12-15-2016

Mixed-Integer Optimization Problems with Applications to Manufacturing Scheduling and Distributed Energy System Operation

Bing Yan
bing.yan@uconn.edu

Mixed-Integer Optimization Problems with Applications to Manufacturing Scheduling and Distributed Energy System Operation

Bing Yan, PhD

University of Connecticut, 2016

Over the past decades, mixed-integer optimization problems have attracted a lot attentions. Two popular topics are the scheduling problems of semiconductor manufacturing, and operation problems of distributed energy systems. On the one hand, the increasing pressure to meet demand is forcing semiconductor manufacturers to seek efficient scheduling methods. On the other hand, with world's increasing energy demand and growing environmental concerns, efficient utilization of energy is essential. Lithography, with a limited number of expensive resources, is a major bottleneck in memory chip manufacturing. Because of its complex characteristics and large sizes of practical problems, developing effective scheduling approaches is challenging. In this thesis, a mixed-integer linear formulation is established for high-volume and low-variety manufacturing through novel resource-based modeling instead of traditional lot-based. To solve this problem efficiently by branch-and-cut, a two-phase approach is established based on convex hull analysis.

The solution methodology for litho machine scheduling can also be used for other mixed-integer linear problems such as distributed energy system (DES) operation. Energy demands and energy supplied by different devices are characterized by different levels of quality, which is measured by exergy in thermodynamics. Exergy is destroyed in various processes, with limited amount of exergy in fossil fuels, it is therefore important to match demand and supply in quantity and quality to avoid exergy waste. Flexible DESs provide a desirable infrastructure. An exergy-based optimization approach is therefore developed for DES operation to reduce energy costs and the exergy losses by considering the whole energy supply chain from energy resources to user demands. To capture the complicated interactions among energy devices and capture the exergy loss of each energy device, exergy networks are established with detailed device and

water network models. The mixed-integer problem is efficiently solved by our latest surrogate Lagrangian relaxation with branch-and-cut. With renewables, a similar DES operation problem is considered to minimize energy and emission costs. To overcome the difficulty caused by the intermittent nature of renewables, PV uncertainties are modeled by a Markovian process. For effective coordination, other devices are modeled as Markov processes with states depending on PV states. The entire problem is stochastic and Markovian, and solved by branch-and-cut. To take capital and maintenance costs into account in the long run, the design problem is also considered to decide device sizes with given types. To evaluate the lifetime cost including the reliability cost under different types of grid connection, a linear model is established. By selecting a limited number of possible device size combinations, exhaustive search is used to find the optimized design.

Mixed-Integer Optimization Problems with Applications to Manufacturing Scheduling and Distributed Energy System Operation

Bing Yan

B.S., Renmin University of China, Beijing, China, 2010

M.S., University of Connecticut, Storrs, CT, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2016

APPROVAL PAGE

Doctor of Philosophy Dissertation

Mixed-Integer Optimization Problems with Applications to Manufacturing Scheduling and Distributed Energy System Operation

Presented by

Bing Yan, B.S., M.S.

Major Advisor _____
Peter B. Luh

Associate Advisor _____
Krishna R. Pattipati

Associate Advisor _____
Peng Zhang

University of Connecticut
2016

ACKNOWLEDGMENTS

First, I would like to express my deepest gratitude to my adviser, Prof. Peter. B. Luh for his continuous support of my doctoral study at the University of Connecticut. His guidance, encouragement, and immense knowledge helped me in all the time of research and paved the way for my successful dissertation.

I would also like to express my sincere thanks to Prof. Krishna R. Pattipati and Prof. Peng Zhang for joining my advisory committee and providing useful suggestions.

I gratefully acknowledge Prof. Sung-Yeul Park, Prof. Ali Bazzi, Prof. Yang Cao, Prof. Laurent D. Michel, and Prof. Bing Wang of University of Connecticut; Prof. Vincenzo Naso and Prof. Nicola Bianco of Università degli Studi Federico II; Dr. Luigi Mongibello and Dr. Giorgio Graditi of ENEA Research Center; Mr. Joey Chang, Mr. Simon Wang, and Mr. Hsin Yuan Chen of Inotera Memories Inc.; Mr. Guy Warner and Mr. Alan McDonnell of Pareto Energy, Ltd.; Dr. Zhongxue Gan, Dr. Chen Song, and Mr. Chenhui Dong of ENN Group; Dr. Khosrow Moslehi, Dr. Xiaoming Feng, Dr. Chien-Ning Yu of ABB; and Dr. Eugene Litvinov, Dr. Tongxin Zheng, Mr. Izudin Lelic, Dr. Feng Zhao, Dr. Jinye Zhao, and Dr. Dane A. Schiro of ISO New England, for insightful discussions and invaluable comments on our collaborated research works. My sincere thanks to Mr. Joey Chang and Dr. Chen Song for providing me internship opportunities for beneficial industry experience, and Prof. Vincenzo Naso and Prof. Nicola Bianco for inviting me to Università degli Studi Federico II as a visiting student.

I am most grateful to Dr. Marialaura Di Somma of the Department of ENEA Research Center for her innovative ideas and essential contributions on developing optimization models during our collaborated research works.

I would like to thank all my labmates for the pleasant study and research experience. The discussions and exchanges of knowledge enriched my experience.

And last but not least, I would like to thank my husband Mr. Zhiheng Xu, my parents Dr. Jinwu Yan and Mrs Hong Xiao, other family members, and friends for their love and support throughout my PhD study and my life.

Table of Contents

List of Figures

List of Tables

Publications Related to this Thesis

Journal Articles

[1] **B. Yan**, H. Y. Chen, P. B. Luh, S. Wang, and J. Chang, "Litho Machine Scheduling With Convex Hull Analyses," *IEEE Transactions on Automation Science and Engineering*, Vol.10, Issue 4, pp. 928-937, Oct., 2013.

[2] **B. Yan**, M. Di Somma, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Exergy-based Operation Optimization of a Distributed Energy System through the Energy-supply Chain," *Applied Thermal Engineering*, Vol. 101, pp. 741-751, 2016.

[3] **B. Yan**, H. Fan, P. B. Luh, K. Moslehi, X. Feng, C.N. Yu, M. A. Bragin, and Y. Yu, "Grid Integration of Wind Generation Considering Remote Wind Farms: Hybrid Markovian and Interval Unit Commitment," to appear in *IEEE/CAA Journal of Automatica Sinica*, accepted in Aug., 2016.

[4] **B. Yan**, P. B. Luh, G. Warner, and P. Zhang, "Operation Optimization for Microgrids with Renewables," *IEEE Transactions on Automation Science and Engineering*, accepted in Nov., 2016.

[5] M. Di Somma, **B. Yan**, N.Bianco, G. Graditi, P. B. Luh, L. Mongibello, and V. Naso, "Operation Optimization of a Distributed Energy System Considering Energy Costs and Exergy Efficiency," *Energy Conversion and Management*, Vol. 103, pp. 739-751, Oct. 2015.

[6] M. Di Somma, **B. Yan**, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Multi-objective Operation Optimization of a Distributed Energy System for a Large-scale Utility Customer," *Applied Thermal Engineering*, Vol. 101, pp. 752-761, 2016.

[7] B. Sun, P. B. Luh, Q. Jia, and **B. Yan**, "Event-based Optimization within the Lagrangian Relaxation Framework for Energy Savings in HVAC Systems," *IEEE Transactions on Automation Science and Engineering*, Vol.12, Issue 4, pp. 1396-1406, Aug., 2015.

Conference Proceedings

[8] **B. Yan**, H. Y. Chen, P. B. Luh, S. Wang, and J. Chang, "Optimization-based Litho Machine Scheduling with Multiple Reticles and Setups," in *Proceeding of IEEE 7th Conference on Automation Science and Engineering*, pp. 575 - 580, Trieste, Italy, Agu., 2011.

[9] **B. Yan**, H. Y. Chen, P. B. Luh, S. Wang, and J. Chang,"Optimization-based Litho Machine Scheduling with Load Balancing and Reticle Expiration," in *Proceeding of IEEE 8th Conference on Automation Science and Engineering*, pp. 114 - 119, Seoul, South Korean, Agu., 2012. (Recommended for *IEEE T-ASE*)

[10] **B. Yan**, P. B. Luh, B. Sun, C. Song, C.Dong, Z. Gan, and L. D. Michel, "Energy-efficient Management of Eco-communities," in *Proceedings of IEEE 9th Conference on Automation Science and Engineering, Madison*, pp. 106 - 111, WI, USA, Agu., 2013.

[11] **B. Yan**, P. B. Luh, M. A. Bragin, C. Song, C. Dong, and Z. Gan, "Energy-efficient Building Clusters," in *Proceedings of IEEE 10th Conference on Automation Science and Engineering*, pp. 966 - 971, Taipei, Taiwan, Aug., 2014.

[12] **B. Yan**, M. Di Somma, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Exergy-based Operation Optimization of a Distributed Energy System through the Energy-supply Chain," in *Proceedings of ASME-ATI-UIT 2015 Conference*, Naples, Italy, May, 2015. (Selected as best papers, and recommended for *Applied Thermal Engineering*)

[13] **B. Yan**, P. B. Luh, E. Litvinov, T. Zheng, D. Schiro, Mikhail A. Bragin, F. Zhao, J. Zhao, and I. Lelic "Unit Commitment with Generation-dependent Ramp Rates and Reserve Requirements," submitted to *2017 IEEE Power and Energy Society General Meeting*, Chicago, Illinois, July 2017.

[14] M. Di Somma, **B. Yan**, P. B. Luh, M. A. Bragin, N. Bianco, G. Graditi, L. Mongibello, and V. Naso, "Exergy-efficient Management of Energy Districts," in *Proceedings of 12th World Congress on Intelligent Control and Automation*, China, Jun., 2014.

[15] M. Di Somma, **B. Yan**, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Multi-objective Operational Optimization of a Distributed Energy System for a Large-scale Utility Customer," in *Proceedings of ASME-ATI-UIT 2015 Conference*, Naples, Italy, May, 2015. (Selected as best papers, and recommended for *Applied Thermal Engineering*)

[16] M. Di Somma, **B. Yan**, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Influence of Energy Quality Management on CO_2 Emissions in Operation Optimization of a Distributed Energy

System," in *Proceedings of 5th edition of International Conference on Clean Electrical Power*, Taormina, Italy, Jun., 2015.

[17] M. Di Somma, **B. Yan**, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Optimal Sizing Design of Distributed Energy Systems through Exergy and Cost Assessment," in *Proceeding of 8th International Conference on Applied Energy*, Beijing, China, October, 2016. (Recommended for *Applied Energy*)

[18] B. Sun, P. Luh, Q. Jia, and **B. Yan**, "Event-based Optimization with Non-stationary Uncertainties to Save Energy Costs of HVAC Systems in Buildings", in *Proceeding of IEEE 9th Conference on Automation Science and Engineering, Madison*, pp. 436 - 441, WI, USA, Agu., 2013.

Chapter 1

Introduction

1.1 Motivations

Over the past decades, mixed-integer optimization problems have attracted a lot attentions. Among them, two popular topics are the scheduling problems of semiconductor manufacturing, and operation and design problems of power systems. The increasing pressure to meet demand is forcing semiconductor manufacturers to seek efficient scheduling methods. Lithography, with a limited number of expensive resources and the re-entrant nature of the fabrication processes, is a major bottleneck [1]. Lithography is the process of transferring circuit patterns to the surface of a wafer by selectively exposing light through a reticle, where a wafer is developed layer by layer. Before processing a specific layer, a machine and a corresponding reticle need to be set up. In addition, reticles need to be recalibrated after processing a certain number of layers, and sometimes a set of layers must be processed on the same machine. Because of these processing requirements and the large sizes of practical problems, developing effective scheduling approaches is challenging. In addition, the mixed-integer problem is believed to be NP hard.

The solution methodology for litho machine scheduling problems can also be used for other mixed-integer linear problems such as operation problems of distributed energy systems. Energy demands such as electricity and space heating are characterized by different levels of quality. Energy supplied by devices such as combined heat and power and heat pumps with different energy resources (e.g., natural gas and

electricity) also has different levels of quality. In thermodynamics, such quality is measured by exergy, which is destroyed in various processes [2-7]. With limited amount of exergy in fossil fuels, it is important to match demand and supply in both quantity and quality to avoid waste of exergy and improve sustainability. Distributed Energy Systems (DESs), where energy is made available close to end-users, provide a unique opportunity to show the benefits of the exergy analysis. The problem is challenging in view of the complicated interactions among devices and the modeling of exergy losses.

With world's increasing energy demand and growing environmental concerns, efficient utilization of energy is essential for sustainable living, especially renewable energy. Reliable and flexible microgrids, which can operate under the grid-connected mode and can also turn into an islanded mode [8, 9], provide a promising opportunity and a desirable infrastructure. In microgrids, different distributed energy devices, such as gas turbines, photovoltaic panels, and natural gas boilers, generate and store different types of energy such as electricity, steam, and hot/chilled water to satisfy time-varying electricity and thermal demand. They should be coordinated through daily operation to reduce the energy cost and greenhouse gas emissions. Optimized microgrid operation, however, is challenging because of the intermittent nature of renewables. To consider capital and maintenance costs in the long run, microgrid design (device types and sizes) is also critical. The design problem is also challenging since the problem complexity increases exponentially as the problem size increases, and energy resources (e.g., solar irradiance), fuel prices, and load are uncertain. In addition, the reliability costs, i.e., costs of microgrid protection devices and costs of unserved load when there is no power supply, are hard to estimate.

1.2 Major Contributions

To overcome the above difficulties, this dissertation develops three novel approaches as follows.

1. *Litho machine scheduling for high-volume and low-variety manufacturing.* Lithography, with a limited number of expensive resources and the re-entrant nature of the fabrication processes, is a major

bottleneck. Because of its complex characteristics and the large sizes of practical problems, developing effective scheduling approaches is challenging. In this work, a mixed-integer linear problem is established with novel modeling of resource setups, reticle expirations, and future stacking layer load balancing. To solve this NP problem efficiently, a two-phase approach is established by using branch-and-cut with convex hull analysis.

2. *Exergy-based operation optimization of distributed energy systems.* The second topic is to develop an exergy-based operation optimization approach by considering the whole energy supply chain from energy resources to user demands of a distributed energy systems. To capture the complicated interactions among energy devices and capture the exergy loss of each energy step/device, exergy networks are established with detailed device models and water network models. A multi-objective mixed-integer problem is formulated to reduce energy costs and the exergy losses at the conversion step. By solving the problem with our latest surrogate Lagrangian relaxation and branch-and-cut, the operators can choose the operation strategy from the Pareto frontier based on costs, essential in the short run, and sustainability, crucial in the long run.

3. *Operation and design optimization of microgrids*. The last topic is to develop mathematical formulations and optimization methods for operation and design of microgrids. The operation problem is to commit and dispatch distributed devices with renewable generation to minimize energy and emission costs while meeting forecasted energy demand. To overcome the difficulty caused by the intermittent nature of renewables, PV uncertainties are modeled by a Markovian process. For effective coordination, other devices are modeled as Markov processes with states depending on PV states. The entire problem is stochastic and Markovian. This combinatorial problem is solved by branch-and-cut. Beyond energy and emission costs, the design problem is to decide device sizes with given types to minimize the lifetime cost while satisfying energy demand. To evaluate the lifetime cost including the reliability cost and the classic components such as capital and fuel costs, a linear model is established.

By selecting a limited number of possible device size combinations, exhaustive search is used to find the optimized design.

1.3 Organization of this Thesis

The rest of this thesis is organized as follows. Chapter 2 introduces litho machine scheduling for high-volume and low-variety manufacturing. Chapter 3 presents exergy-based operation of distributed energy systems. Chapter 4 discusses operation and design optimization of microgrids.

References

[1] E. Akcali, K. Nemoto, and R. Uzsoy, "Cycle-Time improvements for photolithography process in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 14, No. 1, pp. 48-56, 2001.

[2] ECBCS - Annex 49 - Low Exergy Systems for High Performance Buildings and Communities, homepage. Available <http://www.ecbcs.org/annexes/annex49.htm>.

[3] D. Schmidt Low exergy systems for high performance buildings and communities, Energy and Buildings 41 (2009) 331-336.

[4] J. Szargut International progress in second law analysis, Energy 5 (1980) 709-718.

[5] IEA/ECBCS Annex 37, Low Exergy Systems for Heating and Cooling (2003).

[6] D. Schmidt Design of low exergy buildings – method and a pre-design tool, *The International Journal of Low Energy and Sustainable Buildings* 3 (2004) 1–2.

[7] J. Szargut, D.R. Morris, F.R. Stewerd Exergy analysis of thermal, chemical and metallurgical processes (1988) New York: Hemisphere.

[8] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids," *IEEE Power and Energy Magazine*, Vol. 5, Issue 4, pp. 78-94, 2007.

[9] F. Katiraei and M. R. Iravani, "Power management strategies for a microgrid with multiple distributed generation units," *IEEE Transactions on power systems*, Vol. 21, Issue 4, pp. 1821-1831, 2006.

Chapter 2

Litho Machine Scheduling with Convex Hull Analyses

The increasing pressure to meet demand are forcing semiconductor manufacturers to seek efficient scheduling methods. Lithography, with a limited number of expensive resources and the re-entrant nature of the fabrication processes, is a major bottleneck. This chapter presents a litho machine scheduling formulation for high-volume and low-variety manufacturing over a day, with novel modeling of resource setups, reticle expirations, and future stacking layer load balancing. The problem is believed to be NP hard. After linearization and simplification, it is solved by using the branch-and-cut method by exploiting problem linearity. Near-optimal solutions for practical problems, however, are still difficult to obtain efficiently. Through detailed analyses, it was discovered that the convex hull of the problem is difficult to delineate and many low-efficient branching operations are needed. A two-phase approach is therefore established. In the first phase, a simplified problem with certain complicating constraints dropped is efficiently solved by exploiting linearity to reduce ranges of decision variables. The problem with the full set of constraints is then solved in the second phase with a much reduced decision space. Numerical testing shows that this two-phase approach can generate near-optimal schedules within reasonable amounts of computation time. This two-phase approach is generic, and will have major implications on other semiconductor scheduling problems and beyond.

2.1 Introduction

Lithography is the process of transferring circuit patterns to the surface of a wafer by selectively exposing light through a reticle. During this process, a wafer is incrementally developed layer by layer in lots (Shr, et al., 2008), where different products require different sets of layers to be completed. Lithography, with a limited number of expensive resources and the re-entrant nature of the fabrication processes, is a major bottleneck in semiconductor manufacturing (Akcali, et al., 2001). The increasing pressure to meet demand is forcing manufacturers to seek efficient scheduling methods.

In a fab, litho machines are generally unique, and reticles are usually divided into groups based on which product\layer they process. One machine usually requires one reticle to process a layer. Before processing a specific layer, a machine and a corresponding reticle need to be set up, and excessive setups are costly and undesirable. During processing, a lot needs a certain amount of time to be completed. In addition, reticles need to be recalibrated after processing a certain number of lots. Reticles in the same group therefore should not expire simultaneously to avoid reticle shortage. For certain products, a selected set of layers (stacking layers) must be processed on the same machine for precision fabrication. The load on machines processing stacking layers need to be balanced to prevent future overload or starvation. In our problem, products have high volume and low variety, and a daily target is assigned to each product/layer. Therefore there is no need to number and distinguish each lot. The problem is to allocate machines and reticles over a day to meet the daily targets.

As will be reviewed in Section 2.2, reticle expiration was rarely addressed in the literature. Also, most papers focused on balancing the current load, and rarely discussed the effect of machine assignments on future load through stacking layers. In this work, a formulation for litho machine scheduling over a day is established with novel modeling of resource setups, reticle expirations, and future stacking layer load in Section 2.3. It contains four major sets of constraints regarding resource capacities, processing times, maximal numbers of lots scheduled and setups. Since a setup time is generally much shorter than the

corresponding lot processing time, the time of setup is ignored and the number of setups is considered. To simplify the formulation and to reduce the number of setups, it is assumed that all the lots assigned to a machine to process a particular layer within the day will be processed under one setup. The objective function is to meet targets, balance future load, avoid simultaneous reticle expirations, and avoid excessive setups. Future stacking layer load can be adjusted through proper machine assignments, and simultaneous reticle expirations can be avoided by spacing out expiration dates through proper reticle assignments. The problem formulated above is linear and believed to be NP hard.

The problem is solved by using the branch-and-cut method in Section 2.4 after certain constraints are simplified without sacrificing optimality. Branch-and-cut is powerful for certain classes of mixed-integer linear optimization problems, and is easy to code by using commercial solvers. In the method, the integrality-relaxed problem is solved first by using a linear programming method. If all integer decision variables are integers, the solution is optimal to the original problem. If not, valid cuts are added trying to obtain the convex hull. The idea is that once the convex hull is obtained, all integer decision variables of the linear programming solution are integers and optimal to the original problem. The process of obtaining the convex hull, however, is problem dependent, and can itself be NP hard. Low-efficient branching operations may then be needed. We found that near-optimal solutions for practical problems are difficult to obtain efficiently. Through detailed analyses, it was discovered that the convex hull is difficult to delineate because of certain complicating constraints. A two-phase approach is therefore developed. In the first phase, a simplified problem with those complicating constraints dropped is efficiently solved to establish ranges of decision variables. The problem with the full set of constraints is then solved in the second phase with a much reduced decision space.

The methods have been implemented by using IBM ILOG CPLEX, and three examples are presented in Section 2.5. Numerical results show that the two-phase approach can generate near-optimal schedules

within much reduced computation time than the one-phase approach. More importantly, this approach is generic, and will have major implications on other semiconductor scheduling problems and beyond.

2.2 Literature Review

Developing effective scheduling approaches for semiconductor manufacturing is challenging because of its complex re-entrant characteristics and the large sizes of practical problems. Approaches for litho machine scheduling including heuristic rules and mathematical programming will be reviewed.

Heuristic rules with simulation techniques

Heuristic rules for litho machine scheduling are briefly reviewed with simulation techniques used to valid them in most papers. A mixed-integer model for short-time capacity scheduling was developed in Toktay and Uzsoy (1998), and the objective is to maximize throughput and the total amount of WIP processed at each workcenter including lithography. Fast heuristics were presented for computation efficiency. However, only a single product type was considered. A lot scheduling problem with capacity scheduling and lot sequencing sub-problems was discussed in Akcali and Uzsoy (2000). Greedy heuristics were used to solve the problem, and a simulation model of a wafer fabrication facility was used to examine the effects of this method on lithography. For simplicity, the processing time of each layer required by each lot was assumed to be identical on all machines. A method for load balancing in the lithography area based on the greedy algorithm was discussed in Mönch, et al. (2001). A detailed simulation model was developed. To improve load balancing in the lithography area, the lot assignment was decided at the time when the lots were released. Three dispatching rules and four bottleneck scheduling rules for lithography were studied in Lee et al. (2002), and the objective is to maximize the production volume. Some lot scheduling rules were also developed for WIP balancing, and combinations of these rules were tested for various performance measurements. This study was extended in Akcali et al. (2005), where machines were eligible to process a specified subset of operations, and a setup was required when an operation was changed. The

focus was on allocating the capacity to available jobs rather than making sequencing decisions. A number of heuristic algorithms and a greedy randomized adaptive search procedure were presented. A model that characterized the lithography process was developed in Arisha and Young (2004), and dispatching rules for mask change reduction and setup reduction were studied. Dispatching strategies for regular lots and priority lots were investigated in Yugma et al. (2007) to decrease cycle times and increase the number of daily moves. To balance load, the "Resource Schedule and Execution Matrix" model was presented in Shr et al. (2008), and the lot with the largest wait steps was assigned to the litho machine with the smallest load. For simplicity, it was assumed that each lot had the same process steps and quantity, and each layer had the same processing time. With heuristic rules, schedules can be efficiently obtained, but it is difficult to find or know the optimal rules. Also, simulation can be time consuming.

Mathematical programming

Mathematical programming methods including Lagrangian relaxation, branch-and-bound and branch-and-cut that have the capability to solve our problem are reviewed in this subsection. Lagrangian relaxation is a popular method for mathematical programming. A real-time scheduling and dispatching framework was developed in Chang and Liao (1994) for a semiconductor fab including lithography. The problem was solved by using Lagrangian relaxation and network flow techniques without considering setups. Lagrange relaxation was also used to solve a lot scheduling problem with aggregated process steps for high variety and low volume fabrication in Liao et al. (1996). Only problems with short planning horizons (e.g., one shift to one day) were considered due to complexity issues.

Branch-and-bound has also been used. A production control method was investigated in Vargas-Villamil and Rivera (2000), and it was applied to discrete event reentrant semiconductor manufacturing lines for scheduling. A tradeoff was made between production rate and cycle time for overall optimality. A mixed-integer stochastic programming model for capacity planning under demand uncertainty was developed in Barahona et al. (2005). Cutting planes and a heuristic approach were used to improve

computation efficiency of branch-and-bound. Still, computation efficiency remained to be challenging for problems with larger numbers of scenarios and long periods.

Branch-and-cut has now been widely used. A WIP balancing concept was presented in Chung and Jang (2009), and the bottleneck machines were divided into different load levels for higher throughput. The mixed-integer formulation was solved by using CPLEX to decide the quantity of lots to be processed on litho machines. It was believed that the model with lot precedence constraints would require longer computation time. Branch-and-cut was also used to solve a single machine and multiple-lot-per-carrier (front-opening unified pod) scheduling problem in Sarin et al. (2012), and the objective was to minimize the sum of lot completion time. All carriers were assumed identical, and the processing time per wafer was assumed the same. The method could solve a problem at the root node itself, while it could not solve large-sized instances. It can be seen that for the papers with branch-and-cut, how to improve computation efficiency is a major challenge.

