l^{max} are the lower and upper bounds for the power consumption rates of the potlines; τ_l is the critical time duration for minimum energy usage requirement to keep the temperature above safety boundary, and E_l^{τ} is the corresponding energy required for each potline.

The aluminum production revenues of the potlines are approximated by quadratic functions, and the corresponding parameters are listed in Table 3.4. In Table 3.4, the potline's profit per MW of regulation provision are also listed as b_l. The regulation following profit b_l is the net profit per MW (mileage compensation minus moving cost) for following the regulation command.

Table 3.4: Quadratic Parameters for Production Profit

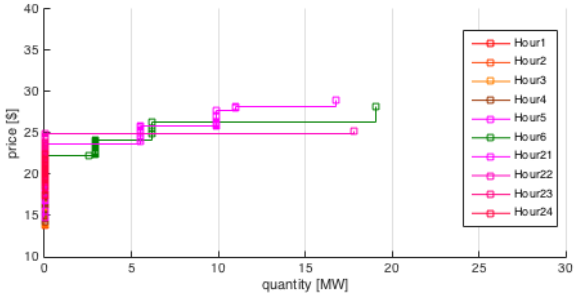
l	a_l^2 [\$/MW^2]	a_l^1 [\$/MW]	b_l [\$/MW]
1	-0.2	60	1
2	-0.4	80	1

As discussed in Section 3.3.1, we employ the Multiple Quantile Graphical Model to represent the distributions of the hourly prices conditioned on their nearby prices as well as other exogenous factors including weather and time information. The model is able to capture the joint distribution of the hourly prices as well as the conditional relationships among prices. After learning the parameters for the MQGM from historical data, we utilize Gibbs sampling to generate price curve samples, which serve as the price scenarios for our stochastic programming optimal bidding model in Section 3.3.2. As shown in Fig. 3.14, the price scenarios generated by Gibbs sampling constitute a very good representation of the possible future prices, which will be used for the case study. All the prices correspond to a bus node around Madison WI (phi_ALTE.CC.RIVS1) in the MISO. For this case study, 1000 Gibbs samples are generated and we randomly choose 25 samples as the price scenarios. Note that a larger number of scenarios can represent the future prices more accurately and more safely, though it may increase the computation time for the optimization problem. In Fig. 3.14, the green dashed lines represent the sampled price scenarios. The probabilities of all scenarios are assumed to be equal.

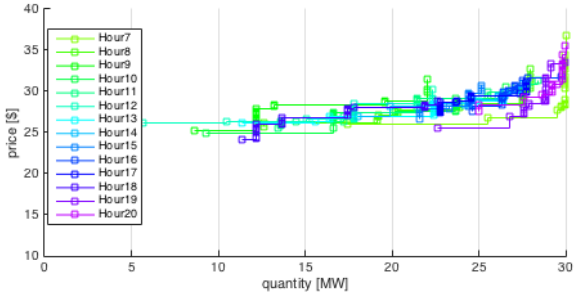
Simulation Results

The optimal bidding problem for our case study is a quadratic programming problem, and we use MATLAB's *quadprog* ("interior-point-convex" algorithm) to solve it. The computation is very fast and the problems can be solved with one minute. From the optimization results, the hourly

bidding curves for each of the 24 hours of the next operating day are generated for the energy market, as displayed in Fig. 3.15. Compared with the energy prices in Fig. 3.14, we observe that the plant is very conservative in selling energy during the price valley hours (e.g. offer to sell no more than 20 MW of energy at the beginning and end of the day), and are aggressive in selling energy during the price peak hours (e.g. offer to sell 30 MW of energy for as high as \$35/MWh during the middle of the day). The bidding curves make sense as it is wise to focus on aluminum production when the energy price is lower and sell extra energy back to the market when the energy price is higher.



(a) Valley hours



(b) Peak hours

Figure 3.15: Hourly bidding curves for energy.

With the bidding curves displayed above, we can obtain the final market shares for our load by comparing the bidding curves with the final clearing prices of the day, as we also have the actual prices for the day under study, i.e. the red solid lines in Fig. 3.14. According to the final clearance, the total profit of the day is \$159,648.4 (energy \$10,744.0, regulation \$3,560.1, and industrial production \$145,344.3). By looking at the revenue composition, we learn that the industrial production still contributes the most to the total revenue, but energy and regulation also contribute a significant amount of revenue. If we instead optimize the productions only by using the point-wise prediction, i.e. the black curves in Fig. 3.14, and calculate the market clearance

by comparing with the final prices, then its economic performance is worse because we cannot hedge the price uncertainties. For example, the electricity market revenues (energy and regulation) obtained by these two approaches over two months are compared in Fig. 3.16. The average revenues over these days are $6507.0 by using prediction and $9862.4 by using scenarios, respectively.

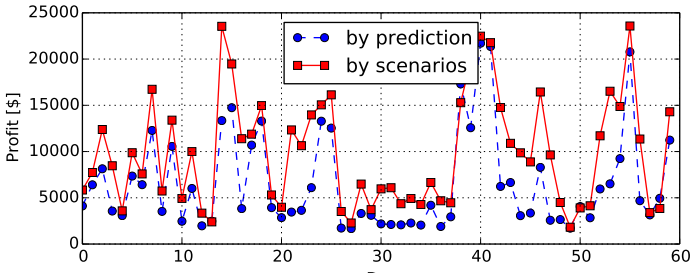


Figure 3.16: Electricity market revenues comparison.

According to the market clearance results, the finalized energy to sell and regulation to provide by the plant are plotted in Fig. 3.17; the hourly power consumption rate of each potline is also plotted. When the energy prices are lower (e.g. the first 5 hours), the aluminum smelter focuses on production - the potlines operate at their highest power consumption level, and the load offers zero regulation as the high loading level leaves no room for regulation provision - its power consumption rate cannot move up to follow the regulation command. While during the peak hours with higher energy prices (e.g. during the middle of the day), the plant is more active in selling to the market, and it offers around 25 MWh energy and 20 MW regulation during these hours. Fig. 3.18 illustrates the clearing process for hour 1 and hour 12, where we compared the bidding curve in that hour with the actual price in the same hour. According to the examples here, the cleared energy to sell in these two hours are around 34 MW and 58 MW, respectively, which correspond to the energy to sell in these two hours in Fig. 3.17.

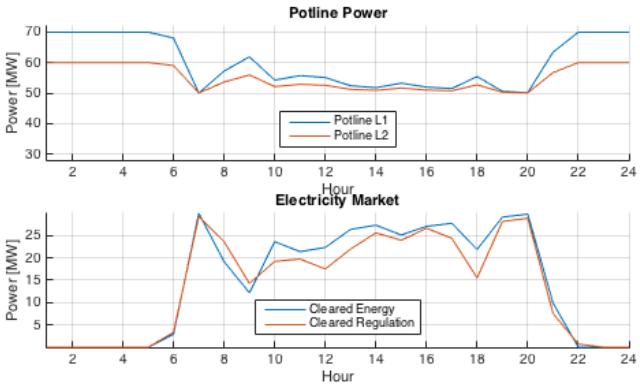


Figure 3.17: Cleared energy and regulation.

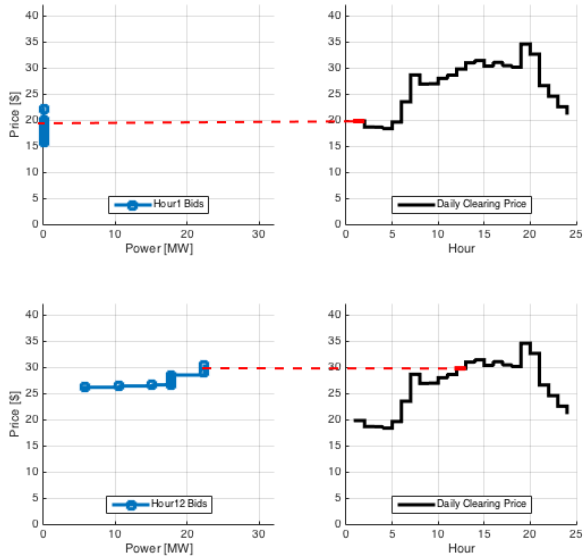
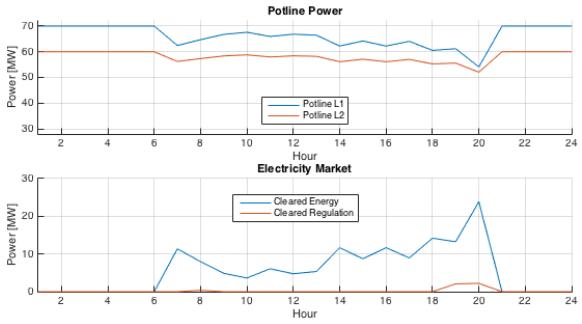


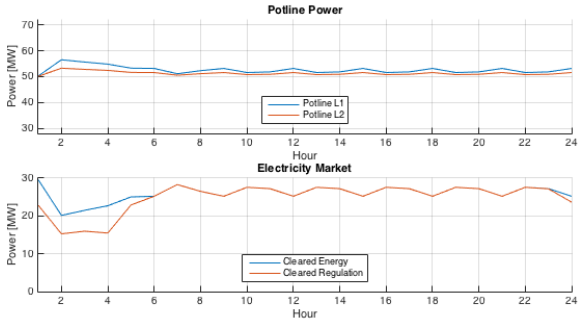
Figure 3.18: Clearing process for hour 1 and 12.

If we manually adjust the regulation profit/cost, then the final clearing results will change as expected. We adjust the net profit for regulation provision per MW to be -15 and 15 $/MW, respectively, solve the stochastic programming problems to get the bidding curves, then obtain the final clearing results by comparing with the actual price curves. The final clearing results are displayed in Fig. 3.19. According to Fig. 3.19(a), the load seldom provides regulation if the cost to provide regulation is too high. While in Fig. 3.19(b), it is pushing regulation provision to the largest extent as the regulation mileage revenue is highly profitable - the potlines are operating at their middle power consumption rate for maximum regulation availability - to move the farthest both up and down. The bidding curves generated by our model are displayed in Fig. 3.20, in which we observe that the bidding curves in Fig. 3.20(b) are dedicated to selling around 25 to 30 MW

energy. In other words, the plant wants to operate near its middle loading level, thus it has the maximum regulation availability to move its power consumption up and down.

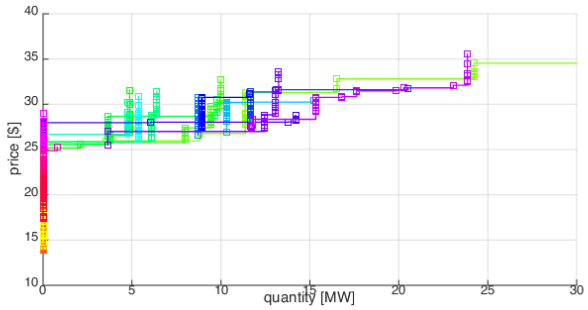


(a) high regulation cost

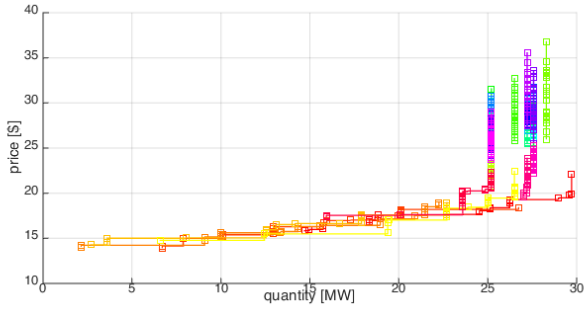


(b) high regulation profit

Figure 3.19: Cleared results: top b_l = -15 \$/MW, bottom b_l = 15 \$/MW



(a) high regulation cost



(b) high regulation profit

Figure 3.20: Bidding curves: top b_l = -15 $/MW, bottom b_l = 15 $/MW

Chapter 4

Handling Complexity of Process

Industrial plants such as steel manufacturers are important to our society and its infrastructure development, they also have great potentials to participate in demand response programs because of the intensive usage of electricity. However, the chemical processes are usually complex and the corresponding scheduling is difficult, which discourages these industrial plants to fully participate in the electricity market. In this chapter, we aim to handle the complexity and develop scheduling methods for steel manufacturers to fully take advantage of the demand response programs to lower their electricity cost and improve their operation efficiency.

The description of the steel manufacturing process is available in Chapter 1, Section 1.2.2. We rely on the Resource Task Network (RTN) modeling framework to represent the steel plant, which is described in Section 4.1. Based on the resource task network framework, we have investigated the scheduling problem of steel plants with controllable transformers in [100], which is presented in Section 4.2. The relevant computational issues have also been studied in [24], in which we propose approaches to make the computations in the steel plant scheduling more tractable, as described in Section 4.3. As presented in Section 4.4, we have also derived an RTN-based scheduling model in [101] for the steel plant to optimize its production activities such that it can benefit the most from electricity markets, both energy and spinning reserve markets.

4.1 Resource Task Network Modeling Framework

We address the scheduling problem by using a discrete-time RTN formulation [102–105]. The RTN model involves two types of nodes: resources and tasks. The resource nodes represent all entities that are relevant to the process such as raw materials, intermediate and final products, equipment, and utilities. The task nodes include all tasks that take place during the production, such as a chemical process or the transportation of material. The task can change the amount of the

product in a resource node and/or the status of the equipment (occupied or available). Resources are necessary to promote state changes or tasks' execution. For example, a certain task can only start if there is enough input material and equipment available. The network connecting these two types of nodes and the interaction parameters on the network describe the detailed interactions between resources and tasks.

The steel manufacturing plant considered in this chapter has parallel units for each of the four stages but the proposed methods can also be applied to a plant with any number of units and any number of stages. Generally speaking, a more rigorous model which represents the process more accurately requires a larger number of variables, resulting in a scheduling problem which is difficult to solve. Among the RTN models presented in [45], the Aggregated Equipment Resource and Simple Transfer Tasks (Basic RTN) model is the best selection for electricity cost minimization because of the reduced computational complexity and the negligible differences in the final solution compared to the solutions of more rigorous models which take a significantly longer time to run. We briefly provide an overview of this model in this subsection and then present the extension to include the flexibility provided by OLTCs.

The Basic RTN assumes that the processing abilities of the units for the first three stages are almost the same, i.e. the equipment units in the same stage have the same power consumption and the same processing time for each heat. Thus, the parallel units for the first three stages are assumed identical to simplify the problem. In contrast, casters need to be considered individually due to their different processing and setup times, and all the heats belonging to the same campaign group should be processed in the same caster sequentially.

The Basic RTN model optimizes the electricity cost by shifting the time of production activities. Thereby, the EAF melting power is always equal to the equipment nominal value. The parallel units for the first three stages are assumed identical, while the casters are considered individually because different casters are designed for casting specific products. The model also overlooks the differences in transfer times between units from two successive stages: parallel units in the same stage might be located far away from each other, so that the transportation time between successive stages actually depends on the exact locations of the specific unit u in the previous stage and the unit u' in the next stage; the transportation time also depends on the transportation mode, e.g. by train or by crane. For instance, if there are 3 units for the first stage and 4 units for the second stage, then there are at least 12 possible transportation times. The Basic RTN assumes the transfers are independent of the units' locations, which is a simplification that might lead to an under- or over-estimation of the actual transfer time. With these assumptions, the Basic RTN achieves a relatively simple model with fast computation and reasonable results.

4.1.1 Resource Task Network

The Basic RTN is illustrated in Fig. 4.1. The tasks considered are the four main production stages, i.e. melting in the EAF, decarburization in the AOD, refining in the LF, and casting in the CC, as well as the transfer tasks between the stages.

Each task is indexed by i and the binary variable $N_{i,t}$ assigns the start of task i to time point t, i.e. $N_{i,t} = 1$ indicates that task i begins at time slot t. For the operational tasks of the first three stages and all the transfer tasks, there is one task for every heat h. For instance, if there are H heats that have to be produced, then we have H melting tasks to schedule: one melting task per heat. The tasks for the first three stages are therefore denoted by i_{E_h}, i_{A_h}, and i_{L_h}, respectively. Similarly, we denote the transfers by i_{EA_h}, i_{AL_h}, and i_{LC_h}.

Unlike the first three stages where we do not distinguish between parallel units, the casters are treated individually. Hence, we need to assign each casting job to a specific caster. A casting task is denoted by $i_{C_{g,u}}$ with a pair of indices (g, u), where g stands for the casting campaign group and u stands for a specific caster. This is because task $i_{C_{g_1,u_1}}$ is different from task $i_{C_{g_1,u_2}}$, e.g. these two tasks' durations are not equal due to the different setup times of the two individual casters. So, we have to consider both $i_{C_{g_1,u_1}}$ and $i_{C_{g_1,u_2}}$ in the problem formulation in order to take into account all possible caster assignments. Of course, casting g_1 will be assigned to just one caster. Consequently, only one of these two tasks will be scheduled, while the other one never takes place.

The transfer task is forced to be executed immediately after the completion of its preceding processing task, which is generally required in the steel manufacturing process. While on the other hand, after transferred to the next stage, heats may need to wait before the following processing step for the equipment to become available or for lower electricity prices.

The resources considered are equipment, electricity, intermediate products, and final products. As the intermediate products are transferred from one stage to the next, each intermediate product needs to be associated with the location where this heat of metal currently is. For example, EA_h^d denotes the intermediate product of the specific heat h located at the transfer destination (superscript d), i.e. the inlet of AOD, which is waiting to be processed by AOD; while EA_h^s denotes the same intermediate product located at the transfer start (superscript s) which is waiting to be transferred. The sets of resources considered in the model are processing units (EAF, AOD, LF, and CC), electric energy (EL), intermediate products (EA^s, EA^d, AL^s, AL^d, LC^s, and LC^d), and final product (H). Nonnegative continuous variables $R_{r,t}$ represent the value of resource r ($\forall r \notin EL$) at time t. For instance, $R_{EAF,t} = 1$ means there is one EAF available at time slot t. Nonnegative continuous variables $\Pi_{EL,t}$ are used to summarize the energy usage over all tasks in time slot t.

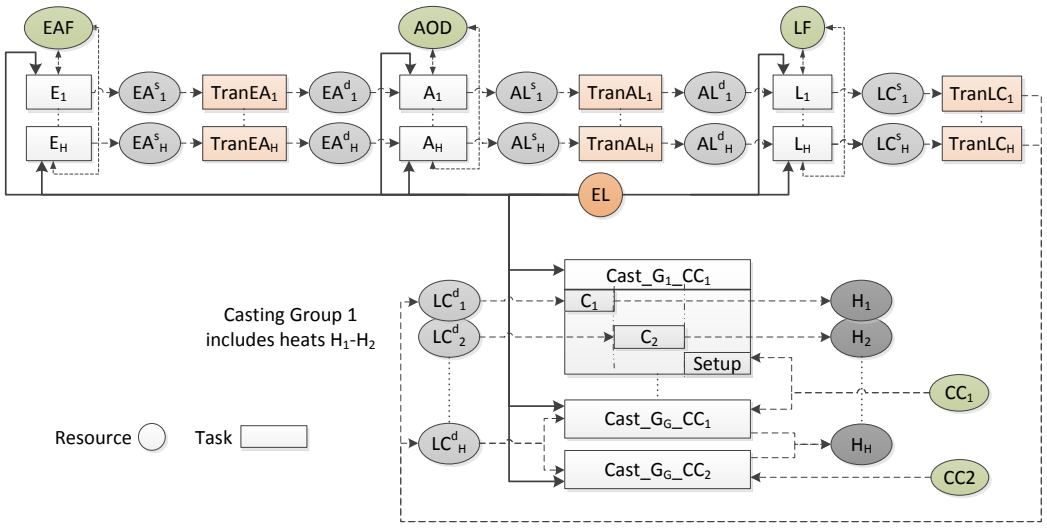


Figure 4.1: Resource task network for a steel plant.

The network flowchart in Fig. 4.1 indicates how each task interacts with each resource. Processing tasks interact with electricity resources continuously, but interact with other resources discretely. Continuous interaction means that the task consumes or generates the resource continuously during the processing time of the task. For example, for simplification we assume that the melting task consumes electric energy continuously during the entire task. While discrete interaction means that the interaction only takes place at very distinct time points during the task. For example, the melting task occupies one EAF at the very beginning of the task and only returns the EAF at the end of the melting process. The melting task may last for several time slots, but it only interacts with the resource EAF in two time slots.

The detailed interactions are described by interaction parameters. Interaction parameter $\mu_{r,i,\theta}$ quantifies how much of resource r is consumed or generated by task i at the relative time slot θ - the time slot that is θ slots after the start of task i. The interaction parameters for the melting task with its interactive resources are illustrated in Fig. 4.2. There are three different time references: $Time$ stands for the actual hour of the day; t is the index for the uniform time grid; the relative time index θ is related to the start of the task. Discrete-time formulation assumes that the task can only start at the beginning of the time slot, but can end anywhere within the time slot. The slots' width of the time grid in Fig. 4.2 is $\delta = 30$ minutes. The time duration for melting is assumed to be 80 minutes. Hence, the melting spans 3 ($\lceil 80/30 \rceil$) time slots. Note that the melting is completed before the last time slot ends. This rounding error due to discrete-time formulation might underestimate the potential flexibility gained from rescheduling. Using a finer time grid can

alleviate this issue but increases the computational complexity.