2.3 Problem Formulation

As reviewed in Subsection 2.2, reticle expiration was rarely addressed in the literature. Also, most papers focused on balancing the current load, and rarely discussed the effect of machine assignments on future load through stacking layers. A novel formulation for litho machine scheduling over a day is established in this subsection. It contains four major sets of constraints as presented in the first four subsubsections. The objective function is to meet targets, balance future load, avoid simultaneous reticle expirations, and avoid excessive setups as discussed in subsubsection 2.3.5. To solve the problem by using branch-and-cut, a linear formulation is needed.

2.3.1 Resource capacity constraints

Consider a fab with M litho machines indexed by m and R reticles indexed by r as resources. There are K discrete time slots indexed by k within a day. In the fab, P types of products with index p are processed,

and each requires L types of layers with index l.

For one machine or reticle, there are only three statuses, processing, idle, and unavailable. To obtain these statuses, a set of binary variables with machine, reticle and time indices is used here. Based on the formulation in Wang and Luh (1996), the key decision variables are defined as follows:

$$\delta_{mr}(k) = \begin{cases} 1, & \text{if machine } m \text{ is combined with reticle } r \\ & \text{to process a layer at time slot } k; \\ 0, & \text{otherwise.} \end{cases}$$

Machine capacity, reticle capacity, machine-reticle matching, and resource maintenance constraints are described as follows.

1) *Machine capacity constraints*

One machine requires only one reticle to process a layer at any time slot, i.e.,

$$\sum_{r} \delta_{mr}(k) \le 1, \forall k, \forall m. \tag{2.1}$$

2) *Reticle capacity constraints*

Likewise, one reticle requires only one machine to process a layer at any time slot, i.e.,

$$\sum_{m} \delta_{mr}(k) \le 1, \forall k, \forall r. \tag{2.2}$$

3) *Machine-reticle matching constraints*

Litho machines are generally unique, and machine m cannot be combined with the reticles in set S_m^{RN} to process layers, i.e.,

$$\delta_{mr}(k) = 0, \forall k, \forall m, and\ r \in S_m^{RN}. \tag{2.3}$$

4) *Resource maintenance constraints*

One machine is not available during maintenance, i.e.,

$$\delta_{mr}(k) = 0, k \in [b_m^M, c_m^M], m \in S^{MM} and \forall r, \tag{2.4}$$

where S^{MM} is the set of machines that need to do maintenance within the day, and $b_m{}^M$ and $c_m{}^M$ are the beginning time and the completion time of maintenance on machine m.

The modeling of reticle maintenance is similar.

2.3.2 Processing time requirements

As mentioned earlier, the time of setup is ignored and the number of setups is considered since a setup time is generally much shorter than the corresponding lot processing time. To simplify the formulation and to reduce the number of setups, it is assumed that all the lots assigned to a machine to process a particular layer within the day will be processed under one setup. As shown in Figure 2.1, machine m_1 with reticle r_1 processes N_1 lots under one setup in Case 1. Sometimes, there may be an unfinished lot on the machine at the beginning or end point of the day as shown in Case 2 and Case 3, respectively. In Case 4, both of these two situations occur.

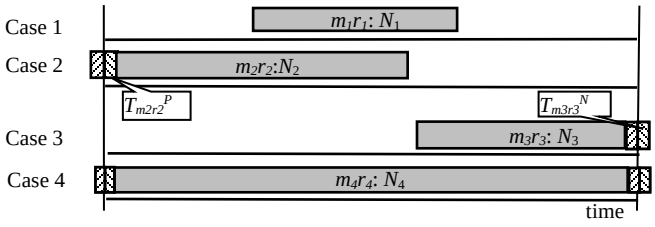


Figure 2.1. Four situations of processing.

In general, let N_{mr} denote the number of lots scheduled on machine m and reticle r within the day, and $N_{mr}{}^{UB}$ denote its upper bound. Let $T_{mr}{}^P$ denote the time required to complete the unfinished lot left over from the previous day on machine m and reticle r, and $T_{mr}{}^N$ denote the time required to complete the unfinished lot left for the next day. The value of the first variable is known, and the second variable is an integer decision variable. For machine m with reticle r, N_{mr} times of processing time T_{mr} must be assigned, and if the last time slot is involved, one unfinished lot can be left. The four cases mentioned above can be combined together as follows:

$$\sum_k \delta_{mr}(k) - T_{mr}^P + T_{mr}^N = N_{mr} \times T_{mr}, \tag{2.5}$$

$$0 \le N_{mr} \le N_{mr}^{UB}, N_{mr}^{UB} = (K - T_{mr}^P) \div T_{mr} + 1, \forall r, \forall m. \tag{2.6}$$

The unfinished lot from the previous day is assumed not to be included in the total number of lots scheduled within the day, while the lot left for the next day is.

In addition, the time required to process the unfinished lot left for the next day TmrN must be smaller than the corresponding lot processing time, i.e.,

$$0 \le T_{mr}^N \le T_{mr} - 1, \forall r, \forall m. \tag{2.7}$$

If the last time slot is not involved, every lot must be finished, therefore TmrN must be zero, i.e.,

$$\text{if } \delta_{mr}(K) = 0 \text{ then } T_{mr}^N = 0, \forall r, \forall m. \tag{2.8}$$

The above constraint is logical, but it is easier to be linearized together with Eq. (2.7) as follows:

$$0 \le T_{mr}^N \le \delta_{mr}(K) \times (T_{mr} - 1), \forall r, \forall m. \tag{2.9}$$

If $\delta_{mr}(K) = 0$, $0 \le T_{mr}^N \le 0$, T_{mr}^N must be zero; if $\delta_{mr}(K) = 1$, $0 \le T_{mr}^N \le T_{mr} - 1$. Therefore the set of linear constraints (2.9) satisfies both constraints (2.7) and (2.8) above.

2.3.3 Maximal number of lots scheduled constraints

If there is extra capacity beyond the total target, machines and reticles will be scheduled to process layers with high priorities because more reward is assigned as will be discussed in the objective function. However, this may lead to imbalance among layers of the same product because of the layer by layer process. To avoid this, the number of lots with layer *l* of product *p* to be processed should be under its upper bound, i.e.,

$$N_{pl} \le T_{pl}^{UB}, \forall p, \forall l. \tag{2.10}$$

In the above, T_{pl}^{UB} is the upper bound for layer *l* of product *p*, and this set of parameters are calculated offline based on heuristic rules (e.g., 1.2 times of target T_{pl}). The number of lots with layer *l* of product *p* to be processed within the day is denoted by N_{pl}, and this integer dependent variable can be derived from Nmr as follows:

$$N_{pl} = \sum_{m} \sum_{r \in S_{pl}^{R}} N_{mr}. \tag{2.11}$$

In the above, S_{pl}^{R} denotes the set of reticles that process layer *l* of product *p*.

2.3.4 Setups-related constraints

When one litho machine switches from processing one layer to another layer, the machine and a corresponding reticle need to be set up. Since one reticle can only process one particular product\layer, a layer process switch on the litho machine can be treated as a reticle switch. In addition, a setup time is generally much shorter than the corresponding lot processing time, the time of setup is ignored and the number of setups is considered. In this way, the number of resource setups can be modeled as the number of reticle changes. The key issue here is how to find the beginning and completion points of machine and reticle combinations. The two situations of one machine completes combining with one reticle and begins to combine with another reticle are shown in Figure 2.2 below.

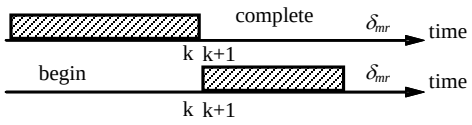


Figure 2.2. Completion and beginning points of processing.

It can be seen that when the values of $\delta_{mr}(k)$ and $\delta_{mr}(k+1)$ switch from 1 to 0, machine *m* completes the combination with reticle *r*; when the values switch from 0 to 1, machine *m* begins the combination with reticle *r*. To get a linear formulation, a new set of binary decision variables $\{y_{mr}(k)\}$ is used as follows:

$$y_{mr}(k+1) \geq \delta_{mr}(k+1) - \delta_{mr}(k),\ y_{mr}(k+1) \geq \delta_{mr}(k) - \delta_{mr}(k+1),\ \forall m, \forall r, 1 \leq k \leq K-1. \tag{2.12}$$

Since the purpose is to reduce the number of setups and $\{y_{mr}(k)\}$ is only shown in objective function, when $\delta_{mr}(k+1) - \delta_{mr}(k) = \pm 1$, $y_{mr}(k+1) = 1$; when $\delta_{mr}(k+1) - \delta_{mr}(k) = 0$, $y_{mr}(k+1) = 0$.

The beginning and completion points occur in pairs, the second half of (2.12) is therefore used for practical problems.

2.3.5 Objective function

The objective function has four terms, to meet targets, balance future load, avoid simultaneous reticle expirations, and avoid excessive setups as presented below.

1) *Meeting targets*

If machine assignments exceed the target, it is expressed as certain reward in the objective function; if not, it is expressed as corresponding penalty. Since different layers have different priorities, different weights are assigned. Let W_{pl}^{R} and W_{pl}^{P} denote the reward and penalty weights for layer l of product p. To check whether the assignments meet the daily target, a piecewise function is used as shown in Figure 2.3 below.

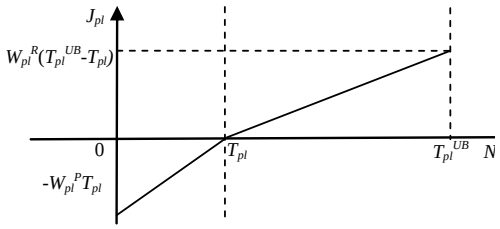


Figure 2.3. The meeting target term of the objective function.

The upper bound and lower bound of N_{pl} is T_{pl}^{UB} and 0, and three break points are 0, T_{pl} and T_{pl}^{UB}. Based on the special ordered set techniques (Beale and Forrest, 1976), the above formulation can be fully linearized.

2) *Future load balancing*

During lithography, a selected set of layers must be processed on the same machine for precision fabrication. For example, stacking layers A, B and C (a stacking group) must be processed on the same machine as shown in Figure 2.4. If Machine 1 is assigned to process layer A within the day, those lots will come to Machine 1 for layers B and C in the future.

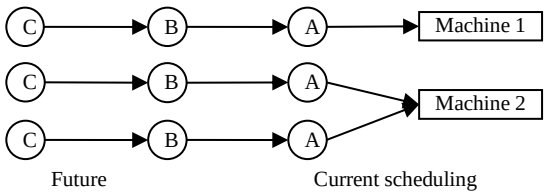


Figure 2.4. Stacking layers.

To avoid overload or starvation, the load on machines should be balanced. This is importance in view of the re-entrant nature and the presence of stacking layers. Of particular interest is to balance the future stacking layer load for a specific day, e.g., the day that the next layer will most likely to come back. Consider an example for a stacking group with layers A and B, and layer A is to be scheduled. The date that layer B will come back is probabilistic. This cycle time distribution can be obtained from historical date. Then the expected load for the day with the highest probability that layer B will come back can be calculated based on the current and past assignments. One of the performance measures is the load difference between a machine load and the average load for layer B. The above will be made specific next.

Let $WIP_{md}^{A\text{-}B}$ denote the number of lots whose layer A was processed d days ago on machine m, with the associated stacking layer B to be processed after a cycle time on the same machine. Let $P(CT_d^{A\text{-}B})$ denote the probability that the cycle time is d days from layer A to B. The total future load $LT_m^{A\text{-}B}$ with layer B to be processed in the Tth day on machine m can be calculated as follows:

$$LT_m^{A-B} = N_m^A \times P(CT_T^{A-B}) + \sum_d WIP_{md}^{A-B} \times P(CT_{d+T}^{A-B}). \tag{2.13}$$

In the above, N_m^A denotes the number of lots whose layer A is scheduled to be processed on machine m within the day, T denotes the cycle time from layer A to B with the largest probability. Then let $LA^{A\text{-}B}$

denote the average load of machines for layer B. Load difference $LD_m^{A\text{-}B}$ between total future load $LT_m^{A\text{-}B}$ and the average $LA^{A\text{-}B}$ on machine m can be described as follows:

$$LD_m^{A-B} = LT_m^{A-B} - LA^{A-B}. \tag{2.14}$$

In the above, $LA^{A\text{-}B}$ is approximated based on WIP and the target of layer A, it therefore contains no decisions. Summation of the load difference is minimized as a part of the objective function.

3) *Reticle expiration*

Reticles need to be recalibrated after processing a certain number of lots. Reticle remaining lifetime is used here to measure how many lots one reticle can process before next recalibration. Ideally the expiration dates of reticles in the same group should be equally spaced with the same time interval to avoid simultaneous reticle expirations. The remaining lifetime difference between two reticles can be adjusted as the expected time interval selected based on heuristic rules within every reticle group through proper reticle assignments. Let R_r^0 denote the remaining lifetime of reticle r before scheduling, representing how many lots reticle r can process before next recalibration. Similarly, let R_r denote the remaining lifetime of reticle r after scheduling. Their relationship can be easily obtained from

$$R_r = R_r^0 - \sum_m N_{mr}. \tag{2.15}$$

If the remaining lifetime difference between two reticles is larger than the expected interval after scheduling, it is expressed as certain reward in the objective function; if not, it is expressed as corresponding penalty. This term is linearized similarly to the first term. Since only the difference of two reticles that have the closest remaining lifetimes is reasonable and useful, a sequence will be established for every reticle group based on their remaining lifetimes. Each reticle will be assigned a ranking number, the smaller the number, the longer the remaining lifetime. An important assumption here is that the rank does not change before and after scheduling.

4) *Summary*

The modeling of the number of resource setups has already been discussed in subsection D, and the total number of setups is considered in the objective function.

In sum, the objective function with the above four terms to be minimized is described as follows:

$$\sum_{p}\sum_{l}\left(-W_{pl}^{P}\times\min(N_{pl}-T_{pl},0)-W_{pl}^{R}\times\max(N_{pl}-T_{pl},0)\right)$$
$$+W^{L}\times\sum_{g\in S^{SG}}\sum_{m\in S_{g}^{MS}}\left|LD_{gm}\right|+\sum_{p}\sum_{l}\sum_{r_1\in S_{pl}^{R},r_2\in S_{pl}^{R},r_1\neq r_2,r_2.No-r_1.No=1}$$
$$\left(-W^{RP}\times\min(R_{r_1}-R_{r_2}-G_{pl},0)-W^{RR}\times\max(R_{r_1}-R_{r_2}-G_{pl},0)\right)+W^{R}\times\sum_{m}\sum_{r}\sum_{k}y_{mr}(k). \quad (2.16)$$

In the second term, W^L denotes the weight for future stacking layer load balancing, S^{SG} denotes the set of stacking groups, and S_g^{MS} denotes the set of machines in stacking group g. In the third term, G_{pl} denote the expected expiration interval for reticles that process layer l of product p, and W^{RR} and W^{RP} denote the reward and penalty weights. In the last term, W^R denote the weight for avoiding excessive resource setups. The absolute values can be linearized similarly to (2.12). The above formulation is linear, and *max* and *min* are kept here for simplicity. The problem formulated is believed to be NP hard.

2.4 Solution Methodology

The problem is solved by using the branch-and-cut method by exploiting problem linearity after simplification as presented in subsubsection 2.4.1. Near-optimal solutions for practical problems, however, are still difficult to obtain efficiently as compared with the required time. The reason is that the convex hull of the problem is hard to delineate as explained with a small example in subsubsection 2.4.2. To improve computation efficiency, a two-phase approach is therefore developed in subsubsection 2.4.3. The convex hull of the first phase is analyzed with the same small example in subsubsection 2.4.4.

2.4.1 Branch-and-cut method

The problem is solved by using the branch-and-cut method. Mixed-integer linear programming problems are usually difficult to solve because a set of decision variables are restricted to integer values. Branch-and-cut is powerful for certain classes of mixed-integer linear optimization problems, and is easy to code by using commercial solvers. In the method, the integrality-relaxed problem is solved first by using a linear programming method. If all integer decision variables are integers, the solution is optimal to the original problem. If not, valid cuts that do not cut off any feasible integer solutions are added trying to obtain the convex hull (the smallest convex set that contains all feasible integer solutions in the Euclidean space). The idea is that once the convex hull is obtained, all integer decision variables of the linear programming solution are integers and optimal to the original problem. The process of obtaining the convex hull, however, is problem dependent, and can itself be NP hard. Low-efficient branching operations may then be needed on the variables whose values in the optimal relaxed solution violate their integrality requirements. The objective value of current optimal relaxed solution is a lower bound, and can be used to quantify the quality of a feasible solution. The optimization stops when CPU time reaches the pre-set stop time or the relative gap falls below the pre-set stop gap (CPLEX User's Manual).

For the problem formulated above, although it is linear, the convex hull is still difficult to obtain. Processing time requirements with multiple decision variables ($\delta_{mr}(k)$, N_{mr}, $T_{mr}{}^{N}$) might increase the difficulty of obtaining the convex hull because of complicating interactions among decisions. To overcome this, $T_{mr}{}^{N}$ is removed, and two sets of decisions are left. Then by relaxing the integrality requirements on N_{mr}, the processing time requirements (2.5) are modified as follows:

$$N_{mr} = \left(\sum_{k} \delta_{mr}(k) - T_{mr}^{P} \right) \div T_{mr}, \forall r, \forall m. \tag{2.17}$$

In the above, N_{mr} is not a decision variable and may not be an integer. However, the integer part of N_{mr} still represents the number of lots scheduled on machine m and reticle r within the day, and the

remaining fractional part can be used to derive T_{mr}^{N}. Optimality is therefore not affected. Based on the scheduling results, N_{mr} might need to be adjusted manually.

2.4.2 Convex hull analyses for the one-phase model

After the simplification above, near-optimal solutions for practical problems, however, are still difficult to obtain efficiently as will be shown in Section V. To overcome this difficulty, the convex hull is analyzed here. The problem now has three major sets of constraints: resource capacity, maximal number of lots scheduled and setup-related constraints since the processing time requirements were simplified as expressions (2.18). With $\delta_{mr}(k)$ represented by x_j ($1 \le j \le n = M \times R \times K$), resource capacity constraints (2.1) to (2.4) can be expressed as $\sum_j a_j x_i \le a_0$ where a_0 and a_j are positive integers and $x_j = 0$ or 1. Each of these constraints is a facet of the convex hull based on the proof in Balas (1975) and Balas and Zemel (1978). Through detailed analyses, it is discovered that the difficulty of obtaining the convex hull is caused by the setup-related constraints. To demonstrate this, convex hulls of the problems without and with these setup-related constraints are analyzed and compared through a simple example in this subsubsection.

Consider a simple example with two machines and two reticles. The first two sets of constraints mentioned above are considered in the first problem, and all of the three sets are considered in the second problem. Through the analyses and comparison of the two convex hulls, it is discovered that if the polyhedron formulated by all the constraints is simple and the convex hull can be easily obtained from the polyhedron, the problem can be efficiently solved; otherwise, low-efficient branching operations are needed. For visualization purpose, certain decision variables are fixed to present this intuitively.

In this example, two layers (l_1 and l_2) are to be scheduled on two machines (m_1 and m_2) with two reticles (r_1 for l_2 and r_2 for l_1). For simplicity, the total number of time slots is two and the processing time of both layers is one time slot. The targets are two lots for the first layer and one lot for the second layer.

In the first problem, the objective function is to meet targets. The problem is solved by using branch-and-cut. It is hard to visualize the convex hull and the optimal solutions because of the high dimensionality. For visualization purpose, δ_{m2r1} (1) and δ_{m2r1} (2) are selected to plot 2-D Figure 2.5 below with other decision variables fixed at their values in the optimal solution. An optimal solution is A (0, 1) (or B), which can be directly obtained from the convex hull ABC.

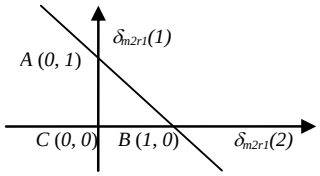


Figure 2.5. The feasible region and convex hull of the problem without setup-related constraints.

Since the future stacking layer load balancing and reticle expiration related constraints are only shown in the objective function, practical problems without setup-related constraints and term in the objective function can be efficiently solved.

In the second problem with setup-related constraints (2.12), the objective function is to meet targets and avoid excessive resource setups. The problem is also solved by using branch-and-cut, and δ_{m2r1} (k) and y_{m2r1} are selected to plot 3-D Figure 2.6 below. After relaxing the integrality requirements, all decision variables can take any value within [0, 1], and the optimal relaxed solution is D (0.5, 0.5, 0). All constraints formulate this polyhedron ACBD, and the convex hull ABC cannot be obtained by adding cuts on the feasible region. This difficulty is caused by the interactions among decisions in the setup-related constraints. It can be seen that two of the three values in the optimal relaxed solution are non-integers. To get the optimal solution A (1, 0, 1) (or C), low-efficient branching operations need to be performed on the first two variables of the relaxed optimal solution.

Generally for a problem with M machines, R reticles, and K time slots, the total number of setup-related constraints is $2MR(K\text{-}1)$. All constraints formulate a complicating polyhedron, and the convex hull

is difficult to obtain. Because of these complicating setup-related constraints, near-optimal solutions cannot be efficiently obtained.

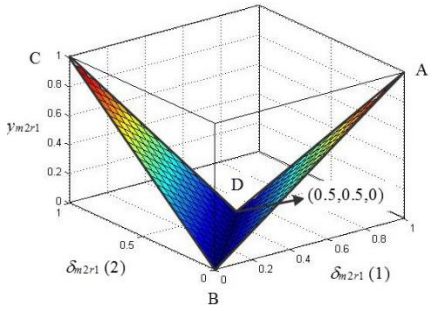


Figure 2.6. The feasible region and convex hull of the problem with setup-related constraints.

2.4.3 A two-phase approach

Based on the above analyses, a two-phase approach is developed to improve computation efficiency. In the first phase, a simplified problem without the complicating constraints is efficiently solved to establish ranges of decision variables. The problem with the full set of constraints is then solved in the second phase with a much reduced decision space.

In the first phase, the setup-related constraints are removed. After taking the setup-related term out of the objective function, many reticles might be assigned to a machine, leading to many excessive resource setups. To avoid this, the total number of reticles assigned is minimized as a part of the objective function. To check whether reticle r is assigned to machine m within the day, a new set of binary variables $\{d_{mr}\}$ is used:

$$d_{mr} \le N_{mr} \le N_{mr}^{UB} \times d_{mr}, \forall m, \forall r. \tag{2.18}$$

In the above equation, if $d_{mr} = 1$, reticle r is assigned to machine m; if $d_{mr} = 0$, otherwise. The last term in the objective function (16) needs to be revised correspondingly. The main decision variables are still $\delta_{mr}(k)$. Since there could be many excessive setups, $\delta_{mr}(k)$ cannot tell when the layers should be processed. Dependent variables N_{mr} are used as schedules instead.

The problem formulated above is solved by using branch-and-cut by exploiting linearity to reduce ranges of decision variables, and the solutions can be obtained fast as will be shown in Subsection 2.5.

In the second phase, the ranges of decision variables are reduced by either fixing certain variables or by restricting the ranges of others based on first phase results. For example, assignments of machines and reticles to layers are fixed. In doing so, optimality might be affected as will be discussed in Subsection 2.5. As another example, the range of the number of lots scheduled on a machine with a certain reticle is restricted. The magnitude of a range is selected based on testing results as a trade-off between solution quality and computation efficiency. If this range is large enough, optimality will not be affected.

The objective function is the same as that of the one-phase approach. The decisions are $\delta_{mr}(k)$ (only for machines and reticles assigned in the first phase), which mean when the layer should be processed.

The problem formulated above is solved by using branch-and-cut. During branching operations, computation time is reduced in a major way since the ranges of decision variables are much reduced. The problem can therefore be solved faster than the one-phase approach as will be shown in Subsection 2.5.

2.4.4 Convex hull analyses for the two-phase model

To compare with the one-phase approach, the same simple example is analyzed. In the first phase, the problem is formulated with resource capacity (2.1) and (2.2), maximal number of lots scheduled (2.10) and (2.11), and reticle assigned detection constraints (2.18). The objective is to meet targets and reduce the number of reticles assigned. It is discovered that the number of non-integer values (= 1) in the optimal relaxed solution is smaller than that (= 2) of the one-phase approach, implying fewer branching operations and faster termination. In the second phase, the problem is formulated with resource capacity (2.1) and (2.2), setup-related (2.12), and variable range restriction constraints. The objective is to meet targets and avoid excessive resource setups. This problem can be efficiently solved because the decision space (= 6) is smaller than that (= 12) of the one-phase approach.

The problem is solved by using branch-and-cut. To compare with the one-phase approach, $\delta_{m2r1}(k)$ and d_{m2r1} are selected to plot 3-D Figure 2.7 below. For the relaxed problem, all feasible solutions are in the polyhedron ABCDE, and an optimal relaxed solution is B (1, 0, 0.5) (or D). The convex hull ACE cannot be obtained by adding cuts on the feasible region, and an optimal solution is A (1, 0, 1) (or E). It can be seen only d_{m2r1} is a non-integer in the optimal relaxed solution and branching operation needs to be performed only on one variable.

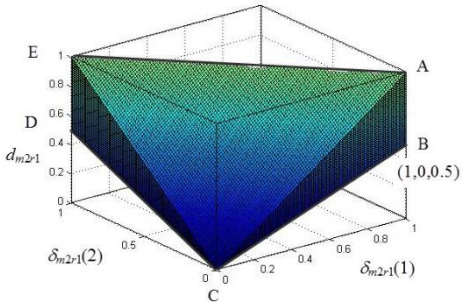


Figure 2.7. The feasible region and convex hull for the first phase of the two-phase model.