In Fig. 4.2, we assume that the melting task of heat h, i_{E_h}, starts at $t = 3$. This task interacts with resources EAF, EL, and EA_h^s. At the very beginning, the task reduces EAF by one as it uses the operation unit; after the completion of the melting process, EAF is increased by one as the EAF is freed up. Also, EA_h^s is increased by one to promote the execution of the following transfer; the melting consumes electric energy continuously during its entire duration. Note that the energy consumption of the last melting time slot is less than the previous slots because the task actually completes before the end of that slot.

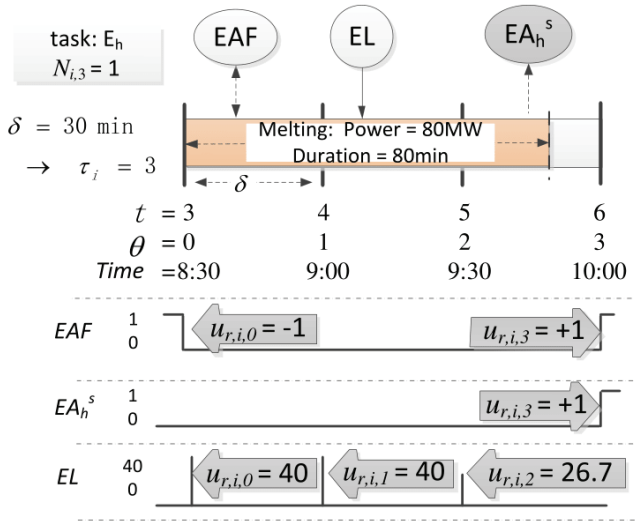


Figure 4.2: Illustration of interaction parameters for a melting task.

4.1.2 Mathematical Formulation

In this section, we use the RTN model, formulate the dependencies mathematically and integrate it into an optimization problem to determine the daily schedule of a plant. The objective is to minimize the electricity cost, and the time-based electricity prices are assumed to be known ahead of time. In a time-of-use (TOU) pricing system, these hourly prices can be obtained easily; in a wholesale market, price forecast methods [106, 107] can be utilized to provide the forecast prices as the price signal for our model; in a tiered pricing system, we need to design an objective function with the detailed parameters in the tiered pricing. Since our focus here is on modeling and optimization, we assume the price signal is known. The formulations in this section are based on [45], and have been updated for simplified presentation.

Resource Balance

Resource evolution over the time horizon is managed by the excess resource balance, as in

$$R_{r,t} = R_{r,t-1} + \sum_{i} \sum_{\theta=0}^{\tau_i} \mu_{r,i,\theta} N_{i,t-\theta} + \pi_{r,t} \quad \forall r \notin EL, t \tag{4.1}$$

in which the value $R_{r,t}$ of resource r at time slot t is equal to its previous value at $t-1$ adjusted by the amounts generated/consumed by all relevant tasks. The above constraint applies to all the resources described in Section 4.1.1 except for the electricity resource.

Nonzero interaction parameters $\mu_{r,i,\theta}$ imply interaction and task i only interacts with resource r when the task is ongoing. In other words, the interaction takes place at time slot t only if the task i starts θ earlier than t ($N_{i,t-\theta} = 1$) with θ being less than τ_i - the duration of task i. Equipment maintenance can be modeled by adding parameters that influence the excess value of equipment in the balancing equation. For example, adding $\pi_{CC_1,5} = 1$ on the right side of (4.1) means caster CC_1 needs to be maintained and cannot be used at time slot 5.

Meanwhile, the electricity usage is calculated as

$$\Pi_{EL,t} = \sum_{i} \sum_{\theta=0}^{\tau_i} \mu_{EL,i,\theta} N_{i,t-\theta} \qquad \forall t \tag{4.2}$$

where $\Pi_{EL,t}$ is equal to the electric energy usage by all possible tasks at time slot t; the right side of (4.2) first sums over all tasks and then for each task, it sums over all possible starting times of task i for which task i would still be running at time t.

Task Execution

Operational constraints (4.3), (4.4), and (4.5) are used to ensure that tasks are executed the proper number of times. The constraints

$$\begin{aligned} \sum_{t} N_{i_{E_h},t} &= 1 \qquad \forall h \\ \sum_{t} N_{i_{A_h},t} &= 1 \qquad \forall h \\ \sum_{t} N_{i_{L_h},t} &= 1 \qquad \forall h \end{aligned} \tag{4.3}$$

ensure that all heats are processed only once by the first three stages. For the casting stage, we distinguish between individual casters and we need to assign one caster for each job. If we have C individual casters and G groups of heats, then the number of possible casting tasks is $C \times G$. But not all casting tasks are actually being executed. Each group of heats should be executed only

once by any unit u from the available casters CCs. This is reflected in the following constraint

$$\sum_{u \in CCs} \sum_{t} N_{i_{C_{g,u}},t} = 1 \qquad \forall g \tag{4.4}$$

Similarly, the intermediate products should only be transferred once between each of the stages, as defined by

$$\begin{aligned} \sum_{t} N_{i_{EA_h},t} &= 1 \qquad \forall h \\ \sum_{t} N_{i_{AL_h},t} &= 1 \qquad \forall h \\ \sum_{t} N_{i_{LC_h},t} &= 1 \qquad \forall h \end{aligned} \tag{4.5}$$

Transfer Time

The transfer task is forced to be executed immediately after the completion of its preceding processing task, which is common in the steel manufacturing process. This requirement is embodied by enforcing

$$\begin{aligned} R_{EA_h^s,t} &= 0 \qquad \forall h, t \\ R_{AL_h^s,t} &= 0 \qquad \forall h, t \\ R_{LC_h^s,t} &= 0 \qquad \forall h, t \end{aligned} \tag{4.6}$$

The variable $R_{EA_h^s,t}$ is either 0 or 1, and if it is equal to 1, then it indicates that the intermediate product EA_h^s is waiting at time slot t. The above constraint requires that there is no waiting time for any of the intermediate products with superscript s.

The transfer time of the intermediate products are assumed to be w_{EA}, w_{AL}, and w_{LC}, which are independent of the specific heats and the operation units. The maximum allowable transportation time W_{EA}, W_{AL}, and W_{LC} are also defined which makes sure that the cooling effect during transportation does not adversely affect the processing of each heat in the next stage, as the products' quality may be compromised by low temperature and would have to be compensated by expensive reheating. The transfer time constraints are therefore given by

$$\begin{aligned} \delta \sum_{t} R_{EA_h^d,t} + w_{EA} &\leq W_{EA} \qquad \forall h \\ \delta \sum_{t} R_{AL_h^d,t} + w_{AL} &\leq W_{AL} \qquad \forall h \\ \delta \sum_{t} R_{LC_h^d,t} + w_{LC} &\leq W_{LC} \qquad \forall h \end{aligned} \tag{4.7}$$

in which δ is the time slot width. The $\sum_t R_{EA_h^d,t}$ is the total time slots that intermediate product EA_h waits before entering into the next stage. The constraint states that for each intermediate

product, the transfer time plus the waiting time should be upper bounded to prevent adverse cooling effects.

Product Delivery

The final products should be deliverable by the end of the time horizon, which is enforced by

$$R_{H_h,T} = 1 \qquad \forall h \tag{4.8}$$

in which T is the last time slot of the time horizon.

Objective

The overall objective of the optimization is to minimize the total energy cost as given by

$$\min \quad \sum_{hr} price_{hr} \sum_{t \in T_{hr}} \Pi_{EL,t} \tag{4.9}$$

The hourly electricity prices $price_{hr}$ are given as inputs and T_{hr} is the set of time slots in hour hr. Here it should be noted that while we only optimize the electricity costs, all other production requirements are enforced through constraints.

4.2 Modeling Controllable Transformers

In EAF-steel manufacturing, most of the electric energy is consumed by the furnaces in the melting stage. Hence, adding flexibility to the power consumption of this stage has the highest impact on the overall electricity cost. In existing steel plant scheduling literature like [45], the EAFs' power consumption rate and processing time are treated as constant parameters, and only the starting times of the melting tasks are shifted in time to provide operational flexibility. However, according to practical operation, it is also possible to adjust the furnaces' power consumption rate very quickly by controlling the on-load tap changers (OLTCs), which gives opportunities to further increase the flexibility of the steel plant's energy management. Hence, our goal is to incorporate this flexibility into the RTN model and exploit it to further reduce the electric energy cost of the steel manufacturing plant.

In [100], both of the proposed Multiple Melting Modes and Arbitrary Flex Melting enable the steel plants to participate more actively in the electricity market by exploiting the EAFs' capability to adjust their power consumption rate through controlling the OLTCs. In the Modes model, we assume that the OLTC setting and therefore the melting power for each heat can be chosen from a set of modes, and this setting does not change until the melting of this heat completes. Hence, we

have increased the number of tasks in the formulations, though only one mode should be chosen for melting each heat. While in the Flex model, we further extend the EAFs' flexibility and assume that the transformers' OLTC setting can be adjusted for every single time slot, thus the EAF power rate can change at any time during the melting of each heat. The consequence of allowing for this flexibility is that the time duration of the melting task and the melting power cannot be covered by a fixed set of options, and we need extra binary variables to denote the end of the melting tasks; we also need to introduce continuous variables $P_{h,t}$ to denote the melting power of the melting task for heat h at time slot t.

4.2.1 Multiple Modes Melting

In the following, we integrate the flexibility of EAFs provided by the controllability of the transformer taps into the model described in the previous section. In this section, we limit such flexibility by making the following modeling assumptions:

- the OLTC setting and therefore the melting power for each heat can be chosen from a set of modes. This setting does not change until the melting of this heat completes.
- for each of these modes, the melting task of each mode fully spans the entire required time slots.

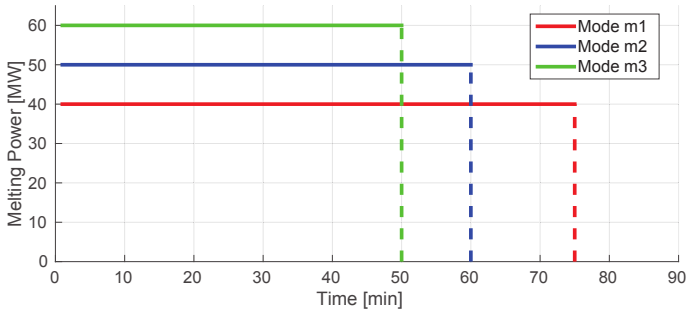
The second assumption makes it convenient to consider ancillary services such as spinning reserve for future research, as it has to be guaranteed that the service can be provided during the entire time slot.

Resource Task Network

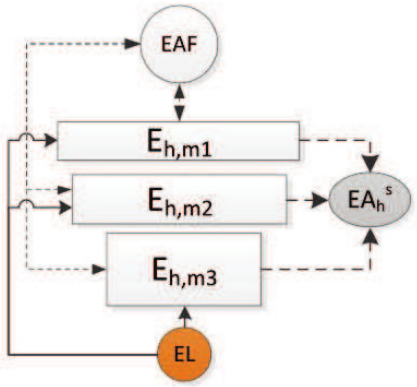
Suppose the nominal melting power of the EAF is P. We assume that the melting power can be adjusted between P^L and P^U from its nominal value P. Note that P, P^L and P^U, as parameters of the EAF, are the same for all the heats. While on the other hand, suppose the nominal melting time for heat h is w_h, which depends on the specific heat. Then according to the nominal case, the electric energy needed to melt heat h is equal to $w_h \cdot P$, and we assume that this amount of energy does not change when we adjust the melting power between P^L and P^U. Hence, the melting time of heat h is bounded between $w_h^L = w_h P/P^U$ and $w_h^U = w_h P/P^L$. The time slots spanned by the melting task of heat h is then bounded by $\tau_h^L = \lceil w_h^L/\delta \rceil$ and $\tau_h^U = \lfloor w_h^U/\delta \rfloor$[1]. We assume that we can change the OLTC settings so that the melting time of heat h is $\tau_m \delta$, where τ_m is a

[1]Generally $\lceil w_h^L/\delta \rceil$ is smaller than $\lfloor w_h^U/\delta \rfloor$; if not, try to reduce the value of δ or formulate the melting modes differently.

integer between τ_h^L and τ_h^U, and the corresponding melting power is $Pw_h/(\tau_m\delta)$. In other words, there are $M_h = \tau_h^U - \tau_h^L + 1$ melting modes to choose from. For each mode, the melting duration τ_m is known, then the melting power and the OLTC setting can be calculated accordingly. In the example illustrated in Fig. 4.3, processing heat h in stage EAF has three melting options, i.e. power consumption rates at 60, 50, and 40 MW and lasting for 50, 60, and 75 minutes, respectively; the updated RTN graph for the modes modeling is displayed in Fig. 4.3(b). Note that the areas under these three lines are the same and are equivalent to the total energy needed for melting the steel scrap. With the extra flexibilities given by the modes modeling, the plant operator can choose among different power consumption curves therefore has more options in minimizing its energy cost as well as helping the power grid to balance demand and supply.



(a) Melting options



(b) Updated RTN

Figure 4.3: Melting in stage EAF with 3 modes.

Unlike the Basic RTN in the previous section, the melting tasks in the Multiple Modes Melting are denoted by $i_{E_{h,m}}$ in which h stands for the heat and m represents the melting mode. Hence, we

have increased the number of tasks compared to the Basic RTN. The resources and the interaction parameters are still the same, except that we now need to update the interaction parameters for each melting task $i_{E_{h,m}}$.

Mathematical Formulation

The formulations (4.1) - (4.9) still apply, except that the melting execution constraint in (4.3) needs to be replaced by the following constraint which incorporates the choice of the melting mode.

Melting Mode Choice

Only one mode should be chosen for melting each heat h, i.e.

$$\sum_{m=1}^{M_h} \sum_{t} N_{i_{E_{h,m}},t} = 1 \qquad \forall h \tag{4.10}$$

hence, only one $i_{E_{h,m}}$ from all possible modes $m = 1, \cdots, M_h$ should actually take place.

Demand Charge As mentioned above, the objective for the Multiple Modes Melting is still the minimization of the total energy cost as given in (4.9). However, the modeling methods proposed in this section are also able to consider the peak demand charge if that is being imposed. In order to take the peak demand into account, we can add a continuous variable P^k to denote the peak demand over the considered horizon, and then we can include this peak demand P^k in the minimization objective with the penalty price, $price_{dc}$ (\$/MW), as its coefficient. Since we already have the energy usage $\Pi_{EL,t}$ for every time slot, we can model the peak demand through the following constraint (given that we are minimizing P^k):

$$P^k \geq \Pi_{EL,t}/\delta \qquad \forall t \tag{4.11}$$

in which δ is the length of the time slot. In the objective function, we add the term $price_{dc} \cdot P^k$ to reflect the demand charge [108]. In the short-term scheduling problems considered in this section, the plants optimize over a single day whereas demand charges are usually only charged for the one single maximum power consumption over the entire month. We can take this into account by setting the lower bound of P^k to be P^k_{max}, with P^k_{max} being a constant equal to the maximum power so far over all days in the ongoing month. The demand charge discussed here applies to all the three models presented in this section.

4.2.2 Arbitrary Flexible Melting

In this section, we further extend the EAFs' flexibility. Compared with Section 4.2.1, here the EAFs are even more flexible by making the following assumption:

- the transformers' OLTC setting can be adjusted for every single time slot, thus the EAF power rate can change during the melting of each heat.

Resource Task Network

The consequence of allowing for adjustment of the melting power during the melting process is that the time duration of melting is not directly associated with a given melting power any more, but generally varies between τ_h^L and τ_h^U. This means that we need extra variables to denote the end of the melting tasks. However, since the heats are still required to be transferred immediately after having been processed, the end of the melting equals the start of the succeeding transfer. Thus, we use i_{EA_h} to denote the end of melting.

Furthermore, since the power consumption rate of the melting process is assumed to be adjustable, we cannot connect the melting task to the electric energy resource by fixed parameters. Hence, we introduce variables $P_{h,t}$ to denote the melting power of the melting task for heat h at time slot t. Accordingly, we remove the connection between the melting task and the electricity resource, which means that $\Pi_{EL,t}$ now only sums the energy consumption for the last three stages. The updated interaction parameters are illustrated in Fig. 4.4. Note that resource EAF interacts with both the melting task and the transfer task. The other tasks and resources remain the same as in Section 4.1.

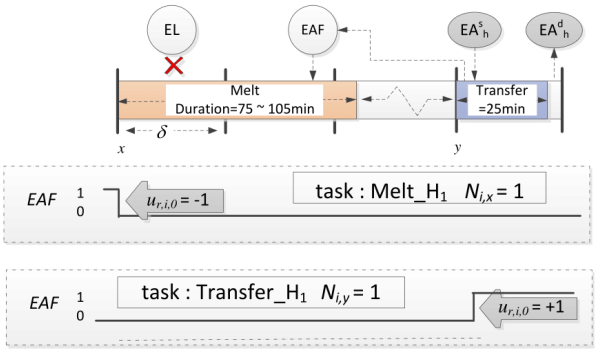


Figure 4.4: Illustration of interaction parameters for a melting task for arbitrary flexible melting.

Mathematical Formulation

Equations (4.1)-(4.8) still apply except for the parameter updates. In addition, we need the following constraints to enable the flexible scheduling.

Melting Duration Bounds All melting tasks should be completed within given bounds, i.e.

$$\sum_{t'=t+\tau_h^L}^{t+\tau_h^U} N_{i_{EA_h},t'} \geq N_{i_{E_h},t} \qquad \forall h \tag{4.12}$$

Keep in mind that the start of the transfer equals the end of melting. Hence, the equation states that if melting task i_{E_h} starts at time slot t ($N_{i_{E_h},t} = 1$), then the transfer i_{EA_h} must start between time slots $t + \tau_h^L$ and $t + \tau_h^U$.

Melting Power Bounds The melting power rate of the EAFs are constrained by the lower and upper bounds P^L and P^U as defined by

$$P^L \cdot S_{h,t} \leq P_{h,t} \leq P^U \cdot S_{h,t} \qquad \forall h \tag{4.13}$$

where $S_{h,t}$ is the melting status: $S_{h,t} = 1$ indicates the melting of heat h is taking place during t; $S_{h,t} = 0$ indicates that heat h is not in the melting stage at time t.

Melting Status Evolution The melting status evolves according to

$$S_{h,t} - S_{h,t-1} = N_{i_{E_h},t} - N_{i_{EA_h},t} \qquad \forall h, t \tag{4.14}$$

with initial condition $S_{h,0} = 0$. The evolution of the variables and their dependencies are visualized in Fig. 4.5: a change in the melting status is initiated by the start of the melting task and the start of the transfer. It is also worth to emphasize that variable $S_{h,t}$ is modeled as continuous variable to reduce the computational burden, but constraint (4.14) ensures it to be binary if all the other constraints hold.

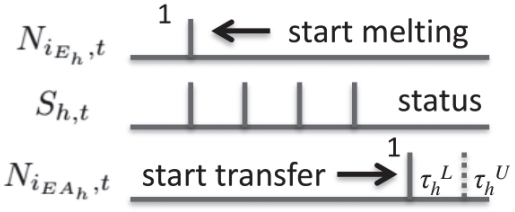


Figure 4.5: Relation of melting status, start of melting and transfer.

Melting Energy Requirement The total energy needed for melting heat h is assumed constant and can be calculated according to the nominal case. It is enforced by including the following constraint

$$\sum_t P_{h,t} \cdot \delta = P \cdot w_h \qquad \forall h \tag{4.15}$$

which states that the summation of the consumed energy over the time horizon is equal to the product of nominal power and nominal melting time.

Objective Again, the objective is to minimize the total energy cost which is now defined as

$$\min \sum_{hr} price_{hr} \sum_{t \in T_{hr}} (\Pi_{EL,t} + \sum_{h} P_{h,t} \cdot \delta) \tag{4.16}$$

As stated in Section 4.2.2, $\Pi_{EL,t}$ only sums the last three stages' energy usage, the melting energy usage, i.e. $P_{h,t}$ times the duration δ, needs to be considered additionally in the objective function. The scheduling model in this section minimizes the objective function (4.16) while subject to constraints (4.1)-(4.8) and (4.12)-(4.15).

4.2.3 Case Study

In this section, we carry out case studies to demonstrate the effectiveness of the proposed models. We consider the daily scheduling problem for an electric arc furnace steel plant.

Problem Parameters

The layout of the steel plant and the corresponding parameters are taken from the example in [45]. In particular, the plant considered has two EAFs, two AODs, two LFs and two CCs. Each heat belongs to a particular casting campaign group as given in Table 4.1. The nominal power consumptions are provided in Table 4.2, where the power consumption is independent of the specific heat. The nominal processing times are shown in Table 4.3, where different heats require different processing times. Combining Table 4.2 and 4.3, we observe that around 90% of the total electric energy consumption takes place at EAFs. The transfer times w_{EA}, w_{AL}, and w_{LC} are 10, 4, and 10 minutes for the three between-stage transfers successively; the maximum waiting time W_{EA}, W_{AL}, and W_{LC} are 240, 240, and 120 minutes, which are higher than practical values but help to provide more flexibilities in scheduling for testing purpose. The caster setup times are 70 minutes for CC_1 and 50 minutes for CC_2. The hourly-based electricity prices are given in Fig. 4.6. Recall that in a wholesale market, a price forecast is needed to provide the expected hourly prices. Note that the first hour in Fig. 4.6 is not necessarily 00:00-01:00 in the day, as we need to consider workers shift time and products' due time; that is why the hourly electricity prices are lower during the middle of the time horizon. Generally speaking, the larger the difference between peak price and non-peak price, the more benefits our methods bring to this industrial load.

For scheduling with flexible EAFs, the melting power rate is assumed to be adjustable between 75% to 125% of the nominal value. The melting times are changed accordingly, e.g. the heats with nominal processing time equaling 80 minutes now can be melted by a duration between 64 minutes

and 106.7 minutes. With this assumption, the steel plant can obtain an extra flexibility of 80 MW (80MW·50% · 2) for the hours when both furnaces are operating. The required energy for melting each heat is set as the product of the nominal power multiplied by the nominal processing time. In the simulations, we do not take demand charges into account because, as already mentioned, the case study spans one day which is the most reasonable time range for this application given the need for price predictions and also for practical operational reasons. Meanwhile, it can be assumed that for the majority of the months, the demand charges are equal to the demand charge price times the plant's power capacity, i.e. the total power consumption of all the equipment. This is because a well utilized plant often needs to use all the equipment concurrently.

Table 4.1: Steel heat/group correspondence [45]

group g	G_1	G_2	G_3	G_4	G_5	G_6
heats	H_1-H_4	H_5-H_8	H_9-H_{12}	$H_{13}-H_{17}$	$H_{18}-H_{20}$	$H_{21}-H_{24}$

Table 4.2: Nominal power consumptions [MW][45]

$d_{h,u}$	EAF_1	EAF_2	AOD_1	AOD_2	LF_1	LF_2	CC_1	CC_2
$power_{h,u}$	85	85	2	2	2	2	7	7

Table 4.3: Nominal processing times [min][45]

$d_{h,u}$	EAF_1	EAF_2	AOD_1	AOD_2	LF_1	LF_2	CC_1	CC_2
H_1-H_4	80	80	75	75	35	35	50	50
H_5-H_6	85	85	80	80	40	40	60	60
H_7-H_8	85	85	80	80	20	20	55	55
H_9-H_{12}	90	90	95	95	45	45	60	60
$H_{13}-H_{14}$	85	85	85	85	25	25	70	70
$H_{15}-H_{16}$	85	85	85	85	25	25	75	75
H_{17}	80	80	85	85	25	25	75	75
H_{18}	80	80	95	95	45	45	60	60
H_{19}	80	80	95	95	45	45	70	70
H_{20}	80	80	95	95	30	30	70	70
$H_{21}-H_{22}$	80	80	80	80	30	30	50	50
$H_{23}-H_{24}$	80	80	80	80	30	30	50	60

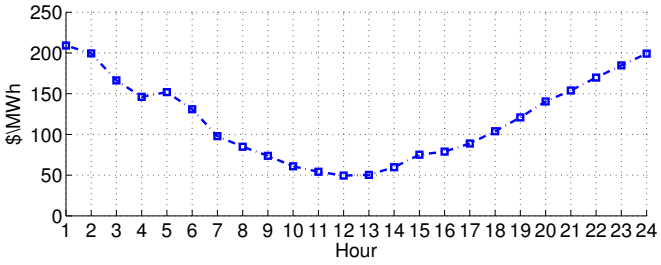


Figure 4.6: Hourly electricity price[45]

Scheduling Results

The optimal scheduling results of the three RTN models described above are given in Table 4.4. Different numbers of heats for daily scheduling are considered to simulate different production profiles for the steel plant. The more heats, the higher is the productivity of the plant, i.e. the higher is the amount of manufactured steel, but the less is the flexibility due to reduced free capacity. Obviously, model complexity and computation difficulty are directly related to the number of heats. Generally speaking, a larger number of heats to produce results in a more complex scheduling problem which is more difficult to solve, as the problem size depends on the number of heats. In Table 4.4, the column *Heats* lists the number of heats; the column *Model* compares the three models in which *Basic* stands for Basic RTN, *Modes* for Multiple Modes Melting, and *Flex* for Arbitrary Flexible Scheduling; the next three columns list the problem size - the number of binary variables, the number of total variables, and the number of constraints; the column *MIP* presents the final integer objective function value - the value of the objective function with the final integer (feasible) solution; the column *GAP* displays the relative objective gap, which is the relative distance between the best integer objective (by a feasible integer solution) and the objective of the best bound remaining (not necessarily an integer solution); the column *CPU* gives the final computation time by the solver. The maximum computation time is set to 7200s and the relative optimality tolerance is 10^{-6}. All of the models are implemented in Matlab and are solved by TOMLAB/CPLEX on a linux 64 bit machine.

Table 4.4: Energy cost minimization with $\delta = 15$min

Heats	Model	# bin	# var	# con	MIP(k$)	GAP	CPU(s)
4	Basic	2496	6048	3397	26.239	0	0.3
	Modes	3264	6816	3397	25.972	0	0.3
	Flex	2496	6816	4917	25.858	0	1.7
8	Basic	4992	11232	6122	60.173	0	0.8
	Modes	6528	12768	6122	57.501	0	1.1
	Flex	4992	12768	9162	57.332	0	31.1
12	Basic	7488	16416	8847	104.301	0	2
	Modes	10176	19104	8847	100.061	0	24
	Flex	7488	18720	13407	99.990	1.97%	7200
17	Basic	10560	22848	12253	171.615	0	4
	Modes	14208	26496	12253	159.454	0	170
	Flex	10560	26112	18713	160.896	3.72%	7200
20	Basic	12480	26784	14297	222.427	0	9
	Modes	16704	31008	14297	204.611	0	37
	Flex	12480	30624	21897	211.459	9.00%	7200
24	Basic	14976	31968	17022	299.782	0	320
	Modes	19968	36960	17022	277.283	0	83
	Flex	14976	36576	26142	287.077	11.36%	7200

From the results, we make the following observations:

- The flexibility increases the computation difficulty. For most cases, the computation times for *Modes* are larger than *Basic*. For $Heats = 12, 17, 20$, the *Flex* model does not converge to the optimal integer solution within two hours of computation.
- The flexibility reduces the electricity cost. For all cases, the final objective values of *Modes* are less than *Basic*. For $Heats = 4, 8, 12, 17$, the *Flex* model achieves the best integer solution; for $Heats = 20, 24$, the *Flex* model does not perform better than *Modes* due to computational difficulties.

The computation difficulty arises from the model's complexity and the tightness of the formulations: a large number of extra constraints are needed to represent the EAF's flexibilities for the model *Flex*; the poor tightness from the formulations such as Eq.(4.13) deteriorate the quality of the relaxed solution.

The equipment occupancy charts for scheduling 24 heats by models *Basic* and *Modes* are displayed in Fig. 4.7, in which different heats are represented by different colors. We can observe that the solution is valid: each heat is processed sequentially by all four stages; each group of heats form a campaign and are casted continuously; there is no conflict in equipment assignment, i.e. each equipment is occupied by one single task for every time slot. It also demonstrates that the RTN model is able to generate detailed and practical schedules which can be clearly understood by the steel plant operators. Besides, compared with the scheduling results by the model *Basic*, the melting durations according to the model *Modes* are shorter, and the melting schedule wisely skips the locally high price in hour 5.

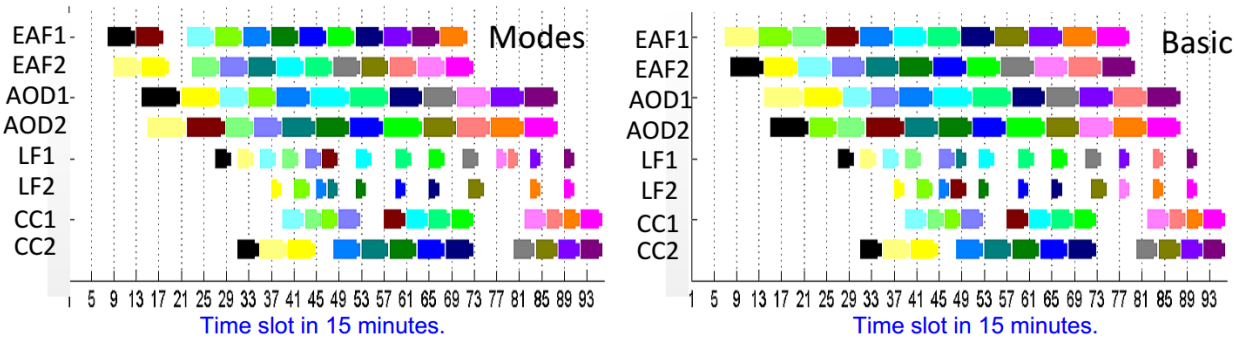


Figure 4.7: Equipment occupancy for 24 heats.

The hourly energy consumptions corresponding to the optimal schedule of 12 heats from the three models are compared in Fig. 4.8. We observe that as the flexibility increases, more of the energy is consumed during the price valley.

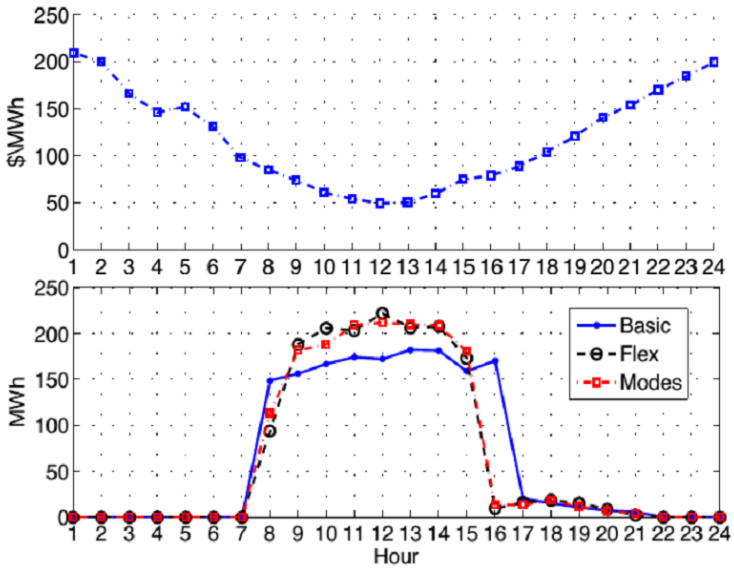


Figure 4.8: Hourly energy consumptions for scheduling 12 heats.

As demonstrated by the case studies, the two proposed models achieve remarkable savings and could encourage steel plants to participate more actively through demand response. In particular, the Multiple Melting Modes model provides a good trade-off between enabling the exploitation of the flexibilities given by the OLTCs and computational complexity.

The above analysis is studied for the comparison among the three optimal scheduling methods which considers the energy price information. Traditionally, the scheduling objective in such a capital-intensive industry is to minimize the make-span of the production, i.e. to produce as much as possible. This is because the plant operators want to fully utilize the plant facilities, whose cost is an important part of the investment. Nowadays, with the development of smart grid, the industrial plants such as steel plants find opportunities in reducing energy cost by actively participating in the electricity markets, the scheduling objective tends to account for the energy cost - which is the objective function in this chapter. To have a understanding of how much the optimal scheduling with consideration of electricity price can save, the following table lists the energy cost from the traditional scheduling.

Table 4.5: Energy cost from traditional scheduling.

Heats	4	8	12	17	20	24
Energy Cost (k$)	92.340	165.045	213.409	251.696	271.961	326.590

Comparing the above table with the MIP column in Table 4.4, we observe that the difference in electricity cost between these two optimization criteria is very big. Of course, the cost difference

heavily depends on the prices over the day. If the price curve is flatter, then the difference will be smaller.

4.3 Computational Approaches for Efficient Scheduling

As seen from previous discussions, the scheduling of steel plants is very complex and the involved computations are intense. In this section, we focus on these difficulties and try to improve the computations in the following two ways: from the modeling aspect, we add additional constraints as cuts to reduce the search space for the MIP problem, and from the algorithmic aspect, we design a tailored branch and bound algorithm that utilizes the knowledge from steel manufacturing to reduce the branching complexity. As our focus is on the computation, in this section we will study the BasicRTN scheduling problem which is presented in Section 4.1; however, the mathematical notations have been slightly changed here for better descriptions of algorithms.

4.3.1 RTN and Mathematical Formulations

The RTN of the considered scheduling problem is the same as Fig. 4.1. The set of resources is denoted by $\mathbb{S}$, i.e. $\mathbb{S} = \{\text{EAF, AOD, LF, CC}\} \cup \{\text{EA}_h^{\text{s}}, \text{EA}_h^{\text{d}}, \text{AL}_h^{\text{s}}, \text{AL}_h^{\text{d}}, \text{LC}_h^{\text{s}}, \text{LC}_h^{\text{d}}, \text{H}_h | h \in \mathbb{H}\} \cup \{\text{EL}\}$ with $\mathbb{H}$ as the set of heats to produce. Intermediate products at different locations (start or destination of the corresponding transfer) are treated as different resources and are specified by superscripts s or d, respectively. For example, AL_h^{d} represents the intermediate product between stage AOD and LF that has already been transferred to the LF stage and is waiting to be processed.

We use a discrete time grid with uniform slot width of t_0 and we use T to denote the total number of time slots. A matrix $Y \in \mathbb{R}^{|\mathbb{S}| \times T}$ is used to denote the available amounts of these resources at all time slots, in which $|\mathbb{S}|$ is the size of $\mathbb{S}$. Each element in Y is a continuous variable, $y_{s,t}$, which represents the available amount of resource s at time t. For example, $y_{\text{EAF},t} = 3$ means there are three furnaces available at time slot t; $y_{\text{EL},t} = 50$ MWh means 50 MWh of electric energy is used by the steel plant during time slot t. Due to their physical meanings, most $y_{s,t}$ can actually only take discrete values such as 0, 1, or 2. However, they are modeled as continuous variables since a larger number of discrete variables generally leads to a problem that is more difficult to solve. As discussed later the constraints in the optimization model will enforce these variables to take discrete values.

There are seven kinds of tasks in total: four operational tasks at each of the four stages and three transfer tasks between the stages. For all kinds of tasks except for the casting task in the last stage, the number of tasks is equal to the number of heats to produce; these tasks are denoted by the task type with the corresponding heat as subscript, e.g. E_h stands for the melting of heat

h in the EAF stage, and EA_h (without superscript s or d) denotes the transfer of heat h between stage EAF and AOD. Meanwhile, the casting task is denoted as $\mathrm{C}_{g,u}$ which stands for the casting of group g by caster unit u. As mentioned before, the tasks in the CC stage are executed by group instead of by heat. Besides, since generally different casters are designed for casting different slabs, we need to specify the caster for the casting task. In other words, C_{g_1,u_1} is different from C_{g_1,u_2}, e.g. their processing durations might be different due to the different casters. We use $\mathbb{K}$ to denote the set of tasks, i.e. $\mathbb{K} = \{\mathrm{E}_h, \mathrm{EA}_h, \mathrm{A}_h, \mathrm{AL}_h, \mathrm{L}_h, \mathrm{LC}_h | h \in \mathbb{H}\} \cup \{\mathrm{C}_{g,u} | g \in \mathbb{G}, u \in \mathbb{CC}\}$, with $\mathbb{G}$ and $\mathbb{CC}$ as the set of casting campaign groups and available casters, respectively.

The starting times of all the tasks are denoted by a $|\mathbb{K}|$-by-T binary matrix X, in which $|\mathbb{K}|$ is the size of $\mathbb{K}$. Each element of X is a binary variable, $x_{k,t}$, which represents whether task k starts at time slot t. For example, $x_{\mathrm{E}_h,t} = 1$ means the processing of heat h in stage EAF starts at time slot t; only one out of $x_{E_h,t}, t = \{1, \ldots, T\}$ is non-zero since this task only takes place once.

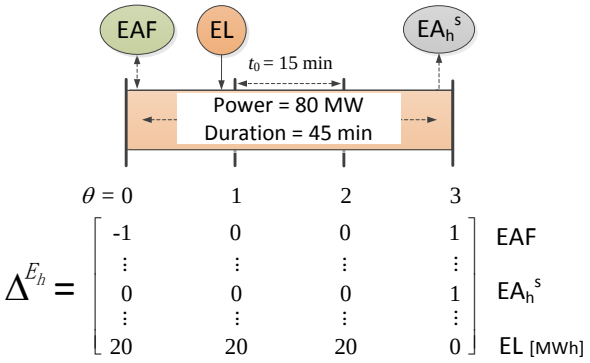


Figure 4.9: Illustration of interaction parameters for a melting task.

In Fig. 4.1, the networks of how each task interacts with each resource are represented by arrows. For each task $k \in \mathbb{K}$, its interaction parameter Δ^k is a $|\mathbb{S}|$-by-$(\tau_k + 1)$ matrix that quantifies how much task k consumes/generates of each of the resources as it proceeds, in which τ_k denotes its duration as a number of time slots. For example, its element $\Delta^k_{s,1}$ quantifies the interaction between task k and resource s at the beginning of the first time slot during this task, and Δ^k_{s,τ_k+1} quantifies the interaction at the end of the last time slot. A zero element means that there is no interaction, and Δ^k is very sparse as a task typically only interacts with a few resources. To better understand the interaction parameters, an example for a melting task is given in Fig. 4.9: the duration of the melting task is 45 minutes and the time slot width is $t_0 = 15$ minutes. This task interacts with resources EAF, EL and $\mathrm{EA}^{\mathrm{s}}_h$, hence its interaction parameter matrix are all zeros except for these

three rows. At the beginning of the task, it uses one furnace so it reduces EAF by one. Meanwhile at the end of the task, it releases that furnace hence EAF is increased by one; $\mathrm{EA}_h^{\mathrm{s}}$ is also increased by one as it has just been generated. Besides, the melting task consumes electric energy at every time slot within its duration.

For the presentation of the following computational approaches, we summarize the relevant formulations in the following.

Resource Balance

The resource balance equation describes the interaction between each resource and its relevant tasks, as in

$$y_{s,t} = y_{s,t-1} + \sum_{k\in\mathbb{K}} \sum_{\theta=0}^{\tau_k} \Delta_{s,\theta}^k \cdot x_{k,t-\theta} \quad \forall s \in \mathbb{S}_{\neg\{\mathrm{EL}\}}, \forall t \tag{4.17}$$

in which the value of resource s at time slot t is equal to its previous value at $t-1$ adjusted by the amounts generated/consumed by all the tasks, and $\mathbb{S}_{\neg\{\mathrm{EL}\}}$ stands for the set of all the resources except EL. Only nonzero $\Delta_{s,\theta}^k$ implies actual interaction. Besides, the interaction occurs at time slot t only if task k starts θ earlier than t ($x_{k,t-\theta} = 1$), with $\theta \leq \tau_k$. Equation (4.17) enforces the continuous variable $y_{s,t}$ to only take integer values, because: (1) the interaction parameters $\Delta_{s,\theta}^k$ for these resources are integers, (2) $x_{k,t-\theta}$ are binary variables, and (3) the initial values for the resources are integers (zero or the number of available equipment).

Similarly, the electric energy usage of the steel plant is calculated as

$$y_{\mathrm{EL},t} = \sum_{k\in\mathbb{K}} \sum_{\theta=0}^{\tau_k} \Delta_{\mathrm{EL},\theta}^k \cdot x_{k,t-\theta} \qquad \forall t \tag{4.18}$$

where $\Delta_{\mathrm{EL},\theta}^k$ is the electricity used by task k at the θ-th time slot within its execution.

Task Execution

The following constraints make sure that each heat is processed exactly once by all types of tasks within the scheduling horizon, as in

$$\begin{aligned} &X_{\mathbb{K}_{\neg\mathbb{CC}}} \mathbb{1}_T = \mathbb{1} \\ &\mathbb{1}' X_{\mathbb{C}_u} \mathbb{1}_T = 1 \quad \forall u \in \mathbb{CC} \end{aligned} \tag{4.19}$$

in which $X_{\mathbb{K}_{\neg\mathbb{CC}}}$ denotes X without the rows involving the casting tasks; $X_{\mathbb{C}_u}$ denotes the rows of X corresponding to casting tasks by caster u; $\mathbb{1}_T$ means a vector of 1s with length T and $\mathbb{1}$ stand for vectors of 1s with appropriate lengths.

Of course, only one of C_{g_1,u_1} and C_{g_1,u_2} will take place as group g_1 should be casted exactly once.

Waiting Time

In steel plant operations, the transfer tasks are usually enforced to be executed immediately after the completion of its preceding processing task. This can be enforced by setting

$$Y_{\mathrm{EA^s}} = \mathbf{0} \tag{4.20}$$

in which $Y_{\mathrm{EA^s}}$ stands for the rows of Y corresponding to the intermediate products $\mathrm{EA^s}$; $\mathbf{0}$ is a zero matrix with the same dimensions as $Y_{\mathrm{EA^s}}$. The above constraint implies that all intermediate products $\mathrm{EA^s}$ do not stay at any time slot. Similar constraints apply for the intermediate products $\mathrm{AL^s}$ and $\mathrm{LC^s}$.