In the second phase, based on the results from the first phase, the number of decisions is 6 ($\delta_{m2r1}(k)$, $\delta_{m1r2}(k)$, y_{m2r1} and y_{m1r2}) as compared with 12 in the one-phase model. The problem can be therefore solved faster than the one-phase approach.

For the problem with M machines, R reticles, and K time slots mentioned in subsection B, the number of reticle assigned related constraints (18) in the first phase is only MR as compared with $2MR(K-1)$ (12) in the one-phase model. The polyhedron formulated by constraints in the first phase is simpler than that of the one-phase model. The total number of branching operations needed is smaller, and the problem can be efficiently solved to establish ranges of decision variables. In the second phase, the problem can be efficiently solved with a much reduced decision space. To quantify the solution quality, lower bounds should be obtained from the one-phase approach since the objective values of optimal relaxed solutions of the two-phase approach might not be lower bounds to the original problem. For this simple example, the final gap turns out to be zero, implying that the optimal solution has been obtained.

2.5 Numerical Results

The methods presented above have been implemented by using the optimization package IBM ILOG CPLEX Optimization Studio V 12.2. Testing has been performed on a PC with 1.60GHz Intel (R) i7 CPU and 4G RAM, and three examples are presented. The first small example is to demonstrate schedules of small problems obtained by using the one-phase approach can be duplicated and obtained by hands. The second medium-sized example is to compare the statistics of one-phase and two-phase approaches. The third practical example is to compare solution quality and computation efficiency of these two approaches, demonstrating that the two-phase approach can generate near-optimal schedules within much reduced computation time.

2.5.1 Example 1: Testing of the one-phase approach with a small problem

This small example is to demonstrate schedules of small problems obtained by using the one-phase approach can be duplicated and obtained by hands. In this example, three layers of one product are to be scheduled on three machines with five reticles in 102 time slots. The information is as follows:

Layers: L_A and L_B (stacking group), and L_C;

Machines: M_1, M_2 and M_3 with total future stacking layer load of 6983, 4490 and 2044 respectively;

Reticles: R_A, R_B, R_{C1}, R_{C2}, and R_{C3} with remaining lifetime of 89572, 88308, 8442, 79059 and 58732 respectively;

Target: 600 for L_A, 650 for L_B and 600 for L_C.

The problem is solved in 3 seconds with a relative gap of 5% by using the one-phase approach, and the Gantt chart is shown in Figure 2.8 below.

The schedule meets all targets of three layers without excessive resource setups. It can be seen that layer L_A is all scheduled on Machine M_3 since it has lowest total load before scheduling, and it also shows

the stacking layer load will be balanced in the next few days. In addition, the remaining lifetime difference between R_{C1} and R_{C2} becomes larger as shown in Figure 2.9, which avoids simultaneous expirations of the three reticles that process layer L_C. The above schedule can be duplicated and obtained by hand, and also shows that our method can satisfy the objective of this litho machine scheduling problem.

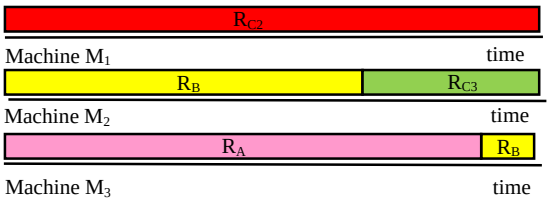


Figure 2.8. Gantt chart of schedule results.

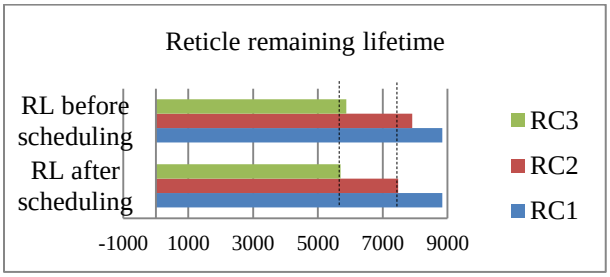


Figure 2.9. Reticle remaining lifetime results.

2.5.2 Example 2: Testing of the one-phase and two-phase approaches with a medium-sized problem

This medium-sized example is to compare the statistics of one-phase and two-phase approaches. In this example, four layers of one product are to be scheduled on five machines with ten reticles in 200 time slots. For simplicity, the objective function is to meet targets and avoid excessive resource setups in the one-phase model and in the second phase of the two-phase model; to meet targets and reduce the number of reticles assigned in the first phase of the two-phase model.

The testing results are shown in the following Table 2.1.

From the results, it can be seen that the number of variables in the first phase of the two-phase model is around 44% of that in the one-phase model, and the number of constraints is about 14%. In the second phase, the numbers of variables and constraints are about 10% of those in the one-phase model. In addition,

the time on branching is 4.5 seconds and 2.6 seconds in the first and second phases of the two-phase model, as compared with 87.4 seconds in the one-phase model. The total CPU time on pre-processing, cutting and branching required by the two-phase approach is 15 seconds as compared with 123 seconds required by the one-phase approach. Although the final gap is about 2% higher than that of the one-phase approach, the computation efficiency is much improved. Both of the scheduling results meet the targets, and there are five and six setups in the one-phase and two-phase results, respectively.

Table 2.1. Testing results of Example 2

	One-phase model	The first phase of two-phase model	The second phase of two-phase model
CPU time	123.03 s	10.02 s	4.77 s
Relative gap	4.32%	0.45%	0.47%
Number of variables	19,951	10,051	2,031
Number of constraints	22,904	3,104	2,352
Objective value	837	852	
Final relative gap	4.32%	6.19%	
Node processed	16	31	10
Root relaxation time	3.12 s	0.14 s	0.11 s
Number of cuts	1	241	368
Time on cuts	29.5 s (24%)	4 s (40%)	0.3 s (6%)
Time on branching	87.4 s (71%)	4.5 s (45%)	2.6 s (55%)

2.5.3 Example 3: Testing of the one-phase and two-phase approaches with a practical problem

This practical problem is to compare solution quality and computation efficiency of one-phase and two-phase approaches. In this example, seven layers of one product are to be scheduled on 11 machines with 71 reticles in 411 time slots (one day). The objective function of the second phase here is to increase the

number of lots processed, finish targets as soon as possible, and reduce the number of reticles assigned, and reduce the number of resource setups. For simplicity of implementation, future stacking layer load balancing and reticle expiration terms are removed as they have already been mostly satisfied in the first phase. Because of this, the final cost is obtained by plugging the solution into the original objective function (2.16).

The testing results are shown in the following Table 2.2.

Table 2.2. Testing Results of Example 3

	One-phase model	The first phase of two-phase model	The second phase of two-phase model
Stop time	300 s	120 s	240 s
Stop gap	5 %	0.1%	0.5%
CPU time	279 s	25 s	2 s
Relative gap	4.3%	0.1%	0.1%
Objective value	86362.5	87281.2	
Final relative gap	4.3%	5.4%	

This practical problem is solved in 5 minutes with a gap of 4.3% by using the one-phase approach. This computation time is still long as compared to the required time, 2 minutes. By using the two-phase approach, the same problem is solved within 30 seconds. Although the final gap is 5.4% and 1.1 % higher than that of the one-phase approach, it is still acceptable. The schedule obtained from the two-phase approach meet all targets without excessive resource setups except for one layer, because there are no available reticles for this layer. In terms of future stacking layer load balancing, the total load among machines from stacking layer A to C before and after scheduling is compared in Figure 2.10 below. It can be seen that the future load is nearly balanced through scheduling. In addition, the remaining lifetime difference between two reticles that have the closest remaining lifetimes in the same group is moving toward the expected interval after scheduling, which avoids simultaneous reticle expirations.

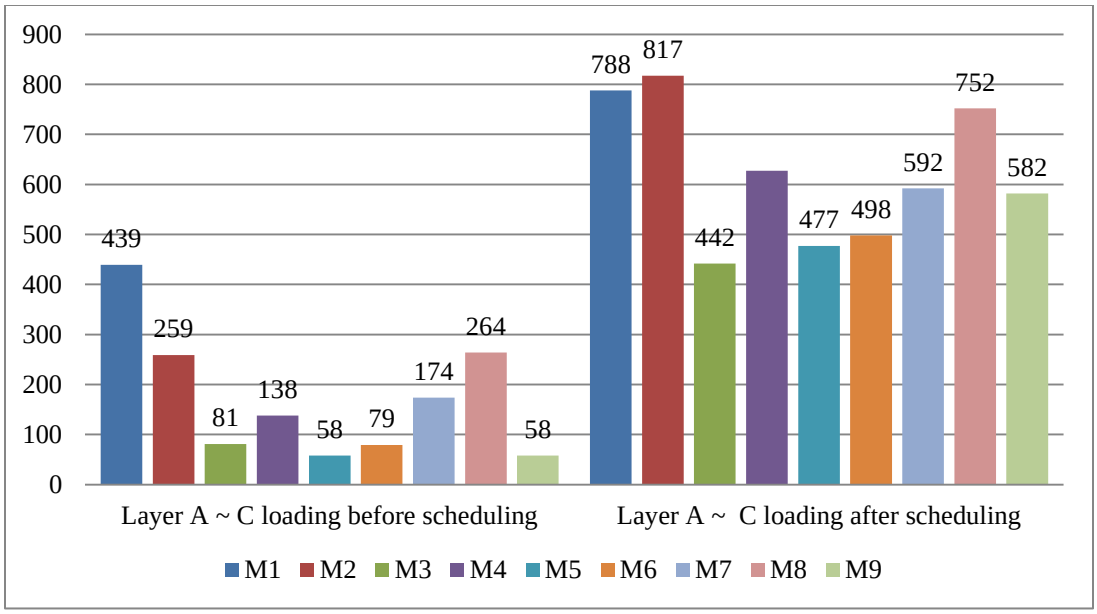


Figure 2.10. Future stacking layer load results.

2.6 Conclusion

In this chapter, a novel mathematical formulation for litho machine scheduling over a day with resource setups, reticle expirations and future stacking layer load balancing is established. The problem is solved by using branch-and-cut by exploiting problem linearity. To improve computation efficiency, a two-phase approach is developed. In the first phase, a simplified problem with certain complicating constraints dropped is efficiently solved to establish ranges of decision variables. The problem with the full set of constraints is then solved in the second phase with a much reduced decision space. Numerical testing shows that the two-phase approach generates near-optimal schedules within reasonable amounts of computation time.

With minor changes in the formulation, our method is also used for real-time rescheduling every ten minutes. Furthermore, this two-phase approach is generic, and will have major implications on other semiconductor scheduling problems and beyond.

References

[1] E. Akcali, K. Nemoto, and R. Uzsoy, "Cycle-Time improvements for photolithography process in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 14, No. 1, pp. 48-56, 2001.

[2] E. Akcali and R. Uzsoy, "A sequential solution methodology for capacity allocation and lot scheduling problems for photolithography," *Proceedings of 26th IEEE/CPMT International, Electronics Manufacturing Technology Symposium*, pp. 374 – 381, Santa Clara, CA, USA, 2000.

[3] E. Akcali, A. Üngör, and R. Uzsoy, "Short-term capacity allocation problem with tool and setup constraints," *Naval Research Logistics*, Vol. 52, No. 8, pp. 754–764, 2005.

[4] A. Arisha and P. Young, "Intelligent simulation-based lot scheduling of photolithography toolsets in a wafer fabrication facility," *Proceedings of 36th Conference on Winter Simulation*, pp. 1935-1942, 2004.

[5] E. Balas, "Facets of the knapsack polytope," *Mathematical Programming*, Vol. 8, No. 1, pp. 146-164, 1975.

[6] E. Balas and E. Zemel, "Facets of the knapsack polytope from minimal covers," *SIAM Journal on Applied Mathematics*, Vol. 34, No. 1, pp. 119-148, 1978.

[7] F. Barahona, S. Bermon, O. Günlük, and S. Hood, "Robust capacity planning in semiconductor manufacturing," *Naval Research Logistics*, Vol.52, No. 5, pp. 459-468, 2005.

[8] E. M. L. Beale and J. J. H. Forrest, "Global optimization using special ordered sets," *Mathematical Programming*, Vol. 10, No. 1, pp. 52-69, 1976.

[9] S. C. Chang and D. Y. Liao, "Scheduling flexible flow shops with no setup effects," *IEEE Transactions on Robotics and Automation*, Vol. 10, No. 2, pp.112-122, 1994.

[10] J. Chung and J. Jang, "A WIP balancing procedure for throughput maximization in semiconductor fabrication," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 22, No. 3, pp. 381-390, 2009.

[11] IBM, ILOG CPLEX 12.2 User's Manual, 2010.

[12] Y. H. Lee, J. Park, and S. Kim, "Experimental study on input and bottleneck scheduling for a semiconductor fabrication line," *IIE Transactions*, Vol. 34, No. 2, pp. 179–190, 2002.

[13] D. Y. Liao, S. C. Chang, K. W. Pei, and C. M. Chang, "Daily scheduling for R&D semiconductor fabrication," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 9, No. 4, pp. 550-561, 1996.

[14] L. Mönch, M. Prause, and V. Schmalfuss, "Simulation-based solution of load-balancing problems in the photolithography area of a semiconductor wafer fabrication facility," *Proceedings of 33rd Conference on Winter Simulation*, pp. 1170-1177, 2001.

[15] S. C. Sarin, L. Wang, and M. Cheng, "A single-machine, single-wafer- processing, multiple-lots-per-carrier scheduling problem to minimize the sum of lot completion times," *Computers & Operations Research*, Vol. 39, No. 7, pp. 1411-1418, 2012.

[16] A. M. D. Shr, A. Liu, and P. P. Chen, "Load balancing among photolithography machines in the semiconductor manufacturing system," *Journal of Information Science and Engineering*, Vol. 24, No. 2, pp. 379-391, 2008.

[17] L. B. Toktay and R. Uzsoy, “A capacity allocation problem with integer side constraints,” *European Journal of Operational Research*, Vol. 109, No. 1, pp. 170-182, 1998.

[18] F. D. Vargas-Villamil and D. E. Rivera, “Multilayer optimization and scheduling using model predictive control: application to reentrant semiconductor manufacturing lines,” *Computers and Chemical Engineering*, Vol. 24, No. 8, pp. 2009-2021, 2000.

[19] J. Wang and P. B. Luh, “Scheduling of a machining center,” *Mathematical and Computer Modeling on Recent Advances in Discrete Event Systems*, Vol. 23, No. 11-12, pp. 203-214, 1996.

[20] C. Yugma, R. Riffart, S. Dauzere-Peres, P. Vialletelle, and F. Buttin, “A dispatcher simulator for a photolithography workshop,” *Proceedings of IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, Stresa, Italy, 2007, pp. 100-104.

Chapter 3

Exergy-Based Operation Optimization of a Distributed Energy System through the Energy-Supply Chain

Developing sustainable energy systems is crucial in today's world because of the depletion of fossil energy resources and global warming problems. Application of exergy principles in the context of energy supply systems may achieve efficient energy-supply chains and rational use of energy in buildings. This chapter presents an exergy-based operation optimization of a distributed energy system by considering the whole energy-supply chain from energy resources to user demands. The problem is challenging in view of the complicated interactions of devices and the modeling of exergy losses. To capture these complicated interactions, energy networks are established with exergy losses modeled at the energy conversion step, which accounts for the largest part of the total exergy loss in the whole energy-supply chain. A multi-objective mixed integer programming problem is formulated. The problem is efficiently solved by the novel integration of surrogate Lagrangian relaxation and branch-and-cut. The Pareto frontier, including the best possible trade-offs between the economic and exergetic objectives, is obtained by minimizing a weighted sum of the total energy cost and total exergy loss occurring at the energy conversion step. Results demonstrate that the use of high-quality energy resources is reduced by the reduction of exergy losses, leading to sustainability of energy supply systems.

Nomenclature

A area (m^2)

c	constant in Eq. (3.41) (kWh/$)
c_p	specific heat of water (kJ/kg/K)
$Cost$	total energy cost ($)
d	stepsize in Eq. (3.47)
DR	maximum ramp-down rate (kW)
ex_{NG}	specifical chemical exergy of natural gas (kWh/Nm^3)
$\dot{E}$	electricity rate (kW)
$\dot{E}x$	exergy rate (kW)
$\dot{E}xloss$	exergy loss rate (kW)
$Exloss_{conv}$	total exergy loss at energy conversion step (kJ)
F_{obj}	objective function
F_q	Carnot factor
$\tilde{g}$	surrogate subgradient vectors
$\dot{G}$	natural gas volumetric flow rate (Nm^3/h)
$\dot{H}$	heating rate (kW)
$\dot{I}_T$	total solar irradiance (kW/m^2)
L	Lagrangian function
$\tilde{L}$	surrogate dual value
LHV_{NG}	natural gas lower heat value (kWh/Nm^3)
$\dot{m}$	mass flow rate (kg/h)
m	mass (kg)
k	iteration
P_{gas}	natural gas price ($/Nm^3)
P_{grid}	electricity price ($/kWh)
q	dual function
$\dot{Q}_{GT}^{exgas}$	heat rate made available by the exhaust gas (kW)
R	energy generation rate (kW)
S	energy source input rate (kW)
t	time (h)
T	temperature (K)
UR	maximum ramp-up rate (kW)
x	binary decision variable
y	all the decision variables in Eq. (3.42)

Greek Symbols

Δt	length of the time interval (h)
ε_{gen}	exergy efficiency of electricity generation
ς	exergy factor
η	efficiency
λ	Lagrangian multipliers
μ_{GT}	percent heat loss rate of the gas turbine
ξ	gas turbine exhaust fraction
ω	weight in Eq. (3.41)

Superscript/Subscripts

0	reference
$boil$	boiler
BP	bypass

build	building
buy	bought
cold	cold
coll	collector
dem	demand
DHW	domestic hot water
e	electricity
ED	energy device
exgas	exhaust gas
NG	natural gas
grid	power grid
GT	gas turbine
hex	heat exchanger
HP	heat pump
HRB	heat recovery boiler
in	input
max	maximum
min	minimum
out	output
r	return
s	supply
SH	space heating
ST	solar thermal
sto	thermal storage

Acronyms

CHP	combined heat and power
DES	distributed energy system
DHW	domestic hot water
SH	space heating

3.1 Introduction

Developing sustainable energy systems is becoming more and more important in today's world because of the depletion of fossil energy resources and the related global warming problems. Therefore, high-quality energy carriers, such as fossil fuels and electricity, should be efficiently used [1]. Buildings are responsible for more than 40% of the total final energy consumption on a worldwide scale [2]. A significant share of this energy consumption is for Space Heating (SH) and Cooling, and Domestic Hot Water (DHW) demands. These are low-quality energy demands because of the associated temperatures required. However, thermal demands in buildings are commonly met by high-quality energy resources. There is a great potential in

energy-management of energy supply systems to attain efficient energy-supply chains and rational use of energy in buildings [3].

Current analyses and optimization methods for energy-management of energy supply systems do not distinguish different qualities of energy flows. In thermodynamics, the quality of an energy carrier is measured by exergy. Exergy is defined as the maximum theoretical work that can be obtained from an energy flow, as it comes to the equilibrium with the reference environment [1, 3 - 7]. The concept of exergy was introduced in building efficiency studies by international research projects, such as ECBCS Annex 37 [5], and Annex 49 [1]. Several studies on the exergy analysis of energy supply systems for the building environment are also found in recent years [6, 8 - 10].

A Distributed Energy System (DES) is an energy system where energy is made available close to energy end-users [11]. DESs provide a unique opportunity to show the benefits of the exergy analysis for preserving high-quality energy resources, since several energy devices convert a set of primary energy carriers (e.g., electricity, solar energy, natural gas) with different energy quality levels to satisfy end-user demands with different quality levels. In terms of DESs, most of the studies in the literature are focused on the operation optimization of DESs to reduce energy costs [12 - 14], which is essential in the short run. The optimized operation strategies of a DES were obtained in a previous work [15] to reduce the total daily energy cost and increase the total exergy efficiency. For simplicity, the total exergy input to the DES and the total exergy output required to meet the energy demands were considered instead of the exergy input and output of each energy device in the energy-supply chain.

This chapter presents an exergy-based operation optimization of a DES through the energy-supply chain from energy resources to user demands (electricity, SH and DHW demands are considered), without neglecting the energy costs. The main goal is to obtain the optimized operation strategies of the DES to reduce the total energy costs and the total exergy loss occurring at the energy conversion step, which accounts for the largest part of the total exergy loss in the whole energy-supply chain. By reducing these

exergy losses, the use of high-quality energy resources can be reduced, leading to sustainability of supply energy systems.

The optimization problem is challenging since several energy devices convert a set of input energy carriers, such as natural gas, electricity, and solar energy, into output energy carriers, such as heat and electricity, with complicated interactions among them; the exergy of thermal energy is directly related to the temperature and the mass flow rate of the corresponding energy carrier; and the problem is nonlinear. To capture the complicated interactions, energy networks are established from energy resources to user demands, based on the physical structure of the energy- supply chain. Exergy losses are then modeled for the energy devices at the conversion step based on the networks to make visible where and how much exergy is lost. A multi-objective mixed-integer problem is formulated. The objective is to minimize a weighted sum of the total energy cost and exergy losses at the conversion step while satisfying given time-varying user demands. Surrogate Lagrangian relaxation and branch-and-cut are integrated in a novel way for a speedy and near-optimal performance. The Pareto frontier, consisting of the best possible trade-offs between the economic and exergetic objectives, is obtained. Results show that the use of high-quality energy resources can be reduced by the reduction of exergy losses, leading to sustainability of energy supply systems.

3.2 Problem Formulation

To match the solution methodology, the surrogate Lagrangian relaxation combined with branch-and-cut method, a separable and linear formulation is preferred to solve the problem efficiently. The energy-supply chain under consideration consists of energy conversion devices, including gas turbine and heat recovery boilers, as the Combined Heat and Power (CHP) system, solar thermal plant, auxiliary natural gas boilers, and heat pump; thermal energy storages, distribution devices (e.g., water pipes) as well as terminal devices (e.g., fan coils for SH) are also considered as shown in Figure 3.1. Electricity is 100% exergy (fully convertible into useful work), while the exergy of thermal energy is directly related to the

temperature and mass flow rate of the corresponding energy carrier. The water networks for space heating and domestic hot water demands need to be established based on the physical structures of water pipes, valves, mixers, etc.

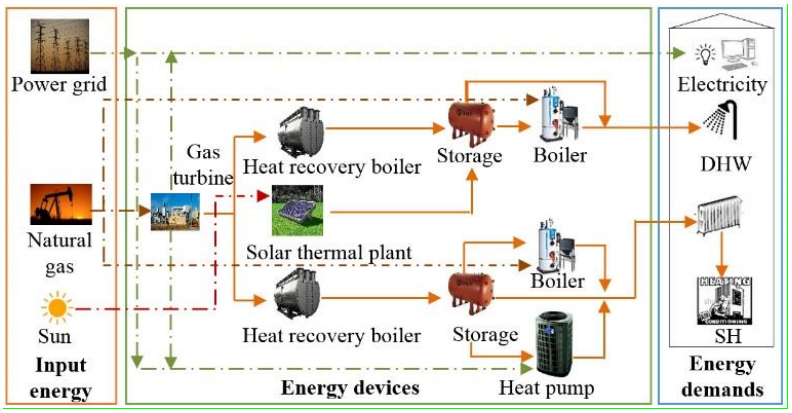


Figure 3.1. Scheme of the energy-supply chain.

The general structure of the energy and exergy modeling and common constraints of energy devices are first described as follows.

Capacity constraints. The energy generation rate (e.g., electricity and heat) of the device, $R_{ED}(t)$, should be within its minimum and maximum values if it is on ($x_{ED}(t) = 1$):

$$x_{ED}(t)R_{ED}^{\min} \le R_{ED}(t) \le x_{ED}(t)R_{ED}^{\max}. \tag{3.1}$$

Ramp rate constraints. The variations in energy generation rates between two successive time intervals should be within the ramp-down, DR_{ED}, and ramp-up, UR_{ED}:

$$-DR_{ED} \le R_{ED}(t) - R_{ED}(t-\Delta t) \le UR_{ED}, \tag{3.2}$$

where Δt is the length of the time interval.

Energy consumption. The input rate of the energy source, $S_{ED}^{in}(t)$, required by the device to provide the output energy is:

$$S_{ED}^{inp}(t) = R_{ED}(t) / \eta_{ED}, \tag{3.3}$$

where η_{ED} is the conversion efficiency of the device.

Exergy loss. The input exergy rate and output exergy rate of the device depend on the type of the energy carrier, and the exergy loss rate will be formulated as the difference of them.

Modeling of the electricity network, water network for space heating, and water network for domestic hot water is presented in Subsubsections 3.2.1, 3.2.2, and 3.2.3, respectively. Objective functions and multi-objective optimization are discussed in Subsubsection 2.4.

3.2.1 Modeling of electricity network

Since electricity is 100% exergy, the exergetic modeling of the electricity network is mainly the energy modeling.

3.2.1.1. Modeling of the gas turbine in the CHP system. The CHP system may consist of multiple gas turbines for electricity and the corresponding heat recovery boilers, using high-temperature exhaust gas to satisfy demands of space heating and domestic hot water [16]. For simplicity, one gas turbine is considered here. The modeling of multiple gas turbines is similar, and the problem complexity may increase as the number of gas turbines increases. Constraints considered for the gas turbine are presented below.

The volumetric flow rate of natural gas, $\dot{G}_{GT}(t)$, required by the gas turbine to provide the electricity rate, $\dot{E}_{GT}(t)$, is:

$$\dot{G}_{GT}(t) = \dot{E}_{GT}(t) / \left(\eta_e LHV_{NG}\right), \tag{3.4}$$

where η_e is the turbine gas-to-electric efficiency and LHV_{NG} is the lower heat value of natural gas.

The heat rate of the exhaust gas recovered from the gas turbine, $\dot{Q}_{GT}^{exgas}(t)$, is:

$$\dot{Q}_{GT}^{exgas}(t) = \dot{E}_{GT}(t)\left(1-\eta_e-\mu_{GT}\right)/\eta_e, \tag{3.5}$$

where μ_{GT} is the fraction of heat lost in the gas turbine.

For the gas turbine, the input energy carrier is natural gas. The specific chemical exergy of natural gas is the maximum work that can be obtained from the substance, by taking it to the chemical equilibrium with the reference environment at the constant temperature and pressure [17]. The exergy input rate of natural gas to the gas turbine, $\dot{E}x_{GT}^{NG}(t)$, is the gas volumetric flow rate consumed, $\dot{G}_{GT}(t)$, multiplied by the specific chemical exergy of natural gas, ex_{NG}:

$$\dot{E}x_{GT}^{NG}(t) = ex_{NG}\dot{G}_{GT}(t). \tag{3.6}$$

The specific chemical exergy of natural gas, ex_{NG}, can be evaluated by multiplying the exergy factor, ς_{NG}, and the lower heat value, LHV_{gas}:

$$ex_{NG} = \varsigma_{NG} LHV_{NG}. \tag{3.7}$$

According to [17], the exergy factor for natural gas is equal to 1.04 ± 0.5%.

The electricity provided by the gas turbine is 100% exergy, and the exergy rate of the output electricity is:

$$\dot{E}x_{GT}^{e}(t) = \dot{E}_{GT}(t). \tag{3.8}$$

The exergy rate of the output exhaust gas, $\dot{E}x_{GT}^{exgas}(t)$, is calculated by multiplying the energy rate by the related Carnot factor, since the temperature of the exhaust gas is assumed to be constant [1],

$$\dot{E}x_{GT}^{exgas}(t) = \dot{Q}_{GT}^{exgas}(t)F_q(t), \tag{3.9}$$

with the Carnot factor, $F_q(t)$, expressed as,

$$F_q(t) = 1 - T_0(t)/T_{exgas}^{s}, \tag{3.10}$$

which depends on both the temperature of exhaust gas, T^{s}_{exgas}, and the reference temperature $T_0(t)$. By following the dynamic exergy analysis, hourly ambient temperatures are considered as reference temperatures [18].

The total exergy loss rate in the gas turbine is:

$$\dot{E}xloss_{GT}(t) = \dot{E}x^{NG}_{GT}(t) - \dot{E}_{GT}(t) - \dot{E}x^{exgas}_{GT}(t). \tag{3.11}$$

3.2.1.2. Meeting demand. The electricity rate demand, $\dot{E}_{dem}(t)$, and the electricity rate required by the heat pump, $\dot{E}_{HP}(t)$, must be covered by the sum of the electricity rate delivered by the gas turbine, $\dot{E}_{GT}(t)$, and the electricity rate from the grid, $\dot{E}_{buy}(t)$:

$$\dot{E}_{dem}(t) + \dot{E}_{HP}(t) = \dot{E}_{GT}(t) + \dot{E}_{buy}(t). \tag{3.12}$$

In order to consider all the exergy losses at the energy conversion step, the exergy losses occurring in the power generation plants are also included. The exergy efficiency of power generation plants, ε_{gen}, is based on the technologies used in the plants, and the exergy loss rate is [19]:

$$\dot{E}xloss_{grid}(t) = \dot{E}_{buy}(t) / \varepsilon_{gen} - \dot{E}_{buy}(t). \tag{3.13}$$

3.2.2 Modeling of water network for space heating

A typical water network for space heating is shown in Figure 3.2. The exhaust heat recovered in the heat recovery boiler is stored through the heat exchanger in a large water tank, which is used to supply hot water to buildings with a constant mass flow rate. A fully-mixed tank model is assumed for simplicity, where the water in the tank has a uniform time-varying temperature, because of the charge and discharge processes with a given efficiency. As to the water temperature in the tank, there are two cases. If the temperature is higher than the required (assumed constant), the water is directly supplied to the buildings and part of the water is mixed with the return water from buildings in the mixer. After mixing, the temperature of the mixed

water is brought to the required one, and then the water is sent to the terminal devices in buildings. In the second case, if the temperature of the water tank is lower than the requirement, the water is sent to the auxiliary natural gas boiler or to the heat pump, and heated to the required temperature.

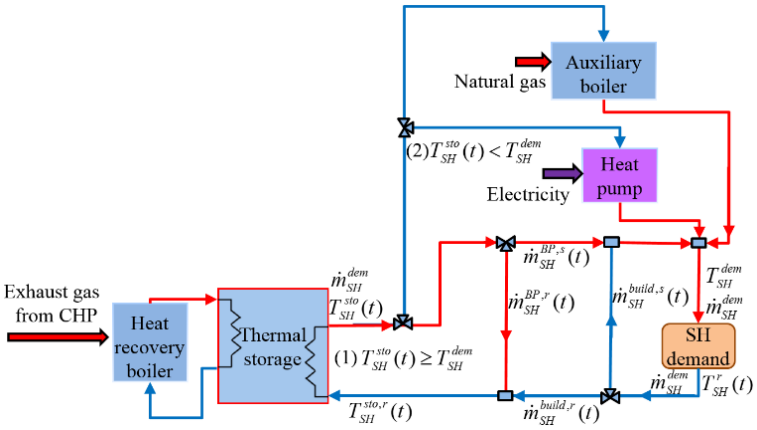


Figure 3.2. Scheme of the water network for space heating.

3.2.2.1. Modeling of the heat recovery boiler. Heat is recovered from the high-temperature exhaust gas in the heat recovery boiler. For simplicity, two heat recovery boilers are considered, for space heating and domestic hot water, respectively. The modeling of more heat recovery boilers is similar, and the configuration of the water network may become more complicated as the number of heat recovery boilers increases. The exhaust gas from the gas turbine is subdivided between the two heat recovery boilers. The sum of fractions of exhaust gas (continuous decision variable) for space heating, $\xi_{SH}(t)$, and domestic hot water, $\xi_{DHW}(t)$, has to be one:

$$\xi_{SH}(t) + \xi_{DHW}(t) = 1. \tag{3.14}$$

The heat rate delivered by the exhaust gas to the heat recovery boiler for space heating, $\dot{H}_{HRB,SH}(t)$, is:

$$\dot{H}_{HRB,SH}(t) = \xi_{SH}(t)\eta_{HRB}\dot{Q}_{GT}^{exgas}(t), \tag{3.15}$$

where η_{HRB} is the heat recovery efficiency of the boiler. The heat balance equation is:

$$\dot{H}_{HRB,SH}(t) = c_p \dot{m}_{HRB,SH}(t)\left(T^s_{HRB,SH} - T^r_{HRB,SH}(t)\right) = \eta_{hex} c_p \dot{m}_{HRB,SH}(t)\left(T^s_{HRB,SH} - T^{sto}_{SH}(t)\right), \quad (3.16)$$

where c_p is the specific heat of water; $\dot{m}_{HRB,SH}(t)$ is the water mass flow rate through the heat exchanger in the storage from the heat recovery boiler (decision variable); $T^s_{HRB,SH}$ and $T^r_{HRB,SH}(t)$ are the temperatures of the water flowing into and out of the heat exchanger, respectively; $T^{sto}_{SH}(t)$ is the temperature of the water in the tank; and η_{hex} is the efficiency of the heat exchanger. The supply temperature and heat exchanger efficiency are assumed known, and the return temperature is a dependent variable.

The exergy input rate to the heat recovery boiler is its fraction of exhaust gas multiplied by the exergy rate of exhaust gas. At the output, the exergy rate of the heat delivered by the heat recovery boiler, $\dot{E}x^{out}_{HRB,SH}(t)$, is related to the mass flow rate and supply and return temperatures:

$$\dot{E}x^{out}_{HRB,SH}(t) = c_p \dot{m}_{HRB,SH}(t)\left[\left(T^s_{HRB,SH} - T^r_{HRB,SH}(t)\right) - T_0(t)\ln\left(\frac{T^s_{HRB,SH}}{T^r_{HRB,SH}(t)}\right)\right]. \quad (3.17)$$

The exergy loss rate in the heat recovery boiler is,

$$\dot{E}xloss_{HRB,SH}(t) = \xi_{SH}(t)\dot{E}x^{exgas}_{GT}(t) - \dot{E}x^{out}_{HRB,SH}(t). \quad (3.18)$$

3.2.2.2. Modeling of the auxiliary natural gas boiler. The auxiliary natural gas boiler converts natural gas into heat for the space heating demand. The natural gas volumetric flow rate required by the boiler to provide the heat rate, $\dot{H}_{boil,SH}(t)$, is given by:

$$\dot{G}_{boil,SH}(t) = \dot{H}_{boil,SH}(t) / \left(\eta_{boil} LHV_{NG}\right), \quad (3.19)$$

where η_{boil} is the combustion efficiency of the boiler. The heat balance equation for the boiler is:

$$\dot{H}_{boil,SH}(t) = c_p \dot{m}^{dem}_{SH}\left(T^s_{boil,SH}(t) - T^r_{boil,SH}(t)\right), \quad (3.20)$$

where:

$$T^{s}_{boil,SH}(t) = T^{dem}_{SH} \quad \text{and} \quad T^{r}_{boil,SH}(t) = T^{sto}_{SH}(t), \tag{3.21}$$

where $T^{s}_{boil,SH}(t)$ and $T^{r}_{boil,SH}(t)$ are the temperatures of the water flowing out and into of the boiler, respectively.

The input energy carrier to the boiler is natural gas. Similarly to the gas turbine, the exergy input rate of natural gas to the boiler, $\dot{E}x^{NG}_{boil,SH}(t)$, is the natural gas volumetric flow rate consumed by the boiler multiplied by the specific chemical exergy of natural gas, ex_{NG}. At the output, the exergy rate of the heat delivered by the boiler, $\dot{E}x^{out}_{boil,SH}(t)$, is evaluated similarly to that of the heat recovery boiler, based on the mass flow rate and supply and return temperatures. The exergy loss rate in the natural gas boiler is:

$$\dot{E}xloss_{boil,SH}(t) = \dot{E}x^{NG}_{boil,SH}(t) - \dot{E}x^{out}_{boil,SH}(t). \tag{3.22}$$

3.2.2.3. Modeling of the heat pump. The heat pump converts electricity into heat for the space heating demand. The electricity consumption, $\dot{E}_{HP}(t)$, of the heat pump to provide the heating rate, $\dot{H}_{HP}(t)$, is:

$$\dot{E}_{HP}(t) = \dot{H}_{HP}(t) / COP_{HP}, \tag{3.23}$$

where COP_{HP} is the coefficient of performance. The heat balance equation for the heat pump is:

$$\dot{H}_{HP}(t) = c_p \dot{m}^{dem}_{SH} \left(T^{s}_{HP}(t) - T^{r}_{HP}(t) \right), \tag{3.24}$$

where:

$$T^{s}_{HP}(t) = T^{dem}_{SH} \quad \text{and} \quad T^{r}_{HP}(t) = T^{sto}_{SH}(t), \tag{3.25}$$

where $T^{s}_{HP}(t)$ and $T^{r}_{HP}(t)$ are the temperatures of the water flowing out and into of the heat pump, respectively.

For the heat pump, electricity is the input energy carrier. The exergy input rate of electricity, $\dot{E}x_{HP}^{e}(t)$, is equal to the electricity consumption of the heat pump. At the output, the exergy rate of the heat delivered by the heat pump, $\dot{E}x_{HP}^{out}(t)$, is evaluated similarly to that of the heat recovery and auxiliary natural gas boilers. The exergy loss rate in the heat pump is:

$$\dot{E}xloss_{HP}(t) = \dot{E}x_{HP}^{e}(t) - \dot{E}x_{HP}^{out}(t). \tag{3.26}$$

3.2.2.4. Modeling of the thermal energy storage system. The energy stored in the water tank at time *t* is affected by: the energy stored at time (*t* - Δ*t*), the heat provided by the heat recovery boiler, and the heat supplied:

$$c_p m_{sto,SH} T_{SH}^{sto}(t) = \eta_{hex} c_p m_{sto,SH} T_{SH}^{sto}(t-\Delta t) + \left[\dot{H}_{HRB,SH} - c_p \dot{m}_{sto,SH}(t)\left(T_{SH}^{sto}(t) - T_{SH}^{sto,r}(t)\right)\right]\Delta t, \tag{3.27}$$

where $m_{sto,SH}$ is the mass of water in the thermal storage and $T_{SH}^{sto,r}(t)$ is the temperature of the return water to the tank.

3.2.2.5. Meeting demand. The above devices are interconnected by the water network through pipes. As mentioned earlier, there are two cases. In the first one, when the temperature of the water in the tank is higher than the required, T_{SH}^{dem}, the water is directly supplied to the buildings, and part of the water is mixed with the return water from buildings before going to the terminal devices inside the buildings, i.e.,

$$\dot{m}_{SH}^{dem} = \dot{m}_{SH}^{BP,s}(t) + \dot{m}_{SH}^{BP,r}(t) = \dot{m}_{SH}^{BP,s}(t) + \dot{m}_{SH}^{build,s}(t) = \dot{m}_{SH}^{build,s}(t) + \dot{m}_{SH}^{build,r}(t) = \dot{m}_{SH}^{BP,r}(t) + \dot{m}_{SH}^{build,r}(t), \tag{3.28}$$

where $\dot{m}_{SH}^{dem}$ is the mass flow rate required to satisfy the space heating demand; $\dot{m}_{SH}^{BP,s}(t)$ is the bypass mass flow rate to be supplied to buildings; $\dot{m}_{SH}^{BP,r}(t)$ is the bypass mass flow rate to be returned to the tank; $\dot{m}_{SH}^{build,s}(t)$ is the return mass flow rate from buildings to be mixed with the water from the storage; and $\dot{m}_{SH}^{build,r}(t)$ is the return water mass flow rate from buildings to the tank. The energy balance in the mixer is expressed by:

$$c_p \dot{m}_{SH}^{dem} T_{SH}^{dem} = c_p \dot{m}_{SH}^{BP,s}(t) T_{SH}^{sto}(t) + c_p \dot{m}_{SH}^{build,s}(t) T_{SH}^{r}(t), \tag{3.29}$$

where $T_{SH}^{r}(t)$ is the temperature of the return water from buildings. After going through buildings to satisfy the space heating demand, the energy balance in the mixer is expressed by:

$$c_p \dot{m}_{SH}^{dem} T_{SH}^{sto,r}(t) = c_p \dot{m}_{SH}^{BP,s}(t) T_{SH}^{sto}(t) + c_p \dot{m}_{SH}^{build,s}(t) T_{SH}^{r}(t), \tag{3.30}$$

where $T_{SH}^{sto,r}(t)$ is the temperature of the return water to the water tank.

The heat balance equation at the demand side is:

$$\dot{H}_{SH}^{dem}(t) = c_p \dot{m}_{SH}^{dem} \left(T_{SH}^{dem} - T_{SH}^{r}(t) \right), \tag{3.31}$$

where $\dot{H}_{SH}^{dem}(t)$ is the heat rate demand of space heating.

In the second case, when the temperature of the water in the tank is lower than the required, the water is sent to the auxiliary natural gas boiler or the heat pump, and heated to the required temperature.

3.2.3 Modeling of water network for domestic hot water

A typical water network of domestic hot water is shown in Figure 3.3. As in the previous subsection, a fully-mixed tank model is assumed. Since the water is used up at the demand side, cold water is continuously supplied to the storage and warmed up by the energy provided by the heat recovery boiler and solar collectors through two heat exchangers in the water tank.

There are two cases. When the temperature of the water in the tank is higher than the required (assumed constant), the water is directly supplied to the buildings and mixed with the aqueduct cold water in the mixer to bring down the temperature to the required one before the terminal use. In the second case, when the temperature of the water in the tank is lower than the required, the water is sent to the auxiliary natural gas boiler, and heated to the required temperature.

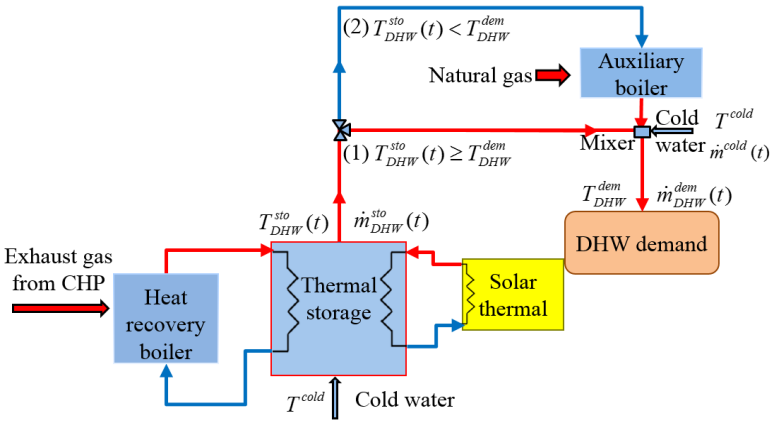


Figure 3.3. Scheme of water network for domestic hot water.

The energy and exergy modeling of the heat recovery boiler, auxiliary natural gas boiler, and thermal storage for domestic hot water is similar to the modeling of the corresponding devices for space heating.

2.3.1. Modeling of the solar thermal plant. The solar thermal plant converts solar energy into heat to meet the domestic hot water demand. The heat rate provided by the solar thermal plant, $\dot{H}_{ST}(t)$, is:

$$\dot{H}_{ST}(t) = \eta_{coll} A_{coll} \dot{I}_T(t), \tag{3.32}$$

where A_{coll} is the collector area, η_{coll} is the collector efficiency, and $\dot{I}_T$ is the total solar irradiance. The heat balance equation for the solar thermal plant is:

$$\dot{H}_{ST}(t) = c_p \dot{m}_{ST}(t)\left(T^s_{ST}(t) - T^r_{ST}(t)\right) = \eta_{hex} c_p \dot{m}_{ST}(t)\left(T^s_{ST}(t) - T^{sto}_{DHW}(t)\right), \tag{3.33}$$

where $\dot{m}_{ST}(t)$ is the water mass flow rate from the solar thermal plant through the heat exchanger in the storage; $T^s_{ST}(t)$ and $T^r_{ST}(t)$ are the temperatures of the water flowing into and out of the heat exchanger, respectively; and η_{hex} is the efficiency of the heat exchanger. The supply temperature is assumed 10 K higher than that of the water in the tank and the return temperature is a dependent variable.

Solar energy from the collectors is considered as a low-exergy source since the solar exergy input rate is evaluated at the output of the solar collector field [20]. Therefore, by following this approach no exergy loss is taken into account.

2.3.2. Meeting demand. Similarly to space heating, the above devices are interconnected by the water network, and there are two cases. When the temperature of the water in the tank, $T_{DHW}^{sto}(t)$, is higher than the required, T_{DHW}^{dem}, the water is directly supplied to the buildings and mixed with the aqueduct cold water before the terminal use, i.e.,

$$\dot{m}_{DHW}^{dem}(t) = \dot{m}_{DHW}^{sto}(t) + \dot{m}^{cold}(t), \tag{3.34}$$

where $\dot{m}_{DHW}^{dem}(t)$, $\dot{m}_{DHW}^{sto}(t)$, and $\dot{m}^{cold}(t)$ are the water mass flow rates to be supplied to terminal users, the hot water mass flow rate taken from the storage, and the cold water mass flow rate from the aqueduct, respectively. The energy balance in the mixer is expressed by:

$$c_p \dot{m}_{DHW}^{dem}(t) T_{DHW}^{dem} = c_p \dot{m}_{DHW}^{sto}(t) T_{DHW}^{sto}(t) + c_p \dot{m}^{cold}(t) T^{cold}, \tag{3.35}$$

where T^{cold} is the temperature of the cold water from the aqueduct. At the demand side, it is assumed that the temperature of hot water is brought down to T^{cold} after terminal use, and the heat balance equation is:

$$\dot{H}_{DHW}^{dem}(t) = c_p \dot{m}_{DHW}^{dem}(t)\left(T_{DHW}^{dem} - T^{cold}\right), \tag{3.36}$$

where $\dot{H}_{DHW}^{dem}(t)$ is the heat rate demand of domestic hot water.

In the second case, when the temperature of the water in the tank, $T_{DHW}^{sto}(t)$, is lower than the required, T_{DHW}^{dem}, the water is sent to the auxiliary natural gas boiler, i.e.,

$$\dot{m}_{DHW}^{dem}(t) = \dot{m}_{DHW}^{sto}(t). \tag{3.37}$$

The water is heated to the required temperature in the natural gas boiler.

For the overall problem, the coupling across the two water networks is that the sum of exhaust fractions for space heating and domestic hot water has to be one (see Eq. 3.14). The coupling across all the three networks is represented by the electricity balance (Eq. 3.12).

3.2.4 Objective functions

The objective is to minimize the total energy cost and the exergy losses at the conversion step. The economic and exergetic objective functions are discussed in Subsubsections 3.2.4.1 and 3.2.4.2, respectively. The multi-objective optimization method to solve the problem is discussed in Subsubsection 3.2.4.3.

3.2.4.1 Economic objective. The economic objective is to minimize the total energy cost, *Cost*, which is the sum of two terms: cost of grid power and cost of natural gas:

$$Cost = \sum_t \Delta t \left(P_{grid}(t) \dot{E}_{buy}(t) + P_{gas} \dot{G}_{buy}(t) \right), \tag{3.38}$$

where $P_{grid}(t)$ is the time-of-day unit price of electricity from the power grid, and P_{gas} is the constant unit price of natural gas. The volumetric flow rate of natural gas bought, $\dot{G}_{buy}(t)$, corresponds to the total consumption requirement of the CHP system and auxiliary natural gas boilers.

3.2.4.2 Exergetic objective. As mentioned earlier, the focus of this work is on the exergy loss at the energy conversion step, which accounts for the largest fraction of the total exergy losses in the energy-supply chain from energy resources to user demands. The total exergy loss at the conversion step, $Exloss_{conv}$, is the sum of the exergy losses of the energy devices at the conversion step over time:

$$\begin{aligned} Exloss_{conv} = \sum_t \Delta t (& \dot{E}xloss_{GT}(t) + \dot{E}xloss_{grid}(t) + \dot{E}xloss_{HRB,SH}(t) + \dot{E}xloss_{HRB,DHW}(t) + \dot{E}xloss_{boil,SH}(t) \\ & + \dot{E}xloss_{boil,DHW}(t) + \dot{E}xloss_{HP}(t)). \end{aligned} \tag{3.39}$$

3.2.4.3 Multi-objective optimization method. With the exergetic objective function formulated in Eq. (3.39) and the economic objective function formulated in Eq. (3.38), the problem has two objective functions to be minimized. To solve this multiple-objective problem, a single objective function is formulated as a weighted sum of the total energy cost, *Cost*, and the total exergy loss at the conversion step, $Exloss_{conv}$:

$$F_{obj} = c(1-\omega)Cost + \omega Exloss_{conv}, \tag{3.40}$$

where the constant c is chosen such that c *Cost* and $Exloss_{conv}$ have the same order of magnitude. The Pareto frontier involving the best possible trade-offs between the two objectives can be found by varying the weight ω in between the interval 0 and 1. The solution that minimizes the total energy cost is obtained when $\omega = 0$, whereas the solution that minimizes the total exergy loss at the energy conversion step is obtained when $\omega = 1$. The above problem is separable, nonlinear and involves both discrete and continuous variables.

3.3 Solution Methodology

To coordinate energy devices with coupling constraints and solve the problem efficiently, our idea is to use multipliers as shadow prices in a decomposition and coordination structure. The surrogate Lagrangian relaxation and branch-and-cut is combined for a speedy and near-optimal performance. The key idea is to relax the constraints that couple across energy devices by Lagrangian multipliers to create subproblems, e.g., the space heating subproblem and the domestic hot water subproblem. Since subproblems are solved individually, their solutions are coordinated through iterative updating of multipliers. To ensure fast convergence, the surrogate subgradient method is used. The key idea is that a proper direction to update multipliers can be obtained without optimally solving all subproblems, which allows more frequent multiplier updating, and reduces zigzagging. However, the convergence requires, in the step sizing process, the knowledge of the optimal dual value, which is unknown in practice. To overcome this, a novel step-sizing formula that does not require the optimal dual value was developed [18]. This is achieved through a constructive process in which distances between Lagrange multipliers at consecutive iterations decrease,

and as a result, multipliers will converge to a unique limit. Because of this, computational efforts are much reduced.

To solve subproblems by using branch-and-cut, which is suitable for mixed-integer linear problems, a linear formulation is needed. Usually, logical constraints can be linearized by introducing new variables and the logarithm function can be linearly approximated within a small range. For other nonlinear terms such as cross product, and square and cube functions, the linearization is not easy. In the framework of surrogate Lagrangian relaxation, solutions from the previous iteration can be used as input data in the next iteration. Therefore the nonlinear terms can be linearly approximated by using the values of the previous solution. The resulting linear problem will be optimized and the previous solution will be updated.

To coordinate energy devices with coupling constraints and solve the problem efficiently, our idea is to use multipliers as shadow prices in a decomposition and coordination structure. Surrogate Lagrangian relaxation and branch-and-cut are combined for a speedy and near-optimal performance [21-23]. After relaxing the coupling constraints, i.e., CHP exhaust gas sharing constraints (Eq. 3.14) by Lagrangian multipliers, the relaxed problem is to minimize the following Lagrangian function, L, as:

$$L(\lambda,y) \equiv c\left(1-\omega\right)Cost(y) + \omega Exloss_{conv}(y) + \sum_{t}\lambda(t)\left(\xi_{SH}(t) + \xi_{DHW}(t) - 1\right), \tag{3.41}$$

subject to Eq. (3.1)-(3.13) and (3.15)-(3.39). In the above, λ represent multipliers relaxing CHP exhaust gas sharing constraints, and y represent all the decision variables.