To simplify the problem, we assume the transfer tasks are independent of the specific heats and the exact locations of the units. The transfer times are denoted as w_{EA}, w_{AL}, and w_{LC}, and the maximum allowable transportation times which prevent adverse cooling effects are denoted as $\bar{w}_{\mathrm{EA}}$, $\bar{w}_{\mathrm{AL}}$, and $\bar{w}_{\mathrm{LC}}$. The maximum waiting time constraint for intermediate products should be satisfied, in order to avoid the expensive reheating:

$$Y_{\mathrm{EA}^{\mathrm{d}}_h} \mathbb{1}_T \leq \frac{(\bar{w}_{\mathrm{EA}} - w_{\mathrm{EA}})}{t_0} \mathbb{1} \tag{4.21}$$

in which $Y_{\mathrm{EA^d}}$ stands for the rows of Y corresponding to intermediate products $\mathrm{EA^d}$ before the transfer. The left side of the constraint corresponds to the number of time slots during which the intermediate product is waiting before being processed. Similar constraints apply for intermediate products $\mathrm{AL^d}$ and $\mathrm{LC^d}$.

Product Delivery

The final products should be available at the end of the time horizon, which is enforced by setting

$$y_{\mathrm{H}_h,T} = 1 \qquad \forall h \tag{4.22}$$

in which $y_{\mathrm{H}_h,T}$ stands for the availability of the final product for heat h at the end of the scheduling horizon.

Objective Function

The objective of the scheduling is to minimize the total electricity cost of the steel production. Given the energy price vector $\lambda_{\mathrm{EL}} \in \mathbb{R}_T$, the overall optimization problem is formulated as

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & Y_{\mathrm{EL}} \lambda_{\mathrm{EL}} \\ \text{subject to} \quad & (4.17) - (4.22) \\ & x_{s,t} \in \{0,1\}, \;\; y_{s,t} \in [0, \bar{y}_s], \;\; \forall s, \forall t \end{aligned}$$

in which $\bar{y}_s$ is the upper bound for the available amount of resource s. For example, $\bar{y}_{\mathrm{EAF}}$ equals to the number of EAF furnaces and $\bar{y}_{\mathrm{EL}}$ equals to the summation of energy usage by all the equipment units in one time slot.

4.3.2 Additional Constraints as Cuts

In steel manufacturing, there are many tasks that are equivalent to each other, e.g. the decarburization of molten metal for two similar batches of products. We can impose an enforced processing order for these tasks, thus the search space of the MIP problem is reduced. Several other types of constraints can also be imposed which reduce the feasible region but with potential sacrifice to the degree of optimality, an example is to restrict the start time of EAF tasks to time intervals that have lower energy price expectations. Here we consider imposing processing orders for the tasks.

In steel manufacturing, the casting sequence for heats belonging to the same casting group are pre-specified - this pre-specified processing sequence results from expert experiences or casting optimization. Intuitively, that processing sequence should also apply to the other three stages. For instance, suppose the casting group G1 consists of heats 1, 2, and 3, and they will be casted one by one sequentially, which means that the intermediate product LC_1 will be casted first. Consequently, LC_1 should be generated first, i.e. heat 1 should be processed before the other two heats in the 3rd stage. Hence, we can define a set of ordered tasks, denoted by $\mathbb{O}$, whose processing sequences are pre-specified. For instance, $\mathbb{O} = \{(\mathrm{E}_1, \mathrm{E}_2), (\mathrm{E}_2, \mathrm{E}_3), (\mathrm{A}_1, \mathrm{A}_2), (\mathrm{A}_2, \mathrm{A}_3), (\mathrm{L}_1, \mathrm{L}_2), (\mathrm{L}_2, \mathrm{L}_3)\}$ for scheduling G1 of heats 1, 2 and 3. The imposed additional constraints (cuts) on processing order enforcement can be written as follows:

$$\sum_{t' \leq t} (x_{k_1,t'} - x_{k_2,t'}) \geq 0 \quad \forall t, (k_1, k_2) \in \mathbb{O} \tag{4.23}$$

in which the ordered tasks set $\mathbb{O}$ considers the processing of heats belonging to the same group for each of the first three stages. In steel manufacturing practice, the above additional constraint is also beneficial as it follows the first-in-first-out principle and reduces the chance of over-waiting and re-heating for the intermediate products.

4.3.3 Tailored Branch and Bound Algorithm

The MIP problems in industrial scheduling are usually solved by commercial solvers [28, 46]. These commercial solvers are powerful, but are designed to handle general optimization problems. We develop a tailored branch and bound algorithm to utilize the special features in steel manufacturing: the heats belonging to the same campaign group are generally processed close to each other.

```
1: procedure TailoredBranchBound
2:     q ← Priority-Queue()                          ▷ pops largest objective first
3:     q2 ← Priority-Queue()                         ▷ pops smallest objective first
4:     q.push(SolveRelaxation({ }))
5:     q2.push(FindIntegerSolutionHeuristics())
6:     while q not empty do
7:         (f, x, y, C) ← q.pop()
8:         q2.push(Rounding((f, x, y, C)))
9:         if q2.first - f ≤ ϵ then
10:            return q2.pop()
11:        else
12:            for C_i in BranchNodes(C) do
13:                q.push(SolveRelaxation(C_i))
14:            end for
15:        end if
16:    end while
17: end procedure
```

Figure 4.10: Tailored branch and bound algorithm

The branch and bound algorithm is presented in Fig. 4.10 in which q and $q2$ are priority queues that store the relaxation solutions at each iteration and the achieved feasible integer solutions, respectively; q provides the lower bounds for the mixed-integer minimization problem, while $q2$ provides the upper bounds. SolveRelaxation(C) is a function that takes input C and solves the relaxation of the original problem plus constraints in C; the relaxation is a linear programming problem and we solve it by using CPLEX's LP solver; the function returns (f, x^*, y^*, C), i.e. the optimal objective value, the optimal values of relaxed integer variables and continuous variables, as well as the corresponding constraints C. The function input C is actually defined as $C = [(a_{k1}, b_{k1}), (a_{k2}, b_{k2}), \ldots, (a_K, b_K)]$, which specifies the start times for each task. For instance, (a_{k1}, b_{k1}) restricts the start time of task $k1$ to be between a_{k1} and b_{k1} - the constraint is actually implemented by setting the upper bounds for $x_{k1,t}$ as zero for t outside of (a_{k1}, b_{k1}) while leaving the upper bounds for other t as one, i.e.

$$\overline{x}_{k1,t} = \begin{cases} 1, & \text{if } a_{k1} \leq t < b_{k1} \\ 0, & \text{otherwise} \end{cases}$$

GetIntegerSolutionHeuristics() is a heuristics method that packs all the tasks to the earliest available equipment units to get a feasible integer solution, which serves as the initial upper bound for the algorithm. Rounding() tries to round the relaxation solution to be integer, and returns the

integer solution if it is feasible.

In order to enforce the heats belonging to the same campaign group be processed close to each other, we term the first heat in each campaign group as the *leader*, and call the other heats belonging to the same group as its *followers*. Similar to the discussions in Section 4.3.2, we require the leader to be processed first, and require its followers to be processed within the time ranges calculated according to their processing time durations. For example, consider group G1 of heats 1, 2 and 3 in the EAF stage (with two EAF units). The leader here is task E_1, and suppose its start time is within (a, b) at a certain node in the branch and bound algorithm. As there are two available furnaces, we require its followers E_2 to be started between $(a, b+\tau_{E_1})$ and E_3 to be started between $(a+\tau_{\mathrm{E}_1}, b+\tau_{\mathrm{E}_1})$. The principle of this requirement is to enforce the offsets and delays as if these followers are packed sequentially to the available equipment units. The relationship of start times for the above example can be described by the following dictionary L that maps the leader to its followers and the corresponding offsets:

$$L[k_{E1}] = \{k_{E2} : (0, \tau_{E1}), k_{E3} : (\tau_{E1}, \tau_{E1})\}$$

1: **function** BranchNodes(C)
2: **if** C == { } **then**
3: **return** $[(0,T), \ldots, (0,T)]$
4: **else**
5: $k^* \leftarrow \arg\max_{k \in L.\mathrm{keys}}(b_k - a_k)$
6: **if** $b_{k^*} - a_{k^*} > \epsilon_d$ **then**
7: $m^* \leftarrow \mathrm{int}(\frac{b_{k^*} - a_{k^*}}{2})$
8: $\{k : (d_a, d_b)\} \leftarrow L[k^*]$
9: $C_1 \leftarrow [\ldots, (a_{k^*}, m^*), (a_{k^*} + d_a, m^* + d_b), \ldots]$
10: $C_2 \leftarrow [\ldots, (m^*, b_{k^*}), (m^* + d_a, b_{k^*} + d_b), \ldots]$
11: **return** $\{C_1, C_2\}$
12: **else**
13: $k^* \leftarrow \arg\max_{k \in \mathbb{K}}(b_k - a_k)$
14: $m^* \leftarrow \mathrm{int}(\frac{b_{k^*} - a_{k^*}}{2})$
15: $C_1 \leftarrow [\ldots, (a_{k^*}, m^*), \ldots]$
16: $C_2 \leftarrow [\ldots, (m^*, b_{k^*}), \ldots]$
17: **return** $\{C_1, C_2\}$
18: **end if**
19: **end if**
20: **end function**

Figure 4.11: Branch by leader heats

With this concept of leader and followers and the restriction on start time described above, instead of branching on the start time intervals for all tasks, we can only branch on the leader tasks and restrict the start times for its follower tasks according to a pre-constructed L. The proposed branching method is described in Fig. 4.11 with parameter ϵ_d as the threshold to switch between branching by leader tasks and branching by all tasks. This method will greatly reduce the complexity of the branching procedure.

4.3.4 Case Study

Numerical studies on the daily scheduling for the same steel plant in Section 4.2.3 are presented in the following to demonstrate the effectiveness of the proposed methods. The hourly energy prices for the case studies are taken from MISO, as displayed in Fig. 4.12.

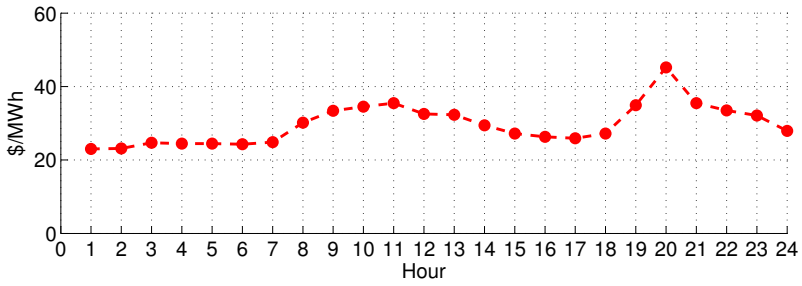


Figure 4.12: MISO hourly prices on 02/06/2014.

Computational Results

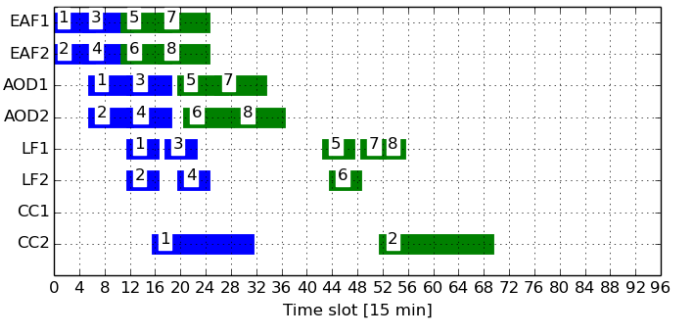
Table 4.6 presents the computational results for the methods proposed in Section 4.3.2 and Section 4.3.3: the column *Groups* gives the campaign groups to produce, e.g. the first row denotes scheduling groups 1 and 2 with the heats as indicated in Table 4.1; the column *c0* stands for solving the original model by CPLEX's MIP solver, while the column *c1* stands for the method proposed in Section 4.3.2 and the column *b1* denotes the tailored branch and bound algorithm proposed in Section 4.3.3 (ϵ_d is chosen as 4); the row *Obj* gives the final objective value of the MIP problem; the row *CPU* shows the computation time for the corresponding test case, where the maximum computation time limit is set to 7200s; the row *lpNum* gives the iteration number of the solving process, where the maximum iteration number is set to 10000. Note that for the test case G1-5, the relaxation solution by CPLEX happens to be a feasible integer solution, hence the corresponding *lpNum* is 0.

Table 4.6: Branch and bound results with $t_0 = 15$min

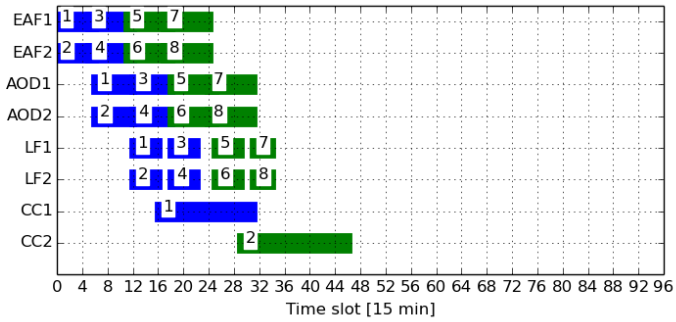
Groups		c0	c1	b1
G1-2	Obj(k$)	24.553	24.553	24.698
	CPU(s)	5.8	3.7	6.2
	lpNum	2460	1985	57
G1-3	Obj(k$)	39.306	39.308	39.665
	CPU(s)	155.4	60.7	50.0
	lpNum	9071	3835	228
G1-4	Obj(k$)	57.857	57.857	58.694
	CPU(s)	60.4	42.7	197.8
	lpNum	3852	2745	280
G1-5	Obj(k$)	69.737	69.737	70.194
	CPU(s)	4.3	7.2	861.0
	lpNum	0	0	478
G1-6	Obj(k$)	86.352	86.352	86.799
	CPU(s)	104.9	80.4	2737.6
	lpNum	3698	2631	725

From the computational results displayed above, we can observe the following: (1) by imposing additional constraints, method c1 reduces the computation time as well as the iteration number, and the final objective values remain the same except for a 0.005% increase for case G1-3; (2) the tailored branch and bound algorithm b1 is able to greatly reduce the iteration number, but the final objective value sees an increase of around 1%. The optimality loss suffered by b1 can be explained by the restriction that we require the heats in the same group be processed close to each other. For example, Fig. 4.13 displays the scheduling results comparison for test case G1-2, where method b1 schedules the LF/CC processing of heats 5, 6, 7 and 8 close to the EAF/AOD stages, which loses the opportunity to utilize the price valley around hours 12-17. Note that the computation time for a single iteration of method b1 is much longer than that of c0 or c1 - this is due to the computation overhead with calling the LP solver, while the solving process of the CPLEX MIP solver has been coherently optimized with many practical techniques. Hence, one has to be careful when comparing CPU time for b1 with any of the other as the time probably could be improved by a more professional implementation of the proposed branch and bound algorithm. Also note that it is possible to achieve the proposed tailored algorithm by using CPLEX MIP solver with priorities setting and additional constraints, where the comparison of CPU time would be fair. For the numerical studies here, it might be more appropriate to focus on iteration numbers for which

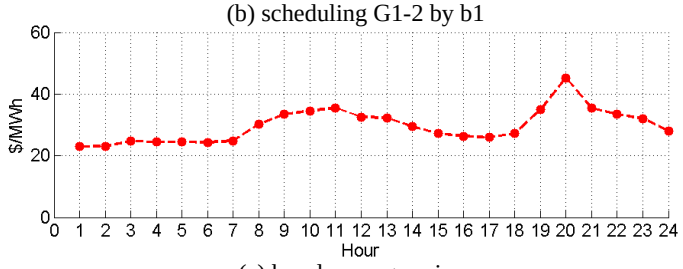
b1 achieves a significant improvement in most of the cases.



(a) scheduling G1-2 by c0



(b) scheduling G1-2 by b1



(c) hourly energy prices

Figure 4.13: Scheduling results comparison.

Another issue we want to emphasize here is the rounding procedure in algorithm b1. The current algorithm rounds each variable to its nearest integer value, which seldom succeeds in yielding a feasible solution, as there are so many constraints on binary variables in the scheduling model. As shown in Fig. 4.14, the upper bound seldom changes along the solving process. The rounding procedure could potentially be improved by taking into account the relationship among these binary variables, e.g. their process-time sequence relationship. A better rounding procedure could potentially further reduces the iteration number.

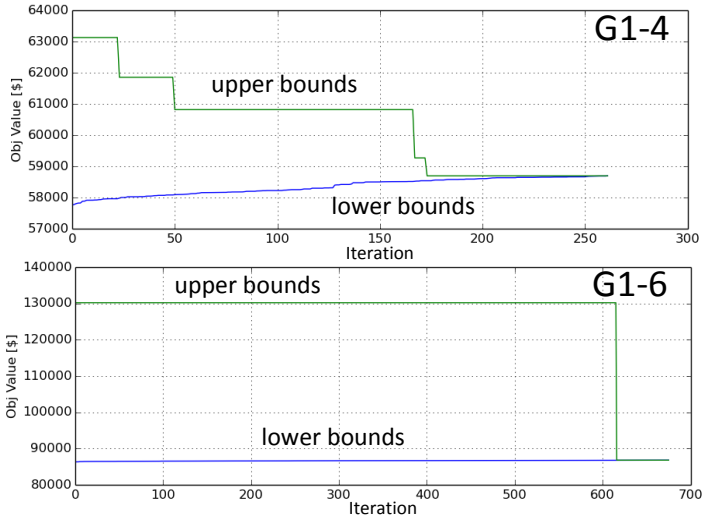


Figure 4.14: Branch and bound iterations.

As demonstrated through numerical studies, the proposed methods show potentials in reducing the computation time and iteration number of the problem considered. The proposed methods are effective in improving the computations, and may play an important role towards developing practical scheduling tools for the steel industry and its demand response.

4.4 Enabling Spinning Reserve Provision

As discussed in Section 4.2, the power consumption rate of the EAFs can be adjusted very quickly by switching the on-load tap changers (OLTCs) of the transformers which supply power to the EAFs. This qualifies the steel plant to be a valid demand response resource for providing spinning reserve. The amount of spinning reserve it can provide depends on the melting power profile, i.e. the power consumption rate of the melting process, and the sustaining (minimal) power the furnace requires to keep the molten metal from solidification.

The payment structures for spinning reserve are different across different electricity markets. In most North American electricity markets, e.g. the Midcontinent Independent System Operator (MISO) where demand response is actively encouraged and where an aluminum smelter (Alcoa's Warrick Operation) has participated as a regulation provider for the first time, spinning reserve is compensated by both reserve capacity and actual allocation. In other words, the spinning reserve provider gets paid for the capacity it has committed to provide independent of if this reserve capacity is dispatched or not; and if it does get dispatched, it receives an additional payment as the allocation compensation. However, the actual dispatch of spinning reserve is very rare. According

to Alcoa's Warrick Operation, the regulation provider we previously mentioned who is also offering spinning reserve to MISO, their so-called interruptible load (i.e. spinning reserve) only gets 55 deployments annually with an average length of 42 minutes, resulting in an actual dispatch rate of only around 0.44% [50]. Even if the capacity payment rate is fairly low, the spinning reserve providers still find it profitable as they earn money simply by standing by and waiting.

In this section, we consider the participation of the steel plant in a day-ahead electricity market with both energy and spinning reserve, from the perspective of the steel plant scheduling. The scheduling horizon is one day. The daily production activities, i.e. the heats to produce, are known ahead according to the business contracts and the long-term scheduling. The hourly prices of the day, both energy and spinning reserve, are assumed to be known ahead: these prices may be part of a given demand response program contract or they could also be obtained by prediction techniques. Given the production activities and their power profile, the steel plant optimizes the scheduling to minimize its net cost - the cost of electric energy minus the revenue from spinning reserve provision. Furthermore, the impact of actual dispatch of spinning reserve is not considered in this daily scheduling problem, as the dispatch rate is very low and it is recommended to be taken into account in a longer term, e.g. weekly or monthly, optimization problem.

4.4.1 RTN and Mathematical Formulations

The RTN for the considered scheduling problem is the same as Fig. 4.1, except that we have added the resource of spinning reserve (SP), as seen in Fig. 4.15, to help accumulate the plant's energy usage and reserve provision. The set of resources is then $\mathbb{S} = \{\text{EAF, AOD, LF, CC}\} \cup \{\text{EA}_h^{\text{s}}, \text{EA}_h^{\text{d}}, \text{AL}_h^{\text{s}}, \text{AL}_h^{\text{d}}, \text{LC}_h^{\text{s}}, \text{LC}_h^{\text{d}}, \text{H}_h | h \in \mathbb{H}\} \cup \{\text{EN,SP}\}$ with $\mathbb{H}$ as the set of heats to produce.

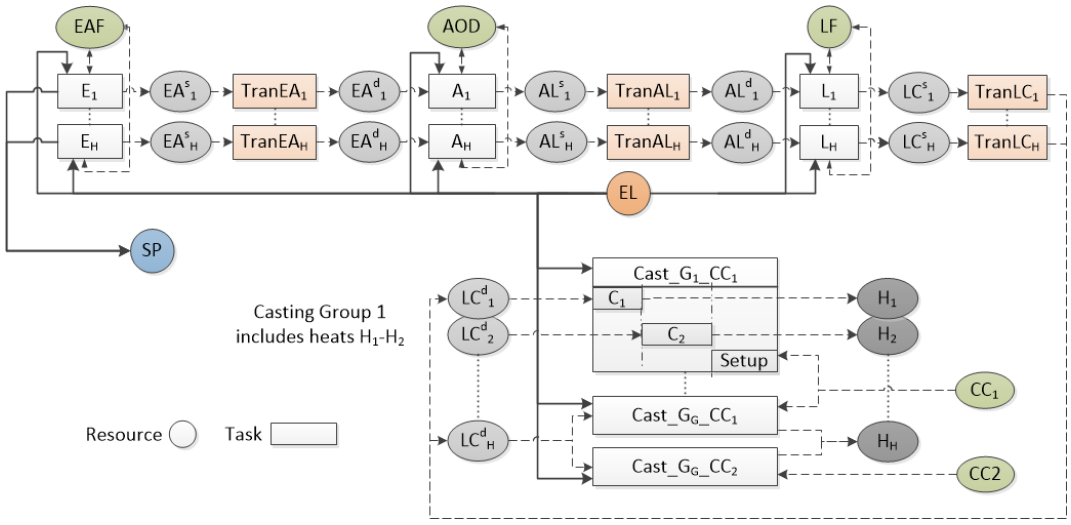


Figure 4.15: Resource task network for a steel plant with spinning reserve provision.