Then, there are two subproblems, e.g., the space heating subproblem and the domestic hot water subproblem (with electricity-related devices). They are solved by branch-and-cut individually.

By solving the relaxed problem, the dual function becomes:

$$q(\lambda) = \min_{y} L(\lambda, \mathrm{y}). \tag{3.42}$$

Instead of obtaining the dual value (3.42), a surrogate dual value is obtained in surrogate Lagrangian relaxation as follows:

$$\tilde{L}\left(\lambda^k, y^k\right) = c(1-\omega)Cost(y) + \omega Exloss_{conv}(y) + \lambda^k \tilde{g}(y^k). \tag{3.43}$$

In the above, λ^k and y^k are multipliers and any feasible solution of the relaxed problem at iteration k, respectively, and $\tilde{g}(y^k)$ are the surrogate subgradient vectors consisting of:

$$\tilde{g}\left(y^k\right) = \xi_{SH}(t) + \xi_{DHW}(t) - 1. \tag{3.44}$$

Since surrogate Lagrangian relaxation does not require the relaxed problem to be fully optimized, surrogate subgradient directions may not form acute angles with directions toward optimal multipliers, which will cause divergence. To guarantee that surrogate directions form acute angles with directions toward the optimal multipliers, the relaxed problem has to be sufficiently optimized, such that surrogate dual values in Eq. (3.43) satisfy the surrogate optimality condition:

$$\tilde{L}\left(\lambda^k, y^k\right) < \tilde{L}\left(\lambda^k, y^{k-1}\right), \tag{3.45}$$

where y^{k-1} is a feasible solution at the iteration k-1.Since the relaxed problem is not fully optimized and subgradient directions do not change much at each iteration, computational requirements and zigzagging of multipliers are much reduced as compared to traditional subgradient methods.

In the method, multipliers are updated as:

$$\lambda^{k+1} = \lambda^k + d^k \tilde{g}(y^k), \tag{3.46}$$

where d^k is the stepsize. It has been proven that the multipliers converge to the optimum if the stepsizes are updated by using the novel step-sizing formula developed in [21].

To solve subproblems by using branch-and-cut, which is suitable for mixed-integer linear problems, a linear formulation is needed [24, 25]. Usually, the logarithm function can be linearly approximated within a small range. For other nonlinear terms such as cross product, the linearization is not easy. In the framework of surrogate Lagrangian relaxation, solutions from the previous iteration can be used as input data in the next iteration. Therefore the nonlinear terms can be linearly approximated by using the values of the previous solution under the monotonic condition as proved in [22]. The resulting linear problem will be optimized and the previous solution will be updated.

3.4 Numerical Results

The method discussed above has been implemented by using the commercial branch-and-cut solver IBM ILOG CPLEX Optimization Studio Version 12.6 on a PC with 2.90GHz Intel (R) i7 CPU and 16G RAM. The targeted end-user is a large hypothetic hotel in Beijing with an area of 30,000 m2. A typical winter day of January is chosen, with one hour as time-step. The input data for the optimization model are first described in Subsubsection 3.4.1. Then, the Pareto frontier is presented and the operation strategies under different weights are discussed for different trade-off points in Subsubsection 3.4.2. The exergy losses of each step in the energy-supply chain obtained by the energy cost minimization and exergy loss minimization are also presented. In addition, the effects of energy resource prices are discussed. Finally, the comparison among different DES configurations is discussed in Subsubsection 3.4.3.

3.4.1 Input data

China is the second largest building energy user in the world, ranked first in residential energy consumption and third in commercial energy consumption [26]. Moreover, in China the application of DESs has been increased rapidly in recent years with the supportive government policies and financial incentives [27]. Therefore, a large hypothetic hotel in Beijing is chose as the targeted end-user. The hourly electricity, domestic hot water and space heating rate demands for a typical winter day of January are taken from a

comprehensive investigation about energy demands of hotels in Beijing [28], and they are shown in Figure 3.4. The time-of-day unit price of electricity from the power grid is also shown in Figure 3.4 [29]. The exergy efficiency of the power generation plant is assumed equal to 0.32, a typical value when electricity is mostly generated by coal-fired thermal power plants as in China. The unit price of natural gas is assumed equal to 0.38 \$/Nm^3 [29], where Nm^3 stands for the volume of gas at 0°C temperature and at 1.013 bar pressure. Its exergy factor is assumed equal to 1.04 [17].

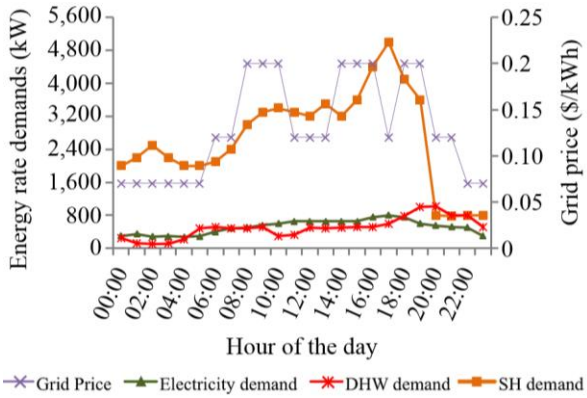


Figure 3.4. Energy rate demands of the hotel and grid price for a typical winter day of January.

To evaluate the heat rate provided by the solar thermal plant, the hourly solar irradiance of a winter day is evaluated as the average of the solar irradiance of corresponding hours of all January days [30]. The sizes of the energy devices and thermal storages as well as the efficiencies assumed in this work are listed in Table 3.1.

Table 3.1. Size and efficiency of energy devices and thermal storages.

Primary energy devices	**Size** (MW)	**Efficiency**	
		Electrical	Thermal
Gas turbine	1.25	0.24	μ_{GT} = 0.080
Solar thermal plant	0.41		0.40
Secondary energy devices	**Size** (MW)	**Efficiency**	
Heat pump	5.0	COP_{SH}^{HP} = 3.0	
Heat recovery boiler SH - DHW	2.4 – 1.1	η_{HRB} =0.80	η_{boiler} = 0.90
Thermal energy storage	**Capacity** (MWh)	**Efficiency**	
SH - DHW	0.064 – 0.24	0.98	

3.4.2 Pareto frontier

Based on the network configurations shown in Figures 3.2 and 3.3, there are non-linearizable logic constraints. The Surrogate Lagrangian relaxation method combined with branch-and-cut is suitable for mixed-integer linear problems. To get a linear problem and test this innovative optimization method, temperatures of the water in the tanks are hypothesized lower than the required temperatures in the numerical testing. This assumption is supported by the fact that, in winter days, solar radiation is in general lower than in summer days, and electricity demand is also lower. As a consequence, water temperature in the storage tanks can be lower than that required for most of the day.

The optimization problem can be solved within several minutes and the Pareto frontier is shown in Figure 3.5. The point marked with *a* is obtained by minimizing the total energy cost, and the daily energy cost is 3,487 $/d whereas the daily exergy losses at the conversion step are 75,459 kJ/d. The point marked with *b* is obtained by minimizing the total exergy losses at the conversion step. The daily energy cost is 3,718 $/d whereas the daily exergy losses are 68,687 kJ/d. The points between the extreme points are found by subdividing the weight interval into 100 equally-spaced points. There are 13 points since some solutions have been found under more than one weight values.

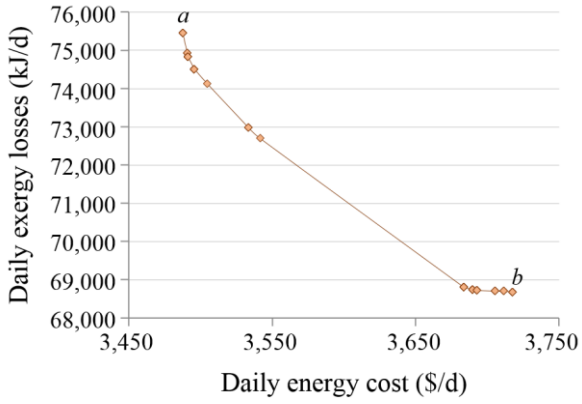


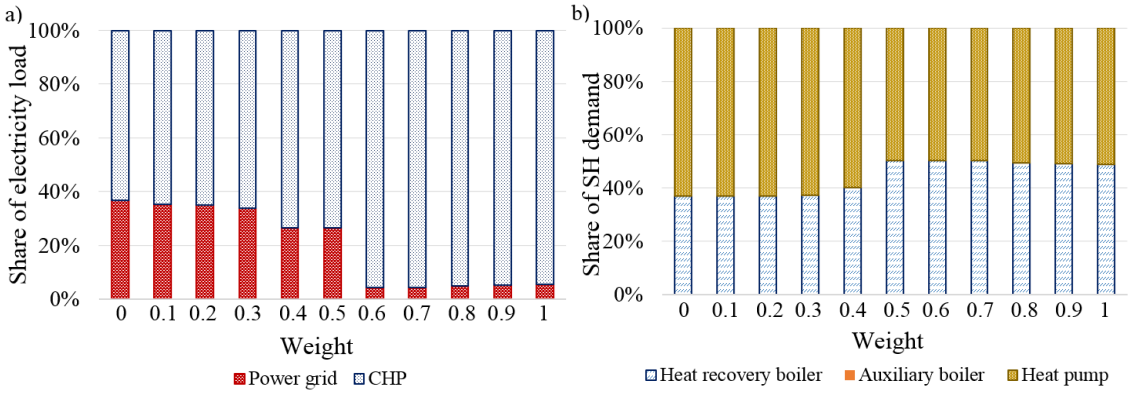
Figure 3.5. Pareto frontier.

Each point on the Pareto frontier corresponds to a different operation strategy of the DES. In order to understand how the operation strategies vary with the weight ω, the optimized operation strategies of the

DES obtained by varying the weight from 0 to 1 with a 0.1 increase, are presented in Figure 3.6. Figure 3.6a shows that, when ω varies from 0 to 1 (from energy cost minimization to exergy loss minimization), the share of the electricity load (sum of electricity demand and electricity required by the heat pump) satisfied by the CHP system increases while the exergy losses reduce. This highlights the essential role of the CHP system in the reduction of exergy losses because of the recovery of waste heat for thermal purposes, leading to efficient use of the high-quality energy resource.

Figure 3.6b shows that from energy cost minimization to exergy loss minimization, the share of space heating demand satisfied by the heat recovery boiler increases, coherent with the increasing use of the CHP (as shown in Figure 3.6a), highlighting the importance of waste heat recovery for the exergetic purpose. The use of exhaust gas for low-exergy thermal demands reduces the exergy losses occurring at the energy conversion step. When ω varies from 0 to 1, the share of space heating demand met by the heat pump exhibits an opposite trend, decreasing with the reduced use of the grid power, as shown in Figure 3.6a.

Figure 3.6c shows that from energy cost minimization to exergy loss minimization, the share of the domestic hot water demand satisfied by the heat recovery boiler increases coherently with the increased use of the CHP system. Conversely, the share of domestic hot water demand satisfied by the auxiliary natural gas boiler reduces, highlighting that combustion processes should be avoided for thermal purposes, thereby reducing the waste of high-quality energy resources.



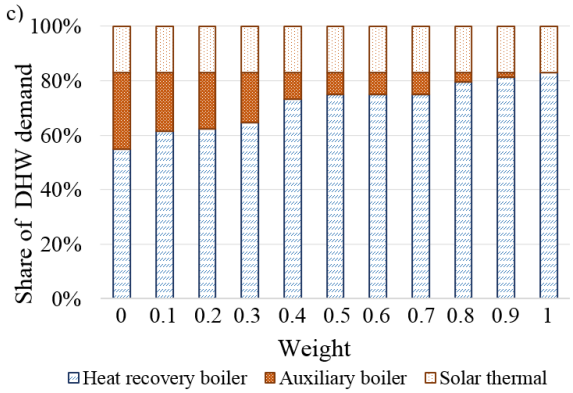


Figure 3.6. Optimized operation strategies of the DES at various trade-off points for a) electricity, b) space heating, c) domestic hot water.

Figure 3.7 shows the exergy losses occurring at the various steps of the energy-supply chain (i.e., conversion, storage, terminal devices, final consumption) under cost and exergy loss (at the conversion step) minimization. Exergy losses occurring in the storages and in the terminal devices as well as exergy of user demands are evaluated according to [1]. In the exergy loss minimization, exergy losses at the conversion step are about 9% lower than those obtained by energy cost minimization. On the other hand, exergy losses occurring in the storages are 22% higher than those obtained under energy cost minimization. This is because, under exergy loss minimization, there is larger use of heat recovery boilers, which charge the storages, than what occurs under energy cost minimization, as shown in Figures 3.6b and 3.6c. This means that the minimization of exergy losses at the conversion step does not guarantee the minimization of exergy losses in the other steps of the energy-supply chain. However, the total exergy loss occurring in the whole energy-supply chain under exergy loss minimization is 7% lower than that obtained under cost minimization.

The optimized operation strategies of the DES may depend on the prices of energy resources. In the problem under consideration (reference case), the price of natural gas is cheaper than that of grid power as in the current Chinese market. In other countries, the opposite may occur. To show how the relative prices of natural gas and grid power affect the optimized operation strategies, the problem is solved with a high natural gas price, 0.57$/Nm3, which is 150% of the original price and higher than that of the grid power.

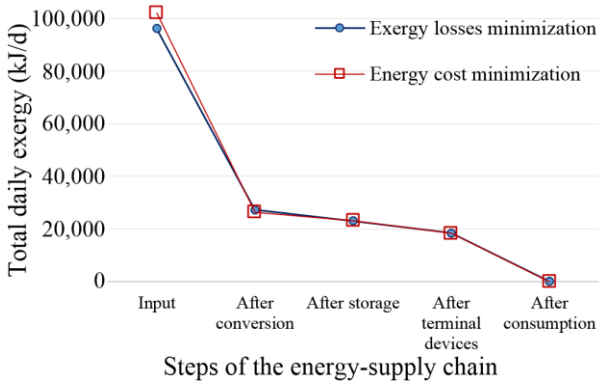


Figure 3.7. Exergy losses of each step in the energy-supply chain under cost and exergy loss minimization

Figure 3.8 shows the optimized operation strategies for electricity at the various trade-off points with a high natural gas price. It is shown that the share of the electricity load satisfied by the CHP system increases from energy cost minimization to exergy loss minimization. However, compared to the reference case in Figure 3.6a, the share of electricity load covered by the CHP system is generally lower when the weight of the economic objective is higher than that of the exergetic one, and almost the same when the weight of the economic objective is lower. In particular, when ω = 0, 0.3 and 0.5, the share of electricity load covered by the CHP system is 46%, 51% and 69% in the new case, respectively, while 63%, 66% and 74% in the reference case, respectively. The lower usage of the CHP system results in lower amount of exhaust gas and consequent higher usage of auxiliary boilers for thermal purposes. This leads to higher exergy losses at the energy conversion step as compared to the reference case.

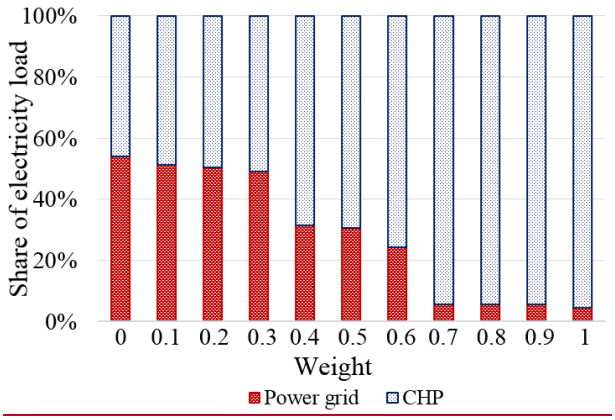


Figure 3.8. Optimized operation strategies at various trade-off points for electricity with high gas price

3.4.3 Configuration comparison

To show how each energy device contributes to the reduction of energy costs and exergy losses, various configurations of the DES are now analyzed. For each configuration, one energy device is taken out of the DES, including the solar thermal plant, auxiliary natural gas boilers, heat pump, and entire CHP system. In addition, a conventional energy supply system is also considered. The grid power is used to meet the electricity demand, the electricity required by an electric heater to satisfy the space heating demand, and the electricity required by an electric boiler to satisfy the domestic hot water demand. All the above configurations are listed in Table 3.2.

Table 3.2. Investigated configurations.

Configuration	**Energy devices taken out of the reference case (Configuration I)**
1	With all devices
2	Without solar thermal plant
3	Without auxiliary natural gas boilers
4	Without heat pump
5	Without CHP system
Configuration	**Conventional energy supply system**
6	All from grid power

The daily energy costs obtained under cost minimization of different configurations are compared in Figure 3.9. Configuration 1 is the reference case, consisting of all energy devices listed in Table 3.1. The reference case shows the best performance in terms of the daily energy costs as compared with the other configurations. For Configurations 2 and 3, the daily energy costs are about 2% larger than those in

Configuration 1. For Configuration 4, the daily energy costs are 18% higher than those in Configuration 1, because of the high conversion efficiency of the heat pump. Configuration 5 excludes the CHP system. The costs are 34% larger than those in Configuration 1, pointing out the essential role of the CHP system in the reduction of energy costs. The worst case is represented by the conventional energy supply system (Configuration 6). The 265% increase in the daily energy costs, as compared with Configuration 1, shows that the energy costs can be strongly reduced by the optimized operation of the DES.

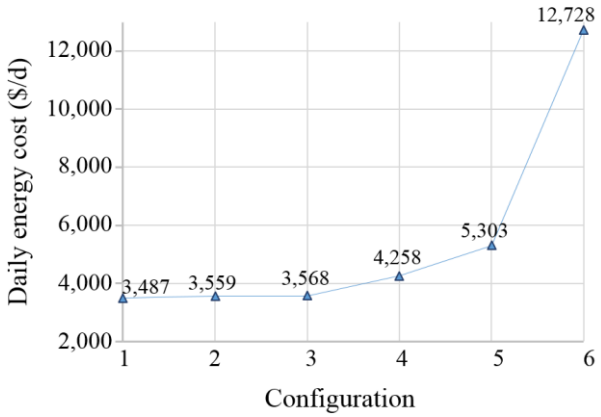


Figure 3.9. Total daily energy costs under energy cost minimization for Configurations 1-6.

Figure 3.10 shows the exergy losses occurring at different steps of the energy-supply chain obtained by the exergy loss (at the conversion step) minimization for the configurations listed in Table 3.2. The reference case (Configuration 1) shows the best performance, also in terms of minimum exergy losses at the conversion step. For Configuration 2, exergy losses increase by 2% as compared with Configuration 1. For Configuration 3, the exergy losses are the same as those in Configuration 1, since in the exergy loss minimization the auxiliary boilers are never used to satisfy the domestic hot water and space heating demands, as shown in Figures 3.6b and 3.6c. The exergy losses for Configuration 4 (without the heat pump), are 22% larger than those in Configuration 1, showing the importance of the heat pump not only in the reduction of energy costs but also in the reduction of exergy losses, thanks to its high conversion efficiency. A 28% increase in the exergy losses is found for Configuration 5, as compared with Configuration 1. Without the CHP system, there is no exhaust gas for heat, and the use of auxiliary boilers increases the exergy losses because of the combustion processes. Finally, similarly to the energy costs, the worst case is

represented by Configuration 6, where the exergy losses are 175% larger than those in Configuration 1. When all the demands are satisfied by electricity, high exergy losses occur, since a high quality energy carrier, electricity, is used to satisfy low-quality thermal demands.

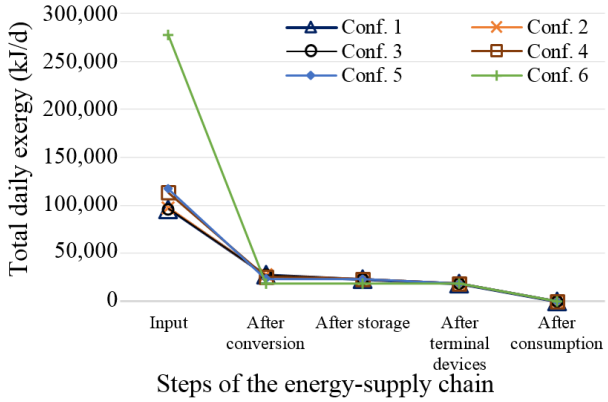


Figure 3.10. Exergy losses of each step in the energy-supply chain under exergy loss minimization for Configurations 1-6.

Exergy losses in thermal storages do not reach the minimum value in the reference case. The minimum value is obtained in Configuration 5, which is 93% lower than that in the reference case. Without the CHP system, the SH storage is not used with consequent zero exergy losses, and the DHW storage is only charged by the solar thermal collectors. Therefore, the total exergy loss is reduced at the storage step. However, the total exergy loss occurring in the whole energy-supply chain reaches the minimum in the reference case as compared with other configurations.

3.5 Conclusion

This work analyzes the exergy-efficient management of the energy-supply chain of a DES to reduce both energy costs and exergy losses at the conversion step. A mixed-integer optimization problem considering several energy devices is formulated and is solved by surrogate Lagrangian relaxation combined with branch-and-cut. The objective is to minimize a weighted sum of energy costs and exergy losses at the conversion step. Numerical results show that when both fossil and renewable energy resources are appropriately combined under the optimized operation of the DES, energy costs and exergy losses can be

reduced. The use of high-quality energy resources can be reduced through the reduction of exergy losses at the energy conversion step, which are the largest part in the whole energy-supply chain, leading to sustainability of energy supply systems. The operators of DESs can choose the operation strategy from the Pareto frontier based on cost and sustainability concerns. Future work may include exergy-efficient management of the various steps of the energy supply chain, from energy resources to user demands, to reduce the total exergy loss of each sub-process.

References

[1] ECBCS - Annex 49 - Low Exergy Systems for High Performance Buildings and Communities, homepage. Available <http://www.ecbcs.org/annexes/annex49.htm>.

[2] Eurostat Statistics Database – Energy statistics – Supply, Transformation, Consumption (2010).

[3] D. Schmidt Low exergy systems for high performance buildings and communities, Energy and Buildings 41 (2009) 331-336.

[4] J. Szargut International progress in second law analysis, Energy 5 (1980) 709-718.

[5] IEA/ECBCS Annex 37, Low Exergy Systems for Heating and Cooling (2003).

[6] D. Schmidt Design of low exergy buildings – method and a pre-design tool, The International Journal of Low Energy and Sustainable Buildings 3 (2004) 1–2.

[7] J. Szargut, D.R. Morris, F.R. Stewerd Exergy analysis of thermal, chemical and metallurgical processes (1988) New York: Hemisphere.

[8] A. Yildiz, A. Güngör Energy and exergy analyses of space heating in buildings, Applied Energy 86 (2009) 1939–1948.

[9] S.C. Jansen, J. Terès-Zubiaga, P.G. Luscuere The exergy approach for evaluating and developing an energy system for a social dwelling, Energy and Buildings 55 (2012) 693-703.

[10] J. Terès-Zubiaga, S.C. Jansen, P. Luscuere Dynamic exergy analysis of energy systems for a social dwelling and exergy based system improvement, Energy and Buildings 64 (2013) 359-371.

[11] A Kari, S Arto Distributed energy generation and sustainable development, Renewable Sustainable Energy Reviews 10 (2006) 539–58.

[12] B. Yan, P.B. Luh, B. Sun, C. Song, C. Dong, Z. Gan, L.D. Michel Energy-efficient management of eco-communities, Proceedings of IEEE CASE; 2013 Aug 17-20; Madison, USA.

[13] X. Guan, Z. Xu, Q. Jia Energy-efficient buildings facilitated by microgrid. IEEE Transactions on Smart Grid 2011;1:466-473.

[14] G. Graditi, M. Ippolito, R. Lamedica, A. Piccolo, A. Ruvio, E. Santini, P. Siano, G. Zizzo Innovative Control Logics for a Rational Utilization of Electric Loads and Air-Conditioning Systems in a Residential Building, Energy and Buildings 102 (2015) 1-17.

[15] M. Di Somma, B. Yan, N. Bianco, G. Graditi, P.B. Luh, L. Mongibello, V. Naso Operation optimization of a distributed energy system considering energy costs and exergy efficiency, Energy Conversion and Management 103 (2015) 739-751.

[16] X.Q. Kong, R.Z. Wang, X.H. Huang Energy optimization for a CCHP system with available gas turbines, Applied Thermal Engineering 25 (2005) 377-391.

[17] K.J. Kotas The exergy method for thermal plant analysis, reprinted. Malabar, FL: Krieger; 1995.

[18] A. Angelotti, P. Caputo. The exergy approach for the evaluation of heating and cooling technologies, first results comparing steady state and dynamic simulations. Proceedings of the 2nd PALENC and 28th AIVC Conference, vol I; 2007 September 27-29; Crete Island, Greece; p. 59-64.

[19] L.M. Ramirez-Elizondo, G.C. Paap, R. Ammerlaan, R.R. Negenborn, R. Toonssen On the energy, exergy and cost optimization of multi-energy-carrier power systems. International Journal of Exergy 2013;13:364-385.

[20] H. Torìo, A. Angelotti, D. Schmidt Exergy analysis of renewable energy-based climatisation systems for buildings: A critical view. Energy and Buildings 2009;41:248-71.M.A.

[21] M.A. Bragin, P.B. Luh, J.H. Yan, N. Yu, G.A. Stern Convergence of the Surrogate Lagrangian Relaxation Method, Journal of Optimization Theory and Applications, 164 (1) (2015) 173-201.

[22] M. A. Bragin, P. B. Luh, J. H. Yan, and G. A. Stern, "An efficient approach for solving mixed-integer programming problems under the monotonic condition," Journal of Control and Decision, 2016, DOI:10.1080/23307706.2015.1129916.

[23] M. A. Bragin, P. B. Luh, J. H. Yan, and G. A. Stern, "Novel exploitation of convex hull invariance for solving unit commitment by using surrogate Lagrangian relaxation and branch-and-cut," in Proceedings of the IEEE Power and Energy Society, General Meeting, Denver, Colorado, 2015.

[24] R. E. Bixby, M. Fenelon, Z. Gu, E. Rothberg, and R. Wunderling, "MIP: Theory and practice – closing the gap." System Modelling and Optimization, pp. 19-49, 2000.

[25] IBM ILOG, "IBM ILOG CPLEX Optimization Studio Information Center," 2013. [Online]. Available: http://pic.dhe.ibm.com/infocenter/cosinfoc/v12r5/index.jsp

[26] Pacific Northwest National Laboratory, China's Building Energy Use: A Long-Term Perspective based on a Detailed Assessment, January 2012. http://www.pnnl.gov/main/publications/external/technical_reports/PNNL-21073.pdf.

[27] J. Han, L. Ouyang, Y. Xu, R. Zeng, S. Kang, G. Zhang. Current status of distributed energy system in China. Renewable and Sustainable Energy Reviews, 55 (2016) 288-297.

[28] Z. Zhou, P. Liu, Z. Li, W. Ni An engineering approach to the optimal design of distributed energy systems in China, Applied Thermal Engineering, 53 (2013) 387-396.

[29] Beijing Municipal Commission of Development & Reform, Current Prices for Public Commodities, 2011, http://www.bjpc.gov.cn/ywpd/wjgl/jgcx/syjg/syjgggspxxjg

[30] ASHRAE International Weather files for Energy Calculations (IWEC weather files). Users manual and CD-ROM, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, USA, 2001.

Chapter 4

Operation and Design Optimization of Microgrids with Renewables

To reduce energy costs and emissions of microgrids, daily operation is critical. The problem is to commit and dispatch distributed devices with renewable generation to minimize energy and emission costs while meeting forecasted energy demand. The problem is challenging because of the intermittent nature of renewables. In this chapter, PV uncertainties are modeled by a Markovian process. For effective coordination, other devices are modeled as Markov processes with states depending on PV states. The entire problem is stochastic and Markovian. This combinatorial problem is solved by branch-and-cut. Beyond energy and emission costs, to consider capital and maintenance costs in the long run, microgrid design is also essential. The problem is to decide device sizes with given types to minimize the lifetime cost while satisfying energy demand, where the complexity increases exponentially with the problem size. To evaluate the lifetime cost including the reliability cost and the classic components such as capital and fuel costs, a linear model is established. By selecting a limited number of possible device size combinations, exhaustive search is used to find the optimized design. Results show that the operation method is efficient in saving cost and computation time, and scalable, and the lifetime cost is reduced by the optimized design. The implications for regulators and distribution utilities are also discussed.

4.1 Introduction

With world's increasing energy demand and growing environmental concerns, efficient utilization of energy is essential for sustainable living, especially renewable energy. Reliable and flexible microgrids, which can operate under the grid-connected mode and can also turn into an islanded mode [1, 2], provide a promising opportunity and a desirable infrastructure. In microgrids, different distributed energy devices, such as gas turbines, photovoltaic (PV) panels, and natural gas boilers, generate and store different types of energy such as electricity, steam, and hot/chilled water to satisfy time-varying electricity and thermal demand. They should be coordinated through daily operation to reduce the energy cost and greenhouse gas emissions. To consider capital and maintenance costs in the long run, microgrid design (device types and sizes) is also critical.

The microgrid under consideration involves different distributed energy devices: Combined Cooling Heat and Power (CCHP), PV panels, natural gas boilers, electrical chillers and batteries, chosen among commonly used devices in practical microgrids. The microgrid operation problem is hierarchical, from unit commitment, economic dispatch, to optimal power flow. Focusing on the first two, the problem under consideration is to commit and dispatch distributed devices to minimize energy and CO_2 emission costs under the grid-connected mode while meeting the forecasted electricity and thermal demand of the following day. While the design problem is to decide device sizes with given types to minimize the lifetime cost while satisfying energy demand.

Optimized microgrid operation, however, is challenging because of the intermittent nature of renewables. In the literature, uncertainties were usually modeled by scenarios in microgrid operation problems. However, it is difficult to select an appropriate number of scenarios to balance modeling accuracy, solution feasibility, and computational efficiency. In this chapter, a mixed-integer model is established from the energy and emission point of view in Subsection 4.3. To avoid the difficulties associated with scenario-based methods, our ideas is to model PV generation by a Markovian process with

the current state summarizing all past information. For effective coordination, other devices are correspondingly modeled as Markov processes with states depending on the states of PV generation. The entire problem is therefore stochastic and Markovian, and this has not been found in the literature for microgrid operation. This combinatorial problem is solved by branch-and-cut.

Optimized design is also challenging since the problem complexity increases exponentially as the problem size increases, and energy resources (e.g., solar irradiance), fuel prices, and load are uncertain. In addition, the reliability costs, i.e., costs of microgrid protection devices and costs of unserved load when there is no power supply, are hard to estimate. In the literature, existing software packages were widely used, with uncertainties addressed by sensitivity analysis. While reliability costs were rarely considered. In this chapter, a linear model is established in Subsection 4.4 to evaluate the microgrid lifetime cost including the reliability cost and the classic components such as capital and fuel costs. The modeling of daily operation is simplified since it is consistent with that in the operation problem. The reliability cost is obtained based on the microgrid configuration and the estimated cost of unserved load during power outages. Based on load profiles, a limited number of possible combinations of device sizes is considered. With heuristic strategies for daily operation, exhaustive search is used to find the optimized design.

In Subsection 4.5, two examples are presented. The first small classroom example is to illustrate the Markov-based modeling of PV generation in operation, and show different components of the lifetime cost in design. The second semi-realistic one is to show that the operation method is efficient in saving cost and computation time, and scalable. It is also to compare lifetime costs of different design configurations and show impacts of uncertain factors in design. The implications of the above models and methods on operation and design of microgrids with renewables for regulators and distribution utilities are discussed in Subsection 4.6.

4.2 Literature Review

To formulate and solve the microgrid operation problem, models and methods provided in the literature are reviewed in Subsubsection 4.2.1. Related works on microgrid design are reviewed in Subsubsection 4.2.2. Since operation and design problems of Distributed Energy Systems (DESs) are similar to those of microgrids, related studies are also involved.

To formulate and solve the microgrid operation problem, models and methods provided in the literature are reviewed in Subsubsection 4.2.1. Related works on microgrid design are reviewed in Subsubsection 4.2.2. Our earlier work is briefly reviewed in Subsubsection 4.2.3. Since operation and design problems of Distributed Energy Systems (DESs) are similar to those of microgrids, related studies are also involved.

4.2.1 Operation of microgrids

The microgrid operation problem is hierarchical, from unit commitment, economic dispatch, to optimal power flow. Many researchers focus on unit commitment and economic dispatch of microgrids or DESs in the literature. Some of them focus on reducing energy costs as the single objective through daily operation [3-11]. For example, a mixed-integer linear model was developed to minimize the daily energy costs of grid power and natural gas for a microgrid while satisfying energy demand [3]. In this model, PV generation was modeled by a deterministic approach without explicitly considering uncertainties, calculated off-line with given parameters and solar irradiation. The battery was modeled by standard dynamics for state of charge without energy losses. The electrical grid was simplified by modeling electricity balance, i.e., electricity generated by the microgrid and bought from the grid equals electricity consumed in the microgrid, where the detailed electrical power models were not considered. The problem was solved by using branch-and-cut, and the impacts of uncertain demand and renewable generation were analyzed by the scenario tree method. In [4], a mixed-integer nonlinear model was developed, where PV and wind uncertainties were modeled by scenarios. It is, however, difficult to select an appropriate number of

scenarios to balance modeling accuracy, solution feasibility, and computational efficiency. Batteries were modeled by standard state dynamics considering charge and discharge efficiencies, and the electrical grid was simplified by modeling electricity balance. A metaheuristic algorithm was used to solve the problem. A decentralized energy management system was developed by Siemens for virtual power plants to minimize overall costs through coordination of distributed generators and energy storage [5]. With simplified models for energy devices, the modeling of renewable generation is based on forecasting and the overall problem is not stochastic.

Beyond considering energy costs as a single objective, multi-objective optimization methods were also developed for microgrid or DES operation by taking other factors such as emissions in to account [12, 13]. In [12], a stochastic model was developed for energy management of microgrids to minimize costs and emissions. In this model, demand and renewable generation uncertainties were modeled by scenarios, and batteries were modeled by standard state dynamics. A teaching-learning-based optimization algorithm was developed to solve the problem. In [14], a deterministic model was developed to minimize power generation cost and to maximize the useful life of batteries without considering renewable generation uncertainties. The problem was solved by a genetic algorithm, and testing was carried out using the actual measured data.

4.2.2 Design of microgrids

In the microgrid/DES operation problem, device types and sizes are given, while operation and maintenance (O&M) costs are not considered. From the long run point of view, determining device types and sizes is also critical. In the literature, mathematical models and optimization methods were developed for optimal design of microgrids or DESs to minimize the total annual cost or lifetime cost [15-24]. Since the design horizon is much longer than the operation horizon, the uncertainties of renewable generation will be averaged out and are usually not considered in the design problem. Also, reliability costs were rarely considered within the design optimization framework.

In the design problem, the operation strategies were usually considered for typical season days, and each day repeats for the entire season [15-19]. To select optimal device sizes based on given types with constant efficiencies, a linear model was developed to minimize the annual cost of a microgrid [15]. The formulation was deterministic, and wind uncertainties were modelled by repeatedly running the deterministic model in a Monte Carlo simulation. The problem was solved by the simplex method. To decide both device types and sizes (with constant efficiencies), a mixed-integer linear problem was developed in [16]. The problem was solved by branch-and-bound combined with the simplex method. Since the energy demand, electricity and gas prices, and the carbon tax rate are uncertain, sensitivity analysis were executed on those factors. To consider the varying device efficiencies with generation levels, a more complex mixed-integer nonlinear model was presented in [17]. Given the nonlinear nature caused by varying efficiencies, after convex underestimation and linearization by introducing new variables, the problem was solved by using branch-and-bound for near-optimal solutions.

In some studies on design, more typical days in a year were considered. For simplicity, heuristic operation strategies were used, where operation optimization was not involved in the design framework [19-24]. To minimize total lifetime costs, exhaustive search with the limited number of possible combinations of device types and sizes was used to find the optimal design in [19-22]. Sensitivity analysis was used to explore the impacts of uncertain factors such as load and fuel prices. With the consideration of multiple objectives, the genetic algorithm was used to find the optimal design of microgrids and DESs (without considering thermal energy) in [23, 24]. In [23], a hybrid PV-wind-diesel microgrid with batteries was considered, a multi-objective model was established to minimize the lifetime cost and emissions. Sensitivity analysis was conducted for inflation of diesel fuel prices, acquisition costs of PV panels, and emissions from PV panels. In [24], a similar model was devolved to minimize lifetime costs, CO_2 emissions and unmet load simultaneously, and the problem was also solved by the same method.

4.2.3 Our previous work

To overcome the difficulties caused by scenario-based approaches, a Markovian approach was developed to solve day-ahead unit commitment problems in our previous work [25]. Without considering transmission capacities, wind generation was aggregated and modeled as a Markov chain, where a state represents the wind generation at a particular hour and captures all the past information. Since the number of states increases linearly with that of hours, the complexity is significantly reduced as compared to scenario-based formulations. The detailed complexity comparison among the deterministic approach, stochastic programming and our approach can be found in Subsection V-D of [25]. The problem was effectively solved by using branch-and-cut. In Example 1 of [25], a small system with two units was tested, and the results show the differences between our approach and stochastic programming. In Example 2, a system with 309 units was tested. The optimization and simulation results demonstrate the computational efficiency, the effectiveness to accommodate high level wind penetration, and the ability to capture low-probability high-impact events of. The approach thus represents a new and effective way to address stochastic problems without scenario analysis. Unlike wind, PV generation has day-night and seasonal patterns.

Recently, we also established a mixed-integer linear model for operation optimization of DESs (without batteries) to minimize the energy cost and increase the total exergy efficiency of a DES [26]. Without considering renewable uncertainties, the deterministic problem was solved by branch-and-cut. To reduce the energy cost and CO_2 emissions, a similar model was developed in [27], and the problem was solved by the same method. In [28], a more complicated mixed-integer model was developed to reduce the energy cost and exergy losses at the energy conversion step, which accounts for the largest part of the total exergy loss in the whole energy-supply chain. The surrogate Lagrangian relaxation method was used to solve the problem.

4.3 The Operation Problem

In this subsection, the operation problem is described in Subsubsection 4.3.1. The mathematical formulation is established in Subsubsection 4.3.2. Based on the problem characteristics, the solution methodology is briefly presented in Subsubsection 4.3.3.

4.3.1 Problem Description

For operation, the microgrid under consideration involves different distributed energy devices as shown in Figure 4.1. The CCHP system consists of multiple gas turbines and heat recovery steam generators, a steam-driven absorption chiller, and a heat exchanger, as sketched inside dashed lines. Electrical load and electricity required by electric chillers can be satisfied by grid power, CCHP, PV panels, and batteries. The microgrid can also sell extra electricity back to the grid. The electric and steam-driven chillers are used for space cooling, while steam and natural gas boilers for space heating. Domestic hot water load can be met by steam through the heat exchanger with sufficient exhaust heat from power generation. From the environmental point of view, combustion of natural gas in CCHP and boilers causes CO_2 emissions.

Based on [15], consider the daily operation of a microgrid over 24 (T) hours with each hour indexed by t ($1 \leq t \leq T$). For devices, their properties such as cost functions and capacities are assumed known. Energy demand including electricity, space heating/cooling, and domestic hot water is also assumed known at hour t. The operation problem is to decide the device operation strategies such as on/off statuses and generation levels to reduce the total energy and emission costs while meeting the given time-varying demand and satisfying individual device constraints.

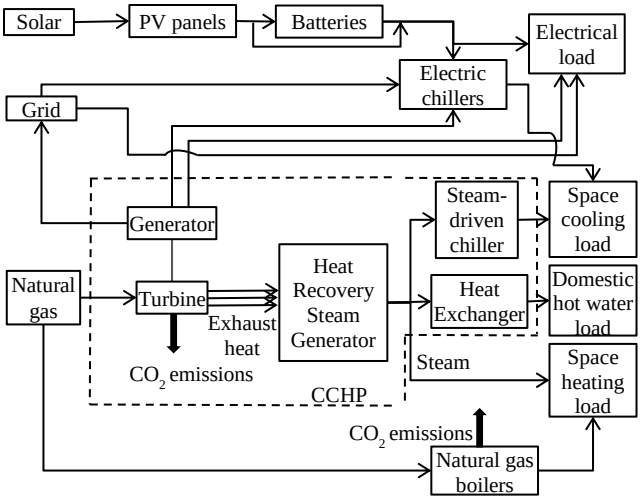


Figure 4.1. Configuration of the microgrid under consideration

4.3.2 Problem Formulation

For microgrid operation under the grid-connected mode, a mixed-integer model is established from the energy and emission point of view. Modeling of devices is presented in Subsubsection 4.3.2.1, and the focus is on PV generation since the intermittent nature of renewables is a major challenge in modeling. System balance is formulated in Subsubsection 4.3.2.2. The objective function is discussed in Subsubsection 4.3.2.3.

4.3.2.1 Modeling of devices

As mentioned earlier, this work is on unit commitment and economic dispatch, and device modeling focuses on on/off statuses and generation levels as in [3-9]. For simplicity, device efficiencies are assumed constant, although they generally depend on generation levels. This fixed-efficiency assumption has often been used in the literature for microgrid design and operation optimization to maintain problem linearity [3, 15]. Modeling of CCHP, boilers, chillers, PV, and battery is presented as follows, and constraints generally include capacity, energy consumption and emissions.

1) Modeling of CCHP [29]

In CCHP, gas turbines are used to meet electrical load by natural gas, while the fossil fuel combustion causes emissions. Then, exhaust heat is recovered in heat recovery steam generators, and the high-

temperature steam could be directly used for space heating, or sent to the absorption chiller and heat exchanger for space cooling and domestic hot water, respectively. Constraints for CCHP are presented below.

Capacity constraints of gas turbines: The generation level of the m^{th} gas turbine $P_m^{GT}(t)$ (continuous decision) should be within its minimum $P_m^{GT,min}(t)$ and maximum $P_m^{GT,max}(t)$ if the device is on (on/off binary decision $x_m^{GT}(t) = 1$), i.e.,

$$P_m^{GT,\min} x_m^{GT}(t) \le P_m^{GT}(t) \le P_m^{GT,\max} x_m^{GT}(t). \tag{4.1}$$

For other devices, this constraint is omitted.

Gas consumption of gas turbines: The amount of natural gas needed in the m^{th} gas turbine $G_m^{GT}(t)$ is calculated as follows:

$$G_m^{GT}(t) = P_m^{GT}(t) / (\eta^{e,GT} HV^{Gas}), \tag{4.2}$$

where $\eta^{e,GT}$ is the gas-to-electric efficiency, and HV^{gas} is the heat value of natural gas.

CO_2 emissions of gas turbines: The amount of CO_2 due to the natural gas combustion in the m^{th} gas turbine $Env_m^{GT}(t)$ is:

$$Env_m^{GT}(t) = G_m^{GT}(t) HV^{Gas} G^{cin}, \tag{4.3}$$

where $G^{cin}(t)$ denotes the carbon intensity of natural gas.

Heat of exhaust gas in turbines: The amount of heat contained in the exhaust gas from the m^{th} gas turbine $Q_m^{GT}(t)$ is:

$$Q_m^{GT}(t) = P_m^{GT}(t) \eta^{th,GT} / \eta^{e,GT}, \tag{4.4}$$

where $\eta^{th,GT}$ is the thermal efficiency of the gas turbine.

Total steam: Steam generated by all steam generators could be directly used for space heating $Q^{Steam\text{-}SH}(t)$, sent to the absorption chiller for space cooling $Q^{Steam\text{-}SC}(t)$, or sent to the heat exchanger for domestic hot water $Q^{Steam\text{-}DHW}(t)$, i.e.,

$$\sum_m Q_m^{GT}(t)\eta^{HRSG} = Q^{Steam-SH}(t) + Q^{Steam-SC}(t) + Q^{Steam-DHW}(t), \tag{4.5}$$

where η^{HRSG} is the energy efficiency of the steam generator.

Heat in the heat exchanger: The amount of heat provided by the heat exchanger for demotic hot water $H^{HE\text{-}DHW}(t)$ is:

$$H^{HE-DHW}(t) = Q^{Steam-DHW}(t)\eta^{HE}, \tag{4.6}$$

where η^{HE} is the efficiency of the heat exchanger.

Cooling in the absorption chiller: The amount of cooling provided by the absorption chiller $C^{SChiller}(t)$ is:

$$C^{SChiller}(t) = Q^{Steam-SC}(t)\eta^{HR,SChiller}COP^{SChiller}, \tag{4.7}$$

where $\eta^{HR,SChiller}$ and $COP^{SChiller}$ denote the heat recovery efficiency and coefficient of performance of the chiller.

2) Modeling of natural gas boilers

Gas consumption of boilers: The amount of gas needed in the n^{th} natural gas boiler $G_n^{boiler}(t)$ is calculated as follows:

$$G_n^{boiler}(t) = H_n^{boiler}(t) / (\eta^{boiler} HV^{Gas}), \tag{4.8}$$

where $H_n^{boiler}(t)$ denotes the heat generation level of the boiler, and η^{boiler} is the efficiency of the boiler. The modeling of CO_2 emissions $Env_n^{boiler}(t)$ is similar to that of gas turbines.

3) Modeling of electric chillers

Electricity consumption of electric chillers: The electricity required by the l^{th} electric chiller $P_l^{EChiller}(t)$ is:

$$P_l^{EChiller}(t) = C_l^{EChiller}(t) / (\eta^{EChiller} COP^{EChiller}), \tag{4.9}$$

where $C_l^{EChiller}(t)$ denotes the amount of cooling generated in the electric chiller, and $\eta^{EChiller}$ and $COP^{EChiller}$ denote the efficiency and coefficient of performance of the chiller.

4) Modeling of PV generation

Unlike wind, an important characteristic of the solar behavior is the day-night pattern, i.e., PV generation is zero when the sun is not shining [30]. A sinusoidal wave with zero values for darkness hours is a good approximation to this behavior [31], while its amplitude and scale factors depend on the hours of sunshine and maximum power output. In addition, PV generation also has a seasonal behavior [30], since the position of the sun influences the incidence of the solar rays on PV panels.

To avoid the computational complexity caused by scenario -based methods as discussed in Section II, a Markov-based model is established to integrate intermittent and uncertain PV generation into microgrids based on our early work for wind [25]. In the model, the PV generation is assumed to be a discrete Markov process, following related real case studies [32, 33]. The capacity of PV generation is evenly divided into *N* intervals, and states are defined as minimum values of each interval, arranged ascendingly. Based on historical data, the probability that the current state is *j* if the previous PV state was *i* can be obtained as follows [34],

$$P_{ij} = \frac{\text{observed transitions from state } i \text{ to } j}{\text{ocurrences of state } i}. \tag{4.10}$$

In this way, the state transition matrix P^{ST} can be established.

To solve the problem for a specific region, the historical PV generation data should be analyzed to determine the number of states N as a balance between modeling accuracy and computation efficiency. The

transition matrix should be also updated by incorporating the latest weather forecast. Because of the seasonal behaviors, a state transition matrix is needed for each season.

Then the probability that the PV generation is P_i^{PV} at time *t*, denoted as $\varphi_i(t)$, is the sum of probabilities at time *t-1* weighted by different transitions:

$$\varphi_i(t) = \sum_{j=1}^{N} P_{ji}\varphi_j(t-1). \tag{4.11}$$

The probabilities of PV generation levels for future time slots can be obtained based on the initial PV generation state and the transition matrix.

The generation level of the *i* interval is denoted as P_i^{PV}. With $\omega^{PV}(t)$ defined as the weight factor to reduce PV generation to zero during darkness hours, the adjusted PV generation $P^{AdPV}{}_i(t)$ is obtained as follows:

$$P_i^{AdPV}(t) = \omega_i(t)P_i^{PV}(t). \tag{4.12}$$

5) Modeling of battery

To capture state dynamics, a simplified battery model is used here, assuming charging/discharging efficiencies are 100%. Battery charge and discharge is extended to depend on PV states. The state of charge at time *t* under PV state *i* is denoted as $P_i^{Bat}(t)$. The standard one dimensional state equation on state of change in the literature is extended to two dimensional on the state of change and PV states as follows,

$$P_i^{Bat}(t+1) = P_j^{Bat}(t) + P_i^{bc}(t) - P_i^{bd}(t), \forall j, \forall i \in \{i \mid \varphi_i(t) \neq 0\}, \tag{4.13}$$

where $\varphi_i(t)$ is the probability that the PV generation is P_i^{PV} at time *t*, which can be calculated as the sum of probabilities at time *t-1* weighted by different transitions. In addition, the battery cannot be charged and discharged simultaneously.

6) Modeling of CCHP, boilers and chillers based on PV states

For effective coordination, other devices are correspondingly modeled as Markov processes with states depending on the states of PV generation. The generation levels of CCHP, boilers and electric chillers, and the amount of grid power are therefore modeled to depend PV states. Take the m^{th} gas turbine in CCHP as an example. For each PV state i, there is a corresponding generation level $P_{m,i}^{GT}(t)$ (continuous decision). The other devices are modeled in a similar way.

4.3.2.2 Modeling of system balance

The electrical grid is simplified by modeling electricity balance as in the literature [3-9], where detailed electrical power models are not considered.

1) Electricity balance: In the microgrid, the summation of electricity generated by PV panels and CCHP, discharged by the battery and bought from the grid equals the summation of electricity demand, electricity consumed by electric chillers, sold to the grid, and stored. Similar to other devices, the amount of electricity from or to the grid is also modeled to depend on PV states for effective coordination. The electricity balance constraint should be satisfied at every hour for each PV state where its probability is nonzero, i.e.,

$$P_i^{AdPV}(t)+\sum_m P_{m,i}^{GT}(t)+P_i^{bd}(t)+P_i^{buy}(t)=P^{dem}(t)+\sum_l P_{l,i}^{EChiller}(t)+P_i^{sell}(t)+P_i^{bc}(t),\forall i\in\{i\mid\varphi_i(t)\neq 0\}. \quad (4.14)$$

In the above, the new decision variables are $P_i^{buy}(t)$ and $P_i^{sell}(t)$, the amount of electricity bought from and sold to the grid, respectively. The demand $P^{dem}(t)$ is assumed given.

2) Thermal balance. For space heating, the summation of heat generated by natural gas boilers and provided by steam equals the demand, i.e.,

$$\sum_n H_{n,i}^{boiler}(t)+Q_i^{Steam-SH}(t)=H^{dem-SH}(t),\forall i\in\{i\mid\varphi_i(t)\neq 0\}, \quad (4.15)$$

where the demand $H^{dem\text{-}SH}(t)$ is assumed given.

Thermal balance for space cooling and domestic hot water is formulated in a similar way.

The entire problem is therefore stochastic and Markovian.