The interaction parameters for a melting task are illustrated in Fig. 4.16. This task interacts with resources EAF, EN, SP and EA_h^s, and its interaction parameter matrix only has four rows with nonzero elements. At the very beginning, the task reduces EAF by one as it uses one furnace. After the completion of the melting process, EAF is increased by one as that furnace is freed up. Also, EA_h^s is increased by one to promote the execution of the following transfer. The melting task consumes electric energy continuously during its entire duration. The sustaining power is assumed to be 48 MW, hence, it can provide 32 MW spinning reserve for each time slot.

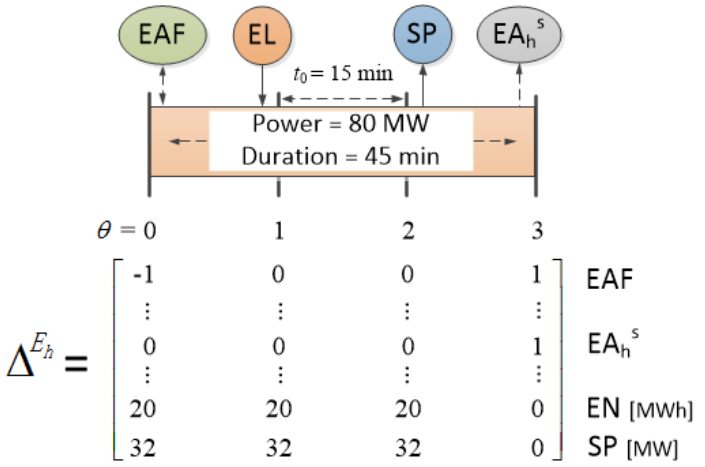


Figure 4.16: Illustration of interaction parameters for a melting task with spinning reserve provision.

The mathematical formulations are the same with Section 4.3.1 except for the addition of spinning reserve resource. We assume that only the EAF can provide spinning reserve, and the

provided spinning reserve should be upper bounded by its availability, as given by

$$y_{\mathrm{SP},t} \leq \sum_{k \in \{E_h | h \in \mathbb{H}\}} \sum_{\theta=0}^{\tau_k} \Delta^k_{\mathrm{SP},\theta} \cdot x_{k,t-\theta} \qquad \forall t \tag{4.24}$$

with $\Delta^k_{\mathrm{SP},\theta}$ denoting the available spinning reserve.

Besides, spinning reserve is traded hourly in most electricity markets. The time slot width in the discrete time formulation is usually smaller than the trading window. Once the reserve provider has committed to the market an hourly quantity, it is obligated to guarantee that amount of reserve for any time slot in that hour. In other words, the time slots belonging to the same hour ($\mathbb{T}_{hr}$) should provide the same amount of spinning reserve, as enforced by

$$y_{\mathrm{SP},t} - y_{\mathrm{SP},t'} = 0 \qquad \forall t, t' \in \mathbb{T}_{hr} \tag{4.25}$$

The objective also needs to consider the revenues from spinning reserve provision. The objective of the scheduling is to minimize the net cost of the steel production, i.e. the electric energy cost minus the spinning reserve revenue. As previously discussed, we do not consider the impact of actual dispatch of spinning reserve, as the actual dispatch is very rare and we assume that there is no actual dispatch in the scheduling horizon. Given the energy and spinning reserve price vectors $\lambda_{\mathrm{EN}}, \lambda_{\mathrm{SP}} \in \mathbb{R}_T$, the overall optimization problem is formulated as

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & Y_{\mathrm{EN}}\lambda_{\mathrm{EN}} - Y_{\mathrm{SP}}\lambda_{\mathrm{SP}} \\ \text{subject to} \quad & (4.17) - (4.22), (4.24), (4.25) \\ & x_{s,t} \in \{0,1\}, \;\; y_{s,t} \in [0, \bar{y}_s], \;\; \forall s, \forall t \end{aligned}$$

in which $\bar{y}_s$ is the upper bound for the available amount of resource s. For example, $\bar{y}_{\mathrm{EAF}}$ equals to the number of EAF furnaces and $\bar{y}_{\mathrm{EN}}$ equals to the summation of energy usage by all the equipment units in one time slot.

4.4.2 Case Study

In this section, we present the study of the daily scheduling for a typical steel plant to demonstrate the effectiveness of the optimal scheduling model.

Problem Parameters

The steel plant layout and parameters are the same as in previous sections. The hourly energy and spinning reserve prices for the case study are taken from MISO, as displayed in Fig 4.17. Note that the spinning reserve (capacity) prices follow the trend of energy prices, but are much lower.

In practice, the hourly prices are obtained either from demand response contracts, e.g. time-of-use pricing programs, or price prediction techniques. For the latter case, the prices are uncertain and are decided by the markets, and we have to rely on price prediction tools. We could simply use the point-wise price prediction or the expected price if its distribution is available.

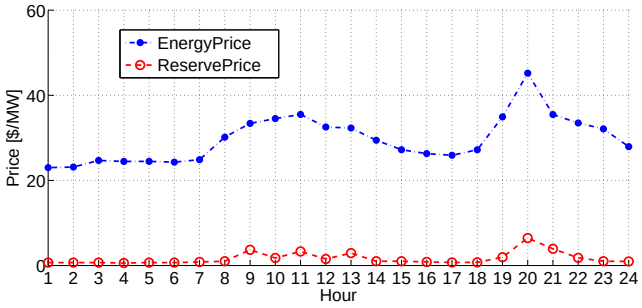


Figure 4.17: MISO hourly prices on 02/06/2014.

Scheduling Results

The optimal scheduling results are given in Table 4.7, in which t_0 is set as 15 minutes. The four rows correspond to different scenarios with respect to how many and which groups are being scheduled and processed in the simulation. The column *Groups* gives the campaign groups to produce, e.g. the first row denotes scheduling group 1, 2, and 3 with the heats as indicated in Table 4.1; *w/o SP* stands for the scheduling model without spinning reserve provision, which corresponds to the above model but without the resource SP and only minimizing electric energy cost; *with SP* is the scheduling model including spinning reserve; the column *Obj* represents the final objective value of the optimization problem; the column *EN* stands for the electric energy cost while the *SP* represents the spinning reserve revenue. All these optimization problems are mixed-integer linear programming problems and we solve them in MATLAB by TOMLAB/CPLEX on a linux 64 bit machine. The relative optimality tolerance is set as 10^{-6}, and all these optimization problems are solved to optimality within three minutes. With spinning reserve participation, the electric energy cost increases a little bit, but the net cost of the steel plant operation is reduced because of the spinning reserve revenues. The decrease for the case studies here are around 1%.

We also set t_0 as 10 minutes to study the scheduling with a finer time grid. The optimal scheduling results are listed in Table 4.8. The time limit for CPLEX is set to 2 hours. For scheduling groups 1-5 under both *w/o SP* and *with SP*, the relative objective gap between the best integer objective (by a feasible solution) and the best bound remaining in the iteration process are 0.02%. Compared with the results in Table 4.7, the final objective values in Table 4.8 are slightly improved because the rounding error due to discrete-time formulation is reduced by using a finer

time grid. However, the computation time in Table 4.8 grows drastically as the number of variables (both integer and continuous) increases by a factor of 1.5.

Table 4.7: Optimization results with $t_0 = 15$min

	w/o SP	with SP		
Groups	Obj(k$)	Obj(k$)	EN Cost(k$)	SP Revenue(k$)
1-3	39.307	39.002	39.321	0.319
1-4	57.824	57.357	57.864	0.507
1-5	69.731	69.157	69.897	0.741
1-6	86.346	85.508	86.474	0.966

Table 4.8: Optimization results with $t_0 = 10$min

	w/o SP		with SP	
Groups	Obj(k$)	CPU Time(s)	EObj(k$)	CPU Time(s)
1-3	39.041	397.7	38.651	739.7
1-4	57.517	637.8	57.009	1094.7
1-5	69.162	7200.0	68.468	7200
1-6	85.228	916.0	84.164	3569.7

The equipment assignment chart for scheduling 24 heats with spinning reserve provision is displayed in Fig. 4.18. The rectangles denote the tasks. Different heats are represented by different colors. From Fig. 4.18 we can observe that the scheduling solution is valid: each heat is processed sequentially by each stage; each campaign group of heats are casted together without any interruption, and there is enough time for caster maintenance between each two campaign groups on the same caster; for any time slot, each equipment is occupied by one single task and there is no conflict in equipment assignment. Figure 4.18 also shows that the RTN model is able to generate detailed and practical schedules that can be easily understood by the steel plant operators. The corresponding spinning reserve provision schedule is displayed in Fig. 4.19. The maximum spinning reserve provided is around 70 MW. The spinning reserve provision cannot always stay at the maximum value due to constraint (4.25). The hourly spinning reserve provision should be less or equal to the available spinning reserve in any time slot of that hour.

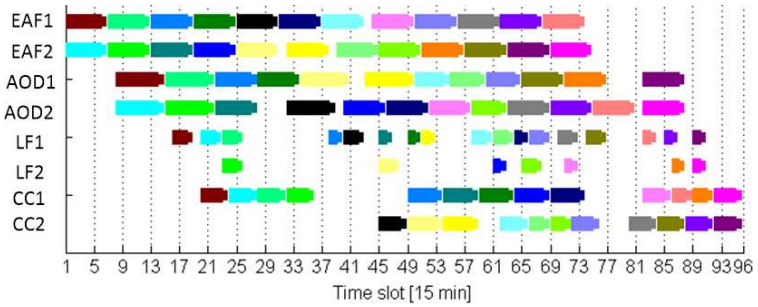


Figure 4.18: Equipment assignment for scheduling 24 heats.

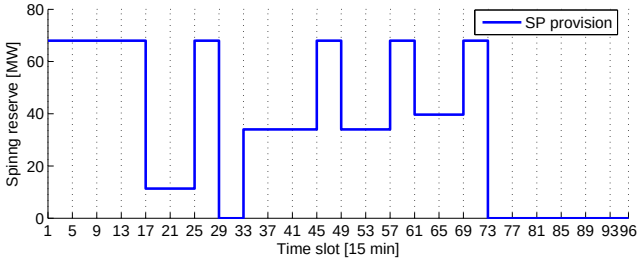


Figure 4.19: Spinning reserve provision from scheduling 24 heats.

As demonstrated by the case studies, the steel plant is able to make use of the electric arc furnaces to offer spinning reserve services to the electricity markets and earn revenues. With the provision of spinning reserve, the steel plant could further lower its operation net cost. The proposed scheduling model can generate detailed and practical production schedules.

Chapter 5

Overcoming Granularity Restriction

A variety of industrial loads can be switched on and off very rapidly, which enables them to change their power consumption rate fast and frequently, e.g. the crushers in the cement crushing industry [48] and the mills in the thermo-mechanical pulp and paper industry [109]. In this chapter, we investigate and develop methods to utilize these industrial loads for the provision of regulation service. Note that the proposed method can also be employed to enable load following.

We are aiming for the cement crushers to provide regulation or load following, which is the most valuable product in the ancillary service markets. The cement crushing process is described in Chapter 1 Section 1.2.3. The regulation signal is in per unit value, i.e. it ranges between -1.0 and 1.0. Suppose the load has committed to provide R MW regulation with a baseline power of B MW, then the regulation command is the RegD signal scaled by R plus B, i.e. the targeted power consumption rate ranges between $B - R$ MW and $B + R$ MW. Hence, the problem to be solved is twofold. First, given a particular B and R, how can the manufacturing plant ensure that it closely follows the given signal? Second, the plant needs to determine the optimal B and R values, such that it can fulfill its regulation commitment at any point in time while not negatively impacting its production. The first problem leads to the proposed MPC Coordination for Hourly Operation of the industrial plant, as presented in Section 5.2. The second problem leads to the proposed Optimal Scheduling for Daily Operation of the industrial plant, in which R and B are determined for every hour of the day, as presented in Section 5.3. The methods and results of this chapter are also presented in [110, 111].

5.1 Energy Storage System Cost Analysis

According to [112], the Lithium-ion (Li-ion) energy storage systems (ESSs) currently cost between \$1,000 and \$2,000/kW (\$350 and \$700/kWh); the average price of the Li-ion battery fell

by 33% between 2010 and 2015, and the cost is expected to fall again, due to the promising sales of electric vehicles and the increasing competition from battery producers such as Tesla, Alevo, Sharp, LG Chem, and Panasonic. The Li-ion technology is expected to dominate the small energy storage (less than 100 MW) market in the following years, as other battery technologies are slightly expensive compared with the Li-ion technology, e.g. the flywheels cost $2,100-2,600/kW, the lead-acid battery systems cost roughly $3000/kW and the NaS battery systems cost $3,500-$6,000/kW [112, 113]. Compressed air energy storage and pumped-hydro storage are the two main players for the bulk storage technologies (larger than 100 MW) [114]. In consideration of its current cost and the future trend, we use $1,000/kW to estimate the cost of a Li-ion ESS and a 3 MW Li-ion ESS roughly costs $3M.

The cement industry has been one of the least profitable businesses, whose net profit margin is only 2.03% [115]. The traditional scheduling strategy, in such a capital-intensive industry, is to produce as much as possible by running all machines at their full capacities, in order to fully utilize the invested equipment and resources. With the falling prices of the cement products, it could be profitable to reduce the production rate, and utilize its capabilities in power response to earn revenues in the electricity markets.

Let us consider a cement plant with two raw mills/crushers, each of which can process 240 tons of raw material per hour with a 4 MW power consumption rate [59]. We will focus on its net cost in the electricity markets per ton of cement produced. The daily market prices, as plotted in Fig.5.1, are taken from a typical day (Jan 30, 2017) in MISO, where the average prices over the day are $29.87/MWh for energy (LMP), $11.36/MW for regulation capacity (REG-MCP), $0.72/MW for regulation mileage (REG-MILEAGE), and $2.45/MW for spinning reserve capacity (SPIN-MCP).

Let us consider three scheduling options as follows. The cement plant is assumed to operate 24 hours a day and 7 days a week.

1. Full Production (8 MW power consumption) The plant produces 240*2*24 = 11520 tons of cement, and pays 29.87*4*2*24 = $5735.04 for energy. Hence, the net cost per ton in electricity markets equals $0.50/ton.

2. Full Production plus Spinning Reserve (8 MW power and 8 MW spinning reserve) Since the spinning reserve dispatch rate is very rare (less than 0.5% according to [50]), we assume there is no actual deployment here. The plant produces 11520 tons of cement, pays $5735.04 in energy, and earns 2.45*4*2*24 = $470.4 from spinning reserve capacity. Hence, the net cost per ton equals $0.46/ton.

3. Half Production plus Regulation (4 MW power and 7 MW regulating reserve) The plant produces 5760 tons of cement, pays 2867.52 in energy, earns 11.36*7*24 = $1908.48 for regulation

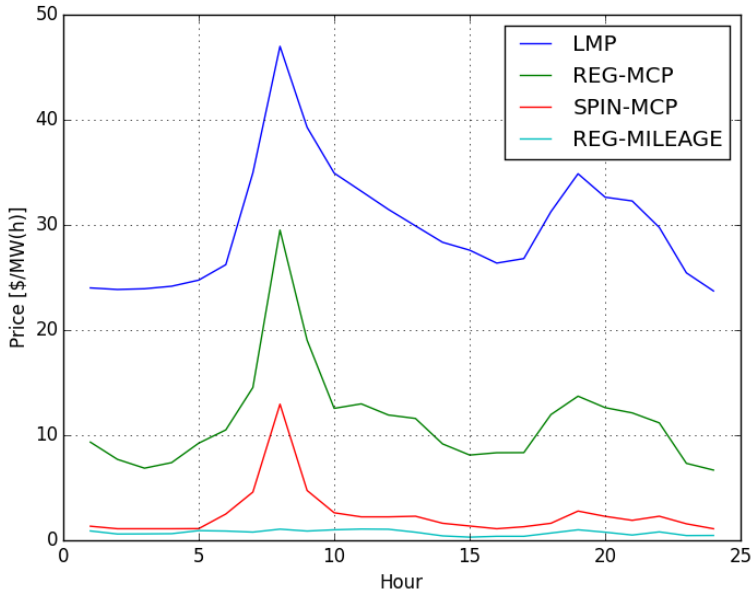


Figure 5.1: Market prices on Jan 30, 2017 in MISO.

capacity provision, and earns 0.72*7*20*24 = $2419.2 for regulation mileage compensation. In calculating the regulation mileage payment, we assume a mild mileage per hour as 20 P.U., which can be referred by Fig. 5.4. Hence, the net cost per ton is roughly $-0.25/ton, i.e. the net profit is $0.25/ton.

By comparing the second (O2) and the third (O3) options, we observe that the saving increase in the electricity markets is $0.71/ton on the typical day, if providing regulation instead of spinning reserve. Hence, the investment in the 3 MW ESS can be covered by producing around 4.23M (3M/0.71) tons of cement, with the assumption that the prices on the typical day can approximate the average prices. In other words, if we choose the (O3) scheduling strategy and let the plant process 240 tons of cement per hour, the ESS can be covered with two years operation.

Of course, the above analysis is not comprehensive enough and it is only meant to illustrate the idea that lowering the production rate and providing regulation could be profitable for the cement plant. Besides, we use the cement plant as an example for the study, yet the proposed methods can be generalized to other loads with similar characteristics, i.e. those loads have machines (e.g. the mills in the paper & pulp industry) that can be switched on/off frequently.

5.2 MPC Coordination for Hourly Operation

In most electricity markets, the market participants bid for their market share for each hour in the following day, and the market share together with the final market price are settled after the bidding process is completed. Hence, in each operating hour, the regulation capacity R and the energy baseline B are pre-determined by the market. The demand response provider is obliged to follow the resulting regulation signal, otherwise, it will be penalized according to market rules. With the settled R and B, the industrial load and the energy storage are coordinated by the MPC method proposed in this section to optimally follow the regulation command with consideration of the machine switching cost, the regulation command violation, and the energy storage level.

As mentioned, the loading units (machines) are switched on/off to follow the regulation command with the support of an on-site energy storage device. For simplicity, we assume that there are M machines which can be switched on and off rapidly, and each machine has a power consumption rate of ρ MW. Note that in practice the power consumption rates for different machines may not be the same, yet the proposed method can be easily extended to consider this deviation from our assumption. It has been demonstrated that stand-alone storage has significant potentials to support the power system operation [116, 117], whereas in our method the storage helps the industrial load to overcome the granularity restriction. We assume that the storage has a maximum energy capacity of E_s MWh and its charging power is bounded by $-P_s$ and P_s MW. To simplify the problem, we further assume that there is no energy loss associated with the charging and discharging processes. Note that the energy loss can be considered easily by extending the formulations.

The real-time coordination framework for hourly operation is illustrated in Fig. 5.2. At each step t, based on historical regulation commands, the predictor outputs the regulation prediction for the next L steps; then the optimal controller optimizes over the number of active machines x_i and the storage charging power y_i for each time step $i = t, \cdots, t + L$ in the MPC horizon, based on the regulation prediction and previous operation records of the machines; after obtaining the optimization results, only the control decision for the current time step t is applied to the industrial load and the energy storage. Then, the horizon is shifted forward by one time step and the optimization is carried out anew.

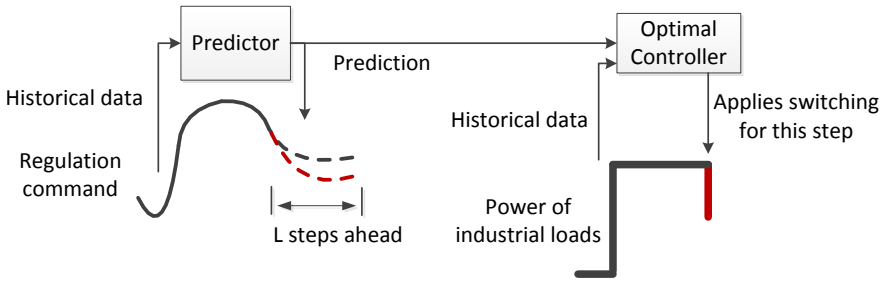


Figure 5.2: MPC coordination framework.

5.2.1 Prediction

Though previous work has demonstrated the advancement of short-term forecasting in power systems [118–120], it is still impossible to accurately forecast the regulation signal over several minutes. However, forecasting its trend with reasonable accuracy over a horizon less than 1 minute is possible, e.g. by using autoregressive-moving-average (ARMA) models. As demonstrated later, this prediction is good enough to coordinate the industrial loads and the energy storage, where the energy storage provides some buffer for prediction errors.