4.3.2.3 Objective function

The objective is to minimize the total daily cost, i.e., energy and emission costs. The energy cost *Cost* consists of three terms, buying natural gas from the station and electricity from the grid, and selling electricity back to the grid, i.e.,

$$Cost = \sum_t \sum_i \varphi_i(t) \left(C^{Gas} \times \left(\sum_m G_{m,i}^{GT}(t) + \sum_n G_{n,i}^{boiler}(t) \right) + C^{Grid,buy}(t) \times P_i^{buy}(t) - C^{Grid,sell}(t) \times P_i^{sell}(t) \right) \cdot \Delta t, \quad (4.16)$$

where $C^{Grid,buy}(t)$ and $C^{Grid,sell}(t)$ denote the unit price of electricity from and to the grid at time t, respectively; C^{Gas} is the unit price of natural gas; and Δt is the time slot length.

To quantify the cost of CO_2 emissions caused by the natural gas combustion in gas turbines and boilers, the carbon tax *CarbonTax* is considered here [35], i.e.,

$$CarbonTax = P^{CTax} \sum_t \sum_i \varphi_i(t) \left(\sum_m Env_{m,i}^{GT}(t) + \sum_n Env_{n,i}^{boiler}(t) \right), \quad (4.17)$$

where P^{CTax} denotes the carbon tax on CO_2 emissions ($/kg). Since the carbon tax associated with grid power generation is already reflected in the grid price, it is not involved here.

Based on the above, the overall objective to be minimized is *Cost* + *CarbonTax*.

4.3.3 Solution Methodology

The problem formulated above is stochastic and linear, and involves both discrete and continuous variables. Branch-and-cut, which is powerful for mixed-integer linear problems, is therefore used to solve the problem. In the method, all integrality requirements on variables are first relaxed, and the relaxed problem can be efficiently solved by using a linear programming method. The solution also provides a lower bound. If the values of all integer decision variables turn out to be integers, the solution of the relaxed problem is optimal to the original problem. If not, valid cuts that do not cut off feasible integer solutions are added,

trying to obtain the convex hull. Once the convex hull is obtained, the values of all integer decision variables in the solution to the relaxed problem are integers, and this solution is optimal to the original problem. If the convex hull cannot be obtained by adding cuts, low-efficient branching operations are needed. Optimization stops when computational time reaches the pre-set stop time or the relative gap (relative difference between the objectives of the optimal relaxed solution and current integer solution) falls below the pre-set gap [36].

4.4 Design Problem

In this subsection, the design problem is described in Subsubsection 4.4.1. The mathematical formulation is established in Subsubsection 4.4.2. The solution methodology are presented in Subsubsection 4.4.3.

4.4.1 Problem Description

The operation problem in Subsection 4.3 is to decide daily operation strategies of microgrids with fixed device types and sizes to reduce the total daily cost in the short run. While the design problem is to decide device sizes with given types to reduce the lifetime cost in the long run. It consists of the reliability cost and the classic components including capital, replacement, O&M, fuel, and emission costs. The O&M, fuel, emission costs are based on daily operation of devices. For simplicity, four typical season days with heuristic operation strategies are considered, while operation optimization is not involved in the design framework.

Consider the design problem for a microgrid over its entire lifetime, N years with each year indexed by t ($1 \leq t \leq N$). Devices include CCHP, natural gas boilers, electrical chillers, PV panels, and batteries, where their properties such as capital costs, lifetimes, and efficiencies are assumed known. Electricity, space heating/cooling, and domestic hot water demand for four typical season days is assumed known.

4.4.2 Problem Formulation

Since the time horizon for design is much longer than that for operation, a linear model is established in this subsection. Modeling of devices is discussed in Subsubsection 4.4.2.1. The focus is on the lifetime cost, and the modeling of daily operation is simplified since it is consistent with that in the operation problem. System balance is briefly presented in Subsubsection 4.4.2.2. The reliability cost is discussed in detail in Subsubsection 4.4.2.3. The objective function is described in Subsubsection 4.4.2.4.

4.4.2.1 Modeling of devices

The device modeling includes four parts, i.e., costs, operation constraints, energy consumption, and emissions. The associated cost includes the capital, replacement, O&M, and fuel costs and carbon tax. Because the device lifetime may not be consistent with the microgrid lifetime, the salvage value, i.e., the remaining value at the end of the project lifetime, is also taken into account. The design problem is usually over 20 years, therefore discounting and inflation has to be considered. Based on the above, the net present cost (NPC) of a device is the present value of all the costs over the project lifetime, minus its salvage value. Consider a gas turbine in CCHP as an example. For illustration purposes, it is assumed that the lifetime of the gas turbine is longer than that of the microgrid, and there are no replacement costs. The net present cost C_{GT}^{NPC} is calculated as follow,

$$C_{GT}^{NPC} = C_{GT}^{Cap} - C_{GT}^{Sal} / (1+i)^{N} + \sum_{t=1}^{T} \left(C_{GT,t}^{O\&M} + C_{GT,t}^{Fuel} + C_{GT,t}^{CTax} \right) / (1+i)^{t}, \tag{4.18}$$

where C_{GT}^{Cap} is the capital cost; C_{GT}^{Sal} is the salvage value; $C_{GT,i}^{O\&M}$ is the O&M cost of the t^{th} year; $C_{GT,i}^{Fuel}$ and $C_{GT,i}^{CTax}$ are the fuel cost and carbon tax based on fuel consumption, respectively; and i is the discount rate. With the given nominal discount rate i' (the rate at which money is borrowed) and the expected inflation rate f, the real discount rate i can be obtained as follows,

$$i = (i' - f) / (1 + f). \tag{4.19}$$

The operation constraints for the gas turbine mainly include the capacity and heat recovery constraints

as in Eq. (4.1) and (4.4), specifying the generation limit and available heat recovered; and energy consumption and emissions as in Eq. (4.2) and (4.3). In addition, the operation lifetime constraint is also considered. The other devices can be modeled in a similar way. The utility grid can also be treated as a device to the microgrid. Its net present cost includes 1) the capital cost as the interconnection cost for the microgrid to connect to the grid (e.g., device and installation costs), 2) the fuel cost as buying electricity from the grid; and 3) the revenue by selling electricity to the grid. PV uncertainties are not considered in the design problem. This is because the uncertainties will be averaged out as the design time horizon is much longer than the operation one.

4.4.2.2 Modeling of system balance

In the design problem, the given electricity and thermal demand has to be satisfied for each time slot of the entire microgrid lifetime as in the operation problem.

4.4.2.3 Reliability cost

The reliability cost includes the capital and replacement costs of fault protection devices such as circuit breakers and fuses to protect the microgrid from the faults coming from the utility grid, and the cost of unserved load during power outages. Here, two types of protection devices are considered. The first type is for synchronized connection with the utility grid such as Current Limiting Protector (CLiP) [37], and the other is for non-synchronized connection such as GridLink [38]. For the first type: if the generators of the microgrid fail, the utility grid provides power immediately; and if the utility grid fails, the protection device trips and the power outage occurs. The cost of unserved load can be calculated as the product of the quantity of unserved load $P^{dem,avg}$ (average load), the interruption duration $T^{Interrupt}$, and estimated interruption cost $C^{Interrupt}$ ($/kW) [39], i.e.,

$$C^{Unserved} = P^{dem,avg} T^{Interrupt} C^{Interrupt}. \tag{4.20}$$

For the second protection device: the power outage only occurs when the generators of the microgrid and the utility grid both fail, whose probability is negligible.

For simplicity, it is assumed that the lifetimes of protection devices are the same as the microgrid's. For illustration purposes, let the generators of the microgrid fail f_1 times per year and the utility grid fail f_2 times per year. Power outages can be categorized into different types according to different causes. For simplicity, they are categorized into major power outages and general power outages, while the former ones have longer restoration time. It is assumed that p (%) of the utility grid power outages are major ones with an average restoration time T_1, and the remaining ones are with an average restoration time of T_2. Let C^1 ($/kW) denote the capital cost of the synchronized protection device, and C^2 ($) and C^3 ($/kW) the replacement costs of fuses and the device. For non-synchronized one, let C^4 ($/kW) denote its capital cost. For simplicity, the cost of other related devices such as transformers, switchgear, circuit breakers and protection relay are not considered. This is justified by that the cost of these devices associated with the synchronized grid-connected microgrid is much higher than that associated with the non-synchronized grid-connected one [40]. With the above data, the reliability costs with the two types of protection devices are compared in Table 4.1.

Table 4.1. Reliability cost comparison

	Synchronized grid-connection	Non-synchronized grid-connection
C^{Cap}	$C^1 \times \sum_m P_m^{GT,Max}$	$C^4 \times \sum_m P_m^{GT,Max}$
$C^{Replace}$	Fuse: C^2 per replacement + Device: $C^3 \times \sum_m P_m^{GT,Max\ x}$	/
$C^{Unserved}$ (per power outage)	$P^{dem,avg} \times (pT_1 + (1 - p)T_2) \times C^{Interrupt}$	/
$C^{Reliability}$ (total)	$C^1 \sum_m P_m^{GT,Max} + \sum_{t=1}^{N} \frac{f_2 \left(c_p^{Replace} + c_p^{Unserved} \right)}{(1+i)^t}$	$C^4 \sum_m P_m^{GT,Max}$

4.4.2.4 Objective function

The objective of the design problem is to minimize the lifetime cost of the microgrid, i.e., sum of net present costs of all devices indexed by d, and the net present reliability cost over the lifetime, i.e.,

$$C^{Lifetime} = \sum_{d} C_{d}^{NPC} + C^{Reliability}. \quad (4.21)$$

4.4.3 Problem Description

The problem formulated above is linear and involves both discrete and continuous variables. Since its complexity increases exponentially as the number of devices sizes increase, only a limited number of possible combinations of device sizes can be considered. Our idea is to select a certain number of choices for each device based on load profiles. By applying heuristic operation strategies for distributed devices and grid power, the total net present cost of devices under different configurations are evaluated. Then the net present reliability cost is estimated on the top of it. In this way, the total lifetime costs of different configurations are obtained. In addition, the impacts of uncertain factors such as fuel price and load growth are analyzed through sensitivity analysis.

4.5 Numerical Results

The method presented above for microgrid operation optimization has been implemented by using IBM ILOG CPLEX Optimization Studio V 12.6.0.0 [36]. The method for design optimization has been implemented by using HOMER Pro [41]. Testing has been performed on a PC with 2.90GHz Intel Core(TM) i7 CPU and 16G RAM. Two examples are presented. The first small classroom example is to illustrate the Markov-based modeling of PV generation in the operation problem, and to show different components of the lifetime cost in design. The second semi-realistic one based on the Kings Plaza microgrid in Brooklyn of New York is to show that the operation method is efficient in saving cost and computation time, and scalable. It is also to compare lifetime costs of different design configurations and show impacts of uncertain factors in design.

4.5.1 Example 1.

In this classroom example, a small microgrid is considered. Devices include PV, CCHP (a gas turbine, a steam generator, a steam-driven absorption chiller, and a heat exchanger), an electric chiller, and a boiler. Electricity and natural gas can be brought from the grid and station with sufficient capacities. The 10-state transition matrix for PV generation is obtained from [29]. The time-varying grid price is taken form [42], where the selling-back price to the grid is set as 90% of the grid price. The natural gas price is obtained from [43]. The carbon intensity of natural gas is based on [44], and the carbon tax is taken from [35, the carbon tax for year 2015]. The results for the operation and design problems are presented in Subsubsections 4.5.1.1 and 4.5.1.2, respectively.

4.5.1.1 Results for the operation problem

The operation problem is for 9 am of a representative summer day in July, where the capacity for the gas turbine is 1,600 kW, and 200 kW for PV. The optimization problem is solved in about 1.5 seconds. For comparison purposes, an isolated DES with the same energy devices and not connected to the utility grid is considered. In addition, a conventional energy system with an electric chiller for space cooling, an electric heater for space heating, and an electric boiler for domestic hot water is also considered, where all types of demand are satisfied by grid power directly or indirectly.

For the microgrid, the total cost is $-73.35, while the energy cost is $-89.84 and the carbon tax is $16.49. The energy cost is negative, which means that the microgrid makes profits by selling electricity back to the grid. For the isolated DES, the total cost is $98.69, while the energy cost is $86.16 and the carbon tax is $12.53. This total cost is much higher than that of the microgrid since it cannot buy electricity from the grid, or sell electricity to the grid. For the conventional energy system, the total cost is $528.68 (no carbon tax), which is much higher than those of the microgrid and DES. This is because during the daytime of summer, the price of grid power is much higher than that of the electricity from CCHP. For these three types of energy systems, the PV generation levels, gas turbine (GT) generation levels, and grid

input under different PV states are shown in Table 4.2. In the table, grid input equals the amount of electricity from the grid minus the amount of electricity to the grid. While the expected device generation levels and grid input of the microgrid and the isolated DES are show in Figure 4.2 below.

Table 4.2. Ex1: PV and gas turbine generation levels, and grid input under different PV states

State	PV gen (kWh)	Microgrid		Isolated DES		Conventional
		GT gen (kWh)	Grid input (kWh)	GT gen (kWh)	Grid input (kWh)	Grid input (kWh)
1	0	1600	-600	1226.43	0	2173.5
2	14.14	1600	-614.14	1218.50	0	
3	28.28	1600	-628.28	1210.57	0	
4	42.43	1600	-642.43	1202.64	0	
5	56.57	1600	-656.57	1194.70	0	
6	70.71	1600	-670.71	1186.77	0	
7	84.85	1600	-684.85	1178.84	0	
8	98.99	1600	-698.99	1170.91	0	
9	113.14	1600	-713.14	1162.97	0	
10	127.28	1600	-727.28	1155.04	0	

Based on Table 4.2, for the microgrid, the gas turbine is always working at the maximum capacity with respect to all PV states. The grid input is negative under all PV states, implying that the microgrid sells electricity back to the grid. When PV generation increases, the amount of electricity to the grid increases since more load is covered by PV. For the isolated DES, there is no grid input. As PV generation increases, the amount of electricity generated by the gas turbine decreases. For the conventional energy system, the grid input is much higher than those of the other two energy systems since all types of demand are coved by grid power.

According to Figure 4.2, for the microgrid, the gas turbine generates more electricity to cover load and sell electricity back to the grid. Space cooling demand is satisfied by the stream-driven chiller with sufficient stream from exhaust heat. For the isolated DES, the gas turbine generates less electricity, just to cover load and electricity required by the electric chiller. Since the steam is not enough, the electric chiller

is used to satisfy the space cooling demand. For domestic hot water, since it can only be met by steam, operation strategies of the heat exchanger are the same for the two energy systems.

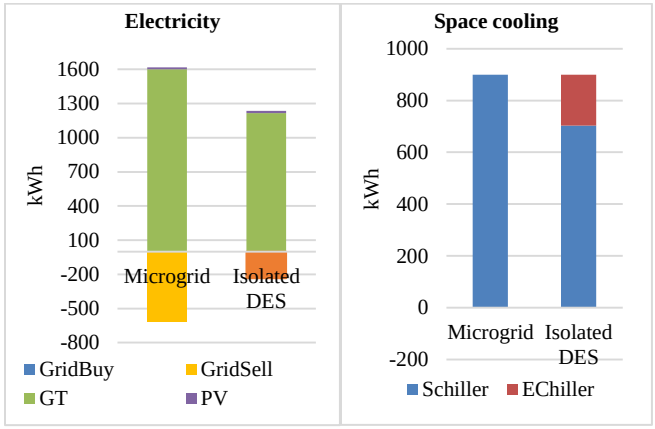


Figure 4.2. Ex1: Expected device generation levels and grid input

4.5.1.2 Results for the design problem

In HOMER Pro, there are no cooling related devices, such as absorption chillers or electric chillers. While the cooling provided by the absorption chiller can be converted to the thermal load required by the chiller, and the cooling provided by the electric chiller can be converted to the electrical load required by the chiller. The capital and O&M costs of absorption and electric chillers are ignored. In this example, the electricity and thermal loads are scaled from the monthly load profiles provide by HOMER Pro. To make the load data more realistic, it is assumed that the load has an 8% day-to-day variation in each month and an 18% time step-to-step variation in each day. Based on the load profiles, the capacity range for the gas turbine is selected from1,600 kW to 2,000 kW with a 50 kW increase. For PV panels, the capacity range is selected from 100 kW to 200 kW with a 20 kW increase. The cost related data of gas-turbines and PV is obtained from [45] and [46], while the energy efficiencies are chosen among typical values. The time-varying grid price and the natural gas price are the same as in the operation problem. The length of the microgrid lifetime is assumed as 20 years. The nominal discount rate i' and the expected inflation rate f are 4.98% and 1.68% (the average values of the monthly interest and inflation rates in the past 6 years [47, 48]), respectively.

Then the real discount rate i is 3.25% based on Eq. (4.19).

To calculate the reliability costs, it is assumed that the microgrid generator fails 6 times per year [49] and the utility grid fails 1.5 times per year [50]. Based on outage records of Northeast Utilities, 38% of the power outage was caused by wind storms with an average restoration time of 8 hours, and the resting has an average restoration time of 2 hours [51]. For comparison purposes, the isolated DES and the conventional energy system mentioned in the operation problem are also considered here. For the isolated DES, the interruption duration is assumed as 4 hours [52].

Total lifetime costs

The total net present costs of devices under different configurations of the microgrid under consideration are evaluated in HOMER Pro, while the net present reliability costs are estimated on the top of it. Then, the lifetime cost of each selected configuration is obtained as shown in Table 4.3, as well as those of the isolated DES and the conventional energy system with specific configurations (the capital and O&M costs of electrical devices in the conventional system are ignored). For illustration purposes, the results for configurations with the gas turbine of 1,600 kW, 1,800 kW and 2,000 kW, and PV of 100 kW and 200 kW are presented.

Table 4.3. Ex1: Total lifetime costs of different microgrid configurations

Type	GT + PV (kW)	Grid-connection	$C^{Cap}+C^{Replace}$ (M$)	$C^{O\&M}$ (M$)	C^{Fuel} (M$)	C^{Grid} (M$)	C^{CTax} (M$)	$C_p^{Reliability}$ (M$)	C^{NPC} (M$)
Microgrid	1600+100	Syn	9.12	2.07	10.31	-11.28	1.5	2.71	14.28
		Non	9.12	2.07	10.31	-11.28	1.5	1.2	12.77
	1600+200	Syn	9.33	2.09	10.31	-11.56	1.5	2.71	14.21
		Non	9.33	2.09	10.31	-11.56	1.5	1.2	12.71
	1800+100	Syn	9.98	2.32	11.56	-13.83	1.68	2.85	14.41
		Non	9.98	2.32	11.56	-13.83	1.68	1.35	12.91
	1800+200	Syn	10.19	2.35	11.56	-14.11	1.68	2.85	14.35
		Non	10.19	2.35	11.56	-14.11	1.68	1.35	12.85
	2000+100	Syn	10.83	2.58	12.81	-16.37	1.86	2.99	14.54
		Non	10.83	2.58	12.81	-16.37	1.86	1.5	13.05
	2000+200	Syn	11.05	2.6	12.81	-16.65	1.86	2.99	14.48

		Non	11.05	2.6	12.81	-16.65	1.86	1.5	12.99
Isolated DES	1600+200	/	7.33	2.07	5.13	0	0.77	2.19	17.33
Conventional system	/	/	0	0	0	20.26	0	1.3	21.56

Each configuration of the microgrid under consideration has a lower lifetime cost than those of the isolated DES and the conventional system, no matter synchronized or non-synchronized grid-connection, although there is a high grid interconnection cost ($2M, scaled from [53]). The conventional system has the highest total lifetime cost. This is mainly because the microgrid can make profits by selling electricity to the grid when the grid price is high. For the same configuration of the microgrid, the lifetime cost with non-synchronized grid-connection is lower than that under synchronized grid-connection, which will be explained later. The best configuration with the lowest lifetime cost is the one with a gas turbine of 1,600kW and PV panels of 200 kW.

Reliability costs

To show how the total reliability costs are calculated in Table 4.3, the details are discussed here. For the microgrid under consideration, each part in the total reliability costs is explained in Table 4.1. For the isolated DES, no protection devices are needed to prevent the faults coming from the utility grid. When the generator goes down and there is no power supply, costs of unserved load occur, which can be calculated by Eq.(4.20). For the conventional energy system, since there are no distributed generators, no protection devices are needed, either. When the grid goes down, there will be costs of unserved load. The reliability costs for the microgrid with a gas turbine of 1,600kW and PV panels of 200 kW under two types of grid-connection, and those of the isolated DES (same device configuration as the microgrid) and the conventional system with no protection devices are compared in Table 4.4 as follows.

Table 4.4. Ex1: Reliability costs for different energy systems

	Synchronized connection	Non-Synchronized connection	Isolated DES	Conventional system
C^{Cap}	79.5 [54] × 1600 = $0.127M	750 [56] × 1600	/	/

		= $1.2M		
$C^{Replace}$	Fuse: $0.045M per replacement [54] Device: 29 [54] × 1600=$0.0464M	/	/	/
$C^{Unserved}$ (per power outage)	518 × (38% × 8+ 62% × 2) ×12.1 ([55])=$0.027M	/	518 × 4 × 12.1 = $0.025M	1164 × (38% × 8+ 62% × 2) × 12.1 = $0.06M
$C^{Reliability}$ (total)	$2.71M	$1.2M	$2.19M	$1.3M

Under synchronized grid-connection, the total reliability cost is $2.71M. Under non-synchronized grid-connection, the total reliability cost is $1.2M. Although the capital cost of the protection device for non-synchronized connection is very high as compared to that for synchronized connection, the cost of unserved load is 0. For the isolated DES, since there is no protection device, its total reliability cost is much lower than that of the microgrid. The conventional system has a higher reliability cost than the isolated DES since the utility grid is assumed to fail more frequently than generators, and the average electrical load of the conventional system is higher as all types of demand are satisfied by grid power.

4.5.2 Example 2.

This example is semi-realistic based on the microgrid of Kings Plaza in Brooklyn, NY, U.S. It is to show that the total energy and emission costs can be reduced by the optimized operation of the microgrid, and to compare lifetime costs of different design configurations and show impacts of uncertain factors in design. In this example, all devices mentioned in Section III are considered. The cost related data is the same as the first example. The hourly electricity, space heating/ cooling, and demand domestic hot water demand of four representative days is built based on [57-60]. For each representative season day, the hourly energy demand is calculated as the average of the energy demand in the corresponding hour of all days in this season. The stop mixed-integer programming (MIP) gap is 0.5% for the operation problem. The results for the operation and design problems are presented in Subsubsections 4.5.2.1 and 4.5.2.2, respectively.

4.5.2.1 Results for the operation problem

In the operation problem, the total capacity for gas turbines is 6,400 kW, 500 kW for PV, and 500 kW for the battery. For a winter day with one hour as a time interval, the daily cost of the microgrid under consideration is 3,583 $/day, while the energy cost is 2,436 $/day whereas the carbon tax is 1,147 $/day. For comparison purposes, an isolated DES with the same energy devices is considered, as well as the conventional energy system mentioned in Example 1. For the isolated DES, the daily cost is 6,264 $/day, while the energy cost is 5,469 $/day and the carbon tax is 795 $/day. For the conventional system, the daily cost is 13,663 $/day (no carbon tax). Among the three energy systems, the microgrid has the lowest daily cost by using grid power or distributed energy devices whatever is cheaper.

To analyze the optimized operation strategies, a particular scenario of PV generation is presented in the following. To demonstrate that the total energy and emission costs can be reduced by the optimized operation of the microgrid, the total costs of each representative season day under the optimized and heuristic operation are compared. The Monte Carlo simulation is also performed on these four days.

Optimized operation strategies for a particular scenario

The particular scenario in the typical winter day is selected, where the PV generation at each time is at state 5. By solving the operation optimization problem for this scenario, the hourly grid power price, electrical load, electricity provided by CCHP, and grid input are shown in Figure 4.3 below.

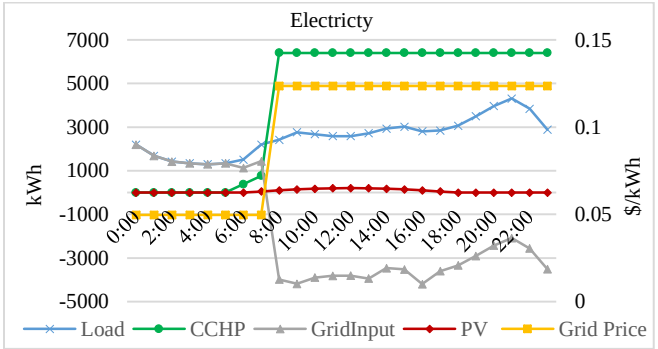


Figure 4.3. Ex2: Hourly grid price, electrical load, electricity provided by CCHP, and grid input

When the grid price is low, e.g., from 0:00 to 5:00, the grid input is positive and the microgrid buys

electricity to cover all the electrical load. During 6:00 to 7:00, CCHP begins to generate electricity and the grid input decreases. When the grid price is high since 8:00, CCHP generates electricity at its maximum capacity to cover the load and to sell to the grid, so the grid input is negative. In addition, PV panels also cover partial electrical load during the day time.

Optimized and heuristic operation strategies

To evaluate the optimized operation strategies, the heuristic operation strategies are also considered. For CCHP, the selected heuristic strategies are as follows: four engines during day time in summer and three in other seasons; and two engines on from 23:00 to 7:00. For battery, it is assumed that it charges during day time and discharges at night under the heuristic operation strategy. For a typical day in each season, the daily costs with the optimized and heuristic operation strategies are compared in Table 4.5. The CPU time is also presented.

Table 4.5. Ex2: Optimized and heuristic operation of the microgrid

Season	Operation strategy	Energy cost ($)	Carbon tax ($)	Total cost ($)	CPU time (s)
Spring	Optimized	2,117	1,097	3,214	8.2
	Heuristic	3,631	1,008	4,639	/
Summer	Optimized	-2,856	1,117	-1,739	9.3
	Heuristic	-2,011	1,241	-770	/
Fall	Optimized	2,532	1,117	3,649	9.5
	Heuristic	4,441	993	5,434	/
Winter	Optimized	2,436	1,147	3,583	8.6
	Heuristic	4,049	1,073	5,122	/

For each season, the daily cost is reduced by more than 40% under the optimized operation as compared with that obtained under the heuristic operation. The relative difference between the daily costs obtained by the optimized and heuristic operation in summer is the largest among the four seasons. This is because the microgrid makes more profits by the optimized operation in summer when the grid price is very high. For this problem, the CPU time is about 9 seconds, while it is 1.5 seconds for the one-hour problem in Example 1. The computational time is nearly linear to the problem size. Therefore the method is efficient

in saving cost and computation time, and scalable for large microgrids.

Monte Carlo simulation

To evaluate the optimization results, 1,000 Monte Carlo simulation runs are performed with 10-state transition matrices for the four typical season days. Modeling accuracy is measured by the absolute percentage error (APE), the ratio of the absolute difference between optimization and simulation costs to the simulation cost. The standard deviation (STD) of scenario costs reflects its variation. Results of the four typical season days are summarized in Table 4.6.

Table 4.6. Ex2: Simulation results for microgrid operation

Season	Simulation cost ($)	Optimization Cost ($)	APE (%)	STD ($)
Spring	3,213	3,214	0.03	1.21
Summer	-1,752	-1,739	0.76	8.75
Fall	3,646	3,649	0.09	1.74
Winter	3,580	3,583	0.10	1.45

It can be seen that the absolute percentage errors are all within 1% for all seasons, demonstrating the modeling accuracy. The APE and STD in summer are the largest among the four seasons, since the variation of PV generation is the largest in summer in this example.

4.5.2.2 Results for the design problem

For the design problem, the cost related data including the reliability costs is the same as in Example 1. The hourly load is obtained based on the four typical days mentioned in the operation problem above. Similar to Example 1, an 8% day-to-day variation in each season and an 18% time step-to-step variation in each day are considered. Based on the load profiles, the capacity range for each gas turbine is selected from 1,600 kW to 3,200 kW with a 400 kW increase. As for PV panels, its total capacity range is selected from 200 kW to 500 kW with a 50 kW increase. The battery is from 200 kW to 500 kW with a 50 kW increase, and its cost related data is taken from [61]. When the grid price is high, grid power is not allowed to charge the battery, and when the price is low, the battery cannot discharge for selling electricity to the grid. The

lifetime of the microgrid is assumed as 20 years. In this subsection, lifetime costs of different configurations are compared. The impacts of uncertain factors, load and fuel prices, are also discussed.

Total costs

By solving the design problem, the total lifetime costs of selected configurations of the microgrid under consideration are presented in Table 4.7. As in Example 1, it is assumed that each microgrid generator fails 6 times per year [49] and the four generators are parallel. According to fault tree analysis, the overall CHP fails 1.64 times per year [62]. The utility grid fails 1.5 times per year [50]. The total lifetime costs of the isolated DES and the conventional energy system with specific configurations (the capital and O&M costs of electrical devices in the conventional system are ignored) are also considered. For illustration purposes, the configurations with a total gas turbine capacities of 6,400 kW, 9,600 kW, and 12,800kW, PV of 200 kW and 500 kW, and battery of 500 kW are selected and presented. In addition, the configuration with a total gas turbine capacity of 12,800kW, PV panels of 500 kW and battery of 200 kW is also presented.

The comparison among the three systems is similar to that of Example 1. The microgrid has lowest lifetime cost, and the conventional system has the highest one. While for the same microgrid configuration, the lifetime cost with non-synchronized grid-connection is lower than that under synchronized grid-connection. Different to Example 1, the higher the total capacity of gas turbines, the lower the total lifetime cost of the microgrid. This is because the additional profits made by selling electricity to the grid exceeds the additional costs (e.g., O&M, reliability costs), opposite to Example 1. In addition, based on the last two microgrid configurations with different battery sizes, the higher the capacity of the battery, the higher the total lifetime cost of the microgrid. In daily operation, the battery can provide electricity when the sun is covered by clouds to dampen the intermittency of PVs. Here the PV generation is deterministic, where its uncertainties are not involved. Also since heuristic operation strategies are considered in the design problem, the economic benefits of batteries are not fully explored. In summary, the best design with the lowest lifetime cost is the configuration with a total gas turbine capacity of 12,800kW, PV panels of 500

kW, and battery of 200 kW.

Table 4.7. Ex2: Lifetime costs of different microgrid configurations

Type	GT + PV +Battery (kW)	Grid-connection	$C^{Cap}+C^{Replace}$ (M$)	$C^{O\&M}$ (M$)	C^{Fuel} (M$)	C^{Grid} (M$)	C^{CTax} (M$)	$C_p^{Reliability}$ (M$)	C^{NPC} (M$)
Micro grid	6400+200 + 500	Syn	30.60	12.44	27.2	-16.52	3.96	8.34	56.8
		Non	31.33	11.67	28.25	-16.52	3.96	4.8	53.82
	6400+500 + 500	Syn	31.07	12.48	27.09	-17.19	3.94	8.34	56.46
		Non	31.07	12.48	27.09	-17.19	3.94	4.8	52.92
	9600+200 + 500	Syn	38.98	15.70	34.00	-36.89	4.95	10.62	53.85
		Non	38.98	15.70	34.00	-36.89	4.95	7.2	50.43
	9600+500 + 500	Syn	39.45	15.77	33.96	-37.67	4.94	10.62	53.51
		Non	39.45	15.77	33.96	-37.67	4.94	7.2	50.1
	12800+200 + 500	Syn	44.86	19.49	42.01	-59.11	6.11	12.89	51.23
		Non	44.86	19.49	42.01	-59.11	6.11	9.6	47.94
	12800+500 + 500	Syn	45.34	19.57	42.00	-59.94	6.11	12.89	50.9
		Non	45.34	19.57	42.00	-59.94	6.11	9.6	47.61
	12800+500 + 200	Syn	45.24	19.46	42.01	-60.40	6.11	12.89	50.27
		Non	45.24	19.46	42.01	-60.40	6.11	9.6	46.98
Isolated DES	12800+500 + 500	/	33.34	5.40	23.62	0	3.44	2.85	61.99
Conventional system	/	/	0	0	0	70.3	0	4.95	75.22

In the operation problem, for the microgrid with a total gas turbine capacity of 6,400kW, PV panels of 500 kW and battery of 500 kW, the daily costs of each seasons are obtained as shown in Table 4.5. With the interest rate mentioned in Example 1and the length of each season, the sum of the total energy cost and carbon tax over the lifetime is approximated as $10.8M and $18.0M under the optimized and heuristic operation. While in the design problem, this number is $13.8M ($C^{Fuel} + C^{Grid} + C^{CTax}$) based on Table 4.7. This implies that the lifetime energy and emission cost is significantly reduced by the optimized operation.

Effects of uncertain factors

To evaluate the effects of load and fuel price growth, sensitivity analysis is performed on two values for the natural gas price and two for the average electrical load. With the four combinations of them, the

lifetime costs for the microgrid with a total gas turbine capacity of 12,800kW, PV panels of 500 kW and battery of 200 kW are compared in Table 4.8 as follows.

Table 4.8. Ex2: Sensitivity analysis

Natural gas price	Average Load	Grid-Cone	Total lifetime cost (M$)	Compared with the nominal (%)
0.27$/m^3	2,479kW	Syn	50.9	/(Nominal)
		Non	47.61	/(Nominal)
	2,727kW (10% increase)	Syn	54.91	7.88
		Non	51.62	8.42
0.297$/m^3 (10% increase)	2,479kW	Syn	55.37	8.78
		Non	51.79	8.78
	2,727kW (10% increase)	Syn	59.4	16.7
		Non	55.83	17.27

Under both synchronized and non-synchronized grid-connection, fuel price growth has a little bit more effects on the total lifetime cost than load growth. This is because the profits made by the microgrid is closely related to the fuel price. Natural gas price almost has the same effects on the lifetime costs under two types of grid connections, while load growth has more under non-synchronized grid-connection.

4.6 Implications for Regulators and Distribution Utilities

The implications of the above models, methods and results on operation and design of microgrids with renewables for regulators and distribution utilities are discussed below.

Historically, electric power distribution companies (DISCOs) have been working as investor-owned regulated monopolies in the United Kingdom (UK) and many states in the US. DISCOs own and operate distribution infrastructures to provide unidirectional delivery of power from upstream merchant generators to downstream consumers. This unidirectional engineering and transactional arrangement is often referred to as a cost-of-service business model. Cost-of-service regulators require DISCOs to approximate the optimal investment and operation of the distribution network using discounted cash flow tools or net present value analysis as standard approaches for investment decision-making [63]. However, once the investment

decision is made, the net present value approach assumes that there is no scope for managers to react to new information, although in practice many investments confer future options and management flexibility. In addition, the net present value approach ignores flexibility with regard to timing of an investment decision. Its static nature means that it systematically undervalues investment opportunities which provide future options. Under certain circumstances, e.g., significant uncertainty and flexibility, the net present value approach can lead to poor policy and investment decisions.

More recently, national regulators in the UK and state regulators in New York and California have begun implementing performance-based regulatory reform to convert DISCOs to a bi-directional two-sided platform business model. This platform enables downstream customers who install more reliable, less expensive and more environmentally sustainable distributed generation to interconnect to and transact with the utility grid. The downstream customers can sell spinning reserves, demand response, power quality services to the distribution network, and buy stand-by power from the network. DISCOs will earn income from the performance of the engineering and transactional platform that they own and manage [64, 65].

Under the two-sided platform business model, most DISCOs in the UK use approaches similar to real options analysis to account for the flexibility of distributed energy resources [65]. Real options analysis (based on Monte Carlo simulations) seeks to value flexibility embedded within the investment option and flexibility of delaying the investment through time [63, pp. 4]. So far, New York and California regulators have persisted in the use of less accurate discounted cash flow techniques from cost-of-service regulation to approximate the resource optimization in two-sided platform business models mandated by performance-based regulatory reform [67, 68].

Beyond the methods mentioned above, some regulators and DISCOs are still looking for more sophisticated optimization tools. The optimization models, methods and results demonstrated in this chapter have shown that mixed-integer programming can deliver accurate optimization results with off-the-shelf computational tools (CPLEX and HOMER Pro), and does not substantially increase the complexity

of the applied use of the tools by regulators and DISCO planners.

4.7 Conclusion

This chapter investigates operation and design optimization of microgrids. From the energy and emission point of view, a mixed-integer model is established for operation. PV uncertainties are modeled by a Markovian process. For effective coordination, other devices are modeled as Markov processes with states depending on PV states. The entire problem is stochastic and Markovian, and solved by using branch-and-cut. For design, a linear model is established to evaluate the microgrid lifetime cost, where the reliability cost is obtained based on the microgrid configuration and the cost of unserved load during power outages. With a limited number of possible combinations of device sizes, exhaustive search is used to find the optimized design. Numerical results show that the operation method is efficient in saving cost and scalable, and the lifetime cost is reduced by the optimized design. The optimization models, methods and results demonstrated in this chapter shows that mixed-integer programming can deliver accurate optimization results with off-the-shelf computational tools, and does not substantially increase the complexity of the applied use of the tools by regulators and DISCO planners.

References

[1] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids," *IEEE Power and Energy Magazine*, Vol. 5, Issue 4, pp. 78-94, 2007.

[2] F. Katiraei and M. R. Iravani, "Power management strategies for a microgrid with multiple distributed generation units," *IEEE Transactions on power systems*, Vol. 21, Issue 4, pp. 1821-1831, 2006.

[3] X. Guan, Z. Xu, and Q. Jia, "Energy-efficient buildings facilitated by microgrid," *IEEE Transactions on Smart Grid*, Vol. 1, Issue 3, pp. 243-252, 2010.

[4] S. Mohammadi, S. Soleymani, and B. Mozafari, "Scenario-based stochastic operation management of MicroGrid including Wind, Photovoltaic, Micro-Turbine, Fuel Cell and Energy Storage Devices," *Electrical Power and Energy Systems*, Vol. 54, pp. 1-7, 2014.

[5] Siemens AG. Virtual Power Plants by Siemens, https://w3.usa.siemens.com/smartgrid/us/en/distributech/Documents/DEMS_VPPs.pdf.

[6] Y. Chen, S. Lu, Y. Chang, T. Lee, and M. Hu, "Economic analysis and optimal energy management models for microgrid systems: a case study in Taiwan," *Applied Energy*, Vol. 103, pp.145-154, 2013.

[7] H. Morais, P. Kadar, P. Faria, Z. A. Vale, and H. M. Khodr, "Optimal scheduling of a renewable microgrid in an isolated load area using mixed-integer linear programming," *Renewable Energy*, Vol. 35, Issue 1, pp. 151-156, 2010.

[8] D. E. Olivares, J. D. Lara, C. A. Canizares, and M. Kazerani, "Stochastic-predictive energy management system for isolated microgrids," *IEEE Transactions on Smart Grid*, Vol. 6, Issue 6, pp. 2681-2693, 2015.

[9] M. Ross, R. Hidalgo, C. Abbey, and G. Joos, "Energy storage system scheduling for an isolated microgrid," *IET Renewable Power Generation*, Vol. 5, Issue 2, pp. 117-123, 2011.

[10] R. Palma-Behnke, C. Benavides, F. Lanas, B. Severino, L. Reyes, J. Llanos, and D. Sáez, "A microgrid energy management system based on the rolling horizon strategy," *IEEE Transactions on Smart Grid*, Vol. 4, Issue 2, pp.996-1006, 2013.

[11] A. Parisio, E. Rikos, and L. Glielmo, "A model predictive control approach to microgrid operation optimization," *IEEE Transactions on Control Systems Technology*, Vol. 22, Issue 5, pp.1813-1827, 2014.

[12] T. Niknam, R. Azizipanah-Abarghooee, and M. R. Narimani, "An efficient scenario-based stochastic programming framework for multi-objective optimal microgrid operation," *Applied Energy*, Vol. 99 pp. 455-470, 2012.

[13] M. Hemmati, N. Amjady, and M. Ehsanc, "System modeling and optimization for islanded micro-grid using multi-cross learning-based chaotic differential evolution algorithm," *International Journal of Electrical Power & Energy Systems*, Vol. 56, pp. 349-360, 2014.

[14] B. Zhao, X. Zhang, J. Chen, C. Wang, and L. Guo, "Operation Optimization of Standalone Microgrids considering lifetime characteristics of battery energy storage system," *IEEE Transactions on Sustainable Energy*, Vol. 4, Issue 4, pp. 934-943, 2013.

[15] A. D. Hawkes and M. A. Leach, "Modelling high level system design and unit commitment for a microgrid," *Applied Energy*, Vol. 86, Issue 8-9, pp.1253-1265, 2009.

[16] H. Ren and W. Gao, "A MILP model for integrated plan and evaluation of distributed energy systems," *Applied Energy*, Vol. 87, Issue 3, pp. 1001-1014, 2010.

[17] K. A. Pruitt, B. J. Braun, and A. M. Newman, "Evaluating shortfalls in mixed-integer programming approaches for the optimal design and dispatch of distributed generation systems," *Applied Energy*, Vol. 102, pp. 386-398, 2013.

[18] Y. A. Katsigiannis, P. S. Georgilakis, and E. S. Karapidakis, "Multiobjective genetic algorithm solution to the optimum economic and environmental performance problem of small autonomous hybrid power systems with renewables," *IET Renewable Power Generation*, Vol. 4, Issue 5, pp. 404-419, 2010.

[19] O. Hafez and K. Bhattacharya, "Optimal planning and design of a renewable energy based supply system for microgrids," *Renewable Energy*, Vol. 45, pp. 7-15, 2012.

[20] M. J. Khan and M. T. Iqbal, "Pre-feasibility study of stand-alone hybrid energy systems for applications in Newfoundland," *Renewable Energy*, Vol. 30, Issue 6, pp. 835-854, 2005.

[21] J. Cotrell and W. Pratt "Modeling the feasibility of using fuel cells and hydrogen internal combustion engines in remote renewable energy systems," NREL/TP-500-34648, National Renewable Energy Laboratory, Sep. 2003.

[22] T. Givler and P. Lilienthal, "Using HOMER software, NRELs micropower optimization model, to explore the role of gen-sets in small solar power systems," NREL/TP- 710-36774, National Renewable Energy Laboratory, May 2005.

[23] R. Dufo-Lopeza, J. L. Bernal-Agustina, J. M. Yusta-Loyoa, J. A. Dominguez-Navarroa, I. J. Ramirez-Rosadoa, J. Lujanoa, and I. Aso, "Muliti-objective optimization minimizing cost and life cycle emissions of stand-alone PV–wind–diesel systems with batteries storage," *Applied Energy*, Vol. 88, Issue 11, pp. 4022-4021, 2011.

[24] R. Dufo-Lopez and J. L. Bernal-Agustin, "Multi-objective design of PV–wind–diesel–hydrogen–battery systems," *Renewable Energy*, Vol. 33, Issue 12, pp. 2259-2272, 2008.

[25] P. B. Luh, Y. Yu, B. Zhang, E. Litvinov, T. Zheng, F. Zhao, J. Zhao, and C. Wang, "Grid integration of intermittent wind generation: a Markovian approach," *IEEE Transactions on Smart Grid*, Vol. 5, Issue 2, pp. 732-741, 2014.

[26] M. Di Somma, B. Yan, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Operation optimization of a distributed energy system considering energy costs and exergy efficiency," *Energy Conversion and Management*, Vol. 103, pp. 739-751, 2015.

[27] M. Di Somma, B. Yan, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Multi-objective operation optimization of a distributed energy system for a large-scale utility customer," *Applied Thermal Engineering*, Vol. 101, pp. 752-761, 2016.

[28] B. Yan, M. Di Somma, N. Bianco, P. B. Luh, G. Graditi, L. Mongibello, and V. Naso, "Exergy-based operation optimization of a distributed energy system through the energy-supply chain," *Applied Thermal Engineering*, Vol. 101, pp. 741-751, 2016.

[29] X. Q. Kong, R. Z. Wang, and X. H. Huang, "Energy optimization model for a CCHP system with available gas turbines," *Applied Thermal Engineering*, Vol. 25, Issue 2-3, pp. 377-391, 2005.

[30] P. S. Perez, J. Driesen, and R. Belmans, "Characterization of the Solar Power Impact in the Grid," in *Proceedings of International Conference on Clean Electrical Power*, pp. 366-371, Capri, Italy, May 2007.

[31] W. Palz, "Photovoltaic power generation," *D. Reidel Publishing Company*, pp. 45, 1982.

[32] P. Poggi, G. Notton, M. Muselli, and A. Louche, "Stochastic study of hourly total solar radiation in Corsica using a Markov model," *International Journal of Climatology*, Vol. 20, Issue 14, pp.1843-1860, 2000.

[33] B. O. Ngoko, H. Sugihara, and T. Funaki, "Synthetic generation of high temporal resolution solar radiation data using Markov models," *Solar Energy*, Vol. 103, pp.160-170, 20014.

[34] C. Weber, P. Meibom, R. Barth, and H. Brand, "WILMAR: A Stochastic Programming Tool to Analyze the Large-scale Integration of Wind Energy," Optimization in the Energy Industry, ch. 19, pp. 437–458, Berlin, Germany, Springer, 2009.

[35] Center for Climate and Energy Solutions, Options and Considerations For a Federal Carbon Tax, http://www.c2es.org/publications/options-considerations-federal-carbon-tax

[36] IBM ILOG CPLEX V 12.1 User's Manual.

[37] G&W, http://www.gwelec.com/current-limiting-protector-clip-p-102-l-en.html

[38] Pareto Energy, http://www.paretoenergy.com/our-tech/

[39] M. A. Ortega-Vazquez, and D. S. Kirschen, "Optimizing the spinning reserve requirements using a cost/benefit analysis," *IEEE Transactions on Power Systems*, Vol. 22, No. 1, pp. 24-33, 2007.

[40] Pareto Energy, "GridLink: A New Power Electronics Interconnection Technology Study Final Report," submitted to NYSERDA, Nov., 2014.

[41] HOMER Energy, http://www.homerenergy.com

[42] ConEd, https://apps.coned.com/CEMyAccount/csol/MSCcc.aspx, https://apps.coned.com/CEMyAccount/csol/MSCcc.aspx

[43] EIA, http://www.eia.gov/dnav/ng/hist/n3020ny3m.htm, the price on 2014, June.

[44] Educogen, The European Educational Tool on Cogeneration, second edition, 2001.

[45] U.S. Energy Information Administration, CHP Database: SENTECH Incorporated - C&I CHP Technology Cost and Performance Data Analysis for EIA, June 2010.

[46] NREL, http://www.nrel.gov/analysis/tech_lcoe_re_cost_est.html

[47] http://cdcloans.com/lender/504-rate-history/504-rate-archive/

[48] http://inflationdata.com/Inflation/Inflation_Rate/HistoricalInflation.aspx

[49] R. K. Wassan, M. A. A. Majid, and A. A. Mokhtar, " Impact of Different Repair Assumptions on Repairable System Risk Assessment," *Research Journal of Applied Sciences, Engineering and Technology*, Issue 7, No. 4, pp. 870-874, 2014.

[50] EIA Annual Electric Power Industry Report, 2015, https://www.eia.gov/todayinenergy/detail.php?id=27892

[51] G. Li, P. Zhang, P. B. Luh, W. Li, Z. Bie, C. Serna, and Z. Zhao, "Risk analysis for distribution systems in the northeast US under wind storms," *IEEE Transactions on Power Systems*, Vol. 29, Issue 2, pp. 889-898, 2014.

[52] UConn, "Assessment of the GridLink Technology," an interim report submitted to Pareto Energy, Jul., 2015.

[53] State of New York Public Service Commission, Proceeding on Motion of the Commission in Regard to Reforming the Energy Vision, 2014, http://documents.dps.ny.gov/public/Common/ViewDoc.aspx?DocRefId=%7BCFC6F0D9-6E64-4196-9AC7-1A199132DF5B%7D

[54] New York State Energy Research and Development Authority, New York Presbyterian Hospital Combined Heat and Power Project: Commutating Current Limiter Electric Power Transmission and Distribution (EPTD) Program, 2010.

[55] M. J. Sullivan, M. G. Mercurio, and J. A. Schellenberg, "Updated Value of Service Reliability Estimates for Electric Utility Customers in the United States," Lawrence Berkeley National Laboratory Research Project Final Report, Jan. 2015, http://eetd.lbl.gov/sites/all/files/lbnl-6941e_0.pdf

[56] Pareto Energy, Microgrids for Data Centers http://www.paretoenergy.com/whitepaperfiles/PresentationParetoEnergyMicrogridsForDataCentersWebPageVersion.pdf

[57] EPBD buildings platform. Country reports, ISBN 2-930471-29-8, 2008.

[58] L. Mongibello, N. Bianco, M. Caliano, and G. Graditi, "Influence of heat dumping on the operation of residential micro-CHP systems," *Applied Energy*, Vol. 160, pp. 206-220, 2015.

[59] A. De Pascale and P. R. Spina, "Guidelines for residential micro-CHP systems design," *Applied Energy*, Vol. 97, pp. 673-685, 2012.

[60] E. S. Barbieri, F. Melino, and M. Morini, "Influence of the thermal energy storage on the profitability of micro-CHP systems for residential building applications," *Applied Energy*, Vol. 97, pp. 714-722, 2012. .

[61] R. Carnegie, D. Gotham, D. Nderitu, P.V. Preckel, "Utility scale energy storage systems, benefits, applications, and technologies," State Utility Forecasting Group, The Purdue University, pp. 48, 2013.

[62] M. Rausand and A. Høyland. System reliability theory: models, statistical methods, and applications. Vol. 396. John Wiley & Sons, 2004.

[63] J. Grayburn, Real Options and Investment Decision Making Consultation, United Kingdom Office of Gas and Electric Markets, London, 2012.

[64] Y. D. Lee and S. Y. Park, "Reactive Power Support Capabilities of Nonsynchronous Interconnection Systems in Microgrid Applications," in *Proceedings of IEEE Applied Power Electronics Conference and Exposition (APEC)*, Long Beach, CA, Mar. 2016.

[65] New York Public Service Commission, Request for Three Regulatory Rulings on a Project Financing and Profit Sharing Plan for Demonstrating a Two-Sided Microgrid Platform Business Model in Brooklyn and Queens, Case 15-E-0250 Proceeding on Motion of the Commission in Regard to Petition of Pareto Energy Ltd. to Implement a Microgrid Business Model as a Least-Cost Resource to Meet Reliability Contingencies and Demand Management Objectives at Consolidated Edison, 27 December 2015.

[66] See https://www.ofgem.gov.uk/publications-and-updates/real-options-and-investment-decision-making for OFGEM's real options proposal and the reaction of gas and electric distribution companies.

[67] California Public Utilities Commission, Assigner Commissioners Ruling on Guidance for Public Utilities Code Section 769 – Distribution Resource Planning, Rulemaking 14-08-013 Order Instituting Rulemaking Regarding Policies, Procedures and Rules for Development of Distribution Resources Plans Pursuant to Public Utilities Code Section 769, 14 August 2014.

[68] New York State Public Service Commission, Order Establishing Benefit Cost Analysis Framework, Case 14-M-0101 Proceeding on the Motion of the Commission in Regard to Reforming the Energy Vision, 21 January 2016.