The prediction of the regulation signal is achieved using an ARMA model. We have trained different ARMA models by the Python Time Series Analysis package. For different training data sets [66], the ARMA(2,1) model achieved the best performance, in terms of the Akaike information criterion (AIC) scores. The ARMA (2,1) model is described by:

$$\omega_t = \phi_1 \omega_{t-1} + \phi_2 \omega_{t-2} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

in which ω_t stands for the regulation signal and ϵ_t stands for the white noise. The auto-regressive parameters ϕ_1, ϕ_2 and moving-average parameters θ_1 are trained and obtained. The regulation prediction mean squared errors by ARMA(2,1) for different prediction horizons are plotted in Fig. 5.3 with comparison to the Persistence Prediction and the Mean Prediction approach. The Persistence Prediction uses the latest available observation as prediction and the Mean Prediction uses the average from all available observations as prediction. According to Fig. 5.3, the ARMA(2,1) model results in a good performance up to horizons of around 1 minute.

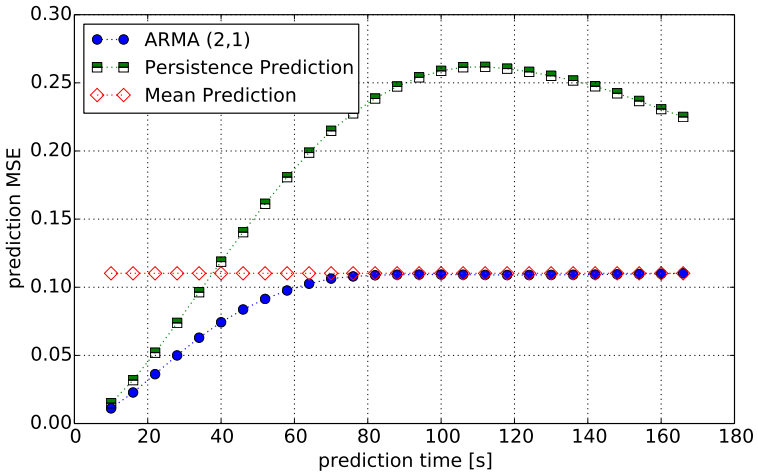


Figure 5.3: Prediction mean squared errors.

5.2.2 Optimal Control

The objective of the optimal control is to provide high quality regulation service at low cost. The decision variables for the optimal control are the number of active machines and the charging power for the storage. The regulation capacity and the regulation baseline, denoted as R and B (MW) respectively, are determined in advance by the schedule optimization approach presented in Section 5.3. Note, as the regulation signal is assumed to be well balanced, i.e. its integral over time is zero, which is the case for RegD signal, the average power consumption of the industrial machines is B MW. This indicates that the throughput from these crushing machines, which is proportional to the energy (MWh) it consumes, is decided by baseline B. The formulations for the optimal control are stated as follows.

Objective

We penalize the regulation violation v_i, the amount of switch actions s_i, and the deviation d of the final storage energy level from the targeted level, as in:

$$\text{minimize} \quad \sum_{i \in \mathbb{L}} (\alpha v_i + \beta s_i) + \gamma d \tag{5.1}$$

in which $\mathbb{L} = \{0, 1, ..., L\}$ is the set of time steps in the current MPC horizon and α, β, γ are the penalty parameters. Different values of the penalty parameters indicate different preferences for the regulation provision. Details of the impact of these parameters are discussed in the case study.

Regulation Violation

Within the MPC horizon at time t, the regulation signal prediction for step i is denoted by $\widehat{\omega}_{t+i}$. According to the regulation prediction, the regulation violation v_{t+i} at the i-th step is defined as:

$$v_{t+i} \geq |B + R\widehat{\omega}_{t+i} - P_m x_{t+i} - y_{t+i}| \quad \forall i \in \mathbb{L} \tag{5.2}$$

in which the first two terms on the right side correspond to the regulation command and the last two terms correspond to the plant power consumption rate. Since we penalize v_{t+i} in the objective function, i.e. α is positive, the above constraint can be formulated as two linear inequality constraints, i.e.

$$v_{t+i} \geq B + R\widehat{\omega}_{t+i} - P_m x_{t+i} - y_{t+i}$$
$$v_{t+i} \geq -B - R\widehat{\omega}_{t+i} + P_m x_{t+i} + y_{t+i}$$

Similar formulations apply to the following two constraints.

Machine Switching

Too much switching of the machines potentially increases degradation and may even damage the machines; that is why we penalize the amount of switch actions in the objective function. The amount of switch actions s_{t+i} at the i-th step is given by:

$$s_i \geq |x_{t+i} - x_{t+i-1}| \quad \forall i \in \mathbb{L} \tag{5.3}$$

in which the right side represents the change in the number of active machines between time steps.

Storage Level Deviation

Another objective is to control the final energy level in the storage by the end of each MPC horizon. Otherwise, if the energy level is near to its full capacity, then there is little room for the storage to contribute to the provision of regulation for the following MPC horizons. This deviation is defined as

$$d \geq |e_{t+L} - \overline{e}| \tag{5.4}$$

in which $\overline{e}$ is the targeted storage level. We usually set $\overline{e}$ equal to 50% of its energy capacity.

Storage Energy Balance

The energy balance for the storage describes the dynamic relationship between stored energy and its charging power, as given by:

$$e_{t+i} - e_{t+i-1} = y_{t+i}\delta \quad \forall i \in \mathbb{L} \tag{5.5}$$

where δ is the length of one time step. In addition, the energy in the storage is constrained by the storage capacity.

Switching Limitation

In practice, the industrial machines cannot be switched on/off without any limitation as the machines could get damaged by too much switching. Hence, we restrict the number of switch actions to be no more than $\bar{s}$ for every successive K steps (typically, $K > L$) for each MPC step t and each time i in the MPC horizon, as given by:

$$\sum_{j=t+i-K}^{t-1} \widetilde{s}_j + \sum_{j=t}^{t+i} s_j \leq \bar{s} \quad \forall i \in \mathbb{L} \tag{5.6}$$

The first term corresponds to the summation of switch actions that already took place before t, and the second term stands for the possible number of switch actions that may take place from t to the i-th step $t+i$. Note that the above constraint applies to each step i within the MPC time horizon, as we require the switching to not violate the bound on the number of switchings for every successive K steps. Consider an example where L is 20 steps and K corresponds to 100 steps, if there are already $\bar{s}$ times of switching between t-90 and t, then for the MPC horizon starting at t, its first 10 steps cannot allow for any switching. Other constraints on switching limitation can be considered in a similar way, e.g. the requirement for the machines to consume a certain minimum amount of energy for a specific number of successive time steps.

Variable Range

The decision variables can take values within the following bounds:

$$x_{t+i} \in \{0, 1, ..., M\} \text{ and } -P_s \leq y_{t+i} \leq P_s \quad \forall i \in \mathbb{L} \tag{5.7}$$

in which x_{t+i} is an integer variable while y_{t+i} is continuous.

To sum up, the MPC recedes forward and at each time step t, it first predicts the upcoming regulation signals, then optimizes (5.1) subject to constraints (5.2)-(5.7), but only applies the optimal decisions at time step t. The resulting optimization problem is a mixed-integer linear programming problem, which can be solved by CPLEX very quickly as the problem size is small.

5.3 Optimal Scheduling for Daily Operation

In Section 5.2, the power baseline B and regulation capacity R are pre-specified for the hourly operation. Before the actual demand response participation, the industrial plant needs to determine

the optimal R and B for each operating hour to maximize its daily revenue. Here, we now consider the daily operation and the goal is to optimally determine the hourly B and R for the load. Since market bidding is not the focus in this thesis, we assume the energy price and regulation price are known. Our scheduling objective is to maximize the cement plant's daily profit which consists of the revenue from industrial production and the revenue from regulation provision minus the cost of energy consumption and the cost of providing regulation. We also need to ensure that the kiln keeps on running at a constant rate continuously and prevent it from turning off, as restarting the kiln is very expensive. As the intermediate product generation rate is proportional to the energy consumption rate of the crushing machines, i.e. the energy baseline B, we need to wisely decide the values of B for each hour in the operating day so that the following processing stages are not interrupted. Note that the proposed optimization model can be extended to consider the market bidding problem, e.g. through stochastic programming with possible price curves as scenarios [22].

Power Baseline

The power baseline B is the base for regulation provision, which equals the load's power consumption rate if no regulation is provided, i.e. without charging/discharging of energy storage nor switching of machines. Hence, B equals the sum of power from a subset of available machines. Consequently, we have a limited choice of values for B. Note that when the load is providing regulation, B is very close to the average power consumption rate, because the regulation signal is assumed to be well balanced (e.g. RegD signal) and therefore its hourly integral is almost zero.

The baseline B is determined by how many crushers are turned on; in case that the machines are not identical with each other, it is also decided by which machines are turned on. For a fixed number of machines, the combinations of their on/off statuses are limited, hence the choices of B are limited.

For each such choice, we term the corresponding machines' statuses as profile $p \in \mathbb{P}$, where $\mathbb{P}$ is the set of all possible profiles, and denote the baseline power as B_p MW. Note that B_p and $\mathbb{P}$ can be obtained as parameters once we know the plant's configuration. We then use binary variable $z_{p,h}$ to denote whether the plant chooses a baseline with profile p, where $z_{p,h} \in \{0, 1\}$ and $h \in \mathbb{H}$ with $\mathbb{H}$ as the set of hours in the scheduling horizon; this is equivalent to using a binary vector of size m to indicate the on/off statuses of m machines, as there is a bijective mapping between the profile p and the binary vector. Since the load chooses only one baseline profile for each hour, we have the following constraint regarding the choice of baseline:

$$\sum_{p \in \mathbb{P}} z_{p,h} = 1 \quad \forall h \in \mathbb{H} \tag{5.8}$$

Regulation Capacity

Unlike the power baseline, the regulation capacity R is continuous; meanwhile, the maximum capacity of R depends on the baseline power B, as B determines the available machines for switching. We use continuous variable $r_{p,h}$ to denote the regulation capacity the load provides given that machine switching profile p is chosen during hour h; for each hour h, only the chosen power baseline's correspondent $r_{p,h}$ is nonzero. Since the energy storage itself is able to provide a regulation of P_s MW, here we consider the case when $r_{p,h}$ is greater than P_s, i.e., we switch machines to provide a larger amount of regulation than the possible amount by merely using the storage. We have the following constraint on the bounds for $r_{p,h}$

$$R_p^{lo} z_{p,h} \leq r_{p,h} \leq R_p^{up} z_{p,h} \quad \forall p \in \mathbb{P} \tag{5.9}$$

where R_p^{lo} and R_p^{up} are bounds associated with baseline profile p. Note that if a baseline profile is not chosen, then its correspondent regulation $r_{p,h}$ is zero because $z_{p,h}$ is zero.

Inventory Stock

Since the production rate of the machines is proportional to their energy consumption rate, the choice of p impacts the intermediate product generation rate. We assume that the following stage consumes the intermediate product at a constant rate τ, and any interruption to the next stage should be avoided as this could be costly. Therefore, there always needs to be enough intermediate product in stock. We use variable q_h to represent the quantity of intermediate product in the inventory after hour h. With the assumption that the production rate is proportional to the power consumption rate, unit conversion can be done and we denote q_h with the unit of MWh for simplicity; similarly, the consumption rate τ of the next stage is also in the unit of MW, and the intermediate production generation rate is B_p MW. This leads to the following constraint regarding the dynamic balance for the inventory stock:

$$q_h + \sum_{p \in \mathbb{P}} B_p z_{p,h} - \tau = q_{h+1} \tag{5.10}$$

The above constraint should always hold in order to keep the next stage running. Besides, the inventory stock is subject to the following bounds:

$$Q^{lo} \leq q_h \leq Q^{up} \tag{5.11}$$

where the parameters Q^{lo} and Q^{up} (also converted to the unit of MWh) correspond to the minimum and maximum amount of intermediate stock allowed in the inventory.

Regulation Cost

The provision of regulation does not come for free as it leads to more switching of the machines and therefore increased equipment degradation. Hence, we need to quantify the hourly regulation cost when optimizing the regulation capacity in the daily scheduling. As discussed in Section 5.2, the regulation provision cost depends on the actual AGC signal trace. However, the relationship between the cost and the AGC signal is nonlinear and complex; besides, the AGC signal itself is uncertain and impossible to predict over a long interval (e.g. more than 5 minutes). To simplify the problem and focus on the daily operation of the industrial plant, we approximate the hourly regulation cost by only considering the cost associated with the switching of machines.

In order to quantify the cost associated with the machine switching, we need to know the switching amount for each hour, but that amount is unknown before the actual hourly operation. However, since the AGC signal is approximately normally distributed, the average hourly switching amount from historical operations can serve as an approximation of future hourly regulation cost in our daily scheduling. As the average hourly switching also depends on the values of B and R, we run numerical simulations with various settings of R and B using historical AGC traces (published by PJM) and record the number of switching for these cases. This gives an approximate average number of switching for each B and R. We then use that average as the approximated switching cost (as a function of B and R) in the daily scheduling problem to determine the optimal B and R for each operating hour. In other words, we assume the approximated switching cost depends linearly on B and R, i.e. given power baseline profile p, the switching cost is assumed to be $C_p^0 + C_p^1 R$; the switching cost increases linearly with regulation capacity R, while the coefficients C_p^0 and C_p^1 are determined by the power baseline. The coefficients can be learned by regression of the historical operation records, which is presented in detail in Section 5.4.3; the linear model assumption is also justified by Fig. 5.10.

To sum up, in consideration of all possible baselines, the hourly switching cost is expressed as follows:

$$C_h = \sum_{p \in \mathbb{P}} (C_p^0 z_{p,h} + C_p^1 r_{p,h}) \tag{5.12}$$

Note that $r_{p,h}$ is zero when $z_{p,h}$ is zero according to Eq. (5.9). The coefficients C_p^0 and C_p^1 can take into account the monetary cost of the switch actions, hence the hourly switching cost C_h is in the unit of \$.

Objective

Suppose the hourly energy price, regulation price, and industrial product market price are known, then the regulation revenue is proportional to R, and both the energy cost and industrial

product revenue are proportional to B. We use $\lambda_{R,h}$ to denote the hourly regulation price, and use $\lambda_{E,h}$ to denote the hourly net energy profit price which equals the unit revenue in industrial production minus the unit cost in energy consumption. Since we assume the kiln operates at a constant speed and hence the final product yield does not change, we only maximize the net revenue from the crushing stage. Then our final optimization problem is:

$$\begin{aligned} \underset{z_{p,h},\ r_{p,h}}{\text{maximize}} \quad & \sum_{h\in\mathbb{H}}(\lambda_{E,h}\sum_{p\in\mathbb{P}}B_p z_{p,h} + \lambda_{R,h} r_{p,h} - C_h) \\ \text{subject to} \quad & (5.8) - (5.12) \end{aligned}$$

which is a mixed-integer linear programming problem. And again we use CPLEX to solve the problem.

5.4 Case Study

5.4.1 Industrial Plant Parameters

For the case study, we consider a cement plant with $M = 4$ crushing machines. These machines can be switched on and off rapidly. The power consumption rate of each machine is either $\rho = 2$ MW (when it is on) or zero (when it is off). The plant has an on-site energy storage device with $E_s = 1$ MWh energy capacity and its maximum charging/discharging power is $P_s = 3$ MW. The crushing machines constitute the first stage of the cement production process, followed by a cement kiln burner which heats the intermediate products to a certain temperature. The kiln burner cannot be interrupted and consumes the intermediate product at a constant rate. The constant consumption rate is equivalent to $\tau = 4$ MW. The intermediate product between the crushing and burning is stored in an inventory, which has a maximum capacity that is equivalent to $Q^{up} = 20$ MWh and its initial stock is 10 MWh.

5.4.2 Simulations of MPC Coordination for Hourly Operation

The simulations of hourly operations are studied to demonstrate the benefits of the proposed MPC coordination approach. As R and B are given values in the hourly operation, here we consider the cement plant providing $R = 5$ MW regulation at a baseline of $B = 2$ MW; as seen later in Section 5.4.4, this setting of R and B corresponds to the scheduling result in Fig. 5.11, e.g. hour 8 and hour 20. The regulation command ranges between -3 MW and 7 MW. Note that this range of the regulation command is 10 MW, which is much higher than that of the energy storage (6 MW). The 20 minutes regulation signal for the simulations is plotted in Fig 5.4, together with the ARMA(2,1) prediction at a few distinct time instances. The length of the time step is $\delta = 2$

seconds, and the prediction horizon is $L = 15$ steps. The penalties α, β, γ in Section 5.2 are set to 10, the targeted final energy is $\bar{e} = 0.5$ MWh, and we require the maximum number of switch actions to be $\bar{s} = 10$ times for every successive 5 minutes, i.e. $K = 150$ steps.

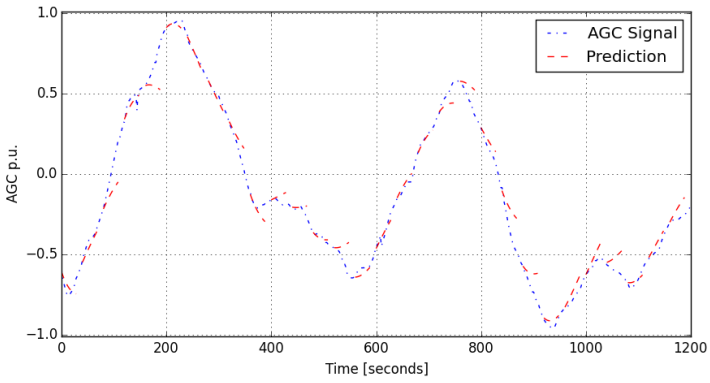


Figure 5.4: Regulation signal (AGC) over 20 minutes and its prediction.

The simulations of hourly operation, i.e. the real-time following of the AGC signal, the switch actions, and the energy storage operation, are plotted in Fig. 5.5. The dashed lines in the middle plot are the bounds for the charging/discharging power of the storage. According to the simulation, the integral of regulation violation over the hour is 0 MWh, i.e. there's no violation at all, the total number of switch actions is 15, and the storage energy level at the end of the hour is 0.64 MWh, i.e the energy deviation is 0.14 MWh. These results demonstrate that the coordination method proposed for hourly operation is able to utilize the advantages of both the industrial loads and the energy storage, and provides high-quality regulation service to support the power system operation.

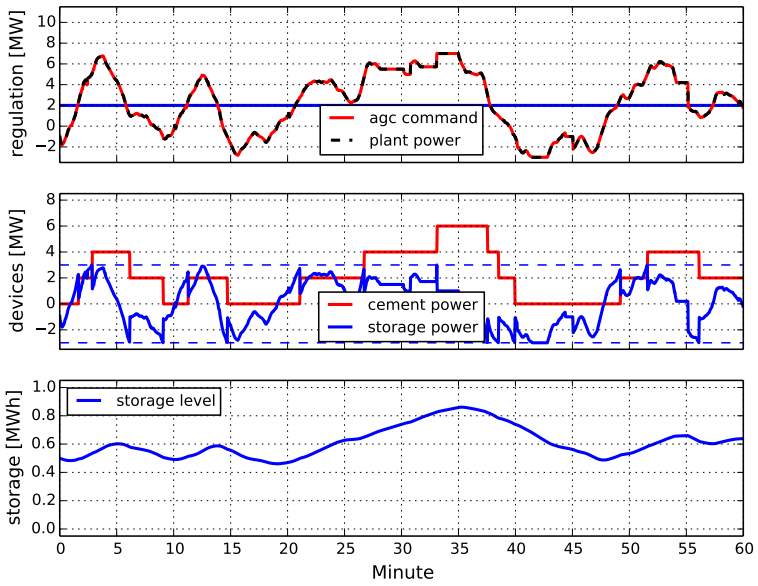


Figure 5.5: Hourly operation simulations with $R = 5$ MW, $B = 2$ MW.

We also study the case of providing an even larger amount of regulation. We consider the cement providing $R = 7$ MW regulation at a baseline of $B = 4$ MW, which corresponds to the scheduling result in Fig. 5.12, e.g. from hour 4 to hour 22. The simulation results are displayed in Fig. 5.6 and all the configurations are exactly the same as the simulation in Fig. 5.5. Over the hour, the integral of regulation violation is 0.001 MWh, the total number of switch actions is 27, and the storage energy level at the end of the hour is 0.60 MWh, i.e the energy deviation is 0.10 MWh. Compared with Fig. 5.5, the machines are switched more frequently and there is some violation of regulation following, which is the cost of increasing the regulation capacity.

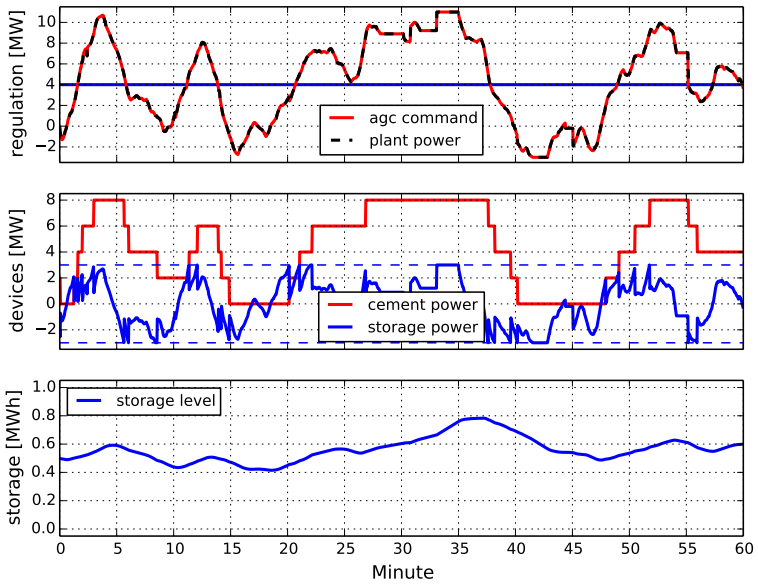


Figure 5.6: Hourly operation simulations with $R = 7$ MW, $B = 4$ MW.

The sensitivity with respect to changes in parameter settings is investigated by simulations with different penalty values. If we impose a stronger switch limitation constraint, e.g. requiring the maximum number of switches to be 3 for every successive 5 minutes, then the switching frequency will decrease, as demonstrated by the simulation results in Fig. 5.7. The total number of switches decreases to 21 times, but the regulation violation increases to 0.25 MWh. If we also increase the penalty on switch actions β, the total number of switch actions is expected to further decrease. For example, increasing β to 100 while keeping all other parameters the same as Fig. 5.7 yields the simulation results in Fig. 5.8, in which the total number of switches further decreases to 19 times, however, there is even more regulation violation, i.e. 0.27 MWh.

In practice, we suggest that the plant operators choose their own penalties according to their preferences. For example, in an electricity market where the regulation quality is highly valued, a higher regulation violation penalty α is recommended; meanwhile, if switching the machines is very expensive, then the operator should use a large switch action penalty β.

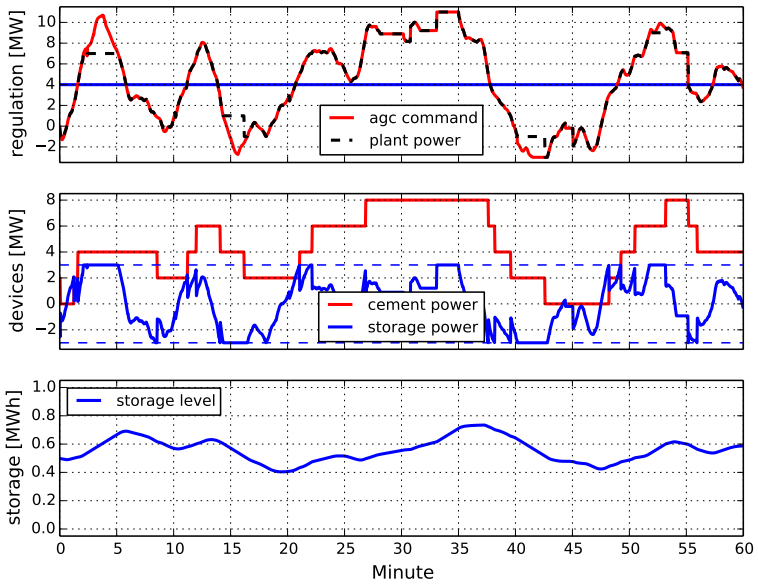


Figure 5.7: Stronger switching limitation with $R = 7$ MW, $B = 4$ MW.

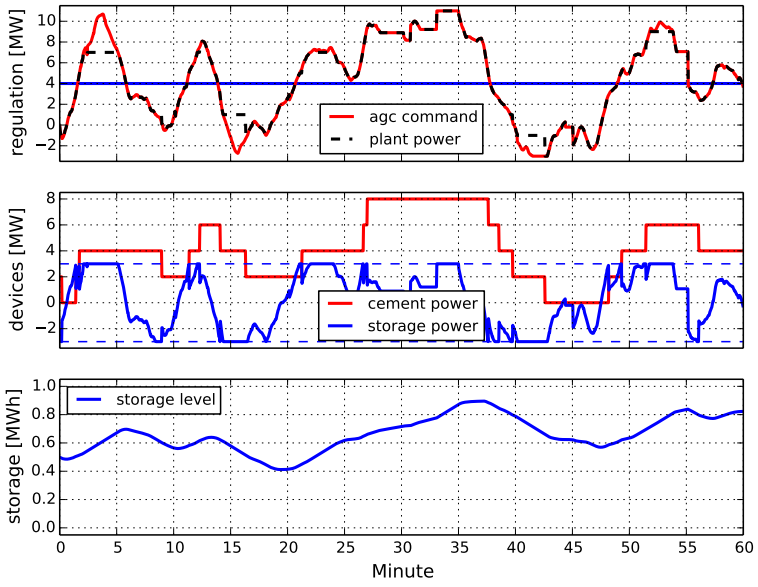
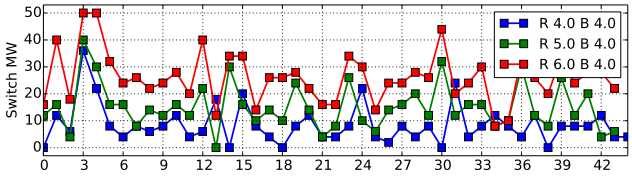


Figure 5.8: Increased penalty on switching with $R = 7$ MW, $B = 4$ MW.

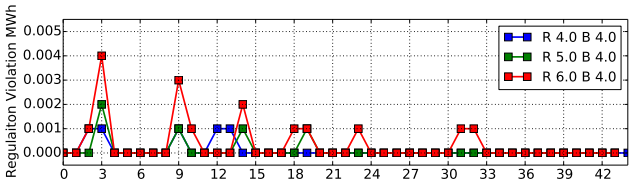
5.4.3 Quantifying Hourly Regulation Cost

In order to optimize the scheduling for the daily operation as in Section 5.3, we need to quantify the hourly cost of regulation provision. As discussed in Section 5.3, we approximate the hourly regulation cost by using the average hourly switching quantities from the records of historical

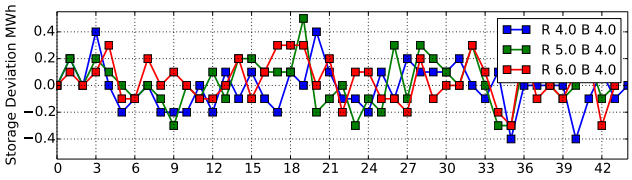
operation.



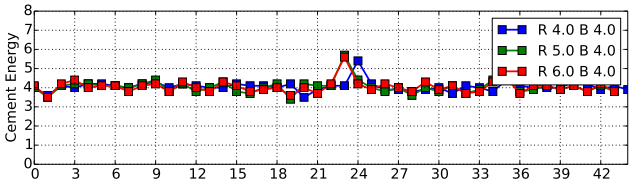
(a)



(b)



(c)



(d)

Figure 5.9: Hourly simulation results over 48 hours.

To obtain the historical regulation cost, we simulate the MPC coordination for each hour over three months by using the historical AGC signal published by PJM. All the three penalties α, β, γ are still chosen as 10. Simulations with different choices of baseline power B and varying regulation capacity R are studied. For the considered plant, we have three baseline powers to choose from, i.e. 2, 4, or 6 MW, leaving at least one machine for switching. The regulation cost given different pairs of (R, B) are obtained for each hour in the historical data. We present the trace of each regulation cost component over two days in Fig 5.9(a), 5.9(b), and 5.9(c); we only present the plots with B = 4 MW here, while the observations are similar when B is 2 or 6 MW. From these plots, we observe that: (1) a larger regulation capacity generally leads to a larger amount of switch actions; (2) the regulation violation is zero for most of the hours, and when it is not zero, the violation

increases with the regulation capacity; (3) the energy storage deviation seems normally distributed and does not necessarily increase with the regulation capacity. The hourly energy consumption by the machines over these two days is plotted in Fig. 5.9(d), which demonstrates that the hourly energy consumption rate is very close to the baseline power.

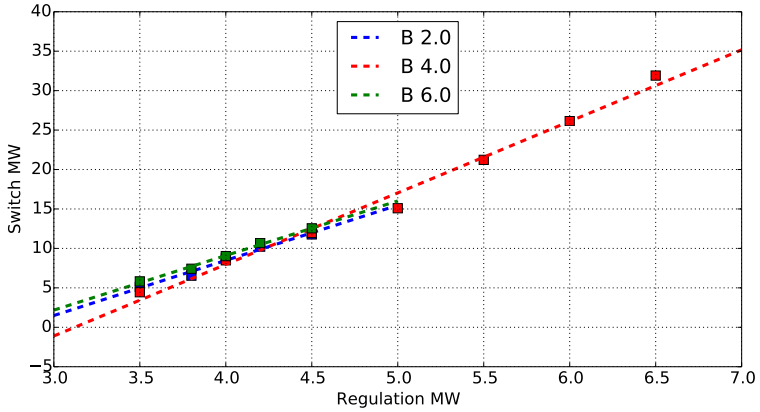


Figure 5.10: Average hourly switch MW and its linear fitting.

In the daily scheduling, we only consider the switching cost. With the simulated hourly switching cost over these three months, we can obtain the average amount of switch actions $\bar{C}_{B,R}$ for any given pair of R and B. This average switch action is displayed in Fig 5.10, where all three possible baseline power are considered. For each baseline power B, we apply linear regression to fit the output C (response/dependent variable) to the input R (explanatory/independent variable). The fitted relationships are also plotted as dashed lines in Fig 5.10. These fitted linear relationships which map R to C under a chosen B provide the regulation cost coefficients C_b^0 and C_b^1 for the daily scheduling.

5.4.4 Simulations of Optimal Scheduling for Daily Operation

Here we apply the optimal scheduling proposed in Section 5.3 to the industrial plant. The hourly electricity energy price and regulation price are taken from historical records from MISO. The profit from cement production is assumed to be \$30/MWh, i.e., for every 1 MWh energy consumption, the machines generate products worth of \$30. The coefficients of hourly regulation cost, C_b^0 and C_b^1, are taken from the regression result in Fig. 5.10, with the assumption that the monetary cost of switching is \$0.5/MW. All these scheduling optimizations can be solved by CPLEX within minutes as the problem size is small. In practice, its computation does not need to be very fast as the problem is intended for day-ahead scheduling which only needs to be solved a few times every day. The daily scheduling is presented in Fig. 5.11. From the result we see that when the energy price

is lower, the baseline power is higher, i.e. the cement plant takes advantage of the lower energy price and consumes more energy by speeding up production; while when the energy price is higher, the baseline power is lower and the cement plant consumes less energy. We also observe that the regulation capacity increases when the regulation price is higher, e.g. around hour 8 and hour 18. If we manually increase the regulation price, e.g. artificially multiply the regulation price by 8, the industrial plant will concentrate on the regulation provision, as in Fig. 5.12. During most of the hours, the baseline power is 4 MW, which guarantees the largest regulation availability.

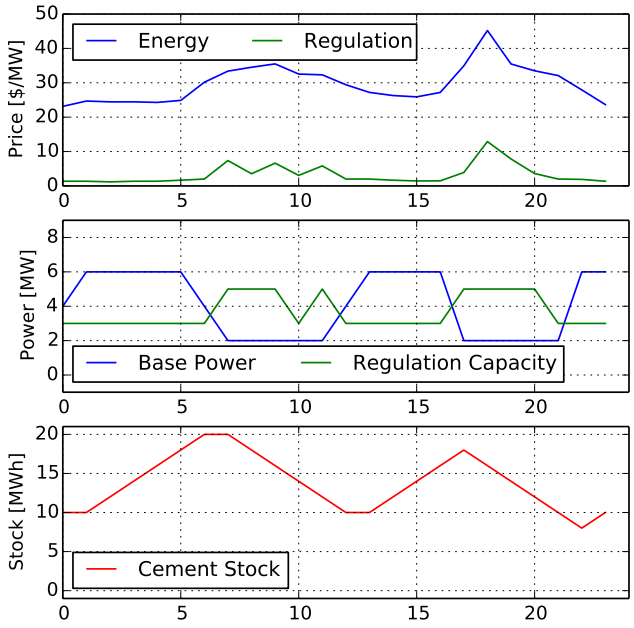


Figure 5.11: Daily scheduling result.

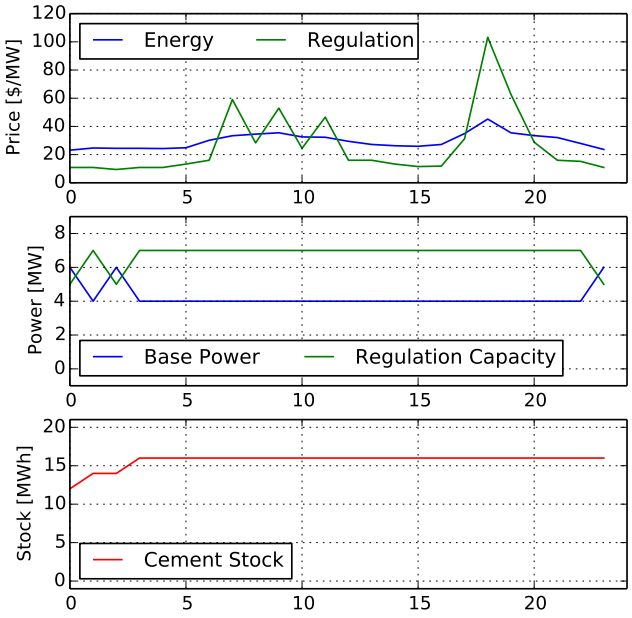


Figure 5.12: Daily scheduling with higher regulation price.

Chapter 6

Conclusion and Future Work

In this closing chapter, the work of this dissertation is summarized and concluded, and the possible directions for future work are discussed.

6.1 Conclusion

In this thesis, we have presented methods to enable industrial loads to support power system operation through demand response. We have identified three challenges in uncertainty, complexity, and granularity for the industrial loads to provide demand response, and have proposed approaches to address these challenges through studies of three representative industries including aluminum smelting, steel manufacturing, and cement crushing.

To account for uncertainty, we build stochastic programming models to optimize the electricity market participation for an aluminum smelter. First, we consider the regulation participation from industrial load and propose an optimal regulation capacity provision model for a given regulation price for an aluminum smelter. The stochastic programming model uses simplified AGC signal scenarios to consider the influence of regulation participation and takes into account the impact of different price settings on the decision of regulation capacity provision. From the case studies, we learn that increasing the compensation for regulation capacity encourages higher capacity provision, while too expensive control cost or too high penalties on non-performance lead to low regulation participation. The simulation results suggest more electricity consumption (and therefore more aluminium production) and lower regulation participation when the profit price is higher. With the demonstrated simulation results, we conclude that the proposed method is able to help the aluminum smelter to optimally decide its regulation capacity.

Next, we study the demand response for aluminum smelters that participate in both energy and spinning reserve day-ahead markets. We propose a stochastic programming model that generates

the day-ahead bidding strategy for the smelters. The model is a mixed-integer linear programming problem which can be solved by commercial solvers very quickly. The inputs to the model are the smelting plant parameters and the price scenarios that represent future price trends. The output of the model are the energy bidding curves and the optimal spinning reserve provision as well as the power consumption levels of the potlines. As discussed in the case studies, the bidding strategy generated by the proposed method is reasonable, and the model can take advantage of the future price trends and arrange the smelting activities to make profits from both electricity markets participation and aluminum production. We conclude that the proposed bidding method can potentially serve as a valuable tool for aluminum smelters to bid in the electricity markets.

Finally, we study how to optimize the participation of industrial demand response resources in day-ahead energy and regulation service markets. A stochastic programming model is proposed which generates the bidding strategy for industrial loads such as the aluminum smelters. We utilize non-parametric Multiple Quantile Graphical Model to represent the distributions of hourly prices and use the Gibbs sampling approach to sample the price curves, which serve as the price scenarios for the stochastic programming model. We also study the performance of the bidding strategy by comparing its revenue from the market clearing with the optimized revenue by using the predicted prices. The case studies demonstrate that the bidding curves generated by the stochastic programming model with sampled price scenarios perform better than the optimization results by merely using the predicted prices, and we conclude that the proposed bidding method does well in addressing the uncertainties in the electricity markets.

To handle the complexity of process, we investigate the scheduling of steel manufacturing with demand response provision from both the mathematical formulation aspect and the solving algorithm aspect. First, we propose the Multiple Melting Modes and Arbitrary Flex Melting to enable the steel plants to participate more actively in the electricity market by exploiting the EAFs' capability to adjust their power consumption rate through controlling the OLTCs. Extended from the Basic RTN which optimizes the schedule merely through arranging the time and sequence of the tasks, the Multiple Melting Modes model enables controlling the transformers at the beginning of each EAF task, while the Arbitrary Flex Melting model allows the control of the transformers at every time slot within the EAF task. In consideration of both optimization benefit and computation cost from the case studies, we conclude that the proposed Multiple Melting Modes model is a good tool for steel plant scheduling, as it provides a good trade-off between enabling the exploitation of the flexibilities given by the OLTCs and computational complexity.

Next, we propose approaches to improve the related computational issues for steel plant scheduling. We try to improve the computations in the following two ways: from the modeling aspect, we add additional constraints as cuts to reduce the search space for the MIP problem, and from the al-

gorithmic aspect, we design a tailored branch and bound algorithm that utilizes the knowledge from steel manufacturing. As demonstrated through numerical studies, adding additional constraints as cuts to the MIP can reduce the computation time significantly, and the proposed tailored algorithm can reduce the iteration number of the solving process by an order but at the cost of slight optimality deterioration. We conclude that the proposed methods have great potential in improving the computations, and may play an important role towards developing practical scheduling tools for the steel industry and its demand response.

Thirdly, since the steel plant is also able to make use of the electric arc furnaces to offer spinning reserve services to the electricity markets and earn revenues, we propose scheduling models to enable the spinning reserve provision. The steel plant is assumed to be a participant in the day-ahead electricity markets, both energy and spinning reserve markets, and we want to optimize its scheduling of production activities to maximize its revenues from the electricity markets. With the provision of spinning reserve, the steel plant could further lower its operation net cost. The numerical analysis of a typical steel plant demonstrates remarkable savings and could encourage the steel plants to participate more actively in the smart grid as demand response resource.

To overcome the granularity restriction, we propose the coordination framework which allows industrial loads to have more options in supporting the power system operation through demand response. Industrial loads, such as the cement crushers that can frequently switch on/off, are able to provide regulation or load following ancillary services, with the help of an on-site energy storage. The power change from the industrial loads serve as the main contributor in the service, while the charging power from the storage is responsible for eliminating the mismatches. The case studies demonstrate the performance of the coordination method, and we conclude that the proposed coordination framework provides more options for demand response participation. The proposed framework also has other potential applications outside of demand response, e.g. the coordination among fast and slow generators and energy storage.

Next, we also study the daily scheduling of the plant and propose a scheduling approach which takes into account the revenues from market participation, the cost of regulation provision, as well as the coupling of the crushing process with other processing stages within the industrial plant. From the case studies, we conclude that the proposed scheduling approach is potentially a great tool for cement plant operators to optimally arrange their production activities with demand response provision.

6.2 Future Work

During the study of this dissertation, we have found several research questions very interesting and important, but cannot deeply explore these directions due to time limitation and the fact that this thesis is focused on demand response of industrial loads. One of these directions is on price forecast in the electricity market. As seen in this thesis, our proposed methods heavily rely on the market price signal. Even though there has already been a lot of research on electricity market price forecast, the performance of the forecast still needs improving. With more and more market data and history records accumulated, it would be interesting to explore this big data set to improve the forecast performance by utilizing methods such as deep learning.

Another interesting direction is to investigate parallel computing. Stochastic programming is adopted in this thesis to develop bidding strategy which uses a set of scenarios to account for the price uncertainty, and generally speaking, it is always better to use more scenarios. However, using more scenarios leads to a larger problem to solve. Parallel computing is very promising to handle a large number of scenarios. Besides, the steel scheduling problem is a very complicated integer programming problem, and the number of branches increases exponentially in the branch and bound solving process. Parallel computing could potentially fix that problem and make it possible to solve a large scheduling problem of very fine time grids within reasonable time. Besides, machine learning techniques such as policy learning have potentials to speed up the steel plant scheduling.

For all the simulations in this thesis, we choose the case parameters to closely reflect the industrial machines and devices in the real world. The intention is to make the proposed methods as practical as possible. However, there is still much work to be done before applying these methods in the field. In the aluminum smelting industry, the complex and nonlinear relationship between the temperature in the smelting pot and its power consumption needs to be considered. In the steel plant scheduling, the power consumption rate of the EAF is not constant; it gradually ramps up from zero and ramps down to zero, and the ramp rate is not infinite, and we need to consider this practical melting curve. In the daily scheduling of the cement plant, we only consider the coupling between the crushers and the kiln, and there is still much more to be considered in scheduling the entire plant. Besides, it is also crucial to study the stability of the MPC method proposed in this thesis. Even though the simulation results look great, the proposed MPC method is a hybrid system that involves discrete decisions and we have not studied its stability. The stability study need to be considered before putting the method into practice.

Bibliography

[1] M. G. Vayá and G. Andersson, "Self scheduling of plug-in electric vehicle aggregator to provide balancing services for wind power," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 886–899, 2016.

[2] H. Marzooghi, G. Verbič, and D. J. Hill, "Aggregated demand response modelling for future grid scenarios," *Sustainable Energy, Grids and Networks*, vol. 5, pp. 94–104, 2016.

[3] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, "Facilitating renewable integration by demand response," in *Integrating Renewables in Electricity Markets*. Springer US, 2014, pp. 289–329.

[4] M. D. Ilić, L. Xie, and J.-Y. Joo, "Efficient coordination of wind power and price-responsive demand - Part I: Theoretical foundations," *IEEE Trans. on Power Systems*, vol. 26, no. 4, pp. 1875–1884, 2011.

[5] A. Negash, T. Haring, and D. Kirschen, "Allocating the cost of demand response compensation in wholesale energy markets," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1528–1535, May 2015.

[6] J. Wang, C. Liu, D. Ton, Y. Zhou, J. Kim, and A. Vyas, "Impact of plug-in hybrid electric vehicles on power systems with demand response and wind power," *Energy Policy*, vol. 39, no. 7, pp. 4016–4021, 2011.

[7] W. Wei, F. Liu, and S. Mei, "Charging strategies of EV aggregator under renewable generation and congestion: A normalized nash equilibrium approach," *To appear in IEEE Transactions on Smart Grid*, 2016.

[8] Y. Yao, W. Gao, J. Momoh, and E. Muljadi, "Economic dispatch for microgrid containing electric vehicles via probabilistic modelling," in *IEEE North American Power Symposium*, 2015.

[9] Z. Chen, L. Wu, and Y. Fu, "Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1822–1831, Dec 2012.

[10] K. Margellos and S. Oren, "Capacity controlled demand side management: A stochastic pricing analysis," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 706–717, 2016.

[11] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 91–99, 2016.

[12] I. Beil, I. Hiskens, and S. Backhaus, "Frequency regulation from commercial building HVAC demand response," *Proceedings of the IEEE*, vol. 104, no. 4, pp. 745–757, 2016.

[13] E. Vrettos and G. Andersson, "Scheduling and provision of secondary frequency reserves by aggregations of commercial buildings," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 850–864, 2016.

[14] F. Pallonetto, S. Oxizidis, F. Milano, and D. Finn, "The effect of time-of-use tariffs on the demand response flexibility of an all-electric smart-grid-ready dwelling," *Energy and Buildings*, vol. 128, pp. 56–67, 2016.

[15] C. Wu, W. Tang, K. Poolla, and R. Rajagopal, "Predictability, constancy and contingency in electric load profiles," in *IEEE Smart Grid Communications*, 2016.

[16] E. C. Kara, Z. Kolter, M. Berges, B. Krogh, G. Hug, and T. Yuksel, "A moving horizon state estimator in the control of thermostatically controlled loads for demand response," in *IEEE Smart Grid Communications*, 2013, pp. 253–258.

[17] R. Yin, P. Xu, M. A. Piette, and S. Kiliccote, "Study on Auto-DR and pre-cooling of commercial buildings with thermal mass in California," *Energy and Buildings*, vol. 42, no. 7, pp. 967–975, 2010.

[18] Z. Zhou, F. Zhao, and J. Wang, "Agent-based electricity market simulation with demand response from commercial buildings," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 580–588, 2011.

[19] Z. Liu, I. Liu, S. Low, and A. Wierman, "Pricing data center demand response," in *ACM International Conference on Measurement and Modeling of Computer Systems*, 2014, pp. 111–123.

[20] A. Wierman, Z. Liu, I. Liu, and H. Mohsenian-Rad, "Opportunities and challenges for data center demand response," in *IEEE Green Computing Conference*, 2014.

[21] X. Zhang and G. Hug, "Optimal regulation provision by aluminum smelters," in *IEEE Power and Energy Society General Meeting*, 2014.

[22] ——, "Bidding strategy in energy and spinning reserve markets for aluminum smelters' demand response," in *IEEE Innovative Smart Grid Technologies Conference*, 2015.

[23] H. Hadera, I. Harjunkoski, G. Sand, I. E. Grossmann, and S. Engell, "Optimization of steel production scheduling with complex time-sensitive electricity cost," *Computers & Chemical Engineering*, vol. 76, pp. 117–136, 2015.

[24] X. Zhang, G. Hug, Z. Kolter, and I. Harjunkoski, "Computational approaches for efficient scheduling of steel plants as demand response resource," in *IEEE Power Systems Computation Conference*, 2016.

[25] X. Zhang, G. Hug, and I. Harjunkoski, "Cost-effective scheduling of steel plants with flexible EAFs," *IEEE Trans. on Smart Grid*, 2016.

[26] L. Merkert, I. Harjunkoski, A. Isaksson, S. Säynevirta, A. Saarela, and G. Sand, "Scheduling and energy–industrial challenges and opportunities," *Computers & Chemical Engineering*, vol. 72, pp. 183–198, 2015.

[27] T. Samad and S. Kiliccote, "Smart grid technologies and applications for the industrial sector," *Computers & Chemical Engineering*, vol. 47, pp. 76 – 84, 2012.

[28] D. Fabozzi, N. Thornhill, and B. Pal, "Frequency restoration reserve control scheme with participation of industrial loads," in *IEEE PowerTech*, 2013.

[29] S. Elmquist, "Alcoa sees aluminum surplus with lower demand growth," April 2015, [Online; posted 13-September-2012]. [Online]. Available: http://www.bloomberg.com/news/articles/2015-04-08/alcoa-profit-beats-estimates-as-aluminum-demand-climbs

[30] I. Harjunkoski and I. E. Grossmann, "A decomposition approach for the scheduling of a steel plant production," *Computers & Chemical Engineering*, vol. 25, no. 1112, pp. 1647 – 1660, 2001.

[31] *PJM Demand Response Programs*, PJM, Sep 2015. [Online]. Available: http://www.pjm.com/markets-and-operations/demand-response.aspx

[32] *ERCOT Demand Response Programs*, ERCOT, Sep 2015. [Online]. Available: http://www.ercot.com/services/programs/load

[33] *Demand Response Programs*, Southern California Edison, 2017. [Online]. Available: https://www.sce.com/wps/portal/home/business/savings-incentives/demand-response/

[34] *Iron and Steel*, United States Environmental Protection Agency, 2016. [Online]. Available: https://archive.epa.gov/sectors/web/html/steel.html

[35] M. D. Galus, S. Koch, and G. Andersson, "Provision of load frequency control by PHEVs, controllable loads and a cogeneration unit," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4568–4582, 2011.

[36] N. Li, L. Chen, and S. Low, "Optimal demand response based on utility maximization in power networks," in *Power and Energy Society General Meeting*, 2011.

[37] M. Roozbehani, M. Dahleh, and S. Mitter, "Volatility of power grids under real-time pricing," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1926–1940, 2012.

[38] P. Finn and C. Fitzpatrick, "Demand side management of industrial electricity consumption: promoting the use of renewable energy through real-time pricing," *Applied Energy*, vol. 113, pp. 11–21, 2014.

[39] Q. Zhang and I. E. Grossmann, "Planning and scheduling for industrial demand side management: advances and challenges," in *Alternative Energy Sources and Technologies*. Springer International Publishing, 2016, pp. 383–414.

[40] ——, "Enterprise-wide optimization for industrial demand side management: Fundamentals, advances, and perspectives," *Chemical Engineering Research and Design*, vol. 116, pp. 114–131, 2016.

[41] P. M. Castro, I. Harjunkoski, and I. E. Grossmann, "Optimal scheduling of continuous plants with energy constraints," *Computers & Chemical Engineering*, vol. 35, no. 2, pp. 372 – 387, 2011.

[42] S. Mitra, I. E. Grossmann, J. M. Pinto, and N. Arora, "Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes," *Computers & Chemical Engineering*, vol. 38, pp. 171 – 184, 2012.

[43] K. Nolde and M. Morari, "Electrical load tracking scheduling of a steel plant," *Computers & Chemical Engineering*, vol. 34, no. 11, pp. 1899 – 1903, 2010.

[44] A. Haït and C. Artigues, "On electrical load tracking scheduling for a steel plant," *Computers & Chemical Engineering*, vol. 35, no. 12, pp. 3044–3047, 2011.

[45] P. M. Castro, L. Sun, and I. Harjunkoski, "Resource–task network formulations for industrial demand side management of a steel plant," *Industrial & Engineering Chemistry Research*, vol. 52, no. 36, pp. 13 046–13 058, 2013.

[46] P. M. Castro, I. Harjunkoski, and I. E. Grossmann, "New continuous-time scheduling formulation for continuous plants under variable electricity cost," *Industrial & Engineering Chemistry Research*, vol. 48, no. 14, pp. 6701–6714, 2009.

[47] M. Paulus and F. Borggrefe, "The potential of demand-side management in energy-intensive industries for electricity markets in germany," *Applied Energy*, vol. 88, no. 2, pp. 432 – 441, 2011.

[48] R. Vujanic, S. Mariéthoz, P. Goulart, and M. Morari, "Robust integer optimization and scheduling problems for large electricity consumers," in *IEEE American Control Conference*, 2012, pp. 3108–3113.

[49] A. Molina-Garcia, M. Kessler, M. C. Bueso, J. A. Fuentes, E. Gomez-Lazaro, and F. Faura, "Modeling aluminum smelter plants using sliced inverse regression with a view towards load flexibility," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 282–293, 2011.

[50] D. Todd, M. Caufield, B. Helms, A. P. Generating, I. M. Starke, B. Kirby, and J. Kueck, "Providing reliability services through demand response: A preliminary evaluation of the demand response capabilities of Alcoa Inc," *ORNL/TM*, vol. 233, 2008.

[51] *Business Practices Manual: Energy and Operating Reserve Markets*, MISO, Feb 2013. [Online]. Available: https://www.misoenergy.org/Library/BusinessPracticesManuals/

[52] T. Andresen, "Storing power in molten aluminum lakes," January 2015. [Online]. Available: http://www.bloomberg.com/bw/articles/2014-11-26/germanys-trimet-aluminium-turns-smelting-tanks-into-batteries

[53] *Demand Response*, Midcontinent Independent System Operator, 2016. [Online]. Available: https://www.misoenergy.org/WhatWeDo/StrategicInitiatives/Pages/DemandResponse.aspx

[54] S. Ashok, "Peak-load management in steel plants," *Applied Energy*, vol. 83, no. 5, pp. 413 – 424, 2006.

[55] A. J. Gonzàlez, R. Castrillón, and E. C. Quispe, "Energy efficiency improvement in the cement industry through energy management," in *Cement Industry Technical Conference, 2012 IEEE-IAS/PCA 53rd.* IEEE, 2012, pp. 1–13.

[56] H. Groenewald, J. Vosloo, and E. Mathews, "Cost-benefit to the cement industry by shifting evening load to off-peak periods," in *Industrial and Commercial Use of Energy Conference (ICUE), 2012 Proceedings of the 9th.* IEEE, 2012, pp. 1–6.

[57] J. Vosloo, I. Kruger, and J. Van Rensburg, "Demand market participation DMP on smaller energy users," in *Industrial and Commercial Use of Energy Conference*, 2012, pp. 1–5.

[58] S. Mitra, I. E. Grossmann, J. M. Pinto, and N. Arora, "Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes," *Computers & Chemical Engineering*, vol. 38, pp. 171–184, 2012.

[59] R. T. Lidbetter and L. Liebenberg, "Load-shifting opportunities for a typical south african cement plant," in *IEEE Industrial and Commercial Use of Energy*, 2011, pp. 17–25.

[60] T. Krause, E. V. Beck, R. Cherkaoui, A. Germond, G. Andersson, and D. Ernst, "A comparison of nash equilibria analysis and agent-based modelling for power markets," *International Journal of Electrical Power & Energy Systems*, vol. 28, no. 9, pp. 599–607, 2006.

[61] S.-J. Deng and S. S. Oren, "Electricity derivatives and risk management," *Energy*, vol. 31, no. 6, pp. 940–953, 2006.

[62] J. Zhong, E. Nobile, A. Bose, and K. Bhattacharya, "Localized reactive power markets using the concept of voltage control areas," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1555–1561, 2004.

[63] E. H. Allen and M. D. Ilić, "Reserve markets for power systems reliability," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 228–233, 2000.

[64] R. Fang and D. J. Hill, "A new strategy for transmission expansion in competitive electricity markets," *IEEE Transactions on power systems*, vol. 18, no. 1, pp. 374–380, 2003.

[65] A. J. Conejo, M. Carrión, and J. M. Morales, "Decision making under uncertainty in electricity markets," 2010.

[66] (2013, Jan) The PJM fast response regulation signal. PJM. [Online]. Available: http://www.pjm.com/markets-and-operations/ancillary-services/mkt-based-regulation/fast-response-regulation-signal.aspx

[67] *Energy & Ancillary Services Market Operations*, PJM, Aug 2013. [Online]. Available: http://www.pjm.com/documents/manuals.aspx

[68] M. H. Albadi and E. El-Saadany, "Demand response in electricity markets: An overview," in *2007 IEEE power engineering society general meeting*, 2007.

[69] "How smart is the smart grid?" July 2010. [Online]. Available: http://www.npr.org/templates/story/story.php?storyId=128365808

[70] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. Springer Science & Business Media, 2011.

[71] I. Grossmann and R. Sargent, "Optimum design of chemical plants with uncertain parameters," *AIChE Journal*, vol. 24, no. 6, pp. 1021–1028, 1978.

[72] M. G. Ierapetritou and E. N. Pistikopoulos, "Novel optimization approach of stochastic planning models," *Industrial & engineering chemistry research*, vol. 33, no. 8, pp. 1930–1942, 1994.

[73] S. Subrahmanyam, J. F. Pekny, and G. V. Reklaitis, "Design of batch chemical plants under market uncertainty," *Industrial & Engineering Chemistry Research*, vol. 33, no. 11, pp. 2688–2701, 1994.

[74] R. R. Iyer and I. E. Grossmann, "A bilevel decomposition algorithm for long-range planning of process networks," *Industrial & Engineering Chemistry Research*, vol. 37, no. 2, pp. 474–481, 1998.

[75] R. Schultz, "Stochastic programming with integer variables," *Mathematical Programming*, vol. 97, no. 1, pp. 285–309, 2003.

[76] S. Ahmed and R. Garcia, "Dynamic capacity acquisition and assignment under uncertainty," *Annals of Operations Research*, vol. 124, no. 1-4, pp. 267–283, 2003.

[77] V. Goel and I. E. Grossmann, "A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves," *Computers & chemical engineering*, vol. 28, no. 8, pp. 1409–1429, 2004.

[78] M. Carrion, A. B. Philpott, A. J. Conejo, and J. M. Arroyo, "A stochastic programming approach to electric energy procurement for large consumers," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 744–754, 2007.

[79] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2197–2206, 2011.

[80] Z. Li and M. G. Ierapetritou, "Capacity expansion planning through augmented lagrangian optimization and scenario decomposition," *AIChE Journal*, vol. 58, no. 3, pp. 871–883, 2012.

[81] A. Shapiro, D. Dentcheva *et al.*, *Lectures on stochastic programming: modeling and theory*. SIAM, 2014, vol. 16.

[82] P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," *Applied Energy*, vol. 96, no. 0, pp. 12 – 20, 2012, smart Grids.

[83] X. Zhang, G. He, S. Lin, and W. Yang, "Economic dispatch considering volatile wind power generation with lower-semi-deviation risk measure," in *Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, 2011, pp. 140–144.

[84] X. Zhang, G. He, K. Liu, J. Li, H. Zhai, and H. Lv., "A coordinated economic dispatch based on lower semi-absolute deviation risk," *Automation of Electric Power Systems*, vol. 36, no. 19, pp. 53–58, 2012.

[85] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton University Press, 2009.

[86] X. Lin, S. L. Janak, and C. A. Floudas, "A new robust optimization approach for scheduling under uncertainty:: I. bounded uncertainty," *Computers & chemical engineering*, vol. 28, no. 6, pp. 1069–1085, 2004.

[87] Z. Li and M. G. Ierapetritou, "Robust optimization for process scheduling under uncertainty," *Industrial & Engineering Chemistry Research*, vol. 47, no. 12, pp. 4148–4157, 2008.

[88] I. E. Grossmann, R. M. Apap, B. A. Calfa, P. Garcia-Herreros, and Q. Zhang, "Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty," *Computers & Chemical Engineering*, vol. 91, pp. 3–14, 2016.

[89] D. Bertsimas and M. Sim, "The price of robustness," *Operations research*, vol. 52, no. 1, pp. 35–53, 2004.

[90] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[91] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, 2006.

[92] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[93] L. T. Biegler, I. E. Grossmann, and A. W. Westerberg, *Systematic methods for chemical process design*. Prentice Hall, Old Tappan, NJ (United States), 1997.

[94] D. L. Applegate, R. E. Bixby, V. Chvatal, and W. J. Cook, *The traveling salesman problem: a computational study*. Princeton university press, 2011.

[95] J. E. Kelley, Jr, "The cutting-plane method for solving convex programs," *Journal of the society for Industrial and Applied Mathematics*, vol. 8, no. 4, pp. 703–712, 1960.

[96] B. C. Eaves and W. Zangwill, "Generalized cutting plane algorithms," *SIAM Journal on Control*, vol. 9, no. 4, pp. 529–542, 1971.

[97] E. Balas, S. Ceria, and G. Cornuéjols, "A lift-and-project cutting plane algorithm for mixed 0–1 programs," *Mathematical programming*, vol. 58, no. 1-3, pp. 295–324, 1993.

[98] N. W. Sawaya and I. E. Grossmann, "A cutting plane method for solving linear generalized disjunctive programming problems," *Computers & chemical engineering*, vol. 29, no. 9, pp. 1891–1913, 2005.

[99] A. Ali, J. Z. Kolter, and R. J. Tibshirani, "The multiple quantile graphical model," in *Neural Information Processing Systems (NIPS)*, 2016, pp. 3747–3755.

[100] X. Zhang, G. Hug, and I. Harjunkoski, "Cost-effective scheduling of steel plants with flexible eafs," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 239–249, Jan 2017.

[101] X. Zhang, G. Hug, Z. Kolter, and I. Harjunkoski, "Industrial demand response by steel plants with spinning reserve provision," in *47th North American Power Symposium*, 2015.

[102] C. C. Pantelides, "Unified frameworks for optimal process planning and scheduling," in *Proceedings on the second conference on foundations of computer aided operations*. Cache Publications New York, 1994, pp. 253–274.

[103] A. Dimitriadis, N. Shah, and C. Pantelides, "Rtn-based rolling horizon algorithms for medium term scheduling of multipurpose plants," *Computers & Chemical Engineering*, vol. 21, pp. S1061–S1066, 1997.

[104] C. T. Maravelias and I. E. Grossmann, "New general continuous-time state- task network formulation for short-term scheduling of multipurpose batch plants," *Industrial & Engineering Chemistry Research*, vol. 42, no. 13, pp. 3056–3074, 2003.

[105] ——, "On the relation of continuous-and discrete-time state–task network formulations," *AIChE journal*, vol. 52, no. 2, pp. 843–849, 2006.

[106] M. Wytock and J. Z. Kolter, "Sparse gaussian conditional random fields: Algorithms, theory, and application to energy forecasting," in *Proc. 30th Int. Conf. Machine Learning*, 2013, pp. 1265–1273.

[107] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.

[108] E. C. Kara, J. S. Macdonald, D. Black, M. Brges, G. Hug, and S. Kiliccote, "Estimating the benefits of electric vehicle smart charging at non-residential locations: A data-driven approach," *Applied Energy*, vol. 155, pp. 515 – 525, 2015.

[109] H. Hadera, P. Wide, I. Harjunkoski, J. Mäntysaari, J. Ekström, G. Sand, and S. Engell, "A mean value cross decomposition strategy for demand-side management of a pulping process," *Computer Aided Chemical Engineering*, vol. 37, pp. 1931–1936, 2015.

[110] X. Zhang, G. Hug, J. Z. Kolter, and I. Harjunkoski, "Model predictive control of industrial loads and energy storage for demand response," in *IEEE PES General Meeting*, 2016.

[111] X. Zhang, G. Hug, and I. Harjunkoski, "Demand response of ancillary service from industrial loads coordinated with energy storage," *IEEE Transactions on Power Systems*, submitted.

[112] N. Ivanova, "Lithium-ion costs to fall by up to 50% within five years," July 2015. [Online]. Available: http://analysis.energystorageupdate.com/lithium-ion-costs-fall-50-within-five-years

[113] N. DiOrio, A. Dobos, and S. Janzou, "Economic analysis case studies of battery energy storage with sam," *National Renewable Energy Laboratory*, 2015.

[114] A. Eller and D. Gauntlett, "Economic analysis case studies of battery energy storage with sam," *International Finance Corporation*, 2017.

[115] "Most and least profitable business types," January 2011. [Online]. Available: https://www.bloomberg.com/news/photo-essays/2011-01-18/most-and-least-profitable-business-types#slide10

[116] S. Chen, T. Zhang, H. Gooi, R. D. Masiello, and W. Katzenstein, "Penetration rate and effectiveness studies of aggregated BESS for frequency regulation," *IEEE Trans. on Smart Grid*, vol. 7, no. 1, pp. 167–177, 2016.

[117] X. Zhang, F. Gao, X. Lv, H. Lv, Q. Tian, J. Ma, W. Yin, and J. Dong, "Line loss reduction with distributed energy storage systems," in *IEEE Innovative Smart Grid Technologies*, 2012.

[118] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation with application to solar energy," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–14, 2016.

[119] X. Sun, P. B. Luh, K. W. Cheung, W. Guan, L. D. Michel, S. Venkata, and M. T. Miller, "An efficient approach to short-term load forecasting at the distribution level," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2526–2537, 2016.

[120] M. Wytock and J. Z. Kolter, "Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields," in *52nd IEEE Conference on Decision and Control*, 2013, pp. 1019–1024.